SOLUTIONS

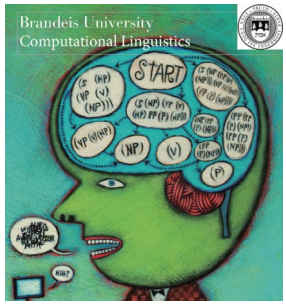
Linguistic Society of America

NAACL

Carnegie Mellon University
Language Technologies Institute

Massachusetts Institute of Technology

MIDDLE TENNESSEE STATE UNIVERSITY



NATIONAL SCIENCE FOUNDATION

Yale University
LUX ET VERITAS

The Fourteenth Annual

North American Computational Linguistics Open Competition 2020

www.nacloweb.org

Invitational Round

March 5, 2020

Serious language puzzles that are surprisingly fun!

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

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

I1.

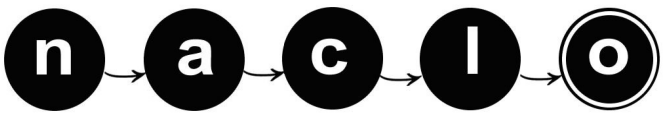
1. S
2. B
3. R
4. J
5. D
6. G
7. F
8. H
9. C
10. E
11. P
12. M
13. O
14. K
15. L
16. T
17. A
18. I
19. N
20. Q

I2.

lõm: air
kãa: kettle/pot
chiãng: city

I3.

beautiful line: *nãeo ngãam*
[Note that the word for “beautiful” is in the description.]



(J) You Made Me See (1/1) [Solution]

J1.

(i) You (s.) heard us.
(ii) They (m.) do not want to heal them (m.) again.
(iii) We did not make you (pl.) visit.

J2.

(i) *aarechaparapiramanahâî isa*
(ii) *nojamachajimari*
(iii) *aapipinimarona*

Explanation

NEG aa-	SUBJ	**stem**	VOL/CAUS -pira -chaji	-ma	REPET -naha	OBJ	3PL.SUBJ -na

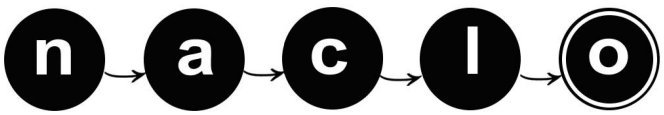
The suffix –*ma* appears in all cases, except when –*pira* by itself.
The subject affixes are as follows:

	Singular	Plural	
1st	*no- / n-*	*a- / aw-*	
2nd	*pi- / p-*	*i- / O-*	*isa*
3rd Masculine	*i- / r-*	*i- / r-*	*isa*
3rd Feminine	*ru- / r-*	*ru- / r-*	*-na*

The object affixes are as follows:

	Singular	Plural	
1st	*-no*	*-î **	
2nd	*-i*	*-î **	*isa*
3rd Masculine	*-ri*	*-na*	
3rd Feminine	*-ro*	*-rona*	

Note: **a-î > âî*



(K) The Pimpled Toad (1/2) [Solution]

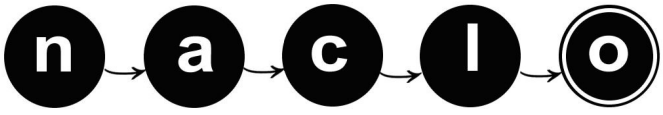
K1.

I-G-E-F-O-P-B-Q-H-C-D-N-J-K-A-L-M

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

A. ***lí-á*** *'crow'*
B. ***já*** *'if'*
C. ***jìeu*** *'toad'*
D. ***lài*** *'so there!'*
E. ***hì*** *'say'*
F. (no words)
G. ***mào*** *'go'*
H. ***áo*** *'water',* ***não*** *'eat'*
I. ***gù*** *'who (is)'*
J. ***dráo*** *'towards / upwards'*
K. ***bào*** *'pool'*
L. ***jáng*** *'stretch'*
M. ***bè-lá-lá á á*** *'all for nothing, ah-ah'*
N. ***jáng*** *'stretch'*
O. ***yā*** *'want',* ***ⁿghãi*** *'flesh'*
P. (no words)
Q. ***ⁿtú-dlĩ*** *'stream'*

K2. ᑕᓗ ᖁₙ, ᖻ᷈ ᒧ° ᒪᵢᵢ ᑐᓗ ᐱ̄ ᒍ" ᑐᓗ ᖶ᷈ ᒣ' ᑕᖶᵣ.



