n a c l o

SOLUTIONS

NACLO thanks the following for their generous contributions:

LSA
Linguistic Society of America

NAACL

Carnegie Mellon University
Language Technologies Institute

Massachusetts Institute of Technology

Brandeis University
Computational Linguistics

MIDDLE TENNESSEE STATE UNIVERSITY

NATIONAL · SCIENCE · FOUNDATION

Yale University
LUX ET VERITAS

The Fifteenth Annual

North American Computational Linguistics Open Competition

2021

www.nacloweb.org
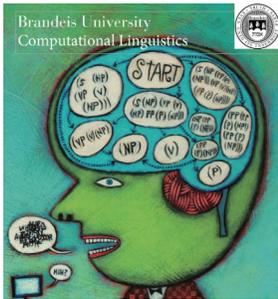
Invitational Round
March 11, 2021

Serious language puzzles that are surprisingly fun!
-Will Shortz, Crossword editor of The New York Times and Puzzlemaster for NPR

(J) A Vintage Sound System (1/1) [Solution]

J1.

(1) *lˤəʔ*
(2) *põk*
(3) *pãt*
(4) *bˤak*
(5) *dzō*
(6) *dā*
(7) *sə*

J2.

1. B
2. F
3. E
4. A
5. D
6. C

J3.

(1) *kjē*
(2) *nō*
(3) *ljẽk*
(4) *pjē*
(5) *bó*
(6) *mà*

J4. The true Classical Era pronunciations are underlined; other options consistent with the data shown are listed after. We did not require students to provide all options for full credit.
<u>ŋaj</u>, ŋe, ŋraj, ŋre, ŋra, ŋa
<u>tˤək</u> [accept also *tˤrək* and several other options]

J5. To convert Classical Era into Medieval Era:

1. Initial consonants:
   r becomes *l*
   l becomes *d*
   r after a consonant disappears
   pharyngealization disappears
2. Vowels:
   firstly, *-aj-* loses its final *-j* (as stated in the introduction)
   if the initial consonant was pharyngealized, *-ə-* becomes *-o*; no change to other vowels
   otherwise, *-a-* and *-e-* become *-je-*, while *-ə-* becomes *-i*
3. Final consonants:
   final *-t*, *-k*, and *-p* cause entering tone
   final *-(C)s* disappears (along with any consonants before it), leaving departing tone (`)
   final *-ʼ* disappears, leaving rising tone (´)
   otherwise, level tone is used (¯)
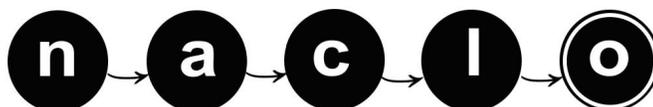
(K) Putting a Place to a Name (1/1) [Solution]

The matching-up part of the problem is complicated by the fact that the names are not directly transliterated. Nevertheless, the usual strategy of starting with the longest name is a good start. Some of the names are somewhat different, and Rabat (Arbat) and Safi (Asfi) could cause a little confusion, as could the fact that there are three different letters transliterated as 'T'; but the two "really quite different" cases are Algiers (Dzhayt) and Casablanca (Anfa), and students could be left with these two, not knowing which is which. The clue for the correct matching is in the introduction, where it is noted that Algeria is called "Dzhayr" in Tamazight. Or they might spot the similarity between Anfa and the last part of (Casabl)anca.

K1.

1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
N.	M.	Y.	H.	G.	F.	X.	Z.	P.	L.	C.	V.	I.
14.	**15.**	**16.**	**17.**	**18.**	**19.**	**20.**	**21.**	**22.**	**23.**	**24.**	**25.**	**26.**
A.	S.	D.	W.	T.	K.	Q.	J.	E.	U.	R.	B.	O.

K2. Africa (the Tamazight name is more directly transliterated as Taferka)

Here's the alphabet in its own sort order, with Latin equivalents. Not all the letters occur in the data:

ⴰ ⴱ ⴲ ⴳ ⴴ ⴵ ⴶ ⴷ ⴸ ⴹ ⴺ
a b b̲ g ḡ dj j d d̲ ḍ d̲

ⴻ ⴼ ⴽ ⴾ ⴿ ⵀ ⵁ ⵂ ⵃ ⵄ ⵅ ⵆ ⵇ ⵈ ⵉ ⵊ
e f k k k̲ h h ḥ kh ʔ q i j j ẓ l

ⵍ ⵎ ⵏ ⵐ ⵑ ⵒ ⵓ ⵔ ⵕ ⵖ ⵗ ⵘ
m n ny ng p u r ṛ gh s s

ⵛ ⵜ ⵝ ⵞ ⵟ ⵠ ⵡ ⵢ ⵣ ⵤ ⵥ
sh t t ch ṭ v w y z z ẓ

(L) Is This Problem Intelligible? (1/6) [Solution]

Background and Notes

First, let's learn a little more about Hawu and Dhao history and culture. It's nice to get some broader context, and it also shows respect for the people who speak and who have historically spoken the languages presented in this problem.

Today the majority of Hawu people practice Protestant Christianity, but until the 1970s most of the Hawu people maintained their ancestral religion, Jingi Tiu, and traditional ways of life. In pre-colonial times, the Hawu people maintained six religio-political domains on the island of Hawu (Savu), including the domain of Seba (after which the modern-day Seba, mentioned in the problem, is named). Each domain had a Council of Priests (of Jingi Tiu). These domains were reorganized during and after the colonial era until a 2001 autonomy law, which the East Nusa Tenggara province of Indonesia used to create new districts that largely reflected the structure of the ancient domains. The Hawu people traditionally transmitted their knowledge orally in a number of ways, for example by naming places in remembrance of key events in Hawu history and naming people by the roles they played in society. The Hawu remember long genealogies over tens of generations that have links at certain stages to the various parts of Hawu island. They recognize two matrilineal lines descended from two sisters as well as patrilineal descent groups and lineages. These genealogies structure Hawu society and are recited during ritual performances; for example, at funerals the reciting of the departed's genealogy memorializes the connection of the departed to his or her ancestors.

