Eighth International Olympiad in Linguistics

Stockholm (Sweden), 19–24 July 2010

Individual Contest Solutions

Problem #1. Rules:

- form 1: ***-mV-*** after the first vowel, whereby ***V*** depends on the vowel in the following syllable (***a*** before ***a***, ***o*** before ***o*** or ***u***, ***e*** before ***i***, ***ö*** before ***ü***);
- form 2:
  - ***-a***, if the stem ends in ***-aR*** or ***-oR***,
  - ***-Ra***, if the stem ends in ***-i***, ***-u*** or ***-ü***,

  where ***R*** is ***l*** or ***n*** if one of these consonants is found in the root, or ***r*** otherwise;
- form 3: form 2 with ***-r-*** after the first vowel, unless ***R*** follows immediately.

Answers:

form 1	form 2	form 3
ɧamerki	*ɧarkira*	
jömölkü	*jölküla*	*jölküla*
qamalq̇al	*qalq̇ala*	
qumorooɋu	*qurooɋura*	*qurooɋura*
somonḳon	*sonḳona*	*sonḳona*

form 1	form 2	form 3
amolq̇ol	***alq̇ola***	*alq̇ola*
emensi	***ensina***	
ɧömörčü	***ɧörčüra***	
čumaraq̇ar		***čuraq̇ara***
ɧamoloɋu		***ɧaloɋula***
ïmankan		***ïnkana***
jemeči		***jerčira***

Problem #2.

- 1–4: ***caa*** 1, ***lue*** 2, ***köni*** 3, ***eke*** 4;
- 5, 10, 15: β-***pi*** $= 5\beta$ $(1 \le \beta \le 3)$;
- 6–9, 11–14, 16–19: α-***ngömen*** $= 5 + \alpha$, α-***ko*** $= 10 + \alpha$, α-***qaihano*** $= 15 + \alpha$ $(1 \le \alpha \le 4)$; ***-e-ko*** $>$ ***-ako***
- 20, 40, 60, 80: γ-***atr*** $= 20\gamma$ $(1 \le \gamma)$; ***caa-atr*** $>$ ***caatr***, ***eke-atr*** $>$ ***ekaatr***
- 21–39, 41–59, ...: Γ ***nge*** $\Delta = \Gamma + \Delta$ $(\Gamma = 20\gamma, 1 \le \Delta \le 19)$.

(a) *caatr nge caako*: **31**, *caatr nge caangömen*: **26**, *caatr nge caaqaihano*: **36**, *ekaatr nge ekengömen*: **89**, *köniatr nge köniko*: **73**, *köniatr nge könipi*: **75**, *köniatr nge köniqaihano*: **78**, *lueatr nge lue*: **42**, *lueatr nge luako*: **52**, *lueatr nge luepi*: **50**.

(b) *köniatr nge eke*: **64** + *caatr nge luepi*: **30** = *ekaatr nge ekako*: **94**
luengömen: **7** + *luako*: **12** = *ekeqaihano*: **19**

(c) 21: ***caatr nge caa***, 48: ***lueatr nge köningömen***, 83: ***ekaatr nge köni***.

Problem #3. : noun, : adjective, : verb (if there is more than one symbol in the word, the mark is placed above the leftmost one).

Pointers (^, ˇ, <, >) are used to refer to specific parts of the symbols.

(a)

	part of speech	composition	meaning
	verb	mouth + nose	to breathe
	noun	water + mouth	saliva
	adjective	circle (sun) + pointer	western
	adjective	activity	active
	noun	body (torso) + 2 pointers	waist
	verb	mouth + (air + outwards)	to blow
	adjective	ill, sick	ill, sick
	noun	mouth + 2 pointers	lips
	verb	eye + (water + downwards)	to weep
	noun	activity	activity
	adjective	heart + upwards	merry

(b)

	part of speech	composition	meaning
	noun	nose	nose
	noun	water	water, liquid
	noun	body (torso) + pointer	neck
	verb	activity	to act, be active
	noun	eye with eyebrow + pointer	eyebrow
	noun	head with neck + pointer	neck

