n → a → c → l → o

The Association for Computational Linguistics
North American Chapter

The Ninth Annual

North American Computational Linguistics Olympiad

2015

www.naclo.cs.cmu.edu

SOLUTIONS

Open Round

January 29, 2015

(A) The Big Dog and the Young Man (1/1)

A1.

a.	en mand	a man	Danish
b.	en ung man	a young man	Swedish
c.	manden	the man	Danish
d.	en ung mand	a young man	Danish
e.	den unge mand	the young man	Danish
f.	mannen	the man	Swedish
g.	den unge mannen	the young man	Swedish
h.	en man	a man	Swedish

In English we put our definite article before the noun as a separate word. But in Danish and Swedish they can add the definite article on to the end of the noun. We know that 'den unge mannen' must be Swedish, as it is the only one which has a definite article on the end of the noun when an adjective is used, like in the example sentences. Therefore, all those sentences based around 'man' rather than 'mand' must also be Swedish, and the others (based around 'mand') must be Danish.

(B) Delphi has the Answers (1/1)

B1.	Ancient Greek	Modern Greek	Toponym
Βλαχια	(a) /blak^{h}ia/	(b) /(β/v)laçia/	Romania[1]
Φλεγεθων	(c) /p^{h}leget^{h}ɔːn/	(d) /flejeθon/	Phlegethon
Βηρυτος	(e) /bɛːrutos/	(f) /(β/v)iritos/	Beirut
(g) Φρ(ει/οι/ι/η/υ)γ(ει/οι/ι/η/υ)α	(h) p^{h}r(ei/oi/i/ɛː/u)g(ei/oi/i/ɛː/u)a	/frijia/	Phrygia
Βαβυλωνια	(i) /babulɔːnia/	(j) /(β/v)a(β/v)ilonia/	(k) Babylonia

Letters are read as follows

Letter (capital/small)	Ancient Greek	Modern Greek
Α/α	a	a
Β/β	b (because Beirut)	v (g->ɣ, d->ð => b->v)
Γ/γ	g	ɟ before e/i, ɣ elsewhere
Δ/δ	d	ð
Ε/ε	e	e
η	ɛː	i
Θ/θ	t^h	θ
ι	i	i
Κ/κ	k	k
λ	l	l
μ	m	m
ν	n	n
ο	o	o
Σ/σ/ς (final position)	s	s
υ	u	i
Φ/φ	p^h	f
Χ/χ	k^h	ç before e/i, x otherwise
ω	ɔː	o
αι	ai	e
ει	ei	i
οι	oi	i
ου	ou	u

[1]The Greek name for Romania is Wallachia.

(C) Aymara Rules (1/1)

C1. *[ke], *[kq], *qi (order doesn't matter)

C2. Counting Mod 2

(D) Elder Futhark Runes (1/1)

D1.

a. Baldur (1) ᛒᚨᛚᛞᚱ

b. Dallinger (8) ᛞᛖᛚᛁᛜᛖᚱ

c. Day (4) ᛞᚨᚷᚱ

d. Earth (7) ᛃᛟᚱᚦ

e. Freya (6) ᚹᚱᛖᛁᛃᚨ

f. Freyr (9) ᚹᚱᛖᛁᚱ

g. Ithun (3) ᛁᚦᚢᚾᚾ

h. Night (5) ᚾᛟᛏᛏ

i. Sun (11) ᛊᛟᛚ

D2.

(2) ᚦᛟᚱ Thor

(10) ᛟᛞᛁᚾ Odin

D3.

a. Tyr ᛏᛁᚱ

b. Ran ᚱᚨᚾ

c. Sif ᛊᛁᚹ

(E) Use the Force (1/2)

E1.

a. you must go
b. I think this one is strong with the force
c. please don't forget to pick up milk before coming home tonight

E2.

a. [Use < [force Luke] the >]
b. < Luke < [the force] Use > > MORE POSSIBLE
c. < Luke < force < the Use > > > MORE POSSIBLE
d. NOT POSSIBLE
e. < [the < Luke force >] Use >

E3. 24 (4!)

