n a c l o

SOLUTIONS

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

(J) Sounds Fishy (1/1) [Solution]

J1.

(a) wiʔu

(b) benɣ

(c) biːɣ

(d) dunʔi

(e) giɣ

(f) hmax

(g) dinx

J2. Likely interpretations:

(a) thank you very much

(b) wash your hands and brush your teeth

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

K1. The correct correspondences are:

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

K2.   a. *hand* = ma'   b. *bad* = way

K3.   a. weep = *sleep* (other word forms acceptable)   b. ma' puk = *finger*

(L) Stopping for a Spell (1/1) [Solution]

L1.

(a) time: TAYM

(b) traded: TREYDUHD

(c) striding: STRAYDIHNG

(d) framing: FREYMIHNG

L2.

stage

L3.

(a) staging: STEYDJIHNG

(b) gaming: GEYMIHNG

L4.

1. C

2. D

3. B
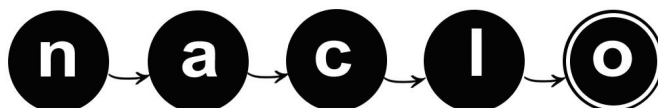
4. A

5. E

6. F

L5.

The sequence of letters that would be pronounced RUHFLEA is: loophole

(M) A Splitting Disagreement (1/1) [Solution]

Q1. Match each algorithm in the table ("Alg A" to "Alg C") to its name ("Baseline", "Pavan's algorithm", or "Arun's algorithm"): **Baseline = A**   **Pavan = C**   **Arun = B**
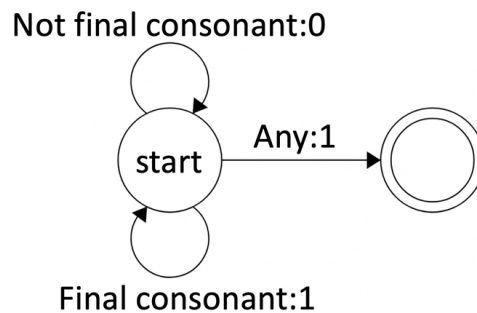
Q2. The output is still the same: [0,0,1,0,0,1,0,0,0,0,0,0,1]
But now the target is: [0,0,1,0,0,1,0,0,0,1,0,0,1]
So we have: TP: 3, FP: 0, FN: 1.
And thus **F1-score of Arun's algorithm = 0.86**

Q3. (a) *Pavan's algorithm:*

Not final consonant:0
Any:1
start
Final consonant:1

(b) *Arun's algorithm: This one is tricky. It needs to handle the scenario where the last V in one V-FC-IC-V template is the first V in another. It also needs to handle cases where the last V in a V-FC-IC-V template is the end of the input; or where V-FC-IC-V is started but not finished, with the sequence ending partway through. Abbreviations used here: Punct = punctuation; FC = final consonant. IC = initial consonant. V = vowel.*
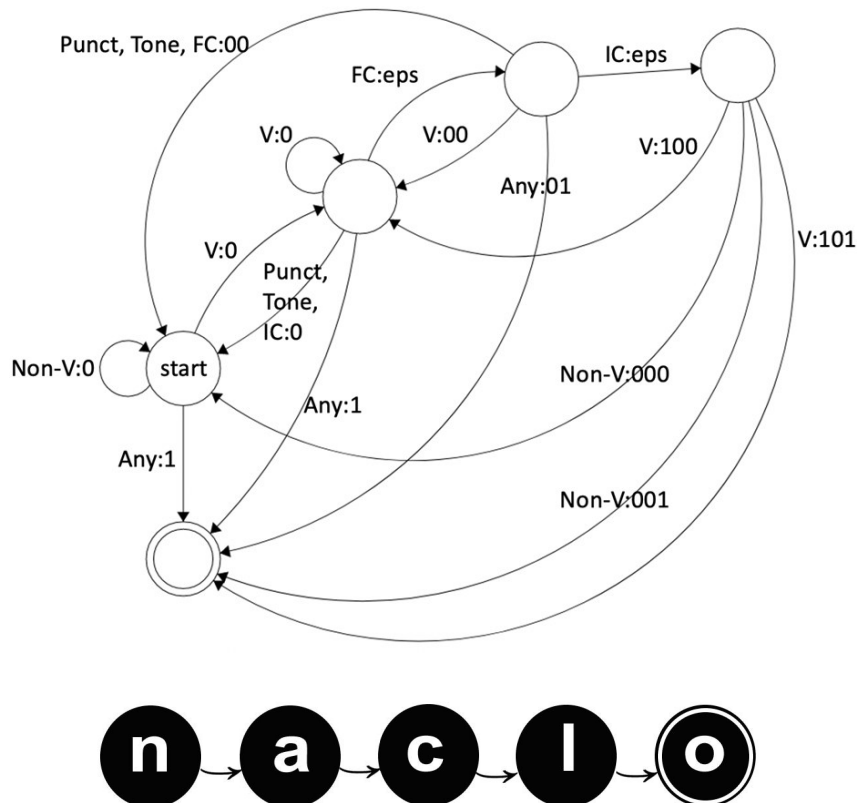
(N) Pseudorandom Numbers (1/2) [Solution]

