naclo

NACLO thanks the following for their generous contributions:

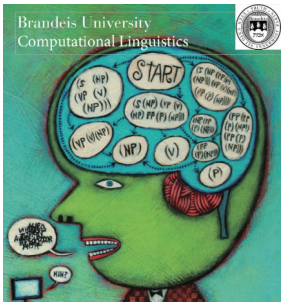
Linguistic Society of America

NAACL

Carnegie Mellon University
Language Technologies Institute

Massachusetts Institute of Technology

MIDDLE TENNESSEE STATE UNIVERSITY



NATIONAL SCIENCE FOUNDATION

Yale University
LUX ET VERITAS

The Sixteenth Annual

North American Computational Linguistics Open Competition

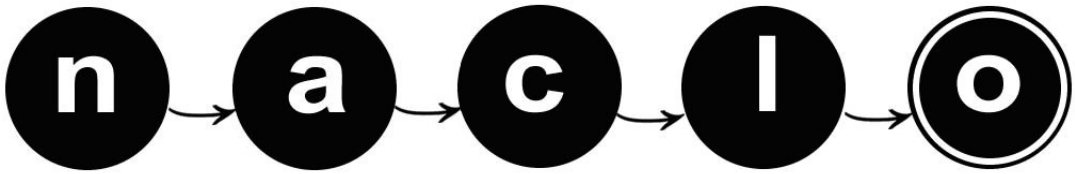
2022

www.nacloweb.org

Invitational Round
March 17, 2022

Serious language puzzles that are surprisingly fun!

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



Welcome to the sixteenth annual North American Computational Linguistics Open Competition! We (the NACLO organizers) are excited for you to participate in this unique event. In order to be completely fair to all participants across North America, we need you to read, understand, and follow these rules completely.

Rules

1. The contest is four hours long and includes nine problems, labeled J 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 & 10.
5. Each problem is worth a specified number of points, with a total of 100 points. **In the Invitational Round, some questions require explanations.**
6. All your answers should be written **clearly** in the Answer Sheets at the end of this booklet. ONLY THE ANSWER SHEETS WILL BE GRADED.
7. Write your name and registration number on each page of the Answer Sheets. Here is an example: Jessica Sawyer #850
8. Some problems are more difficult than others, but all can be solved using ordinary reasoning and some basic analytic skills. You don't need to know anything about linguistics or about these languages in order to solve them.
9. Don't be discouraged if you don't finish everything! If we have done our job well, very few people will solve all these problems completely in the time allotted.
10. **DO NOT DISCUSS THE PROBLEMS UNTIL THEY HAVE BEEN POSTED ONLINE! THIS MAY BE A COUPLE OF MONTHS AFTER THE END OF THE CONTEST.**

<u>Instructions for Virtual Contest Participants</u>

1. Print **one single-sided copy** of the Problems file.
2. Print **two single-sided copies** of the Answer Sheets file.
3. Scan **one copy** of your Answer Sheets to submit for grading. If possible, upload all the pages as one PDF file. Please include all the Answer Sheets pages, even if you left some blank.
4. Be sure your scans are legible before submitting them.
5. If you have technical issues, please ask your facilitator for direction.
6. Please shred this booklet at the conclusion of the contest.

...Oh, and have fun!

NACLO 2022 Organizers

Organizing Committee:

Adam Hesterberg — Massachusetts Institute of Technology
Aleka Blackwell — Middle Tennessee State University
Ali Sharman — University of Michigan
Andrew Tockman — Massachusetts Institute of Technology
Annie Zhu — Harvard University
Ben LaFond — Harvard University
Brian Xiao — Massachusetts Institute of Technology
Cerulean Ozarow — Brown University
Daniel Lovsted — McGill University
David Mortensen — Carnegie Mellon University
Dragomir Radev — Yale University
Duligur Ibeling — Stanford University
Ellina Zhang— University of Toronto
Ethan Chi — Stanford University
Heidi Lei — Massachusetts Institute of Technology
James Hyett — University of Toronto
Jakin Ng— Massachusetts Institute of Technology
James Pustejovsky — Brandeis University
Ji Hun Wang — Stanford University
Ken Jiang — University of Waterloo
Kevin Liang — University of Pennsylvania
Lori Levin — Carnegie Mellon University
Margarita Misirpashayeva — Massachusetts Institute of Technology
Matt Gardner — Allen Institute for AI
Mihir Singhal — Massachusetts Institute of Technology
Nathan Kim — Stanford University
Nathaniel Satriya — University of California, San Diego
Patrick Littell — National Research Council Canada
Pranav Krishna — Massachusetts Institute of Technology
Shuli Jones — Massachusetts Institute of Technology
Skyelar Raiti — University of Michigan
Stella Lau — Massachusetts Institute of Technology
Rui Zhang — Pennsylvania State University
Ryan Chi — Stanford University
Ryan Guan — Stanford University

NACLO 2022 Organizers (cont'd)

Organizing Committee (cont'd)

Tom McCoy — Johns Hopkins University
Tom Roberts — University of California, Santa Cruz
Yilu Zhu — Fordham University

Organizing Committee Co-Chairs:

Kevin Liang — University of Pennsylvania
Shuli Jones — Massachusetts Institute of Technology

Program Committee:

Adam Hesterberg — Massachusetts Institute of Technology
Aleka Blackwell — Middle Tennessee State University
Ali Sharman — University of Michigan
Cerulean Ozarow — Brown University
Daniel Lovsted — McGill University
Dragomir Radev — Yale University
Ethan Chi — Stanford University
Evan Hochstein — Yale University
Jonathan Huang — Massachusetts Institute of Technology
Lori Levin — Carnegie Mellon University
Patrick Littell — University of British Columbia
Pranav Krishna — Massachusetts Institute of Technology
Ryan Chi — Stanford University
Skyelar Raiti — University of Michigan
Tom McCoy — Johns Hopkins University

Program Committee Co-Chairs:

Daniel Lovsted — McGill University
Tom McCoy — Johns Hopkins University

Reviewers and Problem Testers:

Ben LaFond, Jalen Chrysos, Ken Jiang, Matt Gardner, Nathaniel Satriya, William Pan, and others

NACLO 2022 Organizers (cont'd)

USA University Site Coordinators:

Boston Area NACLO Site — Shuli Jones

California State University, Dominguez Hills — Iara Mantenuto

Carnegie Mellon University — John Friday, Lori Levin

College of William and Mary — Dan Parker

Columbia University — Brianne Cortese, Daniel Bauer, Kathy McKeown, Smaranda Muresan

Fort Hays State University — Destiny Gu, Jodi Hill, Sherri Matlock

Georgia Tech — Hongchen Wu

Middle Tennessee State University — Aleka Blackwell

Minnesota State University, Mankato — Dean Kelley, Louise Chan, Rebecca Bates

Montclair State University — Anna Feldman, Jonathan Howell, Lauren Covey

Northeastern Illinois University — Ariana Bancu, Lewis Gebhardt

Ohio State University — Marie de Marneffe, Micha Elsner, Michael White

Planet Word Museum (Washington, D.C. Area NACLO Site) — Emily Gref, Rebecca Roberts

Princeton University — Christiane Fellbaum, Misha Khodak, Oliver Weizel

Rollins College — Margarita Azbel

San Diego State University — Rob Malouf

San Francisco Bay Area NACLO Site — Ethan Chi, Ryan Chi

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

Union College — Kristina Striegnitz, Nick Webb

University at Buffalo — Jeff Good, Leslie Ying, Cassandra Jacobs

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

University of Maryland — Jan Michalowski, Polina Pleshak, Sigwan Thivierge

University of North Carolina at Charlotte — Hossein Hematialam, Kodzo Wegba, Seethalakshmi Gopalakrishnan, Wlodek Zadrozny

