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-I

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:

<table>
<thead>
<tr>
<th>Pomble</th>
<th>Run</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trowby</td>
<td>Mouse</td>
</tr>
<tr>
<td>Ippip</td>
<td>Down</td>
</tr>
<tr>
<td>Foba</td>
<td>Street</td>
</tr>
<tr>
<td>Gorch</td>
<td>Up</td>
</tr>
<tr>
<td>Gwee</td>
<td>Into</td>
</tr>
</tbody>
</table>
B. Learn Yolmo with pleasure (1/1)

B-1 C and D

B-2 ṭánga nyimu

B-3 Ngâ ngâki ádzi nyímu tó säke.
### C. Interstellar First Contact (1/2)

**C-I** 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**

- Ok-voon ororok sprok. Garcia and associates.
- Ok-drubel ok-voon anok plok sprok. Carlos Garcia has three associates.
- Erok sprok izok hihok ghirok. His associates are not strong.
- Ok-voon anok drok brok jok. Garcia has a company also.
- Wiwok farok izok stok. Its clients are angry.
- Lalok sprok izok jok stok. The associates are also angry.
- Lalok farok ororok Lalok sprok izok enemok. The clients and the associates are enemies.
- Lalok brok anok plok nok. The company has three groups.
- Wiwok nok izok kantok ok-yurp. Its groups are in Europe.
- Lalok mok nok yorok ghirok clok. The modern groups sell strong pharmaceuticals.
- Lalok nok crrrok hihok yorok zanzanok. The groups do not sell zanzanine.
- Lalok rarok nok izok hihok mok. The small groups are not modern.

**ARCTURAN**

- At-voon bichat dat. Garcia y asociados.
- At-drubel at-voon pippat rrat dat. Carlos Garcia tiene tres asociados.
- Totat dat arrat vat hilat. Sus asociados no son fuertes.
- At-voon krat pippat sat lat. Garcia tambien tiene una empresa.
- Totat jat quat cat. Sus clientes están enfadados.
- Wat dat kratquat cat. Los asociados tambien están enfadados.
- Wat jat bichat wat dat vat eneat. Los clientes y los asociados son enemigos.
- Lat lat pippat rrat nnat. La empresa tiene tres grupos.
- Totat nnat quat oloat at-yurp. Sus grupos están en Europa.
- Wat nnat gat mat bat hilat. Los grupos modernos venden medicinas fuertes.
- Wat nnat arrat mat zanzanat. Los grupos no venden zanzania.
- Wat nnat forat arrat vat gat. Los grupos pequeños no son modernos.
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 crrok 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 crrok, which had no translation. Crrok, however, turns out to be a natural bridge between nok and hihok, giving you the translation:

wiwok rarok nok crrok 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: bönaenô uatörmêlôn fighs ghreipss painepöl
Colors: braun ghrei pöörpöl töörkois yêlôô
Desserts: aisskriim dzêlôô fôdʒ keik kukis
Games: chêkörs ghalf póökör puul tênöss
Birds: chikôn 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-I

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. mēña
3. mēña go aroke
4. mēña go mēña
5. āēmāēmpoke
6. āēmāēmpoke go aroke
7. āēmāēmpoke go mēña
8. mēña mēña mēña mēña
9. āēmāēmpoke mēña go mēña
10. tipāēmpoke

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).