N1. The nouns that follow one of the common patterns are: **(d) tôoɲ** (*pot*) and **(f) tûuŋ** (*horn*)

N2.

| Root	1st person	3rd person	Translation
(a) lòoj	lwɔ̀ɔɔj	(b) lòooj	*to be different*
(c) càm	(d) càam	cɛ̀ɛm	*to eat*
pèec	pɛ̀ɛɛc	(e) pèeec	*to loot*
wìc	(f) wjɛ̀ɛc	wìic	*to need*
(g) bòk	(h) bwɔ̀ɔk	bòok	*to throw at*

N3. Note that the forms given below are the forms deducible from the data in this problem and the assumption that these nouns conform to one of the common patterns. For some of these nouns, the actual Dinka singular or plural form is different from the deducible one(s).

Singular	Plural	Translation
(a) rjɛ́ɛm	rím	*blood*
(b) wjɛ̀ɛɛl, wîiil	wíil	*bristle*
àɲâaar	(c) àɲɛ́ɛr	*buffalo*
rɛ̀ɛɛc	(d) réec	*fish*
(e) kál, kɔ́l	kàal	*hole in ground*
kók	(f) kwɔ̀ɔk	*hole in tree*
ràaan	(g) rɛ́ɛn, rɔ́ɔn	*person*
(h) lɛ̀ɛɛk, lêeek	léek	*pestle*
ról	(i) rwɔ̀ɔl	*voice*
jìiṯ	(j) jîiṯ	*well*

Explanation

Vowels come in 6 qualities and 3 grades:

Grade 1	i	e	a	ɔ	o	u
Grade 2	i	e	ɛ	ɔ	o	u
Grade 3	jɛ	ɛ	a	a	wɔ	wo
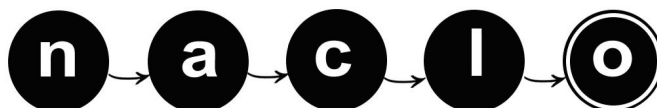
(N) Pseudorandom Numbers (2/2) [Solution]

Singular and plural patterns then correspond by changing the final vowel as follows (singular left, plural right), with subscripts indicating vowel grades:

$$\acute{V}_1 - \grave{V}V_3$$

$$\grave{V}V_1 - \hat{V}V_2$$

$$\grave{V}VV_3 - \acute{V}V_2$$

$$\hat{V}VV_1 - \acute{V}V_2$$

$$\acute{V}V_3 - \acute{V}_1$$

In the verbs there are two (closely related) patterns, shown below (each row its own pattern):

Root	1st person	3rd person
$\grave{V}_1$	$\grave{V}V_3$	$\grave{V}V_2$
$\grave{V}V_1$	$\grave{V}VV_3$	$\grave{V}VV_2$

Height: the Grades (1-3) show patterns with vowel height. All qualities except a / ɛ / a can be described by the rule: Grades 1 and 2 have same height, Grade 3 is lowered. To incorporate the quality a / ɛ / a, the rule can be made more general: from Grade 1 to Grade 2, do not lower (either maintain or increase height); from Grade 2 to Grade 3, lower.