Source: Duggan, G. (2009). The Genealogical Model of Savu, Eastern Indonesia. *Journal of Indonesian Social Sciences and Humanities*, *2*, 163–177. LIPI / KITLV Press. https://web.archive.org/web/20120302193735/http://www.kitlv-journals.nl/index.php/jissh/article/viewFile/URN:NBN:NL:UI:10-1-100162/8444

The Dhao people say their language and culture is similar to that of Hawu island. They recount that the first settlers of their island were three people named Rika, Jote, and Pesa Kèli. Pesa Kèli came from the island of Hawu and brought the indigo plant, dhao, from which Dhao (Ndao) island and the Dhao people get their names. (The Hawu also tell a version of this history.) Traditionally, many Dhao men practiced gold- and silversmithing, traveling around the region to sell their creations. Today, most Dhao men have shifted away from smithing to fishing and local business activities. Women traditionally wove fabrics dyed with a technique known as ikat and are still productive in ikat weaving today. They leave their homes to sell their products, to seek orders for new weavings, or to collect debts from their customers. Thousands of jewels and ikats are produced each year and are traded with neighboring islands. The men tend to leave the island during the dry season to sell jewelry and other products of smithing and the ikat weaving products made by the women. In contrast to the Hawu, the Dhao trace their descent only patrilineally.

Source: Balukh, J. I. (2020). *A grammar of Dhao: An endangered Austronesian language in Eastern Indonesia*. https://scholarlypublications.universiteitleiden.nl/access/item%3A2967898/view

(L) Is This Problem Intelligible? (2/6) [Solution]

Explanation of Grammar

Sound correspondences (identified using the table at the beginning of the problem):

Hawu	Dhao
h	**s**
u	**u**
p	**p** (but **bh** in **subha**)
a	**a**
r	**r**
d'	**d**
e	**a**
b'	**b**
è	**è**
i	**i** (but **e** in **dedha**)
d	**dh**
n	**n**
t	**t**
l	**l**
o	**o/u**
j'	**j'**

Using the differences between Hawu and Dhao shown in the table above, we can figure out which sentences in **L1** are in Hawu and which are in Dhao. For example, **s** is a sound unique to Dhao, so any sentence with an **s** will be in Dhao. Then we can match the six sentences in Hawu and the six sentences in Dhao to their translations and analyze the grammar.

Grammatical descriptions of Hawu and Dhao are in the table on the next page. Each grammar has been simplified according to the data presented in this problem.

n → a → c → l → o
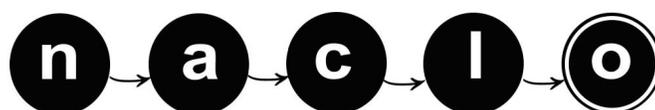
(L) Is This Problem Intelligible? (3/6) [Solution]

Hawu grammar	Dhao grammar
word order: either Verb (Object) Subject (Prepositional Phrase) or V S (O) (PP) **-** in this problem, pronoun objects were only used in VOS order	word order: S V (O) (PP) only

tense/aspect/mood:
    - ta V = non-past tense; otherwise the default interpretation is past tense
        - ke in this problem always appears immediately following **ta** V
    - V … **teruu** = *keep* Ving, *continue to* V
    - b'ole V = *don't* V
	tense/aspect/mood:
	    - no visible tense-marking
	
	    - V … **taruu** = *keep* Ving, *continue to* V
	    - baku V = *don't* V

verbs:
    - pe-V = *to cause to* V (so we call **pe-** a *causative* prefix)
        - this changes an intransitive verb to transitive
            - so, **hewina** = *remember*, **pehewina** = *remind*
            - huti = *spill*, **pehuti** = *spill (something)* (this was tricky!)
	verbs:
	    - pa-V = *to cause to* V
	        - analogous to Hawu
	        - another example:
	            puru = *descend*, **papuru** = *lower (something)*

pronouns:
    - noo = she (or he), 3SG
    - roo = they, 3PL
	pronouns:
	    - na = she (or he), 3SG
	    - ra = they, 3PL

noun phrases:
    - ne marks the object of a transitive verb and the subject of an intransitive verb (this is called *absolutive case*)
        - ne cannot be used with pronouns, unlike **ri**
    - ri marks the subject of a transitive verb (this is called *ergative case*)
    - if there is a possessor, order is Possessed Possessor
        - so, **ngidi dahi**, lit. *edge sea* = *the edge of the sea*
	noun phrases:
	    no visible case-marking
	
	    - if there is a possessor, order is Possessed Possessor
	        - so, **kètu na**, lit. *head 3SG = her head*

prepositional phrases:
    the order is Preposition NounPhrase
	prepositional phrases:
	    the order is Preposition NounPhrase
	
	Note: in Dhao, the word meaning *if* is **ladhe**, the same as the word **ladhe**, meaning *see*. This illustrates the process known as *syntactic change*. The word **ladhe** was probably originally just a verb, but speakers reinterpreted it to also mean *if* and to serve a different grammatical function.

n → a → c → l → o
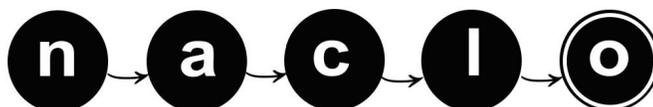
(L) Is This Problem Intelligible? (4/6) [Solution]

Answers to Exercises

L1.
1. Dhao, f
2. Hawu, d
3. Dhao, b
4. Dhao, e
5. Hawu, a
6. Hawu, f
7. Dhao, d
8. Dhao, a
9. Hawu, e
10. Hawu, c
11. Hawu, b
12. Dhao, c