(K) The Pimpled Toad (2/2) [Solution]

K3.

Pollard Script:
• Consonants are written with large letters, vowels with small diacritics.
• The position of the diacritic around the letter gives the tone: directly above = ¯, top right = ´, right = ˜, bottom left = `.
• Nasalization is marked with the letter Ⲥ (*n*) before the consonant.
• The components of compounds are written separately.

Miao Grammar:
• Word order is generally SVO, modifiers after modified
• Zero copula
• Tense not explicitly marked—past tense generally assumed
• *tái* and *tú* both used to mark noun phrases

(L) Shiva Sutras (1/1) [Solution]

L1.

(i) i u ṛ ḷ
(ii) h y v r l
(iii) kh ph ch ṭh th c ṭ t

L2.

(i) ñaM
(ii) aiC
(iii) aL

L3.

A pratyahara consists of the first of the sounds desired, plus the vowel *a* if the first of the sounds desired is a consonant, followed by the capital letter immediately following the last of the sounds desired.

L4.

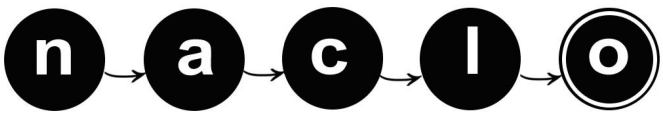
For a member of the set *i u ṛ ḷ*, substitute the corresponding member of the set *y v r l* ('semivowels') when followed immediately by a member of the set *a i u ṛ ḷ e o ai au* ('vowels').

L5.

(i) "for X…"
(ii) "… substitute Y…"
(iii) "… before/when followed (immediately) by Z."

L6.

(i) jagaddhana
(ii) tadduḥkha
(iii) no change



(M) Adjupectiheaval! (1/1) [Solution]

M1.

A is written by a spambot. The use of "yet" here connects two positive adjectives, which although grammatical, does not make sense.

M2.

13. Real
14. Real
15. Spam
16. Spam
17. Real
18. Spam

Each review has the following structure: FOOD ADJ_1 CONJ ADJ_2
where each adjective is either positive (*manis* "sweet", *lezat* "delicious", *menggugah selera* "mouth-watering", *baik* "good", *sehat* "healthy") or negative (*berminyak* "greasy", *hambar* "tasteless", *mahal* "expensive"). We refer to this as the adjective's polarity.

Note also that *tidak* "not" before an adjective negates it.

CONJ may be *dan* "and" or *serta* "and" (which link two adjectives with the same polarity), or *namun* "but" (which links two adjectives with opposite polarities). Incorrect use of these connectors means that the sentence is incorrect.

M3.

The polarity of the three new adjectives introduced here depends on context. Our six food items can be split into two groups from the dictionary entries:

- *Kue*, or soft desserts (*lemang*, *onde-onde*, and *poffertjes*)
- *Kerupuk*, or crackers (*rengginggang*, *rempeyek*, and *kemplang*). These are described as *goreng* ("fried") in the dictionary.

Then *kue* should be soft (*lembut*) and tender (*halus*), while *kerupuk* should be crunchy (*garing*).

28. Spam
29. Spam
30. Real
31. Real

(N) You Can't Handle the Truth (1/2) [Solution]

The 18 total distinct statements come from:

- 2 each from the original three statements (without and with substitution) = 6
- 4 each from the three new identity statements (X = Y, X = X, Y = Y, and Y = X) = 12

N1. For example:
Given a statement "X =/is/was Y" in *TL*,

- replace X with Y in any statement before uttering; or
- replace Y with X in any statement before uttering; or
- both of these.

N2. It depends on the new statement's "form" and how it relates to the old ones. The most basic cases are:

- 1, if new is a non-ID that does not coordinate with any ID
- 2, if new is a non-ID that coordinates with one ID
- 4, if new is ID that does not coordinate
- 5, if new is ID that coordinates with one non-ID

["ID" means a statement of the form, "X =/is/was Y"; non-ID is anything else; "coordinating" means sharing a string]

N3. There is only one correct answer for each:

a. Ruth has 2 + 2 siblings.
b. Ruth knows Ottawa.
c. Ruth prefers the Washington Capitals to the the capital city of Canada Senators.
[note the repeated "the"]
d. Ruth named her stuffed, toy bear after the 26th president of the U.S.
e. Ruth named her stuffed, toy bear the 26th president of the U.S.

