n a c l o

The Association for Computational Linguistics
North American Chapter

Carnegie Mellon

M UNIVERSITY OF MICHIGAN

YAHOO!

The Sixth Annual

North American Computational Linguistics Olympiad

2012

www.naclo.cs.cmu.edu

SOLUTIONS
Open Round
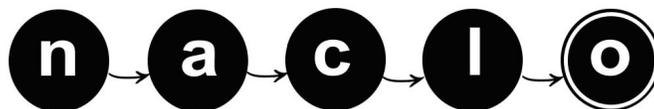## February 2, 2012

A. Cat and Mouse Story (1/1)

A-1

Okay, so my cat ran into the mouse and she ran into the street. She ran down the street and she ran up the street and eventually she ran down the mouse.

The box words translate as follows:

Pomble	Run
Trowby	Mouse
Ippip	Down
Foba	Street
Gorch	Up
Gwee	Into

naclo

B. Learn Yolmo with pleasure (1/1)

B-1 C and D

B-2 ʈánga nyimu

B-3 Ngà ngàki ádzi nyímu tó sàke.

C. Interstellar First Contact (1/2)

C-1 The questions in this assignment are based on examples in Knight (1997). In fact, both Centauri and Arcturan have underlying real world languages, as it turns out Centauri is English and Arcturan is Spanish. The languages are obfuscated to Centauri and Arcturan in order to illustrate how a Statistical Machine Translation (SMT) system must start from scratch, since it has no prior knowledge of how the languages work.

CENTAURI	**ARCTURAN**
Ok-voon ororok sprok.	At-voon bichat dat.
Garcia and associates.	Garcia y asociados.
Ok-drubel ok-voon anok plok sprok.	At-drubel at-voon pippat rrat dat.
Carlos Garcia has three associates.	Carlos Garcia tiene tres asociados.
Erok sprok izok hihok ghirok.	Totat dat arrat vat hilat.
His associates are not strong.	Sus asociados no son fuertes.
Ok-voon anok drok brok jok.	At-voon krat pippat sat lat.
Garcia has a company also.	Garcia tambien tiene una empresa.
Wiwok farok izok stok.	Totat jjat quat cat.
Its clients are angry.	Sus clientes están enfadodos.
Lalok sprok izok jok stok.	Wat dat krat quat cat.
The associates are also angry.	Los asociados tambien están enfadados.
Lalok farok ororok lalok sprok izok enemok.	Wat jjat bichat wat dat vat eneat.
The clients and the associates are enemies.	Los clientes y los asociados son enemigos.
Lalok brok anok plok nok.	Iat lat pippat rrat nnat.
The company has three groups.	La empresa tiene tres grupos.
Wiwok nok izok kantok ok-yurp.	Totat nnat quat oloat at-yurp.
Its groups are in Europe.	Sus grupos están en Europa.
Lalok mok nok yorok ghirok clok.	Wat nnat gat mat bat hilat.
The modern groups sell strong pharmaceuticals.	Los grupos modernos venden medicinas fuertes.
Lalok nok crrrok hihok yorok zanzanok.	Wat nnat arrat mat zanzanat.
The groups do not sell zanzanine.	Los grupos no venden zanzania.
Lalok rarok nok izok hihok mok.	Wat nnat forat arrat vat gat.
The small groups are not modern.	Los grupos pequeños no son modernos.

n a c l o

C. Interstellar First Contact (2/2)

The novel sentence which was offered for translation in English is: "clients do not sell pharmaceuticals in Europe."

Answers
C-1 jjat

C-2 hihok = arrat, yorok = mat

C-3 We need to use the process of elimination, when mapping all the words between the two sentences two words are unaligned, we assume these are translations of each other. Thus, clok = bat.

C-4 Here are the new matches:

crrok	(empty)
kantok	oloat
ok-yurp	at-yurp

"crrok" does not seem to have a Arcturan equivalent, like in English the word "do" is not translated in "do not sell" which simply becomes "not sells" in Spanish. (Or to put it another way, the Centauri word *crrrok* **has** a translation, but it's the "empty" word.)

C-5 *jjat arrat mat bat oloat at-yurp*
Since you cannot deduce with certainty the exact order of the Arcturan sentence, various orders of these words will be accepted.

