nacl o

SOLUTIONS

NAACL YAHOO!

LSA Linguistic Society of America

DARPA

NSF

ARL

Carnegie Mellon University Language Technologies Institute

Brandeis University Computational Linguistics

MICHIGAN

The Eleventh Annual North American Computational Linguistics Olympiad 2017

www.nacloweb.org

Open Round January 26, 2017

Serious language puzzles that are surprisingly fun!

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

(A) A Little Tshiluba (1/1)

A1. a. mbwa
b. muluma uvwa mumona muana.
c. nzolu uvwa mumona bambwa
d. mukulu uvwa mupeta mbuji.
e. tubambuji tuvwa tupeta kamuana.

A2. people

A3. a. The fruit saw the lion.
b. The lion saw the small fruits.

(B) Phở Bar (1/1)

B1. D
B2. I
B3. O
B4. F
B5. G
B6. B
B7. Q
B8. A
B9. C
B10. L
B11. E
B12. P
B13. K
B14. H
B15. R
B16. M
B17. S
B18. N
B19. J
B20. T

(C) LOLWUT (1/1)

C1.

										H	A	H	A	A	A
								R						H	
				L				O						A	
	H	O	H	O		L		T	E	H	E	H	E	H	E
	E			O		U		F				A		A	
	H			L	O	L	O	L	O	L		R			
K	E	K	E			Z				A		R			
	H								M	W	A	H	A		
H	E	H								L		A			
		A										R			
	J	A	J	A	J	A	J	A	J	A		R			

C2. a. LOLOL
b. HA

(D) Let’s Roll! (1/1)

D1.
a. E
b. D
c. A
d. B
e. G
f. C
g. F

D2. 2-5

D3. 3-4

D4. 6-6 (This one is tricky: Students have to notice the rule that the larger number always comes first, or they need to pay attention to the note in Question 1, to rule out 2-6 as a possible answer. The logical other possibility, then, is “two sixes”, i.e. 6-6).

(E) On the Right Track (1/1)

Reading Direction: Tamil is read from Left to Right.

Script Classification: Tamil is an Abugida (Consonant with inherent vowel, only to be modified with strokes to effect changes in vowels)

There are two ways to tackle the problem:

1. There are only three stations within the list that has 3 words in it. They are Choa Chu Kang, Yio Chu Kang, and Ang Mo Kio. Since the former two has "Chu Kang" in it, candidates should be able to pinpoint Ang Mo Kio immediately. Careful observation of "c" will lead to the conclusion of Choa Chu Kang, and by elimination, Yio Chu Kang.

2. There are two stations with a repeated word in it – Bukit Gombak and Bukit Batok. Identifying the appearance of the same word will narrow down the search. Comparing the last word, candidates should be able to notice that the syllable "ba" appeared at the start of "Batok" and in the middle of "Gombak". Matching this piece of information will identify the respective stations.

After all possible deductions are drawn, candidates will be left with City Hall station. This station's name is not transliterated but, instead, translated. The answer for City Hall can be deduced by elimination.

E1. A
E2. N
E3. I
E4. T
E5. E
E6. M
E7. H
E8. L
E9. F
E10. U
E11. D
E12. J
E13. V
E14. Q
E15. W
E16. S
E17. R
E18. P
E19. O
E20. C
E21. B
E22. Y
E23. G
E24. K
E25. X

(F) Transition(al) Numbers (1/1)

F1. a. f 1 q 0
b. i 2 r S c 0
c. o 2 r S h 0
d. n 1 q 0

F2. a. No
b. Yes
c. No
d. Yes
e. Yes
f. No
g. No

F3. *Solution (1)*
Change rule [h] as follows S : eight -> 0
And add new rule S: eigh -> 1,2

Solution (2)
Add new rule: S: eigh -> 1, 2, 3
Add a second new rule: 3: t -> 0

There may be other (less elegant) solutions

(G) Magik Yup'ik (1/2)

There is only one possible type of 3-by-3 magic square, although there are 8 distinct configurations that stems from the one type due to rotation and reflection. These permutations are avoided due to the presence of the Yup'ik cross-number puzzle clues.