Reorganized for reference:

5. **Ta nèru ke noo oro ngidi dahi.**	Hawu
8. **Na kako madhutu sebhe dhasi.**	Dhao
a. She is walking along the edge of the sea.		
11. **Ta nèru ke roo teruu la Hèb'a.**	Hawu
3. **Ra kako taruu asa Sèba.**	Dhao
b. They keep walking to Seba.		
10. **Ta ngède ke ri roo ne kètu noo.**	Hawu
12. **Ra ladhe kètu na.**	Dhao
c. They see her head.		
2. **Pehewina noo ri roo.**	Hawu
7. **Ra pasanède na.**	Dhao
d. They reminded her.		
9. **Ki ta hewina ke ne ina noo, b'ole pekèd'i.**	Hawu
4. **Ladhe ina na sanède, baku pakèdi.**	Dhao
e. If her mother remembers, don't leave.		
6. **Huti ne èi.**	Hawu
1. **Èi suti.**	Dhao
f. The water spilled.		
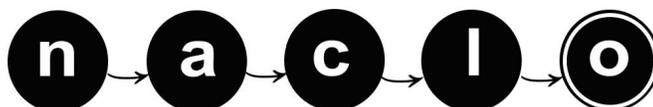
(L) Is This Problem Intelligible? (5/6) [Solution]

L2.

a. **Ra pasanède ina.**   (Dhao)
    They reminded the mother.
b. **Ki ta pedutu ke roo ri ina noo, ta ngède ke noo ri roo.** (Hawu)
    If her mother follows them, they see her.
c. **Pehewina roo ri noo.**   (Hawu)
    She reminded them.
d. **Ladhe na puru, na ladhe sebhe.**   (Dhao)
    If she descends, she sees the edge.
e. **B'ole bèj'i.**   (Hawu)
    Don't sleep/lie down.

Each translation received a score based on (1) translations of words and phrases and (2) syntax.

Accepted variations:
    Any variation in English verb tense was accepted.
    In (a), "mother" and "a mother" were accepted in place of "the mother."
    In (d), "beach" and "shore" were accepted in place of "edge."
    In (d), "goes down" and "comes down" were accepted in place of "descends."

L3.

a. Don't walk to the sea.
        Hawu: **B'ole nèru la dahi.**
        Dhao: **Baku kako asa dhasi.**
b. They keep seeing their mother.
        Hawu: **Ta ngède ke ri roo teruu ne ina roo.**
        Dhao: **Ra ladhe taruu ina ra.**
c. She spilled the water.
        Hawu: **Pehuti ri noo ne èi.**
        Dhao: **Na pasuti èi.**

Each translation received a score based on (1) translations of words and phrases and (2) syntax.

Accepted variations:
    In (b) and (c), the Hawu NPs and continuative aspect marker *teruu* could be permuted with respect to one another (additionally, one NP could move up before *ke*).
    In (b), the Dhao NP *ina ra* and continuative aspect marker *taruu* could be permuted with respect to one another.

L4. madhutu (found in **madhutu sebhe dhasi**, lit. "following the edge of the sea")

Note: ***padhutu** was given half-credit as it is the predicted word based on the Hawu word. ***kaku madhutu** was given half-credit.

n → a → c → l → o

(L) Is This Problem Intelligible? (6/6) [Solution]

L5. Responses received the appropriate score for mentioning strong/weak areas of evidence regardless of whether they were arguing for or against mutual intelligibility. For example, if a response argued that Hawu and Dhao were mutually intelligible, but acknowledged that there were significant differences in word order, this was scored the same (all else equal) as a response arguing that the differences in word order were enough to make Hawu and Dhao not mutually intelligible.

Responses were graded for the ideas presented, not the vocabulary they used. Responses were scored higher for depth of explanation and nuance (for example, acknowledging opposing evidence).

Strong areas of evidence **against mutual intelligibility** included:
- major differences in vocabulary (e.g. there are many non-cognate words)
- word order (e.g. Hawu is verb-initial)
- alignment (e.g. Hawu is ergative)
- Hawu marks agents and objects while Dhao relies on word order
- high degree of difficulty in conversion/translation between languages (as encountered when solving the problem!)

Weak areas of evidence **against mutual intelligibility** included:
- minor differences in vocabulary (e.g. regular sound correspondences with very different sounds)
- isolated examples of grammatical difference
- vague appeals to "grammatical differences" without further explanation

Weak areas of evidence **for mutual intelligibility** included:
- regular sound correspondences (with similar sounds) and shared vocabulary (e.g. pronouns)
- some degree of similarity in sentence structure

The important part was to briefly explain the relevant evidence and recognize which evidence might be more persuasive. For example, it is relevant that the pronouns are similar, but it is also relevant, and more persuasive, that the majority of the words used in the sentences in the problem are extremely different between Hawu and Dhao. As shown above, the strongest evidence suggests that Hawu and Dhao would *not* be very mutually intelligible.

This prediction seems to hold true in the real world. In a 2006 conference paper, Charles Grimes writes that he hasn't met a Hawu speaker who claimed to understand Dhao beyond a few words, and only two Dhao speakers claimed to understand Hawu -- one's mother was a Hawu speaker, and the other had to acquire Hawu to teach on the island of Hawu. At a workshop to train people to write using a computer, "it was immediately obvious to [the Hawu and Dhao speakers] that their languages had more similarities with each other than with other languages around. But they all found it impossible to understand each other in both oral communication and in written form" (p. 18).

He concludes: "How information is strung together in sentences is simply too different" (p. 18).

(M) The Speech Has No End (1/3) [Solution]

Answers

M1.

(17)	**Apo o-na-duhuna babana ae-mi hi-gobu**
	You (pl) will sit down because your (pl) legs are dirty
(18)	**Inapa u-na-towolo apo u-na-mae.**
	If you (sg) stand, you (sg) will stay/be staying.
(19)	**Tewela hi-dewadewa ma natu-m dewadewa-na**
	The children are being good but your (sg) child is always good
(20)	**U-ne-nae nu tahaya daodao-na**
	You (sg) are going to the long path

M2.

(21)	*You (pl) steal a swampy thing.*
	Ginouli bigabiga-na o-danene-ya
(22)	*If the path is long you (sg) will be dirty.*
	Inapa tahaya i-na-dao apo u-na-gobu
(23)	*Limi's difficult child is staying but is not settled.*
	Limi natu-na witewitei-na i-mae ma ega meme-na
(24)	*The children have substance.*
	Tewela ano-hi

M3.

When "child" is used to mean "a young person," it is translated as "tewela."
When "child" is used to mean "(someone's) child", i.e. "son/daughter", then it is translated as "natu" plus a suffix marking the person and number of the "someone" whose child is being referred to:

For "your (sg) child", the Tawala word would be "natum"
For "his/her child", the Tawala word would be "natuna"
For "your (pl) child", the Tawala word would be "natumi"
For "their child", the Tawala word would be "natuhi"

M4.

There are several situations in which Tawala uses reduplication:

A verb (e.g., *peu* "fall") can be reduplicated to form an adjective (e.g., *pipeu* "falling")
A verb (e.g., *witai* "be difficult") can be reduplicated to form a verb whose meaning is durative – that is, ongoing for a long period of time (e.g., *witewitei* "always be difficult")
An adjective (e.g., *gobu* "muddy") can be reduplicated to form a more permanent adjective (e.g., *gobugobu* "stained")
Two specific adjectives, *dewadewa* and *bigabiga*, are always reduplicated.
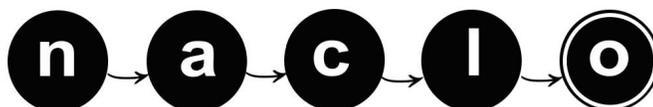
(M) The Speech Has No End (2/3) [Solution]

Explanation of Solution

Syntax: The basic word order is SOV, but subject and object can be omitted if marked on the verb, and adjectives follow nouns. Clauses can be linked by **ma** *and/but* or **babana** *because*. A noun N can be followed by an adjective A to mean "N is A", or a possessed noun N_P to mean "N has N_P". **Ega** *not* precedes any part of the clause it is negating. **Inapa** *if* and **apo** *will* come at the start of the clause.

Nouns: Nouns are not marked for number. They are, however, marked for possessor with the following suffixes:

2s	-m
3s	-na
2p	-mi
3p	-hi

(Technically, these mark inalienable possession.) Note that **natu** means *son/daughter*, so is always possessed; **tewela** means *child* more generally, so is not possessed. **nu** *to* precedes the noun.

Adjectives: Adjectives take the form of reduplicated verbs (see verbs). Both when modifying nouns in a noun phrase or acting as a predicate, they take the possessive suffixes to agree with the noun they modify. They cannot be marked for the potential mood, so cannot act as the predicate in sentences beginning with **apo** or **inapa**.

Verbal Affixes: Verbs mark subject, mood, and object. The subject prefixes are as follows:

2s	u-
3s	i-
2p	o-
3p	hi-

-na- is inserted between the subject prefix and the stem, if the clause begins with **inapa** or **apo**. (This marks the potential mood, but this detail is not necessary for solving the problem). The object suffixes are the same as the possessive suffixes, except that **-ya** marks 3s.

Reduplication: Reduplication has distinct functions, depending on whether the root verb is active or stative. Note that reduplication is unpredictable.

Active verbs reduplicate to form a durative verb, translated as *to be V-ing,* which is still marked as a verb, and so is prefixing. They can also reduplicate to form an adjective, translated as *V-ing*, which are suffixing.

(see table of active verbs on next page)
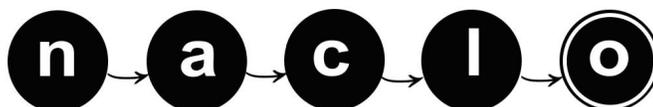
(M) The Speech Has No End (3/3) [Solution]

Active verbs *(see explanation on previous page)*:

Root	Durative	Adjective
-towolo = to stand	**-towotowolo** = to be standing	**towotowolo-** = standing
-dala = to crawl	**-daladala** = to be crawling	**daladala-** = crawling
-danene = to steal	**-danedanene** = to be stealing	**danedanene-** = stealing
-duhuna = to sit	**-duhuduhuna** = to be sitting	**duhuduhuna-** = sitting
-nae = to go	**-nenae** = to be going	**nenae-** = going

Static verbs can only reduplicate to form an adjective. This adjective has a more permanent meaning than the root verb. Two static verbs, **dewadewa** and **bigabiga** are reduplicated in the root form, so do not reduplicate to form an adjective. Thus, the only way to tell whether they are functioning as an adjective or a verb is whether they take a suffix or a prefix.

Root	Adjective
-witai = to be (being) difficult	**witewitei-** = (always) difficult
-gobu = to be dirty	**gobugobu-** = stained
-dao = to be long	**daodao-** = long
-mae = to be staying	**meme-** = settled
-dewadewa = to be (being) good	**dewadewa-** = (always) good
-bigabiga = to be muddy	**bigabiga-** = swampy

How to solve: Most of the problem is fairly routine (e.g. the syntax and affixes). The main difficulty is the reduplication. However, the key is to draw up the following table of how the reduplicated forms of verbs behave:

Redup. form takes both	Redup. form takes suffixes	Always reduplicated
towolo = *stand / is standing*	**witai** = *being difficult / difficult*	**dewadewa** = *being good / good*
dala = *crawl / is crawling*	**gobu** = *is dirty / stained*	**bigabiga** = *muddy / swampy*
danene = *steal / is stealing*	**dao** = *is long / long*	
duhuna = *sit / is sitting*	**mae**= *is staying / settled*	
nae = *go / is going*		

This allows you to see the difference between the active/durative verbs and the stative/permanent verbs, and see that **dewadewa** and **bigabiga** fit in the second category.

