naclo

The Eighth Annual

North American Computational Linguistics Olympiad

2014

www.naclo.cs.cmu.edu

SOLUTIONS
Invitational Round

March 13, 2014

(I) To play or not to play (1/2)

I1.

1.	Atacheza	A	He/she will play
2.	Mlifahamu	I	Y'all understood
3.	Mnapika	J	Y'all cook
4.	Nilicheza	B	I played
5.	Ninapika	C	I cook
6.	Nitapika	D	I will cook
7.	Tulifahamu	H	We understood
8.	Unacheza	K	You play
9.	Utapika	L	You will cook
10.	Wanafahamu	E	They understand
11.	Watapika	F	They will cook
12.	Walicheza	G	They played

I2.

1.	Hakucheza	A	He/she did not play
2.	Hamkupika	H	Y'all did not cook
3.	Hatacheza	C	He/she will not play
4.	Hatapika	B	He/she will not cook
5.	Hatukufahamu	F	We did not understand
6.	Hatupiki	G	We do not cook
7.	Hawafahamu	E	They do not understand
8.	Huchezi	I	You do not play
9.	Sikucheza	D	I did not play

I3.

1.	Hamtakula	G	Y'all will not eat
2.	Hatupi	E	We do not give
3.	Hawakula	B	They did not eat
4.	Hawakupa	C	They did not give
5.	Huchi	H	You do not fear
6.	Mlikucha	F	Y'all feared
7.	Sili	A	I do not eat
8.	Unakucha	I	You fear
9.	Watakupa	D	They will give

n → a → c → l → o
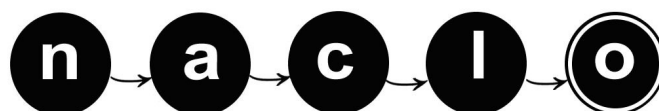
14.

1.	You visit.	Unatembelea
2.	I do not visit.	Sitembelei
3.	Y'all visited.	Mlitembelea
4.	We did not visit.	Hatukutembelea
5.	He/she will visit.	Atatembelea
6.	They will not visit.	Hawatatembelea

7.	You die.	Unakufa
8.	I do not die.	Sifi
9.	Y'all died.	Mlikufa
10.	We did not die.	Hatukufa
11.	He/she will die.	Atakufa
12.	They will not die.	Hawatakufa
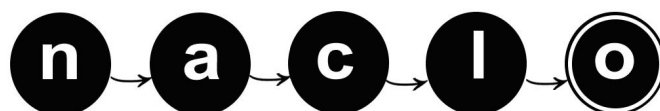
(J) Lexicondensed (1/1)

Section One:

    Task 1:

        The words produced are **cedar**, **beech**, **maple**, and **ginkgo**

        Here is the progression for each entry:

```
cdnrgt->cdnrar->cedar
cdnsks->cdjeejs->beejs->beech
cavrgt->cavre->mavre->maple
cavsks->cakos->cako->cjinkcjo->ginkgo
```

    Task 2:

        The addition to PARTONE is `qcv`.

        The addition to PARTTWO is `bjs`.

        The five new words are **birch**, **cypress**, **poplar**, **hickory**, and **eucalyptus**.

        Here is the progression for each new entry:

```
cdnbjs->cdjirjs->birjs->birch
cavbjs->cavbress->caypress->cypress
qcvrgt->qcvrar->povrar->poplar
qcvsks->qckos->qckory->hickory
qcvbjs->qcvbtus->qcyptus->qcalyptus->eucalyptus
```

Section Two

Spelling Change Rules

```
u s  -> i o t
ian -> 0 || t *
a i -> 0
a n -> 0 || [ l i | e n i ] *
a l i -> u e s e
i -> 0 || [ e | u | i | a n | m ] *
o i -> 0 || c *
i a n -> l e s e || o *
a n d i a n -> i s h
a n -> e s e || [ n | m ] *
n e t h e r l a n d s i a n -> d u t c h
```

NOTE: In the rule "`i -> 0 || [ e | u | i | a n | m ] *`," the letters e, u, i, and m can appear in any order. In addition, the i could be replaced with j.
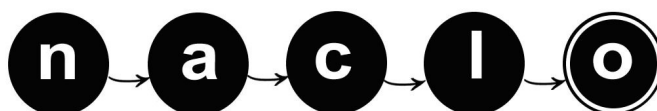
(K) Don't be Ukhrul to a Liver that's True (1/1)