(N) You Can't Handle the Truth (2/2) [Solution]

N4. Answer and explanation were graded together as a pair. Some acceptable answer/explanation pairs for each are:

a. *True* — 2 + 2 evaluates to 4
   [Note that "*Weird—it's unnatural to talk in arithmetic"* is <u>not</u> acceptable here. We know that Alfred judges such statements as "The U.K. contains 2 + 2 countries" as True. We are also told that Alfred and Ruth agree on all judgments.]

b. *False* — Ruth may not "know," *i.e.* be familiar with, Ottawa, so we do not know enough about Ruth and should select *False*.
Weird — the change in sense of "knows" (from "can identify" to "is familiar with") likely confuses Alfred and Ruth and prevents them from interpreting the statement easily.
True — Ruth has heard of Ottawa, since she has feelings about the Ottawa Senators, and this is enough to conclude that she "is familiar with" Ottawa

c. *Weird* — repeated "the" is ungrammatical
Weird — "the the capital city of Canada Senators" is not a real entity, so the statement cannot be interpreted.
False — "the the capital city of Canada Senators" is not a real entity, so Ruth has no feelings about it. We do not know whether Ruth feels net positively or negatively about the Capitals, so we do not know enough about Ruth and should select *False*.
True — "the the capital city of Canada Senators" is not a real entity, so Ruth automatically prefers the (existing) Capitals.

d. *True* — Ruth did indeed name the bear after this man. Whether or not she knew he was the 26th president of the U.S. does not affect the truth of the statement.

e. *False* — Ruth's bear is not named "the 26th President of the U.S."
False — Ruth did not "name" her bear (*i.e.* "designate him to serve as") the 26th President of the the U.S. (we can safely assume, or we don't know enough to say for sure.)

N5. Some ideas:

- recognize compound names like "Ottawa Senators" and avoid internal replacements
  - maybe use a list of these, or search for n-grams in a corpus
- avoid replacements after (mental/indirect speech) words like "knows," "believes," "hopes," etc.
  - store a list of these words in memory
- distinguish "use" from "mention" (*e.g.* in sentence (e), "Theodore Roosevelt" is mentioned)
  - use context (other nearby words) to see, *e.g.* whether a human's name is used to talk about a human (probably "use") or a stuffed toy (probably "mention")
- distinguish "is" (and "was") of identity ("TR was the 26th President") from "is" (and "was") of predication ("TR was a president")

(O) We're Counting on Yoruba (1/2) [Solution]

O1.

a. *àádota*	50	10 off 3(score)
b. *àrúndogórin*	85	5 + 4x20
c. *aárùndilogórin*	75	5 before 4x20
d. *ẹ̀tàdogórun*	103	3 + 5x20
e. *òkándilogóji*	39	1 before 20x2

Note: òkán has not been seen in the data but if you realise there are only words for 1-5 because of the counting back system, it must be 1: we've seen 2, 3, 4, 5.

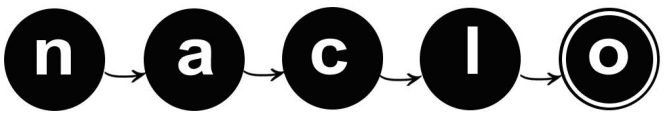
O2.

a. 12	*èjilá*	2 teen
b. 90	*àádorun*	10 before 20x5
c. 57	*ẹ̀ẹ̀tàdilogóta*	3 before 20x3
d. 45	*àrùndogóji*	5 + 20x2
e. 99	*òkándilogórun*	1 before 20x5

Yoruba's counting system is based on 20s, with counting back: the numbers 6, 7, 8, 9 are expressed as 4, 3, 2, 1 before the next ten. 30, 50, 70, 90 are expressed as 10 before the next 20.
In addition, the basic numbers undergo some tone pattern changes.

	Base number	Counting back + *dil*	Counting forward *+ d*	X20 *ogó-*
1	*not shown*	*òkán*	*not shown*	*ogun* * (20)
2	*èji*	*eéjì*	*èjì*	*ji* (40)
3	*ẹ̀ta*	*ẹ̀ẹ̀tà*	*ẹ̀tà*	*ta* (60)
4	*ẹ̀rin*	*ẹ̀ẹ̀rìn*	*ẹ̀rìn*	*rin* (80)
5	*àrun*	*aárùn*	*àrùn*	*run* (100)
Tone pattern	` -	V' `	` `	*ogó* -

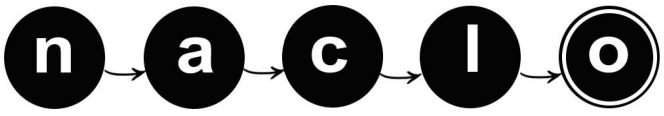
Teens are expressed by adding *lá* to the base number. They are not shown here but there ***are*** base numbers for 6, 7, 8, 9, only used as units. Odd 20s are expressed by *àádo* '10 off'.
The base form of 1 and the form for counting forward, eg 21, 31, do not follow the same tone pattern, and so are not used here. * Likewise 20 *ogun* is not fully regular. But since 20 is a special value, this is not surprising.