C-6 Immediately, you are faced with a dilemma: should you translate *totat* as *erok* or wiwok? Because *wiwok* occurs more frequently and because you've never seen *erok* followed by any of the other words you're considering, *wiwok* seems more likely. (However, admittedly, this is only a best guess, and *erok* will also be accepted.) Next, you consider various word orders. There appears to be no grammatical path through these words. Suddenly, you remember that curious Centauri word *crrrok*, which had no translation. *Crrok*, however, turns out to be a natural bridge between *nok* and *hihok*, giving you the translation:

      wiwok rarok nok crrrok hihok yorok clok.
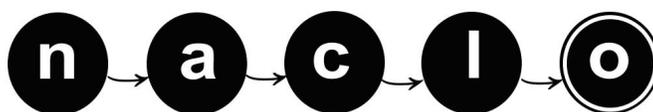
D. All in the Family (1/1)

D-1

No knowledge of language names or families was required. The names below are only shown for completeness.

1. CLQ   Slavic
2. BEFIM Romance
3. J   Basque
4. HO   Baltic
5. DGN   Celtic
6. KP   Finno-Ugric
7. A   English

D-2

Groups 3 and 6 don't belong to this family.

E. Traevölörs Freisbuk (1/1)

Numbers: uön zrii ssikss zöörti höndröd
Fruits: bönaenö uatörmêlön fighs ghreipss painaepöl
Colors: braun ghrei pöörpöl töörkois yêlôô
Desserts: aisskriim dʒêlôô födʒ keik kukis
Games: chêkörs ghalf pôôkör puul tênöss
Birds: chikön dök ghuuss kueil töörki

(or)

Numbers: 3,8,13,25,29
Games: 4,9,21,23,28
Birds: 5,11,18,20,26
Colors: 1,6,12,15,19
Desserts: 7,14,16,22,24
Fruits: 2,10,17,27,30

The actual rendering of the category names is:

nömbörs
fruutss
kölörs
disöörtss
gheims
böörds
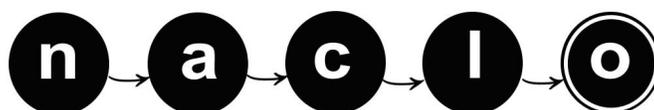
F. Crocodile Bardi (1/1)

F-1 The English words are matched with the Bardi words below:

Cat	Minyaw
Dog	Iila
Horse	Yaawarda
Kangaroo	Boorroo
Man	Aamba
Woman	Oorang
Next to	Bornkony
Behind	Baybirrony
In front of	Alaboor
To the right of	Joorroonggony
To the left of	Aarlgoodony

F-2 The English words are matched with the Bardi words below:

Aalgamadan	East
Alang	South
Ardi	North, Northeast
Baana	East, Southeast
Goolarr	West, Northwest

G. Haitian Creole (1/1)

G-1
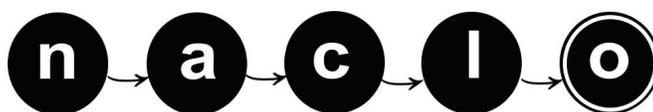
Nan
An
La (the music)
An
La
A (the radio)
Lan
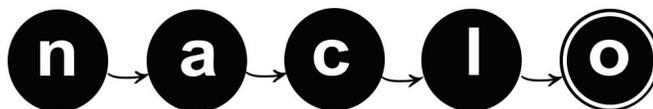An
Kouta a
Kó yo
Sant lan
Liv la

H. Waorani Numbers (1/1)

H-1

1. aroke
2. mẽña
3. mẽña go aroke
4. mẽña go mẽña
5. ãẽmãẽmpoke
6. ãẽmãẽmpoke go aroke
7. ãẽmãẽmpoke go mẽña
8. mẽña mẽña mẽña mẽña
9. ãẽmãẽmpoke mẽña go mẽña
10. tipãẽmpoke

Solution:

Step 1. From b) we can infer that aroke and mena are 1, 2, or 3. Aemaempoke must be 5, 10, or 13 (but 13 is too large, which leaves only 5 and 10 as possibilities).

Step 2. From d) we can see that mena cannot be 1. If aemaempoke is 10, then mena cannot be 2 or 3, therefore aemaempoke is not 10, so it must be 5.

Step 3. Aroke and mena are both 1 or 2, but mena is not 1, so mena is 2 and aroke is 1.

Step 4. Therefore tipaempoke is 10.

Step 5. We still have to account for 3, 4, 6, 7, 8, 9.

Step 6. From c), the squared number cannot be 9, 8, 7, 4, or 3, so it must be 6. Therefore the other two must be 4 and 9 (since 4x9=36). The left hand number is shorter so we call it 4 and the right one is then 9.

Step 7. From a), we are still missing 3, 7, 8. However $z+4=2x6=12$, so $z=8$.

Step 8. The missing numbers now are 3 and 7. We build 3 (2+1) by analogy to 6 (5+1) and we build 7 (5+2) by analogy to 4 (2+2).