Kachai	Tusom	Ukhrul	English
(P) kətʰe	(a) kətʰue	kətʰuj	awaken
(H) kəkʰu	(t) kəkʃi	kəkʰa	bitter
(K) kəkwe	(d) kəkie	kəkaj	break
(N) kəce	(s) makəcuə	kəcuj	burn
(F) kʰəmwe	(e) kʰəŋie	kʰəŋaj	desire
(A) kʰəŋətʰi	(h) kʰantsy	kʰəŋətʰu	exchange
(E) ʔale	(n) luə	luj	field
(L) ʔami	(l) ma	mej	fire
(J) ʔasu	(k) ʃi	sa	flesh/animal
(G) ʔatʰi	(o) ʔətχa	ʔatʰej	fruit
(M) ʔamɐ	(c) mɯ	mi	human
(C) məkʰu	(m) mokʃi	məkʰa	jaw
(S) ʔakʰwe	(i) kʃie	kʰaj	knife
(T) kʰəməni	(g) kʰanny	kʰəmənu	laugh
(D) ʔamətʰen	(f) ʔəntsɯ̃	ʔamətʰin	liver
(O) ʔacu	(q) ci	ca	necklace
(Q) kʰəmɐn	(r) kʰəmɯ̃	kʰəmin	ripe
(R) kətʰi	(b) kətχa	kətʰej	see
(B) kəpʰu	(j) kəpʃi	kəpʰa	seek
(I) kəði	(p) za	tsej	spear
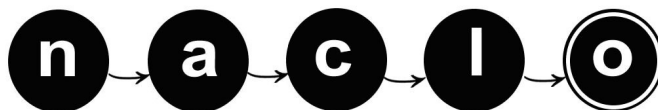
(L) Transducing Runes (1/2)

L1.

B)	start	2	start	start	2	start	start
C)	start	start	1	start	3	start	start
D)	start	3	start	1	start	start	

(a) he#
Start in the "start" state. The edge starting in "start" marked with the character "h" transitions to the "start" state, so that is the second state. Then, the edge starting in "start" with the character "e" transitions to State 1, so that is the third state. Finally, the edge starting in "start" with character "#" returns to the "start" state, so that is the final state.

(b) stash#
Start in the "start" state. The edge starting in "start" marked with the character "s" transitions to State 2, so that is the second state. The edge starting in State 2 marked with character "t" returns to the "start" state, so that is the third state. The edge starting in the "start" state marked with character "a" returns to the "start" state, so that is the fourth state. The edge starting in the "start" state marked with character "s" transitions to State 2, so that is the fifth state. The edge starting in State 2 marked with character "h" returns to the "start" state, so that is the sixth state. Finally, the edge starting in "start" with character "#" returns to the "start" state, so that is the final state.

(c) heath#
Start in the "start" state. The edge starting in "start" marked with the character "h" transitions to the "start" state, so that is the second state. The edge starting in "start" marked with the character "e" transitions to State 1, so that is the third state. The edge starting in State 1 marked with character "a" returns to the "start" state, so that is the fourth state. The edge starting in the "start" state marked with character "t" transitions to State 3, so that s the fifth state. The edge starting in State 3 marked with character "h" returns to the "start" state, so that is the sixth state. Finally, the edge starting in "start" with character "#" returns to the "start" state, so that is the final state.

(d) thee#
Start in the "start" state. The edge starting in "start" marked with the character "t" transitions to State 3, so that is the second state. The edge starting in State 3 marked with the character "h" transitions to the "start" state, so that is the third state. The edge starting in the "start" state marked with character "e" transitions to State 1, so that is the fourth state. The edge starting in the State 1 marked with character "e" transitions to the "start" state, so that is the fifth state. Finally, the edge starting in "start" with character "#" returns to the "start" state, so that is the final state.

n a c l o
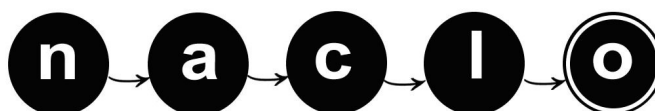
L2. (a) 1   If you see "g" by itself, it should become ✗

(b) 7   If you see "n" followed by "g", it should become ✖

(c) 24   If you see "n" followed by "h", it should be become ⸸ ᚺ

(d) 25   If you see "n" followed by "a", it should become ⸸ ᚠ

(e) 2   If you see "n" followed by "n", it should become ⸸ but nothing else because we don't know if the second "n" will be followed by a "g".

(f) 2   If you see "n" followed by "t", you can write ⸸ , but nothing else because we don't know if the "t" will be followed by a "h".

(g) 3   We know that State 1 corresponds to states where we've seen an "e", but we don't know what will come after (the clue is that there is a edge from State 4 to State 1 with edge e: ᛗ .) If we see an "e" followed by an "n", we should write ᛗ , but nothing else because we don't know if "n" will be followed by a "g".

L3. (a) 40   There are eight possible inputs (aehgnst#), and each of the five states must have an edge to handle any possible input: 5 * 8 = 40

(b) 5   There is a state for each letter that can begin a digraph ("e", "n", "s", and "t") plus one for the start state.