(N) Do I Care about Duikers? (1/1) [Solution]

N1.

(a) bie	(b) bonni	(c) dɛr-bo
(d) foloŋfugri	(e) foloŋfugo	(f) ire
(g) irri	(h) nɔtɪrɪ	(i) nɔtɪɛ
(j) ŋmara	(k) ŋmarrɪ	(l) pɔɔ
(m) taŋa	(n) wɛgɛ	(o) wulo
(p) wuli	(q) yagrɪ	

N2. Dagaare nouns are classified semantically as *inherently plural* or *inherently singular*, based on the most common number in which the noun appears.

Inherently plural nouns include body parts of which humans have multiple (lung, toe, elbow, eye, intestine, cheek), animals that are typically found in large groups or swarms (maggot, bee), and objects that are typically used in multiples (shoes, xylophones, drums).

Inherently singular nouns include body parts of which humans have one (face, stomach) and animals and objects that are not typically found or used in multiples (seed, lizard, donkey, ladder, thorn, duiker, the Moon, roof, mountain, solitary wasp, log, bamboo flute).

Inherently plural nouns are <u>marked</u> in the singular (and <u>unmarked</u> in the plural). Inherently singular nouns are <u>marked</u> in the plural (and <u>unmarked</u> in the singular).

Interrogative form: [root] + -bo

Marked form:

      If [root] ends in ŋ, remove ŋ and add nn*I*   *i.e.*   [root – ŋ] + nn*I*

      If [root] ends in l, add *I*   *i.e.*   [root] + *I*

      Otherwise add r*I*   *i.e.*   [root] + r*I*

      where *I* is:

            i if the last vowel in [root] has advanced tongue root (i, e, u, o)

            ɪ if the last vowel in [root] has unadvanced tongue root (ɪ, ɛ, ʊ, ɔ, a)

Unmarked form: [root] + V

      where V is:

            The last vowel in the root, except:

                  if this vowel is high, use the corresponding mid vowel with all other features the same.

(O) Cameroonian Compounds (1/1) [Solution]

O1.

1) B
2) J
3) N
4) K
5) P
6) M
7) G
8) O
9) F
10) C
11) L
12) H
13) E
14) A
15) D
16) I

O2.

a. R   b. U   c. S   d. W

O3.

Answer: *from*

(P) Family Ties (1/1) [Solution]

The Yanomamö people and language follow a kinship pattern commonly called *Iroquois kinship*, named after the Iroquois (also called Haudenosaunee) people in eastern North America. It is one of several kinship systems used across the world. Iroquois kinship systems employ *bifurcate merging*, in which one's father and father's brother(s) are referred to by the same term, and similarly one's mother and mother's sister(s) are referred to by the same term. Also, a distinction is made between one's *parallel cousins* (children of an aunt/uncle whose gender is the same as their sibling parent) and one's *cross cousins* (children of an aunt/uncle whose gender is opposite of their sibling parent). The merging and distinction of these terms reflect practices of inheritance and marriage in Yanomamö society.

P1.

1) Rerebawa
2) Yarima
3) Mukashe
4) Bushika
5) Davi
6) Okori
7) Krihisiwa (given)

P2.

suaboya refers to a daughter of the mother's brother *or* of the father's sister
soriwa refers to a son of the mother's brother *or* of the father's sister
amiwa refers to a daughter of the mother's sister *or* of the father's brother
eiwa refers to a son of the mother's sister *or* of the father's brother

1/3 partial credit was awarded for answering that *suaboya/soriwa* refers to a daughter/son (respectively) of the mother's brother (only). 1/3 partial credit was awarded for answering that *suaboya/soriwa* refers to a daughter/son (respectively) of the father's sister (only).

2/3 partial credit was awarded for answering that *amiwa/eiwa* refers to a daughter/son (respectively) of the mother's sister (only). 1/3 partial credit was awarded for answering that *amiwa/eiwa* refers to a (direct) sister/brother (respectively).
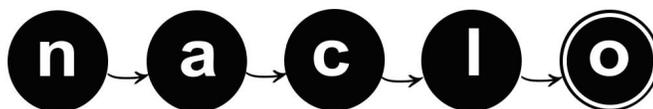
(Q) A Stress Test (1/1) [Solution]

Q1.   **(a)** odd   **(b)** final   **(c)** one

         (d) stressed   **(e)** right

Q2. Responses that received credit included (but were not limited to):

- **Content vs. function words:** function words (e.g., prepositions, pronouns, determiners, conjunctions) are typically unstressed; however, they can be stressed for emphasis.
- **Nouns vs. verbs:** when a two-syllable word can function as a noun or as a verb (e.g., *export*), nouns are stressed as expected (first syllable stressed, second syllable unstressed) whereas verbs get the opposite pattern (first syllable unstressed, second syllable stressed).
- **Emphasis**, largely influenced by context, allows function words to be stressed and can also change which syllable is stressed in a content word. Emphasis can be used to highlight one particular word, or to express a contrast between two words.

Q3.