Dinka also features a distinction between "creaky" and "breathy" vowels, but to simplify this problem, it has not been marked.

Source: Andersen, Torben. "Chapter 9. Number in Dinka," in *Number— Constructions and Semantics: Case studies from Africa, Amazonia, India and Oceania* (Anne Storch and Gerrit J. Dimmendaal, eds.), John Benjamins, 2014.

n → a → c → l → o
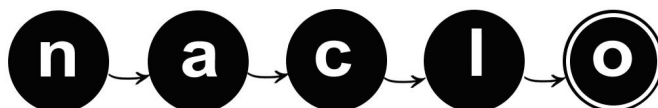
(O) Seeing the Future (1/3) [Solution]

Explanation

The present tense prefixes and suffixes are shown in the table (grey cells are not given anywhere in the problem). Vertical axis is the subject person. Number is not distinctive in the third person.

	Object person				
	1sg	2sg	1pl	2pl	3
1sg		a-		a- -eɫ	a-
2sg	si-		kasi-		hi-
1pl		a- -a		a- -ahaɫ	a- -a
2pl	si- -eɫ		kasi- -eɫ		hi- -eɫ
3	si-	in-	kasi-	in- -eɫ	

From this, we get:

There is a person hierarchy, 1 > 2 > 3, which is seen in the prefixes: whatever argument is higher is expressed. The prefixes are:

1st person singular subject: a-

1st person singular object: si-

1st person plural subject: a-

1st person plural object: kasi-

2nd person subject (singular and plural): hi-

2nd person object (singular and plural): in-

(We don't identify 3rd person prefixes because there is no 3rd person acting on 3rd person example shown in the problem; since it is lowest in the hierarchy we will always use the 1st or 2nd person prefix.)
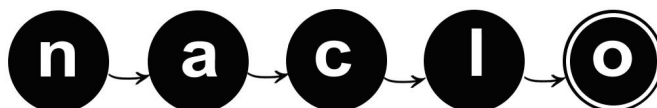
The suffixes are:

1st person plural subject: -a

2nd person plural (subject or object): -eɫ

The 2nd person plural suffix is not affected by the hierarchy. (We don't have enough information yet to say for the 1st person plural subject suffix.)

Morphophonology: -a + -eɫ = -ahaɫ.

(O) Seeing the Future (2/3) [Solution]

In the future, the affixes given in the data pattern as follows (future stem is 'wehnayi'):

	Object person				
	1sg	2sg	1pl	2pl	3
1sg		si-			si-
2sg	si-				
1pl					
2pl					in- -eɫ
3	si-			in- -eɫ	

The main pattern to be deduced from this: the distinct subject and object forms for prefixes merge into one form each (the object form) in the future, i.e., si- takes the place of both si- and a- (1st person singular); in- takes the place of both hi- and in- (2nd person). The distribution of the 2pl suffix -eɫ is unaffected. A few more cells of the table can be filled:

	Object person				
	1sg	2sg	1pl	2pl	3
1sg		si-		si- -eɫ	si-
2sg	si-				in-
1pl					
2pl	si- -eɫ				in- -eɫ
3	si-	in-		in- -eɫ	

The plurals are a bit tricky. The important observation (this is tested in question (e)) is that -eɫ is independent of the subject/object function of the person it pluralizes, while in the first person there is an alternation between kasi- (1st person plural object), where person and number are both expressed in the prefix, and a- -a (1st person plural subject), where two distinct elements appear. Presumably, if 1st person singular and 2nd person are merging into the object form of the prefixes in the future paradigm, 1st person plural will be expressed as kasi- throughout the future paradigm, and there will be no trace of the suffix -a. So the complete table looks like (next page):

(O) Seeing the Future (3/3) [Solution]

	Object person				
	1sg	2sg	1pl	2pl	3
1sg		si-		si- -eɬ	si-
2sg	si-		kasi-		in-
1pl		kasi-		kasi- -eɬ	kasi-
2pl	si- -eɬ		kasi- -eɬ		in- -eɬ
3	si-	in-	kasi-	in- -eɬ	

Morphophonology: 'wehnayi' + eɬ = 'wehnayiweɬ'

Answers

O1.