University of Notre Dame — David Chiang

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

University of Southern California, ISI campus — Jon May

University of Southern Maine — Claire Holman, Dana McDaniel

University of Utah — Aniello De Santo, Justin Nistler, Karen Marsh Schaeffer

University of Washington — Jim Hoard, Joyce Parvi

University of Wisconsin, Milwaukee — Anne Pycha, Gabriella Pinter, Joyce Boyland

Wichita State University — Jill Fisher, Mythili Menon

Yale University — Raffaella Zanuttini

NACLO 2022 Organizers (cont'd)

Canada University Site Coordinators:
McGill University — Lisa Travis, Michael Wagner
Opus Academy — Janette Lim, Lydia Cheng
University of British Columbia — Jozina Vander Klok, Yadong Liu
University of Ottawa — Andrés Pablo Salanova
University of Toronto — Ellina Zhang

Special thanks to:
The hosts of the 130+ High School Sites

Booklet Editor:
Daniel Lovsted — McGill University

US Team Leaders:
Aleka Blackwell — Middle Tennessee State University
Lori Levin — Carnegie Mellon University

Canadian Team Leader:
Daniel Lovsted — McGill University

NACLO Co-Chairs:
Aleka Blackwell — Middle Tennessee State University
Lori Levin — Carnegie Mellon University

Problem Credits:
(J) Harold Somers
(K) Ryan Chi
(L) Tom McCoy
(M) Gordon Chi
(N) Simi Hellsten
(O) Andrés Pablo Salanova
(P) Ali Sharman
(Q) Simi Hellsten
(R) Ethan Chi and David Mortensen

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

(J) Sounds Fishy (1/1) [5 Points]

As a child learns to talk, they "acquire" the sound system of their language bit by bit, with some speech sounds appearing later than others. Unfortunately, some children have difficulties during this process, and may be referred to a speech therapist. One of the first things a therapist does is try to assess the state of the child's sound "system", and they sometimes do this by administering an "articulation test" in which the child is asked to name pictures and in this way pronounce a set of words specifically chosen to profile the child's sound system.

Here are some examples slightly adapted from a genuine case: Scott, a 4-year-old British boy. Each example contains the target word (the word being pronounced) and Scott's pronunciation (see below the tables for an explanation of the unfamiliar symbols). Note that, in this particular case, we are not interested in the vowels, which are all pronounced "correctly".

Target	*Pronunciation*
church	dɜ:x
teeth	di:x
fish	pix
yellow	jewou
stamps	danx
queen	gi:n
clouds	gauɣ
soldier	douɣə
thumb	dum

Target	*Pronunciation*
Christmas	gixməx
pencil	penduw
flower	bauwə
smoke	hmouk
sneeze	hni:ɣ
wings	wiŋɣ
very	bewi
sugar	dugə
monkey	munʔi

Target	*Pronunciation*
plane	bein
spoon	pu:n
toothbrush	du:xbux
birthday	bɜ:xdei
loose	wu:x
feather	peɣə
elephant	ewipənt
bottle	boʔu
string	diŋg

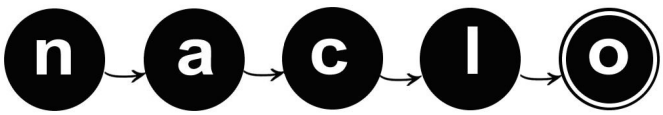
Pronunciation guide:
ʔ is pronounced like the middle sound in "uh-oh"; x like the "ch" in "Bach"; ɣ like x but with vocal cords vibrating; j like the "y" in "yes"; ŋ like the last sound in "sing".
All transcriptions of vowel sounds are the same as the adult target. ə and ɜ are vowel sounds. The symbol : indicates a long vowel. Note that the data comes from a British child, so the r is not pronounced in "soldier", "birthday", etc.

J1. How would you predict that Scott would pronounce the following words? (Your answers should include vowels, but you will not be graded on the specific vowels that you use.)

(a) little (b) friends (c) please (d) chunky
(e) quiz (f) smash (g) shrimps

J2. What do you think Scott is saying here? Give one likely interpretation for each.

(a) danʔ ju bewi mux (b) wox jo: hanɣ an bux jo: di:x



(K) A Tough Word to Swallow (1/1) [15 Points]

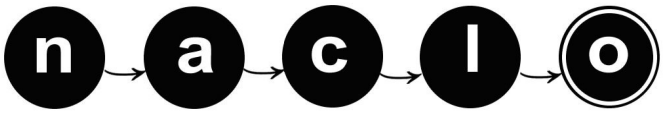
Wik-Mungkan (literally: "to swallow one's words") is a Paman language spoken in Queensland, Australia, by around 1,650 Wik-Mungkan people. On the left below are Wik-Mungkan words and phrases. On the right are their English translations, in a scrambled order.

1. ma' ek	*A. awake*
2. ma' puk pi'an	*B. brave*
3. ma' puuy	*C. crab*
4. ma' thayan	*D. crab shell*
5. mee'	*E. English language*
6. mee' thayan	*F. eye*
7. mee' weep	*G. fingernail*
8. min	*H. fresh water*
9. ngak	*I. good*
10. ngak mee'	*J. handcuffs*
11. ngak min	*K. happy*
12. ngak way	*L. heart*
13. ngangk	*M. law*
14. ngangk ek	*N. sad*
15. ngangk min	*O. shoulder blade*
16. ngangk thayan	*P. sound asleep*
17. ngangk way	*Q. spring (water source)*
18. puuy	*R. strong / firm*
19. puuy ek	*S. thumb*
20. thayan	*T. tired*
21. weep thayan	*U. trustworthy (e.g., with belongings)*
22. wik kiith	*V. undrinkable water*
23. wik thayan	*W. water*

K1. Determine the correct correspondences.

K2. Translate into Wik-Mungkan: a. *hand* b. *bad*

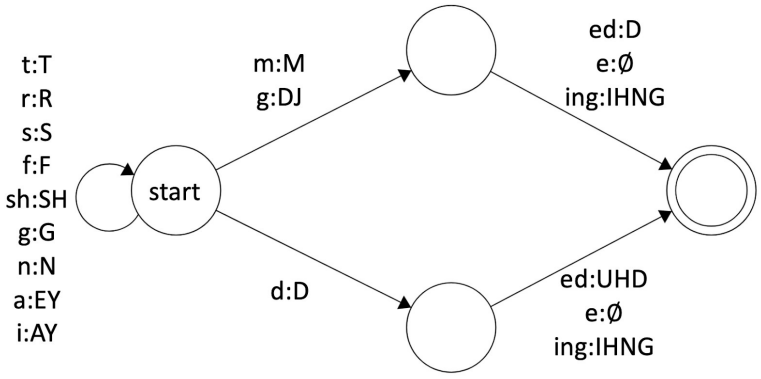
K3. Translate into English: a. weep b. ma' puk



(L) Stopping for a Spell (1/2) [5 Points]

Many types of technology have to convert writing to sounds, a process known as text to speech. For example, a GPS needs to read street names to the person driving the car, and virtual assistants (such as Siri or Alexa) may need to read text from a webpage. An important step in this process is grapheme-to-phoneme conversion: changing a sequence of graphemes (the basic units of writing, such as letters) to a sequence of phonemes (the basic units of speech).

In this problem, we will study finite-state transducers (FSTs), one type of system that can perform grapheme-to-phoneme conversion. Below is an example of an FST:



The FST takes in a sequence of letters (in lowercase, before the colons) and outputs a sequence of sounds (in uppercase, after the colons). The FST starts at the circle labeled "start." When it reads in some lowercase letter(s), it follows the arrow marked with the letter(s) and also outputs the phoneme(s) associated with the letter(s), until the entire input has been used up. For example, given the input "siding," the system would produce SAYDIHNG. Ø is a special symbol which means that no output is produced: for the input "side", the output is SAYD. We need to represent letters and sounds differently from each other because letters can be pronounced differently in different words. For example, the letters ed can be pronounced D (as in "timed") or UHD (as in "sided").