Since there are 9 digits, the numbers can be arranged via this formula:

1 n-1	2 n+4	3 n-3
2 n-2	n	n+2
3 n+3	n-4	n+1

This formula takes into consideration 9 consecutive numbers, as well as the fact that the sum of every row, columns, and diagonals is the same – 3n. Since the sum of each column is 3n, the sum of three columns is 9n. The sum from 1 to 9 is 45. Hence 9n = 45, leaving us with n=5.

G1.

1 4	2 9	3 2
2 3	5	7
3 8	1	6

The answer to Task 1 is as above. Candidates may in the process arrive at the other 7 possible combinations. These permutations will be eliminated when checked against the Yup'ik hints.

Since the question provided 2D as a 3-digit number that starts with 9, and that all numbers in the clues are 3-digit, and that all Yup'ik numbers begin with Yuinaat, candidates have to consider that Yup'ik may be using a base larger than 10. The most common, in fact, is the Vigesimal system, which is base 20. This is also practical based on the background provided in the question – Yup'ik people based their concept of counting on body parts (20 fingers + toes).

(G) Magik Yup'ik (2/2)

Despite the various permutations, candidates should be able to arrive at 2D, which is 951. Further attempts at solving the spelling will reveal that 951 = (20 x 20 x 2) + (20 x 7) + (10 + 1). This is true to the base 20 system. Arriving at this conclusion will reveal:

- Suffix –k: multiply by 2/double/to do with two (a dual number)
- Suffix –t: multiply by more than 2 (a plural number)
- Suffix –q: the root suffix.

If the word ends in suffix –q, it signifies to the candidate that the subsequent number should be an addition and not a multiplication.

Candidates can then work on 3A, which is the next biggest number with 20 x 20 x 2. Hence, the number, which is the second biggest after one starting with "9", should start with "8". This will lead to 3A being 816. Candidates can then associate *akimiaq atauciq* to 16. From the 2D, *atauciq* = 1, hence *akimiaq* is 15. This is a reasonable and valid guess since Yup'ik people pay attentions to numbers based on the hands and feet.

Next, refer to 3D where candidates can make out (20 x (10+?)) + 16. Given that diagonal also sum up to 15, we can gather 2 and then 7 from it being the remaining cell in 3D. Hence 3D is 276 and that it is (20 x (10+3)) + 16.

Knowing that *qula* = 10, *akimiaq* = 15, and that *malruk* = 2, *pingayun* = 3, we will be able to solve most of the magic square. The other number not mentioned is *cetaman* =4.

G2. Yuinaat yuinaq malruk akimiaq atauciq.

1Diagonal is 456 and can be expressed in Yup'ik spelling as (20 x 20) + (20 x 2) + 16. However, notice from other clues that numbers below 800 are not spelled in this manner but more of (20 x 22) + 16. Hence the number will be spelled according to (20 x (20 + 2)) + 16.

(H) Nothing But Net(works) (1/1)

As can be seen pretty quickly from the *a + 2b* example, each new cell in the network is determined by taking the sum across all inputs of each input times the weight on that input's arrow. From the example translations of the translator program, we can learn all the weights. Let's call the four elements of the input *a,b,c,d*. From the examples, we can learn that the four outputs are (from left to right) *(b + c), (a + b + c), (c + d), (a + c)*. (These expressions can be figured out by treating the examples as a system of equations and solving it rigorously, or you can also figure it out just by eyeballing the examples).

More formally, if we treat the Rigelese as a column vector of dimension 4 called *R* and the English as a column vector of dimension 4 called *E*, then we can write *WR = E* Where *W* is a matrix of the weights that can be written as

$$W = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

However, it is by no means necessary to know anything about matrices to solve this--you can easily just use the expressions *(b + c), (a + b + c), (c + d), (a + c)*.

Using either the matrix multiplication approach or the expressions *(b + c), (a + b + c), (c + d), (a + c)*, we can fill out the translation steps to get that the final message is **GOOD LUCK WITH THAT**: