

SOLUTIONS

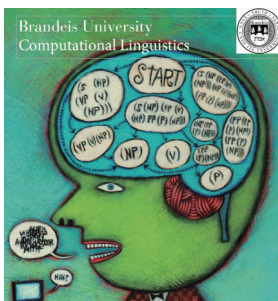


***The Eleventh
Annual
North American
Computational
Linguistics
Olympiad
2017***



Carnegie Mellon University
Language Technologies Institute

www.nacloweb.org



**Invitational
Round
March 9, 2017**

Serious language puzzles that are surprisingly fun!

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

(I) Basque Tasque (1/1)

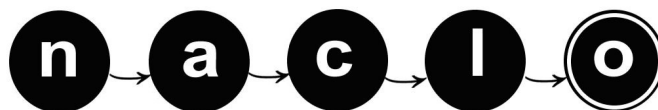
- I1.** a. (E)
b. (J)
c. (H)
d. (M)
e. (L)
f. (K)
g. (C)
h. (A)
i. (G)
j. (F)
k. (D)
l. (I)
m. (B)
- I2.** a. The girl knows the problem.
b. The woman's house/home is in Europe.
c. I have read my father's new book.
d. I will go to the hotel with my father.
- I3.** a. Ardo zuria nahi dut.
b. Ardo zuri berria da.
c. Nere aitaren semea Inglatzeran bizi da.
d. Nere familiak etxe berria erosi du.
- I4.** EXPLAIN YOUR ANSWER (observations that can be drawn from this dataset would be the following, keeping in mind that knowing the linguistic terminology is not expected):
1. Cognates can be used to narrow down the hypothesis space in the matching task.
 2. Basque is an SOV language.
 3. Tense marking is periphrastic (e.g., erosi du = has bought, erori da = has fallen).
 4. Two different sets of auxiliaries are used, one for transitive verbs (erosi du) and one for intransitive verbs¹ (erori da).
 5. The Subjects of intransitive verbs take a zero case marker (absolutive case).
 6. The Subjects of transitive verbs take a -k case marker (ergative case).
 7. The Direct Object of transitive verbs take a zero case marker (absolutive case).
 8. Basque has an inessive case, i.e., in California = Californianan; in the street = kaleanan.
 9. Basque has an allative case, i.e., to the house = etxeran, to the hotel = hotelera.
 10. Basque has a comitative case, i.e., with wine = ardoarekinin, with brother = anaiarekin.
 11. Basque has a possessive case, i.e., father's = aitarenn, girl's = neskaren.
 12. The suffix -a is a singular definite marker which attaches to the last word of the noun phrase (e.g., ardoa = the wine, ardo zuria = the white wine).

¹The language draws a distinction grammatically between unergative and unaccusative intransitive verbs, too, but this problem focuses only on unaccusative intransitive verbs.



(J) The Norwegian Problem / A Norwegian Problem / Norwegian Problems / The Norwegian Problems? (1/1)

- J1.** a. C
b. J
c. E
d. H
e. B
f. A
g. D
h. G
i. F
j. I
- J2.** a. Jenta stanser her.
b. En jente har et hotell.
c. Jeg har hundene.
d. Jenta har hunder.
- J3.** a. The men have apples.
b. The woman does not have the shoes.
c. I do not have the apples.

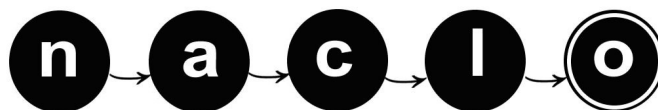


(K) Sentences that go on and on and on and on (1/1)

K1. a. shining in the sky Verb Preposition Article Noun
b. in the state Preposition Article Noun

K2. $1+2+3 = 6$

K3. a. Yes
b. No Preposition Article Noun Preposition Article Noun



(L) The Goat, the Mother, and the Wardrobe (1/1)

- L1.** a. The girls are wearing the shoes.
 b. The son is wearing the trousers.
 c. You (pl) are wearing the shawls.
- L2.** a. Ak^wəmɜ̄wə bʃ^wup’.
 b. Ab ak’asə lʃ^wup’.
 c. Anch^wak^wa ajmsək^wa rəʃop’.

Notes

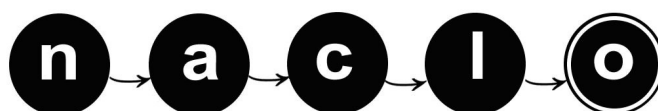
The grammar here is simplified from real Abkhaz grammar. Sentences have the word order SOV, where the subject may be omitted if the agent is indicated by the personal endings. Nouns are sorted according to human/nonhuman, where human nouns take the plural suffix -ch^wa, while nonhuman nouns take -k^wa. Verbs are of the form: subject marker + verb root + p’, where the subject marker varies for person and gender. The subject markers are as follows (not all of the following are deducible from the data):

<i>Person and gender</i>	<i>Subject marker (verb root 1)</i>	<i>Subject marker (verb root 2)</i>
1st person singular	s	sə
1st person plural	ah	ha
2nd person singular	b	bə
2nd person plural	ʃ ^w	ʃ ^w ə
3rd person human masculine singular	i	jə
3rd person human feminine singular	l	lə
3rd person non-human singular	a	a
3rd person plural	r	rə

Challenge 1: unlike the English translation, the Abkhaz contains two different verb roots for ‘wear’, according to where on the body the clothing is worn: -ʃ^wu- is used for clothing worn on the upper half of the body, and ʃo for the lower half. Candidates have to notice this semantic difference, to know which root to use in their Abkhaz sentences in 1.2

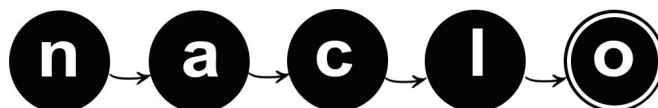
Challenge 2: the singular word for ‘god’, anch^wa, coincidentally looks like a plural (identical, in fact, to the plural ‘mothers’). The third part of 1.2 requires the candidate to realize this and add a real plural ending onto ‘god’, to form the plural anch^wak^wa ‘gods’. Gods are nonhuman (perhaps counterintuitively, given people’s often anthropomorphic idea of gods), so they take -k^wa.

Challenge 3: no word for ‘nanny-goat’ is given, and so candidates should make the step of using the word for ‘billy-goat’ with feminine endings on the verb (which requires noticing that the verb varies according to the gender of the subject).



(M) It will be you who solves this problem! (1/1)

- M1.** a. They will go to the beach.
b. I have hit the ball.
c. The house was built by Bob.
d. The boy (emphatic) spoke to the girl.
e. The bull (emphatic) will chase the boys.
f. I did not go to the courtyard.
- M2.** a. Kua moe rātou ki te hui.
b. Ka haere koe ki Te Maioro Nui Whakaharahara o Haina.
c. Kāore au e moe ki tātahi.
d. Kua horoia te pūru e koe.
e. Nā te tama te mahi i mahi.
f. I haere te wahine ki te whare.



(N) To Be Determined (1/1)

- N1.**
- a. I
 - b. G
 - c. B
 - d. F
 - e. H
 - f. E
 - g. D
 - h. K
 - i. A
 - j. L
 - k. C
 - l. J

Students should realize that there exist two Bulgarian sentences with two words (b, h) and two English sentences with intransitive verbs, having no direct object (G, K). Based on frequency counting, “you” appeared thrice and hence $h = K$. By elimination, $b = G$. From this, we can deduce that the definite article takes on the form of a suffix in Bulgarian.

From the above conclusions, arriving at the remaining answers is relatively straightforward and simple. Based on this, candidates can then observe the following:

(1) Bulgarian uses SVO generally. A Verb-Object inversion (SOV) occurs when:

- a. Object is a personal pronoun (acc.),
- b. Object is a reflexive pronoun.

(2) Bulgarian nouns express definiteness as follows:

- a. Definite article “the” as a suffix,
- b. Indefinite article as a separate word preceding the noun it modifies.

(3) There are potentially three grammatical categories, most likely grammatical gender based on the different variation of the suffix and “a”.

- N2.**
- a. The monkey watched your witch.
 - b. Your camel dressed a girl.
 - c. The wizard cursed himself.
 - d. You hugged the baby.
 - e. You walked.
 - f. You cursed a wizard.

- N3.**
- a. Veshtitsata te obleche.
 - b. Beбето gleda momicheto.
 - c. Maymunata skochi
 - d. Ti pregärna edin sin.
 - e. Tvoiat sin obleche edno bebe.



(O) Common Sense (1/1)

- O1.** (1) she = Vassar, her = Smith
(4) her = Smith
(5) it = Barnard's ricktick
(7) she = Holyoke
(8) it = Holyoke's riplin
(9) she = Vassar
(11) her = Wellesley, it_1 = Barnard's ricktick, it_2 = Wellesley's lapsine
- O2.** From first to last: Barnard, Smith, Vassar, Wellesley, Radcliffe. (Holyoke didn't arrive--no need to mention Holyoke)



(P) Do you see what I see? (1/1)

- P1.** a. you (sg) see us (excl)
b. newa:pama:ehmena:naki
c. newa:pamekwehmaki
- P2.** a. 1st person
b. 2nd person plural
c. inverse 2 over 1 (the prefix marks a 2P object even though the subject is 1P)
d. direct 2 over 1 (the prefix tracks the 2P subject, and the object is 1P)
e. present
f. 3rd person plural



(Q) Pluses and Minuses (1/1)

- Q1.** a. $1/8 + 2/8 = 3/8$
b. $7/3 + 1/6 = 2 \frac{1}{2}$
c. $2/9 + 1/9 = 2/6$
d. $5/5 + 1/7 = 1 \frac{1}{7}$
e. $2/7 + 2/5 = 24/35$
f. $2/6 + 1/9 = 8/18$
g. $1/4 - 3/20 = 1/10$

- Q2.** a. tört
b. on pys
c. čybirgi ýs

Q3. whole (indicates that the number before it is not part of a fraction)

Explanation

The base of the Khakas number system is 10. The numbers from 1 to 10 in Khakas are as follows: pir, iki, ýs, tört, pys, altu, čyti, sygis, toğus, on. The word for 20 is čybirgi.

Numbers above 10 are formed as follows: number_tens number_units. Fractions in Khakas are formed using two different constructs:

- 1) (Denominator + $-nu\eta$ / $-ni\eta^1$) (numerator + $-zi$ / $-i^2$). Here if the base of the numeral ends in $-s^3$, it is voiced. Only then add the appropriate suffix.
- 2) (Numerator) (Denominator + $-luğ$ / $-lig$ or $-nuğ$ / $-nig$ or $-tuğ$ / $-tig^4$)

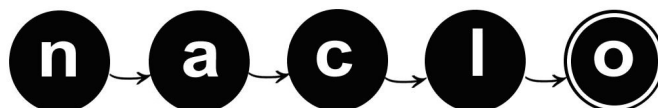
¹This is a possessive ending. In general, except $-nu\eta$ / $-ni\eta$. The ending is $-tu\eta$ / $-ti\eta$. The first pair is used when the base numeral ends in a vowel or voiced consonant, the second when the base ends in a voiceless consonant. Since voicing occurs before adding the ending, the second pair of endings is not observed in this data.

²This is the ending of the possessive form in the third person. Generally endings are $-zu$ / $-zi$ when the base ends in a vowel and $-u$ / $-i$ when it ends in a consonant. Only the former occur in the data.

³In general, when the base of the numeral ends in p / t / s they are voiced, converting respectively to b / d / z . In the data only the $s \rightarrow z$ voicing is observed.

⁴ $-luğ$ / $-lig$ is added to a numeral whose base ends in a vowel or voiced consonant, other than m / n / η , in which case $-nuğ$ / $-nig$ is added; otherwise $-tuğ$ / $-tig$ is added.

Note: The suffixes containing $-u$ are added to the numerals whose last syllable contains $-y$ / $-o$, otherwise use endings containing $-i$.



(R) Text-o-matic (1/1)

R1.

E> → >
Q> → QU> or NQ → NQU
F → V
> → IÈME>
UNIÈME → PREMIER

R2. The simplest solutions for each new case are:

Modify (e.) to:

<UNIÈME → <PREMIER

Add the following rule anywhere:

TS → T

R3. Many solutions are possible. If the solution is based on the rules for R1, then context must be restricted on rules like $F \rightarrow V$ to avoid giving outputs like FINGT in the reverse direction. The cases from R2 must also be handled correctly; $TS \rightarrow T$ is no longer specific enough, since it will produce TSROIS and other errors in reverse. For example:

RE> → R>
ZE → Z
Q> → QU>
F> → V>
-VINGTS → -VINGT
> → IÈME>
<UNIÈME → <PREMIER