L1. What output would the system produce for the following words?

(a) time **(b)** traded **(c)** striding **(d)** framing

L2. Sometimes, when the system reads in a letter, there are two possible paths that it could follow. In such cases, it tries one path and then, if it gets stuck, it backtracks (goes back) and tries a different path until it finds one that works – somewhat like how you might solve a maze.[1] Exactly one of the following three words could potentially force the system to backtrack – which word is it? Answer on your Answer Sheet.

fading, stage, name

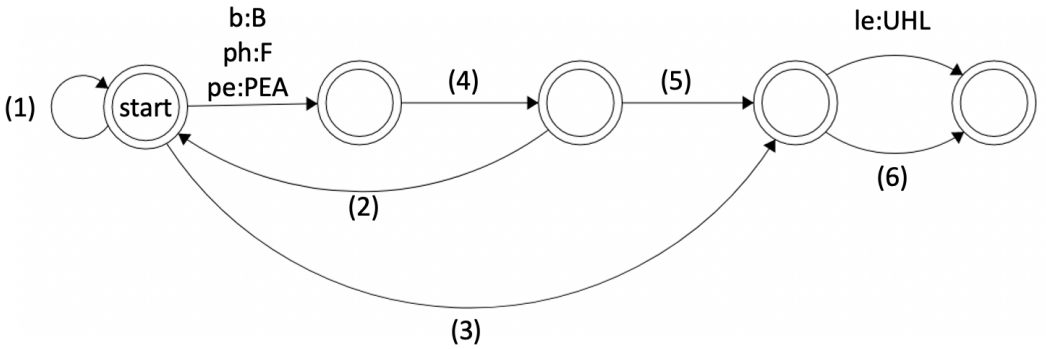
1. It is possible to create an FST that gives more than one output for a given input. However, for all cases used in this contest, a given input will have at most one output.

(L) Stopping for a Spell (2/2)

L3. A path is only valid if it ends at a position with a double circle. With this fact in mind, what output would the system on the previous page produce for the following inputs?

(a) staging **(b)** gaming

L4. Many English words are spelled very strangely. For example, "colonel" is pronounced KUHRNUHL (like "kernel") — there is an R in the pronunciation even though there is no r in the spelling! The FST below is designed to handle some of these exceptions. Match the arrows ((1)-(6)) with their labels ((A)-(F)) so that the system gives the correct outputs for the 5 words listed under the FST.



(A)	(B)	(C)	(D)	(E)	(F)
ui:IH a:EY o:Ø	n:N	c:K o:UH lo:R h:H e:EA t:T	se:Z l:L	p:P	el:UHL

Spelling	Pronunciation
colonel	KUHRNUHL
he	HEA
people	PEAPUHL
phase	FEYZ
built	BIHLT

L5. When using an FST, it is possible to swap what counts as the input vs. the output. In our case, this means that we can provide a sequence of sounds (the symbols to the right of the colons) and have the system produce letters (the symbols to the left of the colons). Since the system is converting sounds into spelling, this process is something like having the system compete in a spelling bee. When you are using the previous FST (the one that handles "colonel"), you try asking it what sequence of letters would be pronounced RUHFLEA. You expect its answer to be "roughly", but instead you get something very different! What sequence of letters does the system say would be pronounced RUHFLEA?

(M) A Splitting Disagreement (1/4) [10 Points]

Note: This problem builds upon the previous problem, (L) Stopping for a Spell, so we recommend solving that one first. There are parts of this problem that you may not be able to solve unless you have first completed (L) Stopping for a Spell.

Thai, the official language of Thailand and a member of the Kra-Dai family, uses a writing system derived from the Old Khmer script. Unlike in English, in Thai writing there are no spaces between words. Thus, **Thai word segmentation** — the task of breaking a piece of Thai text into words — is a complex problem in computational linguistics.

Pavan and Arun are both computer scientists who are trying to develop a word segmentation model for Thai. To evaluate the performance of their model, they use the F1-score which is calculated using the following equation:

$$F_1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

TP represents the number of true positives, FP represents the number of false positives, and FN represents the number of false negatives. A true positive is a case where the correct answer is a positive label, and the model returns a positive label; a false positive is a case where the correct answer is a negative label, but the model returns a positive label; and a false negative is a case where the correct answer is a positive label, but the model returns a negative label.

In order to segment a piece of Thai text into words, Pavan and Arun's models take in the Thai text and assign a label to each character in the text. Specifically, the label should be 1 if the character is the end of a word, or 0 otherwise. As an example, consider the following sentence:

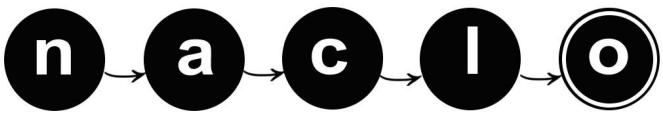
เวลาเท่าไหร่แล้ว? ("What is the time?")

This sentence has 4 words (counting the question mark as a "word"), which we can separate using vertical bars:

เวลา | เท่าไหร่ | แล้ว | ? |

We can further break the sentence into 13 characters. Some characters contain a dotted circle, indicating that the character combines with some other character that goes in the place of the dotted circle:

[เ, ว, ล, า, เ, ท, ◌่, า, ไ, ห, ร, ◌่, แ, ล, ◌้, ว, ?]



(M) A Splitting Disagreement (2/4)

Based on the word boundaries indicated before, a perfect word segmentation output would be:

[0,0,0,1,0,0,0,0,0,0,0,1,0,0,0,1,1]

Initially, Pavan and Arun develop a simple baseline approach. Soon, they learn a bit more about the Thai writing system. Based on this knowledge, they group Thai characters into several groups: They label some as vowels (listed as Vowel at the end of the problem), some as consonants that can appear at the start of a syllable (listed as Initial consonant at the end of the problem), and some as consonants that can appear at the end of a syllable (listed as Final consonant at the end of the problem). Note that some characters appear in more than one category. Using these categories, Pavan and Arun each develop a new algorithm. All three of these algorithms are described below:

Baseline algorithm:

- Label the last character of the sentence with a 1.
- Label all other characters with a 0.

Pavan's algorithm:

- For all characters that appear in the Final consonant list at the end of the problem, label them with a 1.
- Label the last character of the sentence with a 1.
- Label all other characters with a 0.

Arun's algorithm:

- Assign a label of 1 to all characters that satisfy the following criteria:
  - The character is in the Final consonant list
  - The character is preceded by a Vowel
  - The character is followed by an Initial consonant and then a Vowel
  - (In other words, use the label of 1 for a character that fits the FC slot in the following template: V **FC** IC V).
- Label the last character of the sentence with a 1.
- Label all other characters with a 0.

On the next page are 3 Thai sentences along with the F1-score that each algorithm achieves on each sentence. Unfortunately, Pavan and Arun have lost track of which algorithm is which!

(M) A Splitting Disagreement (3/4)

M1. Match each algorithm name from the previous page ("Baseline algorithm", "Pavan's algorithm", or "Arun's algorithm") to its label in the table ("Alg A" to "Alg C"). Answer on your Answer Sheet.