	Direction	Skip?	Iterate Stress?	Unstress Final?	Primary Stress
Mapudungun	L-to-R	skip	every alternating	leave it	leftmost
Maranungku	L-to-R	don't	every alternating	leave it	leftmost
Weri	R-to-L	don't	every alternating	leave it	rightmost
Mansi	L-to-R	don't	every alternating	unstress	leftmost
Warao	R-to-L	skip	every alternating	unstress OR leave it	rightmost
Comalapa Kaqchikel	R-to-L	don't	only the first	leave it	rightmost OR leftmost

Q4.   Selkup assigns stress according to these rules:

1. Assign primary stress to the rightmost syllable that has a long vowel (written as two copies of the vowel in a row).
2. If there are no long vowels, assign primary stress to the first syllable.
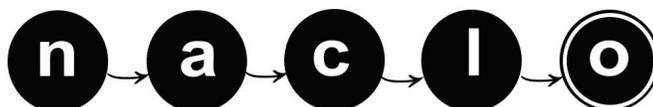
(R) GloVe Compartment (1/3) [Solution]

R1. *Answers:*

1. G	2. K	3. D	4. F	5. H	6. C
7. B	8. I	9. A	10. J	11. E	

Explanation:

The words *man* and *woman* are given as examples. Those 2 words mean the same thing except that they have different genders. If we look at the vectors that are provided for them (which are [0.5, 0.9, 0.3, 0.3] for *man* and [0.5, 0.9, 0.1, 0.5] for *woman*), we see that they have the same first 2 elements, then the last 2 elements differ slightly: the 3rd element is 0.3 for *man* and 0.1 for *woman*, while the 4th element is 0.3 for *man* and 0.5 for *woman*.

We've also been given *daughter* as an example. Based on what we've seen about *man* and *woman*, it seems like a good place to start would be to try to figure out which vector goes with *son*, because we have *daughter* and it seems reasonable to expect that *son* and *daughter* are related in the same way that *man* and *woman* are. We observed that *man* and *woman* have the same first 2 elements as each other, so let's start by assuming that *son* has the same first 2 elements as *daughter*: 0.5 and 0.7. Thus, *son* would be one of the following options:
    [0.5, 0.7, 0.4, 0.1]
    [0.5, 0.7, 0.2, 1.1]
    [0.5, 0.7, 0.4, 0.9]
But which one? We also saw that the 3rd element for *man* was 0.2 plus the 3rd element for *woman.* Thus, perhaps we can expect that the 3rd element of *son* should be 0.2 plus the 3rd element of *daughter*, giving 0.4 – narrowing it down to the first and third of our options. Finally, we saw that the 4th element of *man* was the 4th element of *woman* minus 0.2; so, the 4th element of *son* is probably the 4th element of *daughter* minus 0.2, or 0.3 – 0.2 = 0.1. Thus, we conclude that *son* = [0.5, 0.7, 0.4, 0.1] (2 = K). More generally, we've also figured out that, for a pair of words that differ only in gender, their vectors will differ by [0, 0, 0.2, -0.2].

What other vector pairs do we have that have this difference of [0, 0, 0.2, -0.2]? We have:
    3 = [0.5, 0.9, 0.3, -0.5] and 9 = [0.5, 0.9, 0.1, -0.3]
    11 = [0.5, 0.7, 0.4, 0.9] and 4 = [0.5, 0.7, 0.2, 1.1]
    5 = [0.5, 0.8, 0.9, 1.3] and 7 = [0.5, 0.8, 0.7, 1.5]
    6 = [0.5, 0.8, 0.9, 0.5] and 10 = [0.5, 0.8, 0.7, 0.7]

We also have 4 pairs of male/female words to account for: *boy/girl, king/queen, prince/princess, father/ mother*. As leftovers, we also have the words *person* and *ruler,* and the vectors 1 = [0.5, 0.9, 0.2, 0.4] and 8 = [0.5, 0.8, 0.8, 1.4]. Let's try to match these leftovers up. If you look closely, you'll notice that the vector 1 = [0.5, 0.9, 0.2, 0.4] is exactly halfway between the vector for *man* and the vector for *woman*. Thus, perhaps this vector goes with a word that is in between *man* and *woman* in meaning, which is most likely *person* (rather than *ruler*). This leaves our last leftover, [0.5, 0.8, 0.8, 1.4], to mean *ruler*. Thus, we can match up 1 = *person* = G, and 8 = *ruler* = I.

n a c l o

(R) GloVe Compartment (2/3) [Solution]

Since *person* was halfway between *woman* and *man,* we can expect *ruler* to be halfway between *king* and *queen.* Looking at the pairs of vectors listed above, this would be true if *king* = 5 and *queen* = 7.

If we look at the pairs we have left, we might notice that there is another type of relationship here: not just male/female pairs, but also adult/children pairs. That is, we might expect analogies of the form *queen* is to *princess* as *mother* is to *daughter*. How could our remaining vectors be assigned to make such analogies work systematically with the vectors? Let's take a look again at the pairs we've figured out and the pairs that are left:

Pairs figured out:
    son = [0.5, 0.7, 0.4, 0.1], *daughter* = [0.5, 0.7, 0.2, 0.3]
    man = [0.5, 0.9, 0.3, 0.3], *woman* = [0.5, 0.9, 0.1, 0.5]
    king = [0.5, 0.8, 0.9, 1.3], *queen* = [0.5, 0.8, 0.8, 1.4]
Pairs still left:
    3 = [0.5, 0.9, 0.3, -0.5] and 9 = [0.5, 0.9, 0.1, -0.3]
    11 = [0.5, 0.7, 0.4, 0.9] and 4 = [0.5, 0.7, 0.2, 1.1]
    6 = [0.5, 0.8, 0.9, 0.5] and 10 = [0.5, 0.8, 0.7, 0.7]

Note in particular that, of the pairs still left, we have one pair with 0.9 in the second spot, one with 0.7 in the second spot, and one with 0.8 in the second spot. This could make these pairs match up nicely with the pairs that are already figured out, where the child and adult versions of a word have the same second element. Thus, we can match up that 3/9 = *boy/girl*; 11/4 = *father/mother*; and 6/10 = *prince/princess.* Overall, this gives us that and adult version of a word minus a child version of a word will be [0, 0, 0, 0.8].