(c)

	part of speech	composition	meaning
	noun	air	air
	noun	body (torso)	body (torso)
	verb	upwards	to rise
	noun	circle (sun) + pointer	east
	adjective	heart + downwards	sad

Problem #4. The four polypeptides in the example consist of 24, 10, 3 and 25 amino acids, and the mRNA sequence contains $195 = ((24 + 10 + 3 + 25) + 3) \times 3$ nucleotides. It appears probable that three nucleotides (a triplet) denote one amino acid or are a separator between polypeptides (in reality a signal to terminate synthesis). However, since there are $4^3 = 64$ possible triplets (all but two of which are present in the example) and only 20 different amino acids, some triplets have the same meaning.

	...U...	...C...	...A...	...G...
U...	UUU → *Phe* UUC → *Phe* UUA → *Leu* UUG → *Leu*	UCU → *Ser* UCC → *Ser* UCA → *Ser* UCG → *Ser*	UAU → *Tyr* UAC → *Tyr* UAA → STOP UAG → STOP	UGU → *Cys* UGC → *Cys* UGA → STOP UGG → *Trp*
C...	CUU → *Leu* CUC → *Leu* CUA → *Leu* CUG → *Leu*	CCU → *Pro* CCC → *Pro* CCA → *Pro* CCG → *Pro*	CAU → *His* CAC → *His* CAA → *Gln* CAG → *Gln*	CGU → *Arg* CGC → *Arg* CGA → *Arg* CGG → *Arg*
A...	AUU → *Ile* AUC → *Ile* AUA → *Ile* AUG → *Met*	ACU → *Thr* ACC → *Thr* ACA → *Thr* ACG → *?*	AAU → *Asn* AAC → *Asn* AAA → *Lys* AAG → *Lys*	AGU → *Ser* AGC → *Ser* AGA → *Arg* AGG → *Arg*
G...	GUU → *Val* GUC → *Val* GUA → *Val* GUG → *Val*	GCU → *Ala* GCC → *Ala* GCA → *Ala* GCG → *Ala*	GAU → *Asp* GAC → *Asp* GAA → *Glu* GAG → *Glu*	GGU → *Gly* GGC → *Gly* GGA → *Gly* GGG → *?*

All mRNA sequences start with AUG → *Met*.

AUG UUA ACG UUC UAA AUG UGG GGG GGA CAC CAG
↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓

(a) *Met-Leu-?Thr-Phe* STOP *Met-Trp-?Gly-Gly-His-Gln*. The sequence contains both nucleotide triplets that were absent from the example, so we cannot be sure in the answer, but it will be confirmed when we have solved the problem to the end.

(b) *Met-Lys-Cys-Ile* ← AUG { AAA, AAG } { UGU, UGC } { AUU, AUC, AUA } ($1 \times 2 \times 2 \times 3 = 12$ possibilities).

(c) A root XY is strong if XYA, XYG, XYC and XYU encode the same amino acid (UC, CC, CG, GC). A root is weak if this is not the case (UU, CA, AG, GA).

Problem #5.

Sursilvan	Engadine	
uo	***uo***	before a cluster of ***l*** or ***r*** and another consonant
u	***u***	before ***l*** or ***r*** without another consonant
u	***o***	before ***m***
u	***uo***	before another consonant

(a)

Sursilvan	Engadine	
uolm	*uolm*	elm
stumi	***stomi***	stomach
cuort	*cuort*	short
mund	***muond***	world
fuorcla	***fuorcla***	mountain pass
plumba	*plomba*	tooth filling
mussar	*muossar*	to show
culant	***culant***	generous

(b) ***lavur*** in both dialects.

(c) In Sursilvan (unlike Engadine) the first rule doesn't apply in plural forms. This may mean that it doesn't work if one consonant is part of the stem and the other belongs to the ending, or that the vowel is chosen before the ending is added, or that the vowel in the plural is made to match the vowel in the singular.

(d) 'elms': ***uolms*** (in both dialects).
'angles': ***anguls*** (Sursilvan), ***anguols*** (Engadine).