Thai sentence (with word boundaries added)	Character array	F1: Alg A	F1: Alg B	F1: Alg C
ฉัน | จะ | พบ | เขา | ใน | วันพรุ่งนี้ | "I will meet him tomorrow."	[ฉ, ◌ั, น, จ, ะ, พ, บ, เ, ข, า, ใ, น, ว, ◌ั, น, พ, ร, ◌ุ, ◌่, ง, น, ◌ี, ◌้]	0.29	0.60	0.42
เขา | ให้ | ของขวัญ | ฉัน | "He gave me a present."	[เ, ข, า, ใ, ห, ◌้, ข, อ, ง, ข, ว, ◌ั, ญ, ฉ, ◌ั, น]	0.40	0.67	0.36
ฉัน | พูด | ภาษาไทย | "I speak Thai."	[ฉ, ◌ั, น, พ, ◌ู, ด, ภ, า, ษ, า, ไ, ท, ย]	0.50	1.00	0.60

One challenge for Thai word segmentation is that often there is no single right answer: There can be multiple valid ways to break a Thai sentence into words. For example, consider the 4 Thai lines below. In the final example of the table above, example (1) is treated as a single word (meaning "Thai"). However, it is also valid to treat this as two words, listed in examples (2) and (3). If we use this view, then the final sentence would be segmented as shown in example (4).

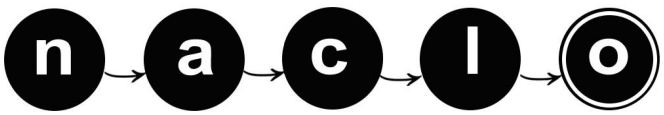
(1) ภาษาไทย "Thai"

(2) ภาษา "language"

(3) ไทย "Thai"

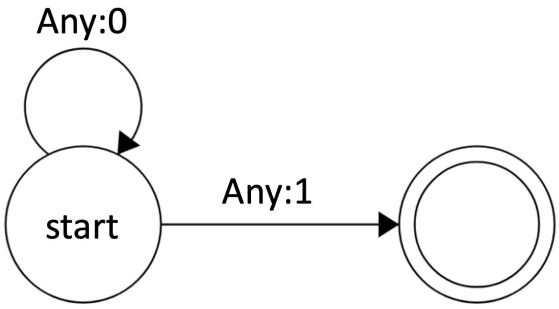
(4) ฉัน | พูด | ภาษา | ไทย |

M2. What F1-score would Arun's algorithm get in the last row of the table above if we used (4) as the intended segmentation, rather than the segmentation shown in the table? If you want, you can leave your answer as a fraction — e.g., 19.4/51.7 — rather than simplifying it into a decimal.



(M) A Splitting Disagreement (4/4)

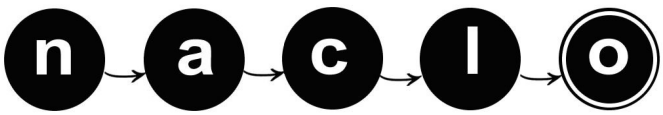
Here is a finite-state transducer that implements the Baseline Algorithm:



M3. Draw **(a)** a finite-state transducer that implements Pavan's Algorithm and **(b)** a finite-state transducer that implements Arun's Algorithm. (See the previous problem, (L), for a definition of finite-state transducers.) It may be helpful to use the category labels listed below; in an FST, one of these category labels can match any character that is a member of that category. For example, in the transducer above, we have used Any to match any single character. If you wish, you can abbreviate these terms — just make sure to include a key for any abbreviations.

List of character categories:

- Punctuation: ?
- Tone: ◌่, ◌้
- Vowel: ◌ั, ◌ู, ◌ี, า, ไ, ใ, ย, เ, แ, ะ, ◌ุ, อ
- Initial consonant: น, พ, ด, ภ, ท, ษ, ว, ล, ร, จ, บ, ข, ง, ญ, ฉ, ห
- Final consonant: น, พ, ด, ภ, ท, ษ, ว, ล, ร, จ, บ, ข, ง, ญ
- Any: Can match any character from any of the 5 categories above
- Not [CATEGORY]: Can match any character that does not belong to the category [CATEGORY], where [CATEGORY] can be replaced with any of the category names above (Punctuation, Tone, Vowel, etc.). For example, Not Vowel would mean any character that is not one of the vowels.



(N) Pseudorandom Numbers (1/2) [15 Points]

Dinka is a Nilotic dialect cluster with about 1.3 million native speakers, mostly ethnic Dinka people in South Sudan. There are several main varieties, but this problem focuses on the Agar dialect.

When linguists first studied the language, they believed that the singular and plural forms of Dinka nouns were completely unpredictable. More recently, however, studying the way that verbs conjugate in Dinka allowed linguists to find patterns in the singular and plural forms. This has allowed many nouns to be grouped according to common patterns, although many remain unexplained.

Below are 22 Dinka nouns, in both singular and plural forms, each of which follow one of the common patterns. The translations have been provided only for interest: they have no bearing on the solution to the problem.

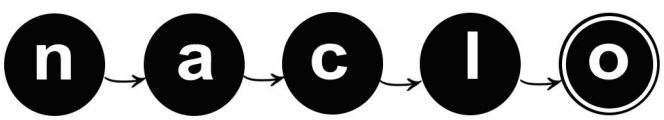
Notes: **ɛ** is the vowel in "bed", and **ɔ** is the vowel in "bought". Dinka has three vowel lengths: short (e.g., **a**), medium (e.g., **aa**), long (e.g., **aaa**); as well as three tones, high (e.g., **á**), low (e.g., **à**), falling (e.g., **â**). **j** and **w** are semivowels, pronounced like the first sounds in "yes" and "with" respectively. **ṭ**, **ḍ**, **ɲ**, **ɟ** and **ŋ** are consonants; how consonants are pronounced is not relevant for this problem. While it is not strictly necessary for solving the problem, it may be helpful to know that vowels can be classified by (among other things) *height*, i.e., how high or low the tongue is in the mouth during their pronunciation. In this problem, **i** and **u** are high, **e** and **o** are high-mid, **ɛ** and **ɔ** are low-mid, and **a** is low.

Singular	Plural	Translation
láj	**làaj**	*animal*
gâaar	**gɛ́ɛr**	*ankle bell*
ŋàaar	**ŋɔ́ɔr**	*bean*
dít	**djɛ̀ɛt**	*bird*
àɟwɔ̀ɔɔŋ	**àɟóoŋ**	*blacksmith*
ḍàaŋ	**ḍɛ̂ɛŋ**	*bow, gun*
gɔ́l	**gàal**	*cowdung fire*
twɔ́ɔŋ	**tóŋ**	*egg*
màac	**mɛ̂ɛc**	*fire*
rúp	**rwòop**	*forest*
àdjɛ́ɛl	**àdíl**	*gazelle*

Singular	Plural	Translation
bôook	**book**	*hide*
ròok	**rôok**	*kidney*
ɲɔ̀ɔk	**ɲɔ̂ɔk**	*louse*
àgâaaɲ	**àgɛ́ɛɲ**	*monitor lizard*
àgɔ̂ɔɔk	**àgɔ́ɔk**	*monkey*
ḍél	**ḍɛ̀ɛl**	*path*
wáal	**wál**	*plant*
ɲêeel	**ɲéel**	*python*
dèeŋ	**dêeŋ**	*rain*
àmàaal	**àmɛ́ɛl**	*sheep*
àtwòoor	**àtúur**	*slime*

Here are some forms of 4 Dinka verbs.

Root	1st person	3rd person	Translation
nɔ̀ŋ	**nàaŋ**	**nɔ̀ɔŋ**	*to have*
kùc	**kwòoc**	**kùuc**	*to not know*
màat	**màaat**	**mɛ̀ɛɛt**	*to smoke*
lɔ̀ɔk	**làaak**	**lɔ̀ɔɔk**	*to wash*



(N) Pseudorandom Numbers (2/2)

N1. Some singular or plural forms of Dinka nouns are given below. Only two of them follow one of the common patterns demonstrated above. On your Answer Sheet, mark which two they are.