(O) We're Counting on Yoruba (2/2) [Solution]

So the given data can be glossed as follows:

èji	2	
ẹ̀rin	4	
àrun	5	
ẹ̀rinlá	14	4 teen
eéjìdilogun	18	2 before 20
ẹ́ẹ́rìndilogóji	36	4 before 20x2
ẹ̀rìndogóji	44	4 + 20x2
àádorin	70	10 off 4(score)
ẹ́ẹ́tàdilogórin	77	3 before 20x4
ẹ̀tàdogórin	83	3 + 20x4



(P) Doubling Up on Nakanai (1/1) [Solution]

P1.

Underlying Form	Reduplicated Form	Meaning
tahalo	***tahalalo***	'man'
sekela	***sekelakela***	'one at a time'
pita	***papita***	'muddy'
bake	***baebake***	(a kind of fish)
loke	***lokeloke***	'to break (a rope)'
voro	***vorovoro***	'to pound'
valolohoka	***valolohokoka***	'to warn someone of trouble'
pasi	***paipasi***	'extremely'
kusa	***kakusa***	'to shout'
bebe	***bebebe***	'butterfly'
hilo	***hililo***	'to see'
sivo	***sosivo***	'to descend'
rabu	***raburabu***	'charred wood'
tarile	***tarilerile***	'tree'
sobe	***soesobe***	'young woman'
vitaumetari	***vitaumetaritari***	'younger sibling'
vituga	***vitatuga***	'to walk'

P2.

Let $C_1V_1(C_2)V_2$ be the last two syllables of the word. Then follow the following rules:

- If C_1 = *h*, insert -V_1C_2- after C_1
- Otherwise, insert before the last two syllables [following the rules in order]:
  - If $C_1V_1 = C_2V_2$, insert -C_1V_1-
  - If there is at least one liquid {*l, r*} in {C_1, C_2}, insert -$C_1V_1C_2V_2$-
  - If $V_1 > V_2$, insert -$C_1V_1V_2$-
  - If $V_2 > V_1$, insert -C_1V_2-

Note that vowels are ranked by sonority as follows: *a > o > e > u = i*

Source: Dimensions of Variation in Multi-Pattern Reduplication (Spaelti, 1997)

(Q) Cut to the Chase (1/1) [Solution]

Q1.

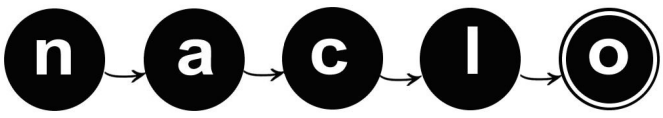
(a) — *á-*
(b) — *áa-*
(c) — *é-*
(d) — *none*
(e) — *-oki*
(f) — *ɔlmʊraní*
(g) — *ɔlmʊráni*

Q2.

A and **F**

Q3.

	Sentence	Translation
A	*ádúŋ ɔlmʊraní ɔlcɛtá*	An ungrammatical sentence.
B	*ádúŋokí ɔlmʊráni ɔlcɛtá*	I cut the tree for the warrior.
C	*ádúŋ ɔlmʊráni*	I cut the warrior.
D	*áaduŋokí ɔlmʊráni*	He/she/it cut the warrior for me.
E	*áaduŋokí ɔlcɛtá*	He/she/it cut the tree for me.
F	*áadúŋ ɔlmʊráni*	An ungrammatical sentence.
G	*áadúŋ ɔlmʊraní*	The warrior cut me.
H	*édúŋ ɔlmʊraní ɔlcɛtá*	The warrior cut the tree.
I	*édúŋokí ɔlmʊráni ɔlcɛtá*	He/she/it cut the tree for the warrior.
J	*édúŋokí ɔlmʊraní*	The warrior cut him/her/it for him/her.
K	*édúŋ ɔlmʊráni*	He/she/it cut the warrior.



(R) The Obviative Solution (1/3) [Solution]

R1.