R2. *Answers:*

12. N	13. Q	14. L or U	15. O	16. R
17. M	18. U or L	19. T	20. P	21. S

Explanation:
For R1, we figured out 2 things that will be important here:
    (a) When words form an analogy "*A* is to *B* as *C* is to *D*", their vectors will have a relationship of the form $A - B = C - D$.
    (b) If a word *Y* is in between words *X* and *Z* in meaning, then the vector for *Y* will be halfway between the vectors for *X* and *Z*.

Here, there is one more complicating factor: 2 of our words (*second* and *third*) have 2 different relevant definitions. Let's pretend that these relevant definitions were 2 different words: $second_{time}$ meaning "one sixtieth of a minute" vs. $second_{list}$ meaning "after first"; and $third_{fraction}$ meaning "one third" vs. $third_{list}$ meaning "after second." Then fact (a) gives us:
    $second_{time} - clock = millibar - barometer$
    $second_{list} - two = first - one$
    $third_{fraction} - three = half - two$
    $third_{list} - three = first - one$
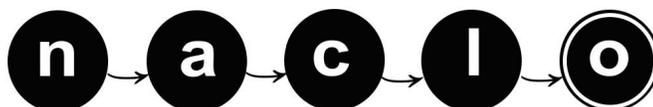
(R) GloVe Compartment (3/3) [Solution]

Solving for each *second* or *third* word gives:

$second_{time}$ = millibar − barometer + clock
$second_{list}$ = first − one + two
$third_{fraction}$ = half − two + three
$third_{list}$ = first − one + three

Based on fact (b) (and also on the hint given in R3), let's now assume that the vector for *second* will be halfway between the vectors for $second_{time}$ and $second_{list}$, and similarly for *third*. That is:

second = 0.5($second_{time}$ + $second_{list}$)
third = 0.5($third_{fraction}$ + $third_{list}$)

Plugging in from above gives:

second = 0.5(millibar − barometer + clock + first − one + two)
third = 0.5(half − two + three + first − one + three)

Now, with some guess and check, we can find that there are only two ways to match the vectors up to the words that will satisfy both of these equations; either of these ways counts as a correct solution.

R3. *Answer:*
Third in a list:
[0.4, -0.2, -0.8, 0]
One third, the fraction:
[0, -0.2, 0.2, -0.4]

Explanation:
Once we've solved R2, we can use the equations given for $third_{fraction}$ and $third_{list}$ to figure out what these two vectors would be.

R4. *Answer:*
(a) is nurse, (b) is doctor.

Explanation:
Words encode gender-related properties as follows: For a pair of words that are identical except for gender (e.g., *woman* and *man*), the vector for the female word minus the vector for the male word will equal [0, 0, -0.2, 0.2]. (Equivalently, the male word minus the female word will be [0, 0, 0.2, -0.2]). That is, gender is encoded via the difference between 2 vectors having that specific value. As the *doctor/nurse* example shows, these differences do not always reflect true gender differences, but instead sometimes reflect statistical or societal biases in terms of which roles tend to be held by members of particular genders.
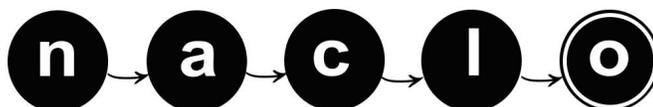
(S) Peace Only (1/4) [Solution]

S1.

    a. I have already forgotten his new cats.

    b. The weaver and the European saw thin mangoes because you hid.

    c. The thieves criticize a crooked dog in the house, don't they?

S2.

    a. **jèmèn tèyⁿín wó èmɛ́∴ wó nìnìwⁿé∴ náŋátì**

    b. **jâm ìjù pěyⁿ kùⁿ dìgɛ́tɔ́ɣɔ̀ là:**

    c. **gùgûn bé kùⁿ∴ fú: páɣárájɛ̀w sábù mǎŋgòlò kùⁿ ñɛ́:tìw**

S3. Observations of Jamsay grammar:

Sentence structure

    Sentences have a basic Subject-Object-Verb structure.

    "[sentence₁] **sábù** [sentence₂]" means "[sentence₁] because [sentence₂]".

    là: (literally meaning "or") at the end of a sentence indicates a confirmation question.

Verbs

    Verbs have an initial root. Affixes are added optionally to reverse meaning and always to convey tense and person of the subject. The order of elements within a verb is:

	(optional) Reversive ("undoes" ROOT's meaning)	*(mandatory)* Tense	*(mandatory)* Person (matches subject)		
ROOT-	-RV-, where: R = r^n if ROOT has a nasal (**n, ŋ, m**), or **r** otherwise, and V = final vowel (including tone) of ROOT	**tì** = simple past		*singular*	*plural*
		jɛ̀ = recent perfect ("have already")	1ˢᵗ:	**-m**	*n/a*
			2ⁿᵈ:	**-w**	*n/a*
		tɔ́ɣɔ̀ = present	3ʳᵈ:	Ø	**-ba**

    Note that the way some verbs take direct and prepositional objects varies between Jamsay and the English translations. For example, the root **sáŋárⁿá-** is translated as "take down the fence around [object]". It is the reversive of **sáŋá-**, meaning "fence in [object]".

n a c l o
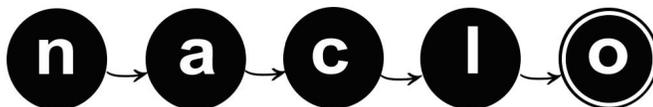
(S) Peace Only (2/4) [Solution]

<u>Noun Phrases</u>

Noun phrases are not marked for case, i.e., the same rules apply to subject and object noun phrases.