	Singular	Plural	Translation
(a)	**àdɛ̀ɛn**		*beautiful one*
(b)	**mìiit**		*firefly*
(c)	**wèeet**		*metal*
(d)		**tôoŋ**	*pot*
(e)		**ɟâak**	*evil spirit*
(f)		**tûuŋ**	*horn*

N2. Assuming that the following verbs conform to one of the common patterns, fill in the blanks in the table below. Answer on your Answer Sheet.

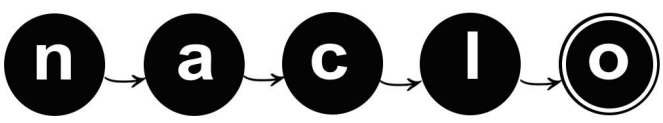
Root	1st person	3rd person	Translation
(a)	**lwɔ̀ɔɔj**	(b)	*to be different*
(c)	(d)	**cɛ̀ɛm**	*to eat*
pèec	**pɛ̀ɛɛc**	(e)	*to loot*
wìc	(f)	**wìic**	*to need*
(g)	(h)	**bòok**	*to throw at*

N3. Below are the singular or plural forms of 10 more Dinka nouns. Assuming that they conform to one of the common patterns, predict the missing forms. If there is more than one possible prediction, give them all. Answer on your Answer Sheet.

Singular	Plural	Translation
(a)	**rím**	*blood*
(b)	**wíil**	*bristle*
àŋâaar	(c)	*buffalo*
rɛ̀ɛɛc	(d)	*fish*
(e)	**kàal**	*hole in ground*

Singular	Plural	Translation
kók	(f)	*hole in tree*
ràaan	(g)	*person*
(h)	**léek**	*pestle*
ról	(i)	*voice*
jìiṯ	(j)	*well*

N4. Explain what you have observed about Dinka nouns and verbs from the data in this problem.



(O) Seeing the Future (1/1) [10 Points]

The Chorote Iyo'awujwa' are a Matacoan people living in the Chaco region of Argentina and Paraguay. A linguist working with one of the varieties of Iyo'awujwa' obtains the following data from a native speaker ("sg." and "pl." mean "singular" and "plural", respectively):

a.	*a'wen*	I see you (sg.), I see him/her/them
b.	*a'wenɬ*	I see you (pl.)
c.	*si'wen*	you (sg.) see me, he/she/they see me
d.	*hi'wen*	you (sg.) see him/her/them
e.	*kasi'wen*	you (sg.) see us, he/she/they see us
f.	*in'wen*	he/she/they see you (sg.)
g.	*in'wenɬ*	he/she/they see you (pl.)
h.	*a'wena*	we see you (sg.), we see him/her/them
i.	*a'wenahaɬ*	we see you (pl.)
j.	*si'wenɬ*	you (pl.) see me
k.	*hi'wenɬ*	you (pl.) see him/her/them
l.	*kasi'wenɬ*	you (pl.) see us

The linguist then starts asking for other tenses. She asks how to say 'you (sg.) are going to see me' and gets the form *si'wehnayi'* from her consultant.

She says to herself, "I got this." She asks her consultant, "Is 'you (pl.) are going to see him/her/them' *hi'wehnayiwɬ*?"

To her surprise, the form she gets is *in'wehnayiwɬ*. The consultant adds the following explanation: "it can also mean 'he/she/they are going to see you (pl.)'; and *si'wehnayi'* can also mean a few other things, by the way: 'I am going to see you (sg.)', 'I am going to see him/her/them', and 'he/she/they are going to see me'."[1]

O1. Translate into Iyo'awujwa':

a. you (sg.) are going to see him/her/them
b. he/she/they are going to see you (sg.)
c. you (sg.) are going to see us
d. you (pl.) are going to see us
e. we are going to see you (pl.)

O2. Describe how to form the Iyo'awujwa' verb meaning "see". Be sure to reference the present ("see") and future ("going to see") tenses in your answer.

1. You may assume that in all cases, *all* the possible translations of a certain Iyo'awujwa' form are given.

(P) Yumology (1/4) [15 Points]

To understand a piece of text, it can be extremely helpful to have some background knowledge about the items discussed in the text: What properties do the items have, and how are they related to each other? This problem deals with the important question of how we can represent such information in a way that a computer can use.

As part of an initiative to increase their nation's health, the Yaldish government has decided to list the mineral potassium (which is abbreviated as K) on their nutrition labels. To ensure proper labeling, the Yaldish Unified Ministry (YUM) maintains a Food Database of Compositions (FDC), but prior to the recent update in requirements, they were not tracking potassium. Obtaining this information for each food listed in the database through lab testing would be time-intensive and costly. The Yaldish have thus hired *NacLabs* to develop a method to supplement the YUM FDC with the English-Language (EL) FDC, which has more complete nutritional information.

The main challenge that *NacLabs* faces is that the food descriptions in the YUM FDC are written in Yaldish. Even though they are also translated into English, the descriptions are not exactly the same as the descriptions of similar foods in the EL FDC (which are described only in English). The following demonstrates these kinds of differences:

Closest matches in YUM FDC, English translations (left) and EL FDC (right):

Chuck roast, uncooked, minced	*Beef, ground, 20% fat, raw*
Puréed vine tomatoes, pasteurized and packaged	*Tomato sauce, canned*

Furthermore, not all foods in the YUM FDC are listed in the EL FDC.

Taking these limitations into account, *NacLabs* has developed an algorithm that automatically fills in potassium for YUM foods. On the next two pages are the YUM FDC (containing the automatically-estimated K values), the EL FDC, and a set of food classification charts. Within each FDC, foods are classified based on four facets plus a fifth "extra facet." The food classification charts illustrate relationships between some of the facets. If you are unfamiliar with any of the food terms in the EL FDC, see the glossary on Page 4 of this problem.

P1. Two foods in the EL FDC are missing part of their description (**(a)** and **(b)**). On your Answer Sheet, fill in the missing information. Word order does not matter as long as the desired meaning is clear.

P2. Three foods in the YUM FDC are missing their "Estimated K mg/100g" values (**(c)**, **(d)**, and **(e)**). On your Answer Sheet, fill in the missing values. Note that the "extra facet" is not involved in determining these values.

(P) Yumology (2/4)

EL FDC

EL ID	Description	K mg/100 grams	Facets	Extra facet
E01	Apple, raw, with skin	107	B1245; C0121; E0151; F0003	A2003
E02	Pineapple rings, homemade, oven-dried from fresh, unsweetened	778	B1484; C0126; E0133; F0013	A2001
E03	Applesauce, canned baby food, unsweetened, no ascorbic acid	74	B1245; C0126; E0215; F0013	A2003
E04	Beet greens, raw	762	B1423; C0240; E0151; F0003	A2003
E05	Bacon	565	B1136; C4545; E0133; F0001	A2003
E06	Bacon, raw	201	B1136; C4545; E0133; F0003	A2003
E07	Bacon, meatless, pan-fried or broiled	170	B1452; C0120; E0133; F0013	A2003
E08	Raisins, golden	746	B1275; C0121; E0151; F0001	A2001
E09	Coconut water, from a coconut	250	B1530; C0339; E0114; F0003	A2003
E10	Beetroot powder, red or golden	2400	B1423; C0140; E1152; F0001	A2001
E11	Pumpkin, canned purée	209	B1534; C0126; E0215; F0013	A2003
E12	Potato **___(a)___**	274	B3544; C0140; E0310; F0013	A2003
E13	Pumpkin **___(b)___**	919	B1534; C0120; E0151; F0013	A2002