E4. 22
This can be done by enumeration or by noting that all permutations of 3 items are Yoda-isms and that placing two adjacent sequential items (i.e. the permutation contains e.g. 23 or 32) yields a sequence of 3 items after the first rule application; there are only 2 permutations of 4 items that don't place two adjacent sequential items (2413 and 3142).

E5.

a. Permutations: 12 Yoda-isms: 12

Since 4! = 24 (from above) and each sequence has a twin (where the first "do" can stand in for the second) we can divide by 2. Any sequence is identical to the one where 1 and 3 are swapped; it can be seen that the two non-Yoda-ism cases are equivalent to Yoda-ism cases so all are Yoda-isms.

b. Permutations: 120 Yoda-isms: 90

5! = 120 Note that any sequence of 4 items in either of the 2 bad-4 patterns (call them a and b) will result in a fail, regardless of where the 5th item is placed. So 52413, 25413....24135. But let the sequence "1235" map to "1234" and then its permutation, "2513" is non-Yoda-ism no matter where the 4 is placed. 5 sets of 4 items x 5 configurations of those 4 (i.e. where to place the 5th) x 2 patterns = 50. But some of those 50 are duplicates. Note:

2413 3142
2513 3152
2514 4152
3514 4153
3524 4253

(the sequence of 4 items in configurations a and b). Each vertical pair contains 2 duplicates. E.g. 2413 and 2513 both yield 25413 and 24513. That's 4 pairs x 2 duplicates x 2 configurations = 16 duplicates. And there are 4 "twins" between configurations a and b; each yields a single duplicate:

2413 - 4153 (24153)
2513 - 4253 (42513)
3524 - 3152 (31524)
3514 - 3142 (35142)

Note 2514 and 4152 are not twins, so 50 - 16 - 4 = 30, which are the non-Yoda-ism permutations.
120 - 30 = 90

(E) Use the Force (2/2)

c. Permutations: 60 Yoda-isms: 52

5!/2 = 60 per above. From the above table note that any sequence of 4 containing both 3 and 4 will have identical patterns that are Yoda-isms (e.g. 2413 = 2314, which is a Yoda-ism, no matter where the 5 is placed). That leaves 2513 and 3152 (2514/4152 are redundant to them). That's 5x2 = 10 cases, but of those 10 the placement of the 4 on either side of the 3 are identical, so eliminate 2. 60 - 10 - 2 = 52.

(F) My Friend Nomura from Osaka (1/1)

	Japanese		English
	Japanese		English
1.	Ota	M.	Big rice-field
2.	Nakayama	I.	Middle mountain
3.	Kigawa	B.	Tree river
4.	Kazan	L.	Fire mount (= volcano)
5.	Murakami	E.	Above the village
6.	Kagawa	N.	Fire river
7.	Ono	H.	Little field
8.	Nomura	D.	Field village
9.	Tanaka	G.	Middle of the rice-field
10.	Sakuragi	K.	Cherry tree
11.	Nihon	O.	Origin of the sun (= Japan)
12.	Osaka	A.	Big slope
13.	Yamazaka	C.	Slope of the mountain
14.	Kawakami	F.	Above the river
15.	Honda	J.	Original rice-field

Hon = origin(al)
Ka = fire
Kami = upper
Kawa/gawa = river
Ki/gi = tree
Mura = village
Naka = middle
Ni = sun
No = field
O = big/little [for information: 'big' has a long vowel, 'little' has a short vowel]
Saka/zaka = slope
Sakura = cherry
San/zan = mount
Ta/da = rice-field
Yama = mountain

(G) Zoink! (1/2)

G1.

a. CORRECT	B, C, E, F
b. WRONG	D, G
c. MAYBE	A, H, I

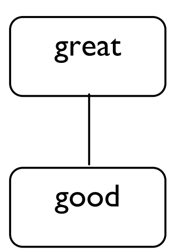
Different people will find different routes through this problem, but the following notes present one route which is hopefully clear and relatively similar for those whose intuitions don't take them straight to the answer.