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

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

c. you (sg.) are going to see us   kasi'wehnayi'

d. you (pl.) are going to see us   kasi'wehnayiweɬ

e. we are going to see you (pl.)   kasi'wehnayiweɬ

(P) Yumology (1/4) [Solution]

Inspired by the Patrice Buche, et al. presentation at International Conference on Biomedical Ontologies (ICBO) 2020 Integrated Food Ontology Workshop (IFOW) titled "A new alignment method based on FoodOn as pivot ontology to integrate nutritional legacy data sources." https://foodon.org/icbo-2020-food-workshop/

P1.
(a) mashed. The facets listed for this food mean "potato, root, chunky semi-liquid, cooked."
(b) seeds, roasted. The facets listed for this food mean "pumpkin, seeds/pods, whole, cooked"

P2.
(c) 2400. The food has the same facets as E10.
(d) 107. The food's nearest neighbor is E01.
(e) 209. Looking just at the facets, this food is equally similar to E03 and E11. However, looking at the facet chart, E11 is much closer.

P3. The algorithm is first to select the best EL food match for each YUM food based on the average of path length from each facet in the YUM food to the corresponding facet in the candidate EL food. One step on the path is either moving between a parent and a child, or between two siblings.
Then, if no ties, get the metric (here, K mg/100 g) from the EL food. If there is a tie, then average the K mg/100 g from the tied matches.

P4. B = food origin. C = part of the thing it originates from. E = texture. F = the extent to which it has been treated with heat. A = whether it is dried.

P5.
(a) B1245 = Apple
(b) B1530 = Coconut
(c) C0240 = Leaves
(d) E0310 = Chunky semi-liquid
(e) F0013 = Cooked
(f) F0001 = Unclear whether cooked or not

P6. Salt. (or anything else that is neither an animal nor a plant)
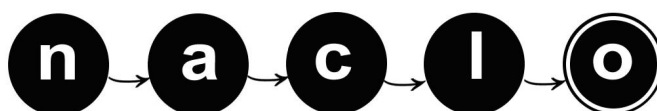
(P) Yumology (2/4) [Solution]

P7. Multiple answers are possible; any answer consistent with all of the facets provided would receive full points.

(a) Y4: Apple slices without skin, air-dried
(b) Y5: Smooth peanut butter. (Also acceptable would be any smooth paste made from cooked nuts, seeds, or pods)
(c) Y6: Chunky peanut butter. (Also acceptable would be any chunky paste made from cooked nuts, seeds, or pods)
(d) Y9: Canned tomato puree. (Also acceptable would be any smooth cooked puree made from any non-root vegetable)

P8. It could help or hurt.
For K mg/100 grams, a powdered product will be much more highly concentrated than the original fresh version, therefore, getting the dried distinction would be important for determining how much potassium there is. However, if the underlying data is sparse in terms of ingredient variety, if using this factor causes a complete switch in plant type, then the K mg per grams may be completely inaccurate, not just miscalibrated by weight. A possible solution to deal with this would be to use a tiered approach, i.e., first find the variety, then from among these, choose the closest to the dried/fresh version.
In general, for other metrics, you may care more about one factor or another, in which case either a factor-weighted or rule-based (tiered) approach would be appropriate. I.e., to find serving size of a liquid, 8 oz may be generalizable, so here variety may matter less, but to find allergen information, variety would basically be the only thing that mattered.

The YUM foods are:

YUM ID	Yum description
Y1	Bacon, raw
Y2	Raw coconut oil
Y3	Pineapple juice, canned
Y4	Apples, air dried w/out skin, slices
Y5	Smooth peanut butter
Y6	Chunky peanut butter
Y7	Beetroot powder, red or golden
Y8	Baked whole apple
Y9	Tomato purée, canned

n → a → c → l → o
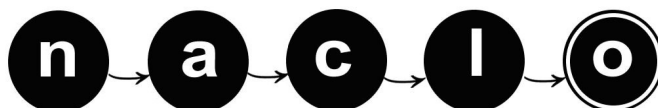
(P) Yumology (3/4) [Solution]

Full graph with annotations

Based (loosely) on LanguaL — The International Framework for Food Description.
https://www.langual.org/default.asp

```
├── A100 Dehydration
│       ├── A2001 Dried/dehydrated (applied to originally watery foods) E02, E08, E10, Y4, Y7
│       ├── A2002 Not applicable/ambiguous (baked goods, seeds, oils) E13, Y5, Y6
│       └── A2003 Not dehydrated/dried E01, E03, E04, E05, E06, E07, E09, E11, E12, Y1, Y2, Y3, Y8, Y9
│
│
├── B100 Animal or plant origin organism
│       ├── B1297 Animal  ## ALL/ONLY ANIMAL PRODUCTS *MADE WITH MEAT* LISTED HERE
│       │       ├── B3391 Mammals
│       │       │       ├── B2002 Goat
│       │       │       ├── B2004 Bovine
│       │       │       └── B1136 Swine E05, E06, Y1
│       │       ├── B2299 Poultry
│       │       └── B2399 Seafood
│       │
│       ├── B1347 Plant
│       │       ├── B1140 Fruit-producing plant   ## ALL/ONLY FRUITS LISTED HERE
│       │       │       ├── B1599 Core or pome fruit
│       │       │       │       ├── B1245 Apple E01, E03, Y4, Y8
│       │       │       │       └── B1344 Pear
│       │       │       │
│       │       │       ├── B1139 Grape family
│       │       │       │       └── B1275 Grapes E08
│       │       │       │
│       │       │       ├── B1539 Stone fruit
│       │       │       │       ├── B1350 Peach
│       │       │       │       └── B1206 Plum
│       │       │       │
│       │       │       └── B1024 Tropical or sub-tropical fruit-producing plant
│       │       │               ├── B4270 Banana
│       │       │               ├── B1249 Papaya
│       │       │               ├── B1530 Coconut E09, Y2
│       │       │               └── B1484 Pineapple E02, Y3
│       │       │
│       │       └──B1141 Vegetable-producing plant  ## ALL/ONLY VEGETABLES LISTED HERE
│       │               ├── B1160 Plants that produce fruits as vegetables
│       │               │       ├── B1534 Squash, gourds, and pumpkins E11, E13
│       │               │       ├── B2530 Tomatoes Y9
│       │               │       └── B3530 Cucumbers
│       │               │
│       │               ├── B1165 Plants that produce commonly used roots and shoots
│       │               │       ├── B1423 Beets E04, E10, Y7