YUM FDC

YUM ID	Estimated K mg/100g	Facets	Extra facet
Y1	201	B1136; C4545; E0133; F0003	A2003
Y2	250	B1530; C0339; E0115; F0003	A2003
Y3	250	B1484; C0339; E0114; F0013	A2003
Y4	107	B1245; C0126; E0133; F0003	A2001
Y5	189.5	B1430; C0120; E0215; F0013	A2002
Y6	170	B1430; C0120; E0310; F0013	A2002
Y7	**(c)**	B1423; C0140; E1152; F0001	A2001
Y8	**(d)**	B1245; C0121; E0151; F0013	A2003
Y9	**(e)**	B2530; C0126; E0215; F0013	A2003

(P) Yumology (3/4)

Food classification charts

B1297	B3391	B2002	
		B1136	
	B2299		
	B3399		
B1347	B1140	B1599	B1245
			B1344
		B1139	B1275
		B1024	B4270
			B1530
			B1484
	B1141	B1160	B1534
			B2530
			B3530
		B1165	B1423
			B3542
		B1170	B3544
		B1263	B1430
			B1452
B1438			

C3001	C0760	C0140	
		C0240	
		C0243	
	C0654	C0120	
		C0128	C0121
			C0126
		C0339	
C4322	C2345		
	C2545	C5545	
		C4545	
		C4546	

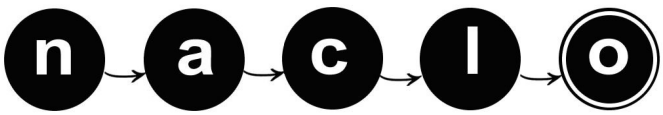
E0130	E0115	
	E0114	
E0103	E0215	
	E0310	
E0152	E0122	E0132
		E0133
		E1152
	E0151	

F0013
F0001
F0003

A2001
A2002
A2003

P3. Briefly describe how the "Estimated K mg/100g" values are determined in the YUM FDC. For this question, you do not need to describe what any specific facets mean. As noted above, your answer to this question should not involve the "extra facets."

P4. Each facet starts with a letter (B, C, E, F, or A). The facets that start with F describe whether the food is cooked. What type of information does each other letter correspond to?



(P) Yumology (4/4)

P5. For each of the following facets, briefly describe what that facet means:

(a) B1245 **(b)** B1530 **(c)** C0240 **(d)** E0310 **(e)** F0013 **(f)** F0001

P6. Name a food ingredient that might have the facet B1438.

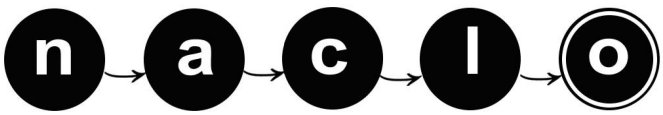
P7. For each of the following YUM IDs from the YUM FDC, give a food description that could be associated with that ID (in the style of the descriptions in the EL FDC). There are many possible answers. For full credit, make sure that your answers cover all of the facets listed with each YUM ID:

(a) Y4 **(b)** Y5 **(c)** Y6 **(d)** Y9

P8. Even though they were not previously using it, NacLabs has decided to now include the "extra facet" in determining the "Estimated K mg/100g" values in the YUM FDC. Will this decision make the estimated values more accurate or less accurate? Explain your answer.

Glossary of food terms:

Apples are a fruit grown on a tree, available in red, green, and yellow varieties.
Applesauce is a dish made of apples (with their seeds and skin removed) blended until smooth.
Ascorbic acid is a chemical used to help preserve foods.
Bacon is a sliced breakfast food, typically made of pork but also available in meatless varieties made out of protein extracted from beans, nuts, grains, etc.
Beet greens are the leaves of a beet plant.
Beetroot is the root of a beet plant.
Broiling is a method of cooking in which the heat source comes from above.
Canning is a food preservation process that involves raising the food to a high temperature and then sealing it in a metal can.
Chuck roast is a type of beef.
Coconut water is a clear liquid found inside coconuts.
Mincing refers to chopping food into very small pieces.
Pasteurization is a process of heating food before packaging it in order to increase its shelf life.
Pan-frying is a method of cooking vegetables and other foods in a pan.
Pineapples are a fruit grown in a shrub.
Potatoes are a root vegetable. They are often served either baked (in which case the whole potato is baked in an oven or microwave) or mashed (in which case the potato is cooked and then pounded with a utensil until it is mostly smooth).
Pumpkins are a type of large orange vegetable that grows on a vine.
Puréeing is the process of blending a fruit or vegetable, often with its seeds and skin removed, into a smooth liquid.
Raisins are dried grapes. They can be dried via heating or by being left out in the air.



(Q) Relatively Speaking (1/1) [15 Points]

Niuean is a Polynesian language spoken by nearly 8,000 people around the world. It is the official language of Niue, a self-governing island in the Pacific, although most speakers of Niuean live in other countries, such as New Zealand.

Below are some sentences in Niuean. For each one, we have also listed one possible translation into English; some sentences have additional possible translations that are not shown. Note that **ā** and **ū** are long vowels, and that **g** represents the *ng* sound in *sing*.

	Niuean	English
1	**Kua kai noa a au.**	*I have only eaten.*
2	**Kua fai fakatino foki ne tā e ia.**	*There have also been pictures that he drew.*
3	**Muhu moa tūmau.**	*There are always plenty of birds.*
4	**Ne fai faiaoga e kāmuta.**	*The carpenter had teachers.*
5	**Kua kitia e ia a au.**	*He has seen me.*
6	**To kai he moa ka holoholo e au e ika.**	*The bird that I will wash will eat the fish.*
7	**Ne totou a Sione.**	*Sione read.*
8	**Tā tūmau e Mele e fakatino.**	*Mele is always drawing the picture.*
9	**Ne kai e ika ne takafaga he tama**	*The fish that the child caught ate.*
10	**To holoholo foki he tama e vaka ne tā he kāmuta.**	*The child will also wash the canoe that the carpenter built.*
11	**To muhu ika a Mele.**	*Mele will have plenty of fish.*
12	**Muhu tama foki e faiaoga ka kitia he moa.**	*The teacher also has plenty of children that the bird will see.*
13	**Fai vaka a Sione ne holoholo e au.**	*Sione has canoes that I washed.*

Q1. Translate the following sentences into English. For sentence (c), there are two possible translations; give them both.

(a) **Fai moa noa.**
(b) **Kua holoholo foki he faiaoga ne takafaga e au a ia.**
(c) **To muhu vaka e tama ka holoholo he moa.**

Q2. Translate the following sentences into Niuean.

(a) *He will also read.*
(b) *Sione has only had fish that the teacher will eat.*
(c) *The teacher that Mele saw built the canoe.*
(d) *There have always been plenty of carpenters.*

Q3. Describe what you have observed about Niuean grammar from the data in this problem.

(R) I Stop Being Afraid of This Problem (1/4) [10 Points]

One common task in developing technologies for human languages is grapheme-to-phoneme conversion (or G2P).[1] In G2P, you convert words written in orthography (the practical writing systems that people use day-to-day) to a standardized phonetic transcription that can be used for recognizing and synthesizing speech, among other things. In this problem, you will help develop G2P for Carib (*Karìnja*), a Cariban language spoken by about 7,400 Carib people in Venezuela, Guyana, Suriname, French Guiana, and Brazil. Below are some words in Carib with their pronunciations and meanings (a guide to the phonetic symbols is provided on the next page). Here are some words to get you started:

Word (orthography)	Phonetic transcription	Meaning
aikuma	/ˈaih.kʲu.ma/	to make something juicy
yrama	/ˈɽa.ma/	I am turned
asaperary	/a.ˈsaʔ.pe.ˈɾa.ɽa/	your cup
õkaikõ	/ˈoŋ.gai.gʲõ/	combs
taweipore	/ta.ˈweih.pʲo.ɽe/	well-lit
saraisarai	/sa.ˈɽai.ʃʲa.ɽai/	the sound of raking
kynapojaton	/kɨ.ˈnaʔ.po.ˈjaʔ.ton/	they feel it
anipynary	/a.ˈniʔ.pʲɨ.ˈna.ɽa/	your love (i.e., the object of your love)
tainawerùke	/ˈtai.nʲa.ˈwe.ɾuh.ke/	having a skin fungus on the hand
asewenàpota	/a.ˈse.we.ˈnah.po.da/	one after another
sapera	/sa.ˈbe.ɾa/	cup
apo	/ˈaʔ.po/	feel
erèny	/e.ˈɾeʔ.nɨ/	nervous motion
ytonoroipory	/ˈto.no.ˈɽoih.pʲo.ɽo/	my *Matayba* tree
sampura	/ˈsam.bu.ɽa/	drum
sukurusaniry	/su.ˈgu.ɽu.ˈsa.ni.ɾi/	of the candy
Sipanijorory	/ʃi.ˈbʲa.ni.ˈjʲo.ɽo.ɽo/	of the Spaniard
yjenarĩkepy	/ˈje.na.ˈɽiŋ.gʲe.bɨ/	I stop being afraid
sikirìma	/ʃi.ˈgʲi.riʔ.mʲa/	to divide into pieces

(Table is continued on the next page.)

1. G2P also showed up in problem (L), Stopping for a Spell. However, solving problem (L) will give you no advantage in solving this problem, and vice versa.

(R) I Stop Being Afraid of This Problem (2/4)

(Table continued from the previous page.)

pẽputu	/ˈpem.bu.du/	dung beetle
aperẽperẽka	/a.ˈbe.ɾem.ˈbe.ɾeŋ.ga/	you are made to flap
enaro	/e.ˈna.ɽo/	governor
ytuwarõ	/ˈtu.wa.ɽõ/	I am forgotten
pirai	/ˈpi.ɾai/	piranha
uwẽposapariky	/u.ˈwem.bo.ˈsaʔ.pa.ˈɽiʔ.kʲɨ/	puffiness of the belly
mòwusa	/ˈmoʔ.wu.sa/	to sharpen
jarã	/ˈja.ɽã/	fence

As you can see, some consonants in the phonetic transcription of Carib are different from those of English. To help you understand how these consonants are pronounced, here is some information on their location in the mouth and manner of articulation (ɽ is pronounced farther back in the mouth than ɾ). "Voiced" and "voiceless" indicate that the vocal cords vibrate and don't vibrate, respectively, during pronunciation. Note also that ʲ after a consonant represents palatalization, or softer articulation, and ˈ before a syllable indicates that it is stressed (i.e., pronounced more emphatically, like the first syllable of the geographical feature "desert", or the second syllable of the food "dessert"). The boundary between syllables is marked with a period.

		Labial (lips)	Coronal (tip of tongue against roof of mouth)	Palatal (middle of tongue against hard palate)	Velar (back of tongue against soft palate)	Glottal (vocal cords)
Nasal	Voiced	m	n		ŋ	
Stop	Voiceless	p	t		k	ʔ
	Voiced	b	d		g	
Fricative	Voiceless		s	ʃ		h
Liquid	Voiced		ɾ, ɽ	j	w	

(R) I Stop Being Afraid of This Problem (3/4)

R1. On your Answer Sheet, fill in the following table to provide rules for the pronunciation of orthographic *y*. Note that for each answer you provide in the "Phoneme" column (i.e., answers **(c)**, **(e)**, **(g)**, and **(h)**), you should answer with a single phoneme (i.e., a single character used in the phonetic transcriptions). Multiple correct answers are possible.

Grapheme	Environment		Phoneme
if *y*	**(a)**	it is silent	—
but if *y*	**(b)**	it is pronounced as	**(c)**
but if *y*	**(d)**	it is pronounced as	**(e)**
but if *y*	**(f)**	it is pronounced as	**(g)**
otherwise	—	y is pronounced as	**(h)**

R2. On your Answer Sheet, fill in the following table to provide rules for the pronunciation of orthographic *p*, *t*, and *k*. For this question, you should ignore palatalization (none of your answers in the "Phonemes" column should include the ^{j} symbol).

Note that for each entry you fill in the "Phonemes" column (i.e., **(c)**, **(e)**, and **(f)**), you should provide three answers of one or more phonemes each, and you should order your answers respectively, separated by commas (i.e., with the pronunciation of *p* first, then *t*, then *k*). Multiple correct answers are possible.

Grapheme	Environment		Phonemes
if *p*, *t*, or *k*	**(a)**	they are pronounced as	h.p, h.t, h.k respectively
but if *p*, *t*, or *k*	**(b)**	they are pronounced as	**(c)**, respectively
but if *p*, *t*, or *k*	**(d)**	they are pronounced as	**(e)**, respectively
otherwise	—	they are pronounced as	**(f)**, respectively

(R) I Stop Being Afraid of This Problem (4/4)

R3. On your Answer Sheet, fill in the blanks in the following table, using the G2P rules you developed based on the data (including, but not limited to, the rules from **R1** and **R2**).

Word (orthography)	Phonetic transcription	Meaning
makopamy	**(a)**	to grow dark
aitopòma	**(b)**	homeless
kerikeri	**(c)**	a species of bird
parimy	**(d)**	son in law of
kurijara	**(e)**	boat
ykurijarary	**(f)**	my boat
tykupimy	**(g)**	what needs to be bathed

R4. On your Answer Sheet, explain the G2P rules you developed for Carib based on the data. You do not need to repeat your rules from **R1** and **R2**.

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

Answer Sheets

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 of the Answer Sheets**, and **turn in all pages of the Answers Sheets** even if you have left some blank .

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

Signature: _____

YOUR NAME: REGISTRATION #

Answer Sheets (1/11)

(J) Sounds Fishy

J1. Write your prediction for Scott's pronunciation of:

(a) little

(b) friends

(c) please

(d) chunky

(e) quiz

(f) smash

(g) shrimps

J2. Give one likely interpretation for each of the following things Scott says:

(a) danʔ ju bewi mux

(b) wox jo: hanɣ an bux jo: di:x

(K) A Tough Word to Swallow

K1. In each box, write the letter of the English word/phrase that corresponds to the Wik-Mungkan word/phrase of that number.

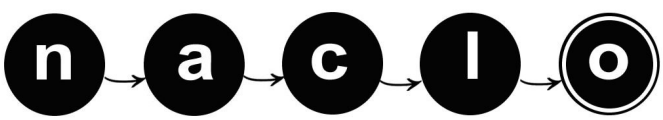
1. 2. 3. 4. 5. 6. 7. 8.

9. 10. 11. 12. 13. 14. 15. 16.

17. 18. 19. 20. 21. 22. 23.

K2. Translate into Wik-Mungkan: a. *hand* b. *bad*

K3. Translate into English: a. weep b. ma' puk



YOUR NAME: REGISTRATION #

Answer Sheets (2/11)

(L) Stopping for a Spell

L1. Give the system's output for each of the following:

(a) time

(b) traded

(c) striding

(d) framing

L2. Exactly one of these words could potentially force the system to backtrack – circle that word:

fading stage name

L3. Give the system's output for each of the following:

(a) staging

(b) gaming

L4. In each box, write the letter (A-F) whose label corresponds to the arrow of that number:

1. 2. 3. 4. 5. 6.