1. Look at the syntax of the phrases 1-17 and reduce the diversity. Each phrase links two adjectives which we can call x and y. Here's a list of the different patterns, distributed into two groups:

$x<y$	$x>y$
x but not y	not x, but just y
x, though not y	not x, just y
not just x but y	
not only x but y	

In all the $x<y$ patterns, adjective x has a weaker meaning than adjective y; for example, in good but not great, 'good' is weaker than 'great', just as it is in good, though not great, not just good but great and not only good but great; so all these different phrases tell you the same thing: 'good' < 'great'. In the $x>y$ patterns, the relation is reversed: not great, but good and not great, just good, so 'great' > 'good'.

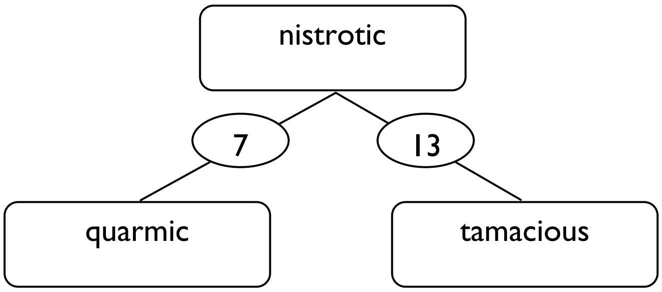
2. Since the relative strength of x and y is all you're interested in, you can replace all the phrases 1-17 by a representation of this relation. To do this you need to remove the difference between $x<y$ and $x>y$, which is just a matter of syntax. You could convert one of these notations into the other, e.g. converting all the $x>y$ patterns into $y<x$ patterns; but at this point it will be more helpful to move towards a network notation which shows the relation more clearly. The obvious iconic ('natural') notation for this relation uses the vertical dimension to link the stronger adjective down to the weaker one below, like this:

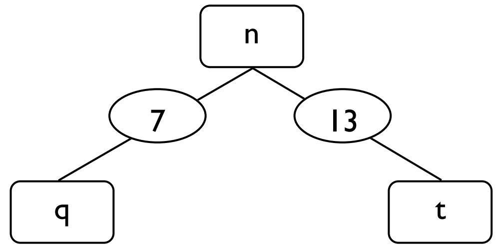


3. Now you work through 1-17 to convert the relations into this notation. But notice that the same adjective can appear in two or more phrases, so you're actually dealing with a complex network of relations rather than a list of isolated relations. Here too your new notation will come in handy because it allows you to write each adjective just once, but to give it more than one link to other adjectives. For instance, nistrotic is linked both to quarmic (phrase 7) and to tamacious (13), as shown below. The diagram on the right takes advantage of a convenient fact: that every adjective begins with a different letter, so you can save writing them

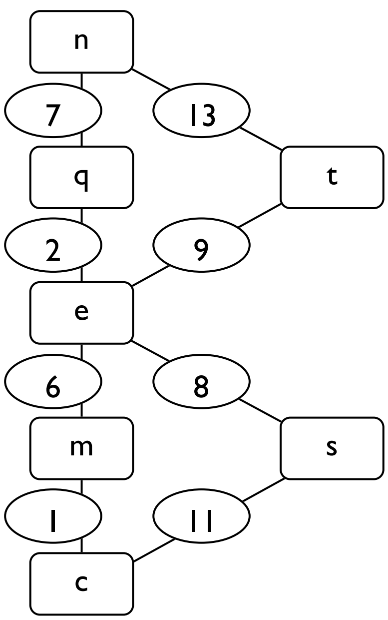
(G) Zoink! (2/2)

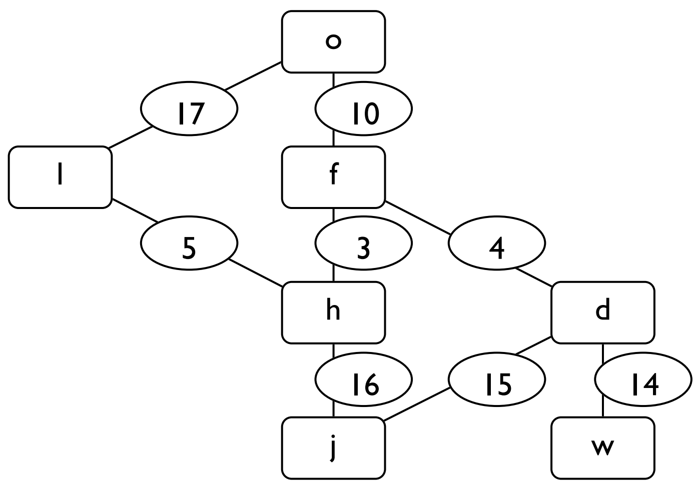
out by just using their initial letter. The number in the ellipse reminds you of the phrase which defines the relation concerned.