Singular	Plural	Obviative Singular	Locative Singular	Meaning
hisei	hiseinoʔ	hisein	hiseinewe ʔ	‘woman’
hotii	hotiiwoʔ	hotiiw	hotiiwowe ʔ	‘car’
nebi	nebihoʔ	hibio	nebihewe ʔ	‘one’s older sister’
neicet	neicetino		neicetine ʔ	‘one’s hand’
nooku	nookuhoʔ	nookuo	nookuhowe ʔ	‘beaver’
hiseeθ	hiseetoʔ	hiseet	**a. hiseetewe ʔ***	‘pine tree’
b. ooθ	ooto		oote ʔ	‘leg’
beiciθ	beicito		beicite ʔ	‘tooth’
coox	**c. cooθoʔ**	**d. cooθ**	**e. cooθowe ʔ***	‘enemy’
ceʔeinox	ceʔeinoθo		ceʔeinoθe ʔ	‘bag’
hinen	hinenino ʔ	**f. hinenin**	**g. hineninewe ʔ**	‘man’
wotoo	**h. wotooho**	**i. N/A**	wotoohe ʔ	‘pair of pants’
j. woθonohoe	woθonohoeno	**k. N/A**	woθonohoene ʔ	‘book’
l. niiʔeihii	**m. niiʔeihiiho ʔ**	niiʔeihiio	**n. niiʔeihiihewe ʔ***	‘eagle’
ceʔibes	ceʔibexo	**o. N/A**	**p. ceʔibexe ʔ**	‘block (of wood)’
benes	**q. benexo**	**r. N/A**	**s. benexe ʔ**	‘arm’
t. nesi	nesiho ʔ	**u. hisio**	**v. nesihewe ʔ**	‘one’s uncle’

*For the cells marked ***, ***hiseetowe ʔ***, ***cooθewe ʔ***, and ***niiʔeihiihowe ʔ*** were also marked as correct, respectively. These forms are generated by the “alternate solution,” under Morphology.

(R) The Obviative Solution (2/3) [Solution]

Stems

Nouns have two "stems," S1 and S2. Rules for going from S1 to S2:

	S1	→	S2
a.	*-V*		*-Vh, -Vn, or –Vx*
b.	*-t*		*-tin*
c.	*-n*		*-nin*
d.	*-ʊ*		*-t*
e.	*-x*		*-ʊ*
f.	*-s*		*-x*

All transformations except for (a.) can be performed uniquely in the direction of S1 → S2, and all can be performed uniquely in the reverse direction of S2 → S1.

For brevity, the same patterns can also be expressed with the following scheme:

s ↔ *x* ↔ *ʊ* ↔ *t* ↔ *tin* (process of phonetic softening or lenition)

n ↔ *nin*

V ↔ *Vh, Vn,* or *Vw*

and the rule:

Locate the stem's ending (as specifically as possible) in the above chart.
Move one step right across a "↔" to convert S1 → S2, and one step left for S2 → S1.

Animacy

Nouns are classified as animate or inanimate semantically — note, body parts are inanimate, while "pine tree" and "car" are animate (all others are as expected).

Morphology

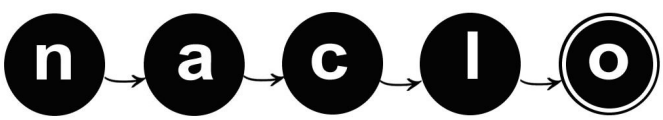
The noun form in each case is:

	singular	plural	obviative singular	locative singular
Animate	S1	S2 + *oʔ*	if S2 *-h,* then S1 + *o* else S2	S2 + *Vweʔ*, where: if root contains *o*, then *V = o* else *V = e**
Inanimate	S1	S2 + *o*	N/A	S2 + *eʔ*

The vowel rule for the locative singular for animate nouns is a kind of vowel harmony.

Note that this rule can be described, consistently with the data, in several ways, including:

- if root contains *e*, *V = e*
- else *V = o*



(R) The Obviative Solution (3/3) [Solution]

*Alternate solution: consistently with the data, the vowel alternation pattern in the locative singular for animate nouns can be explained as:

- *V = e* (for humans)
- *V = o* (for non-humans)

This alternate solution is not an actual phenomenon in Arapaho, but since it fits the data given, it was scored identically, in both parts (R1 and R2), to the vowel harmony solution.

Possession

The prefix “ne-” indicates possession (in English, “one’s…”). In the obviative singular, when this form exists (*i.e.* for animate nouns):

ne- → *hi-*

One way of explaining why this change happens is that the grammatical person is different in the obviative.