│       │               │       └── B3542 Leeks
│       │               │
│       │               ├── B1162 Plants that produce mainly stems, stalks, and leaves
│       │               ├── B1170 Plants that produce mainly roots and tubers
│       │               │       └── B3544 Potatoes E12
```
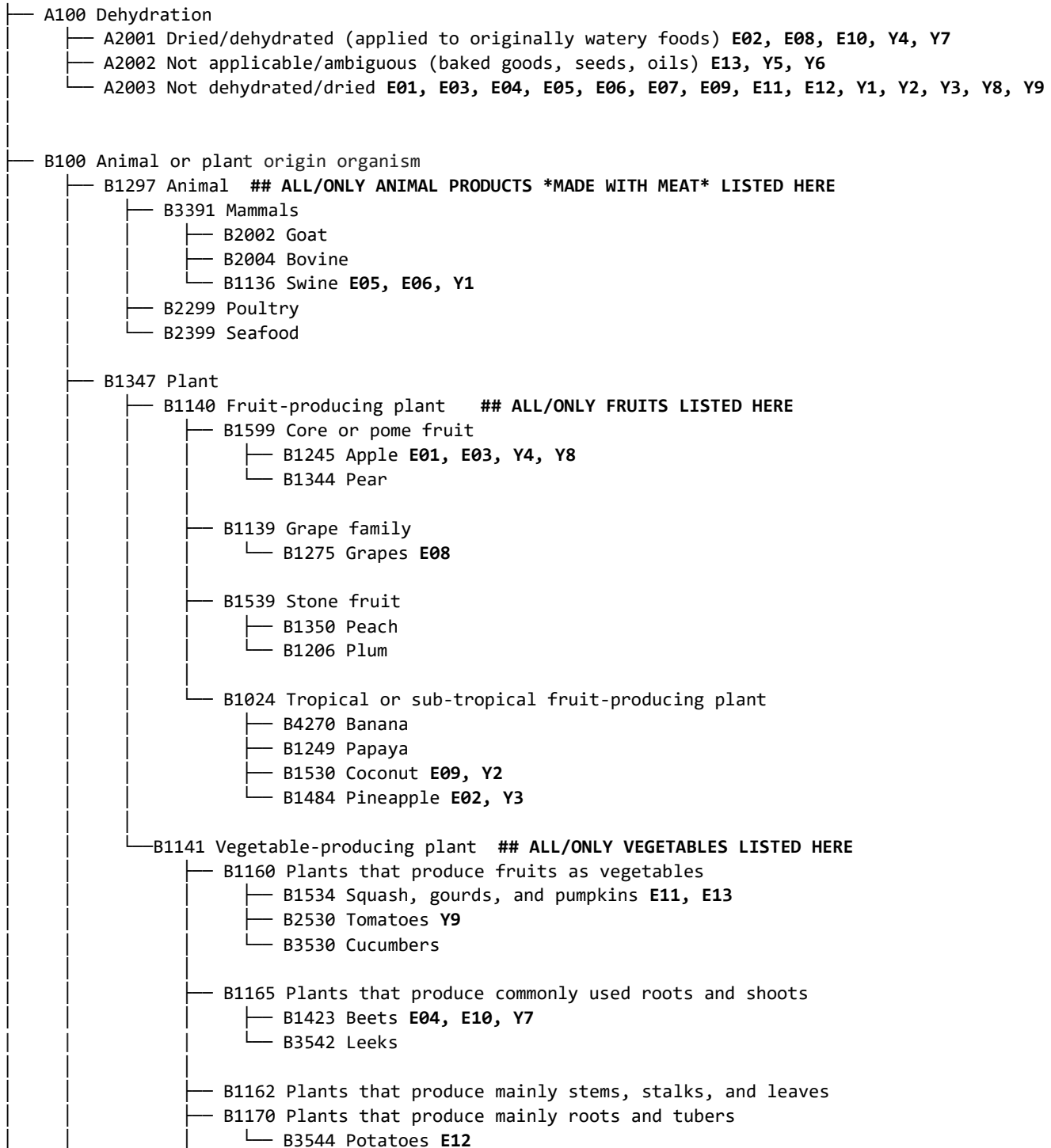
```
│     │       │
│     │       └── B1263 Plants that produce mainly pods and seeds
│     │             ├── B1430 Peanut Y5, Y6
│     │             └── B1452 Soybean E07
│     │
│     └── B1348 Minerals
│
├── C. Part of plant or animal
│     ├── C3001 Part of plant
│     │     ├── C0760 Roots, stems, leaves, flowers
│     │     │     ├── C0140 Roots, tubers, bulbs E10, E12, Y7
│     │     │     ├── C0240 Stems, leaves, stalks E03
│     │     │     └── C0243 Flowers, florets
│     │     └── C0654 Fruit or seed
│     │           ├── C0120 Pod or seeds only E07, E13, Y5, Y6
│     │           ├── C0128 Fruit ## BELOW DISTINGUISHES APPLE (PARTS USED)
│     │           │     ├── C0121 Fruit with peel E01, E08, Y8
│     │           │     └── C0126 Fruit without peel E02, E03, E11, Y4, Y9
│     │           ├── C0339 Juice or liquid extract E09, Y2, Y3
│     │           └── C0329 Peel only
│     │
│     └── C4322 Part of animal
│           ├── C2345 Eggs and dairy
│           └── C2545 Animal body or body part
│                 ├── C5545 Cartilage
│                 ├── C4545 Skeletal meat part or fat E05, E06, Y1
│                 └── C4546 Organ
│
├── E. Physical state, shape, or form
│     ├── E0130 Liquid
│     │     ├── E0115 Liquid, high viscosity Y2
│     │     └── E0114 Liquid, low viscosity E09, Y3
│     │
│     ├── E0103 Semi-solid-liquid ## PEANUT BUTTER (AMBIGUOUS, CAN BE BOTH)
│     │     ├── E0215 Semi-solid-liquid, smooth E03, E11, E13, Y5, Y9
│     │     └── E0310 Semi-solid-liquid, chunky E12, Y6
│     │
│     └── E0152 Solid
│           ├── E0122 Divided or disintegrated
│           │     ├── E0132 Coarsely ground or grated
│           │     ├── E0133 Sliced or diced E02, E05, E06, E07, Y1, Y4
│           │     └── E1152 Powdered E10, Y7
│           └── E0151 Whole E01, E04, E08, Y8
│
└── F. Extent of heat treatment
            ## MUST LIST COOKED OR BE IMPLIED (CANNING, MASHED POTATOES)
      ├── F0013 Fully heat treated  E02, E03, E07, E11, E12, E13, Y3, Y5, Y6, Y8, Y9
      ├── F0001 Whether fully heat treated is not known E05, E08, E10, Y7 ## AMBIGUOUS
      └── F0003 Not heat-treated E01, E04, E06, E09, Y1, Y2, Y4 ## MUST LIST RAW/FRESH, etc.
```
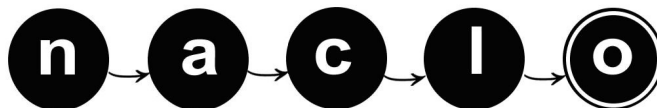
(Q) Relatively Speaking (1/2) [Solution]