Conjunction is marked with dying-quail intonation on each of the two nouns, i.e., "[NP$_1$]∴ [NP$_2$]∵" means "[NP$_1$] and [NP$_2$]". **săy** means "only" and follows its noun phrase (or conjoined noun phrases).

Nouns can be accompanied by an adjective, plural marker, definite marker, and/or quantifier ("all"). The general form of a noun phrase is:

[NP] = [possessive] [noun] [adjective] [plural marker] [definite marker] [quantifier]

> **wó** is the possessive pronoun "his". The possessive pronoun and the definite marker are not used in the same NP.

> Human nouns end in -**n** when singular and -**m** when plural. A separate plural marker is not used with human nouns.

> Adjectives have a root form that is used with non-human nouns. When an adjective accompanies a human noun, it ends in -(**i**)**n** when the noun is singular and -(**u**)**m** when the noun is plural, with the vowel included only if the root ends in a consonant.

> **bé** is the plural marker. It only occurs with non-human nouns and is obligatory to include when the noun is plural.

> **kù^n** is the definite marker, translated as "the". When it is not present, the English article "a" is used if the noun is singular, and no article is used if the noun is plural.

> **fú:** is the quantifier, translated as "all". When it is present, the previous word takes a dying-quail intonation (∴ at end of word).

<u>Tone Rules</u>

(Locative) A noun immediately before the verb, with final tone low (e.g., **úró → úrò**), is a locative, i.e., it indicates a location for the sentence. Locatives are translated with "in the" or "on the", depending on the noun.

When accompanied by an adjective, a noun has all its tones become low (e.g., **mǎŋgòlò → màŋgòlò**).

When a human singular or plural suffix -**in** or -**um** is added to an adjective, a final rising or falling tone in the adjective root "spreads out" over the extra syllable to become low-high or high-low, respectively (e.g., **pěy^n → pèy^{n}ín**, and **gôn → gónìn**).

(S) Peace Only (3/4) [Solution]

<u>Tone Rules: Possible Path to Solution</u>

The rules about tone changes were the most difficult parts of this problem. This section gives one possible approach to figuring out these rules.

The first step in discovering the tone rules is to observe which words have changing tone and which always appear with the same tones. Setting aside dying-quail intonation, nouns and adjectives are the only parts of speech that change tone; verbs, markers/particles, and so on always appear with consistent tone.

Let's start with nouns. We can make a list of nouns that appear (including in the exercises) with different tone configurations (and no other changes, e.g., to the suffix):

> **úró / ùrò / úrò**
> **ójù / ójú**
> **ànà / àná**
> **gùgûn / gùgùn**
> **nìnìw^{n}é / nìnìw^{n}è**
> **ìjù / ìjú**
> **mǎŋgòlò / màŋgòlò**
> **sùrgòn / sùrgôn**
> **ànsà:ràn / ànsá:rán**

Optionally, we can loosen the criteria for our list slightly, and include two examples where instances of a noun differ only by suffix and display a varying tone pattern:

> **jèmèm / jémèn**
> **gùgùy^{n}ím / gùgùy^{n}ìn**

The main regularity we can notice is that most of the nouns here display exactly two patterns. The exception is **uro**, which displays three.

By working through other grammatical features in the problem, we can identify the locative as a distinct function of a noun: the noun appears directly before the verb, without any accompanying words, and is translated to indicate a location (e.g., "in the village"). (See previous page for full explanation.)

Idea: let's set locatives aside for now, and deal with non-locative instances of changing tone patterns. (We might arrive at this idea by wondering why **uro** behaves differently from other nouns, looking at the contexts it appears in, and noticing that one of them is locative. Since we can explain that appearance fairly well, let's focus on what we still can't explain.) If we remove locative instances from the list (and also remove any nouns which, once locatives are removed, only show a single tone pattern) we get:
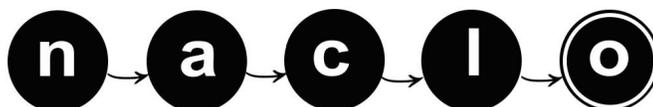
úró / ùrò
gùgûn / gùgùn
nìnìwⁿé / nìnìwⁿè
ìjú / ìjù
mǎŋgòlò / màŋgòlò
sùrgôn / sùrgòn
ànsá:rán / ànsà:ràn
(jémèn / jèmèm)
(gùgùyⁿím / gùgùyⁿìn)

Once locatives are excluded, each noun displays exactly two tone patterns. After some examination, we can see that one of these is always all-low (these have been grouped on the right-hand side of the list above). Next, we follow the all-low pattern, observing the contexts in which it appears, and eventually notice that it is used whenever the noun occurs with an adjective.

We can now arrive at the rule: nouns have a "natural" or default tone pattern that they exhibit when they are not accompanied by an adjective. This tone pattern varies between nouns and cannot be deduced/ recovered from other features of the word. When followed by an adjective, the natural pattern switches to an all-low pattern.

Now we can return to the locative, and compare (what we now believe to be) the natural pattern to the tone pattern in the locative, for all instances of locatives in the problem:

natural	locative
úró	**úrò**
ójú	**ójù**
àná	**ànà**

We can see that the natural → locative change is just the changing of the final tone from high to low.

Note: We don't know what to do when the final tone of the natural pattern is not high – there's not enough evidence in the problem to figure this out. We might even hypothesize (incorrectly, it turns out) that every noun that has a locative form in Jamsay has a final high tone naturally.

Finally, we turn to adjectives. There are no instances where the tone changes without any other change in the word, but we do see cases where an adjective "root" changes its tone when a suffix is added:

dǒŋ → dòŋúm
pěyⁿ → pèyⁿín
gôn → gónìn

From the similarity between **dǒŋ** and **pěyⁿ**, we can generalize to: rising → low-high over the final two syllables. **gôn** shows the opposite pattern: falling → high-low. The two patterns generalize further to the "tone spreading" rule (see above for full explanation).