L5. The sequence of letters that the system says would be pronounced RUHFLEA is:

(M) A Splitting Disagreement

M1. Write a letter (A, B, or C) in each box to match the algorithms in the table to their names:

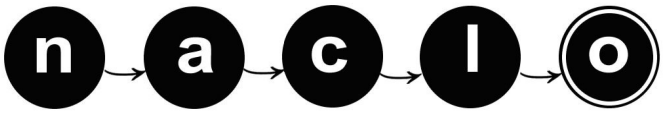
Baseline algorithm:

Pavan's algorithm:

Arun's algorithm:

M2. The F1-score of Arun's algorithm would be:

(continued on the next page)



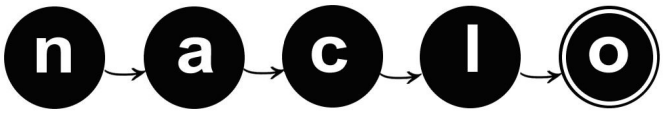
YOUR NAME: REGISTRATION #

Answer Sheets (3/11)

(M) A Splitting Disagreement (continued)

M3. Draw a finite-state transducer that implements: **(a)** Pavan's algorithm:

(b) Arun's algorithm:



YOUR NAME: REGISTRATION #

Answer Sheets (4/11)

(N) Pseudorandom Numbers

N1. Circle the letters of the two forms that follow one of the common patterns:

(a) (b) (c) (d) (e) (f)

N2. Fill in the blanks:

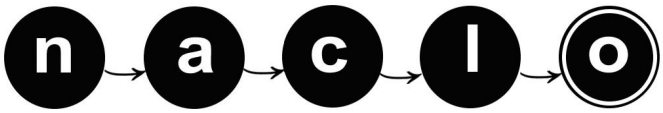
Root	1st person	3rd person	Translation
(a)	**lwɔ̀ɔɔj**	(b)	*to be different*
(c)	(d)	**cɛ̀ɛm**	*to eat*
pèec	**pɛ̀ɛɛc**	(e)	*to loot*
wìc	(f)	**wìic**	*to need*
(g)	(h)	**bòok**	*to throw at*

N3. Fill in the blanks:

Singular	Plural	Translation
(a)	**rím**	*blood*
(b)	**wíil**	*bristle*
àŋâaar	(c)	*buffalo*
rɛ̀ɛɛc	(d)	*fish*
(e)	**kàal**	*hole in ground*

Singular	Plural	Translation
kók	(f)	*hole in tree*
ràaan	(g)	*person*
(h)	**léek**	*pestle*
ról	(i)	*voice*
jìiɪ̩	(j)	*well*

(continued on the next page)



YOUR NAME: REGISTRATION #

Answer Sheets (5/11)

(N) Pseudorandom Numbers (continued)

N4. Explain what you have observed about Dinka nouns and verbs from the data in this problem:

(O) Seeing the Future

O1. Translate into Iyo'awujwa':

a. you (sg.) are going to see him/her/them

b. he/she/they are going to see you (sg.)

c. you (sg.) are going to see us

d. you (pl.) are going to see us

e. we are going to see you (pl.)

(continued on the next page)

YOUR NAME: REGISTRATION #

Answer Sheets (6/11)

(O) Seeing the Future (continued)

O2. Describe how to form the Iyo'awujwa' verb meaning "see":

(P) Yumology

P1. Fill in the blanks **(a)** and **(b)**:

EL ID	Description	K mg/100 grams	Facets	Extra facet
E12	Potato **(a)**	274	B3544; C0140; E0310; F0013	A2003
E13	Pumpkin **(b)**	919	B1534; C0120; E0151; F0013	A2002

P2. Fill in the blanks **(c)**, **(d)**, and **(e)**:

YUM ID	Estimated K mg/100g	Facets	Extra facet
Y7	**(c)**	B1423; C0140; E1152; F0001	A2001
Y8	**(d)**	B1245; C0121; E0151; F0013	A2003
Y9	**(e)**	B2530; C0126; E0215; F0013	A2003

(continued on the next page)

YOUR NAME: REGISTRATION #

Answer Sheets (7/11)

(P) Yumology (continued)

P3. Briefly describe how the "Estimated K mg/100g" values are determined in the YUM FDC:

P4. What type of information does each letter correspond to?

B:

C:

E:

F: whether the food is cooked

A:

P5. For each of the following facets, briefly describe what that facet means:

(a) B1245

(b) B1530

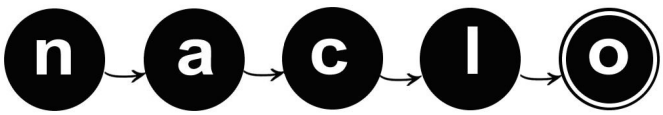
(c) C0240

(d) E0310

(e) F0013

(f) F0001

(continued on the next page)



YOUR NAME: REGISTRATION #

Answer Sheets (8/11)

(P) Yumology (continued)

P6. Name a food ingredient that might have the facet B1438:

P7. Give a food description that could be associated with each ID:

(a) Y4

(b) Y5

(c) Y6

(d) Y9

P8. Will the decision make the estimated values more accurate or less accurate? Explain your answer:

(Q) Relatively Speaking

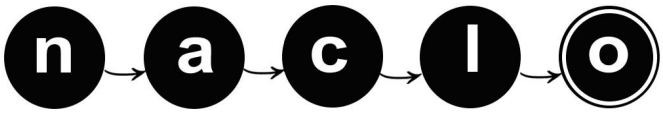
Q1. Translate the following sentences into English. For sentence (c), there are two possible translations; give them both:

(a) **Fai moa noa.**

(b) **Kua holoholo foki he faiaoga ne takafaga e au a ia.**

(c) **To muhu vaka e tama ka holoholo he moa.**

(continued on the next page)



YOUR NAME: REGISTRATION #

Answer Sheets (9/11)

(Q) Relatively Speaking (continued)

Q2. Translate the following sentences into Niuean:

(a) *He will also read.*

(b) *Sione has only had fish that the teacher will eat.*

(c) *The teacher that Mele saw built the canoe.*

(d) *There have always been plenty of carpenters.*

Q3. Describe what you have observed about Niuean grammar from the data in this problem:

YOUR NAME: REGISTRATION #

Answer Sheets (10/11)

(R) I Stop Being Afraid of This Problem

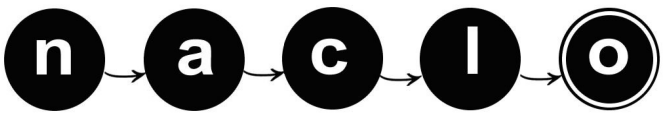
R1. Fill in the blanks (**(a)**, **(b)**, etc.):

Grapheme	Environment		Phoneme
if *y*	**(a)**	it is silent	—
but if *y*	**(b)**	it is pronounced as	**(c)**
but if *y*	**(d)**	it is pronounced as	**(e)**
but if *y*	**(f)**	it is pronounced as	**(g)**
otherwise	—	y is pronounced as	**(h)**

R2. Fill in the blanks (**(a)**, **(b)**, etc.):

Grapheme	Environment		Phonemes
if *p*, *t*, or *k*	**(a)**	they are pronounced as	h.p, h.t, h.k (respectively)
but if *p*, *t*, or *k*	**(b)**	they are pronounced as	**(c)** (respectively)
but if *p*, *t*, or *k*	**(d)**	they are pronounced as	**(e)** (respectively)
otherwise	—	they are pronounced as	**(f)** (respectively)

(continued on the next page)



YOUR NAME: REGISTRATION #

Answer Sheets (11/11)

(R) I Stop Being Afraid of This Problem (continued)

R3. Fill in the blanks (**(a)**, **(b)**, etc.):

Word (orthography)	**Phonetic transcription**	**Meaning**
makopamy	**(a)**	to grow dark
aitopòma	**(b)**	homeless
kerikeri	**(c)**	a species of bird
parimy	**(d)**	son in law of
kurijara	**(e)**	boat
ykurijarary	**(f)**	my boat
tykupimy	**(g)**	what needs to be bathed

R4. Explain the G2P rules you developed for Carib based on the data:

YOUR NAME: REGISTRATION #

Additional Answer Space (1/1)

Clearly indicate which question(s) you are answering on this sheet.