Q1.

(a) **Fai moa noa.**

There are only birds.

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

The teacher that I caught has also washed him.

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

The child that the bird will wash will have plenty of canoes.
The child will have plenty of canoes that the bird will wash.

Q2.

(a) *He will also read.*

To totou foki a ia.

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

Kua fai ika noa a Sione ka kai he faiaoga.

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

Ne tā he faiaoga ne kitia e Mele e vaka.

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

Kua muhu kāmuta tūmau.

Explanation

- The default word order is *Tense + Verb + (Adverb) + Subject + Object*

- Tense markers are as follows:

Past	**ne**
Perfect	**kua**
Present	**Ø**
Future	**to**

(Q) Relatively Speaking (2/2) [Solution]

- Nouns must normally be preceded by a case marker:

	Ergative	Absolutive
Name / Pronoun	**e**	**a**
Else	**he**	**e**

- The ergative is used for a subject if it is followed by an object; the absolutive is used elsewhere.

- Nouns can be followed by a relative clause:

> - Relative clauses are introduced by **ne** if the clause is in the past tense, and **ka** in the future, then follow typical syntax.

- Verbs come in two types: existential verbs, **fai** *be/have* and **muhu** *be/have plenty*; and general verbs.
    - Existential verbs are translated as *there + be* without a subject, and as *have* with a subject.
    - The object is compulsory. The head noun is moved to come directly after the verb, and does not appear with case markers. Relative clauses stay in the object position.
    - Note that existential verbs are always treated as intransitive, i.e., subjects of existential verbs are absolutive.

Massam (2001), *Pseudo Noun Incorporation in Niuean*: http://www.jstor.org/stable/404791.