4. Continuing this process with some juggling, you eventually find that the adjectives actually form two separate and unconnected networks. This means that, between them, these phrases actually define just two scales or dimensions, each with several different grades of increasing strength, and each with some synonyms for the same grade. We don't, of course, know what these scales are – reliability, speed, attractiveness, price or whatever – but we do know about the formal structuring of the different points on the scales. Here are the two networks:





5. And finally to the solution: You represent each of A-I using the same notation, and compare it with the networks.

- Correct: if the adjectives x and y are in the same network, and the higher one is also higher in the network.
- Wrong: if x and y are in the same network but the higher one is lower in the network.
- Possible: if x and y are in different networks.

(H) Phàasàa and Pháasǎa (1/3)

H1.

Shan	Lao	English
ႁၢႆႉ	(a) ຮ້າຍ	bad
မွပ်ႈ	(b) ໝອບ	to crouch
လၢၵ်ႊ	(c) ລາກ	to pull
(d) ပႅတ်ႈ	ແປດ	eight
(e) မႃႈ	ມ້າ	horse
(f) ပိူၵ်ႈ	ເປືອກ	rind

H2.

Consonants: ၽ ງ

Vowels: ႅ ໄ ຼ

Tones: ႊ

Detailed explanation of correspondences between Shan and Lao:

Consonants

Shan	Lao 2	Lao 3,5	Lao end	Sound
မ	ໝ	ມ	ມ	m
ၼ		ນ	ນ	n
င		ງ	ງ	ng
ၽ	ຜ	ພ		ph
ပ			ປ	p

(H) Phàasàa and Pháasǎa (2/3)

Shan	Lao 2	Lao 3,5	Lao end	Sound
တ	ດ	ຕ	ດ	t
ၵ	ກ	ຄ	ກ	k
သ		ຊ		s
ၸ		ຊ		ch/s
ႁ		ຮ		h
ဝ			ວ	w
လ	ດ	ລ		l/l, t

Vowels

Shan	Lao	Sound
◌ၢ	◌າ	aa (before consonant)
◌ႃ	◌າ	aa (before vowel)
◌ိ	◌ິ	i
◌ႅ	ເ	e/iiə
◌ႄ	ແ	ɛ
◌ူ	◌ູ	u
◌ု	◌ຸ	o/u
◌ွ	ອ	ɔ
◌ိူ ◌ိ ◌ိူ	◌ຶ	u

(H) Phàasàa and Pháasǎa (3/3)

Shan	Lao	Sound
◌ူ◌ိ(ဝ်)	ເີອ	ə/ɯɯə
◌ႆ	ໄ	ai
◌ၢႆ	າຍ	aai
◌်	(nothing)	no vowel

Tones

Shan	Lao	Shan Tone Number and Shan/Lao Sound
◌ႇ	no mark, type 2 consonants	2, low/low
◌ႈ	no mark, type 3 consonants	3, low falling/high falling
◌့	◌້	5, mid falling/high falling

How to make words in Shan and Lao:

The word in Shan and Lao consists primarily of a consonant with a vowel mark attached and tone indications. This CVT complex may be followed by a consonant which must take a killer-vowel mark in Shan and does not take anything in Lao.

In Lao tone 2 is indicated by the initial consonant taking a certain form. The consonant takes the same form for both tone 3 and tone 5. However, tone 5 is differentiated from tone 3 by a tone mark appearing above the main consonant.

In Lao final consonants take a specific form as well which is identical to either the tone 2 form or the tone 3/5 form.

Explanation for Supplemental Task:

Assuming that syllables follow a CV(C) format, it makes sense to assume that the main symbols are consonants and the secondary ones are vowels, because of the optionality of syllable-final consonants. The tone marks exhibit the most complicated correspondences and there are only three different tones, so they may be easily recognized as such.