Massam (2000), *VSO and VOS: Aspects of Niuean Word Order*. In Andrew Carnie and Eithne Guilfoyle (eds), *The Syntax of Verb Initial Languages*.

n a c l o

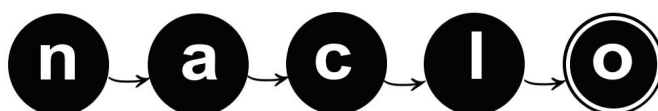# (R) I Stop Being Afraid of This Problem (1/2) [Solution]

Note that there are several correct ways of filling out the tables in R1 and R2; only one is shown here for each question. Also, R1, R2, and R4 were graded together, i.e., for a particular rule, an acceptable answer in any of these places received full credit.

R1. Provide rules for the pronunciation of orthographic *y*:

if *y*	**(a)** is at the beginning of a word	it is silent	—
but if *y*	**(b)** is preceded by an o, in the suffix *-ry*	it is pronounced as	**(c)** o
but if *y*	**(d)** is preceded by an i, in the suffix *-ry*	it is pronounced as	**(e)** i
but if *y*	**(f)** is preceded by an a, in the suffix *-ry*	it is pronounced as	**(g)** a
otherwise	—	y is pronounced as	**(h)** ɨ

R2. Provide rules for the probable pronunciation of orthographic p, t, and k, ignoring palatalization ([j]):

if *p, t*, or *k*	**(a)** are preceded by a stressed syllable with two vowels	they are pronounced as	h.p, h.t, or h.k (respectively)
but if *p, t*, or *k*	**(b)** are preceded by a stress syllable with one vowel	they are pronounced as	**(c)** ʔ.p, ʔ.t, or ʔ.k (respectively)
but if *p, t*, or *k*	**(d)** occur after some other sound	they are pronounced as	**(e)** b, d, or g (respectively)
otherwise	—	they are pronounced as	**(f)** p, t, or k (respectively)

n **a** **c** **l** **o**

(R) I Stop Being Afraid of This Problem (2/2) [Solution]

R3. Fill in the blanks:

Word	Phonetic transcription	Meaning
makopamy	**(a) /maˈgoʔ.pa.mɨ/**	to grow dark
aitopòma	**(b) /ˈaih.tʲo.ˈboʔ.ma/**	homeless
kerikeri	**(c) /keˈriʔ.kʲe.ri/**	a species of bird
parimy	**(d) /paˈɾi.mɨ/**	son in law of
kurijara	**(e) /kuˈɾi.jʲa.ɻa/**	boat
ykurijarary	**(f) /ˈku.ɾi.ˈjʲa.ɻa.ɻa/**	my boat
tykupimy	**(g) /tɨ.ˈguʔ.pi.mʲɨ/**	what needs to be bathed

Explanation

In addition to the phonological/orthographic rules from R1 and R2, there are a number of rules that account for the data in R3:

Stress assignment
- If a word has two syllables, stress the first syllable...
- ...else if the first syllable of a word is heavy (containing a diphthong or a coda consonant, where a diphthong is a pair of vowels acting as one vowel, and a coda consonant is a consonant at the end of a syllable), stress the first syllable and every other following syllable except the final syllable
- ...else stress the second syllable and every other following syllable except for the final syllable.
- Word initial y- is deleted, but still counts as a syllable for purposes of stress assignment.

Diacritics
- The grave accent x̀ indicates a following glottal stop /ʔ/ before a voiced consonant and a glottal fricative /h/ before a voiceless consonant.
- The tilde x̃ indicates a nasalized vowel if there is no following vowel. Otherwise, it indicates a nasal stop (/m/, /n/, /ŋ/) with the same place of articulation as the following consonant.

Assimilation
- There is vowel harmony, as noted in R1. This pattern is actually confined to the possessive suffix *-ry*.
- The vowel /i/ palatalizes the next consonant unless it is <r>.
- <r> is realized as [ɾ] following /i/ or /e/ but is [ɻ] elsewhere.

n a c l o