(c) 45   There are nine possible inputs (adehgnst#), and each of the five states must have an edge to handle any possible input: 5 * 9 = 45.

n → a → c → l → o
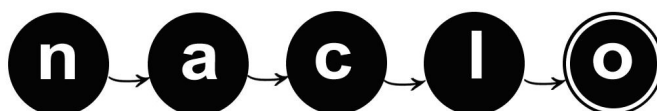
(M) Come to Istanbul (1/2)

From the sentences given, we can determine some facts about Turkish:

- The verb tends to come at the end of the sentence.
- That the verb takes the ending –*iyor*, and after this vary according to who's doing the action (-*uz* for "we", -*lar* for "they", -*sun* for "you", and no further ending for he/she/it).
- When there is a verb of the form "to X", this comes out as X-*mak* or X-*mek*. (So "to come" is *gelmek*, while "to buy" is *almak*.
- That noun endings vary according to plurality (-*lar*/-*ler* for plurals), according to their possessor (-*ım*, -*am* for "my", -*an* for "your", -*imiz* for "our"), and according to their role in the sentence (-*de* for "in X", -*den* for "from", -*a*/-*e* for "to", -*u* for the object of the sentence, and no further ending for the subject of the sentence).

Looking at the sentences in M1 and M2 gives us additional information about word endings. We see familiar suffixes except for their vowels, like –*dan* where we earlier saw –*den*, and –*i* where we previously saw –*u*. We've already seen that the vowels sometimes differ, like how –*mak* and –*mek* both represented "to <verb>". Keeping this in mind helps us work out unfamiliar suffixes like -*umuz*; this is the familiar -*imiz* but with different vowels. (It's possible, from this data, to figure out why the endings take the vowels they do, but this step wasn't necessary to solve the problem.)

Given these insights, we can assign the following meanings to the sentences in M1. (Note that there are several ways to express some of these in English, and many similar sentences were given full or partial credit.)

A. Baban mutlu mu? — Is your father happy? [or: Your father is happy?]
B. "Şehrimize gel" diyoruz. — We say/said "come to our city".
C. Arkadaşım doktor olmak istiyor. — My friend wants to be/become a doctor.
D. Fakir evimi seviyorlar mı? — Do they like my poor house?
E. İstanbul'dan mı geliyorsun? — Are you coming [or: Do you come] from Istanbul?

For part M2, we find a new word ending that we have not before encountered, -*diğ*/-*duğ*. We don't know what this means, but from the endings we can determine something about how it functions. It takes familiar verb roots like *gel-* (say), *ol-* (be), and (*al-*) (buy), but what comes after it are the endings we associate with nouns (possessor endings like –*imiz* and role endings like –*den*). Whatever –*diğ* means, its function appears to be to make meanings like "say", "be", and "buy" into nouns, or something that acts like a noun.

Any answer that gets that function right was accepted, so long as it took into account the other endings as well, but the real answer is that -*diğ*/-*duğ* functions like –*ing* in English, turning a verb "buy" into something like "buying". So *geldiğimde* means "to my coming" and *geldiğimizden* means "from our coming". This gives us meanings like the following for the sentences in M2 (next page):

n → **a** → **c** → **l** → **o**

(M) Come to Istanbul (2/2)

A. Geldiğimde "merhaba" diyorlar.   When I come/came they say/said "hello".   (Literally, "At my coming they say 'hello'.)

B. Baban geldiğimizden mutlu mu?   Is/was your father happy that we came/come/are coming? (Literally: "Is your father happy from our coming?")

C. Fakir olduğunu diyorlar.   They say/said (that) you are poor. (Literally: "They state your being poor.")

D. Aldığın ev büyük mü?   Is the house that/which you bought/are buying big? (Literally: "Is your buying house big?")

E. En mutlu olduğum şehir, Van.   The/a city where I am/ was happiest is/was Van. (Literally: "My happiest-being city is Van".)

F. Fakir olduğumuz halde mutluyuz.   Although we are/were poor, we are/were happy.

This last sentence contains an element *hal* that is also unfamiliar. From context, it probably means "although/despite", "because", "unless", or something similar that would join "We are happy" and "We are poor". In fact, it means something similar to "although/despite".
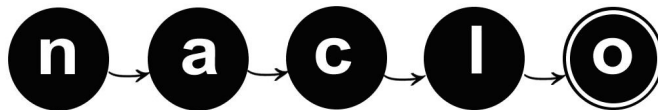
Rows 1-7, columns A-G. Then:

C1 facing west
A5 facing west
B4 facing north
G5 facing south

(O) CCG (1/1)

O1. CCG assigns a category to each word and constructs a parse by combining pairs of categories to form an S. Not all pairs of categories can combine. A pair is allowed to combine if one category (e.g. A) is contained within the category next to it (e.g. B / A) and lies on the side indicated by the slash (\ for left, / for right). When two categories combine, the result is a new category, taken from the left of the slash (B in this example).

O2. There are four categories that 'long' could have that would create a successful parse of 'I enjoy long books':

    1. NP / NP
    2. ((S \ NP) \ ((S \ NP) / NP)) / NP
    3. ((S \ NP) / NP) \ ((S \ NP) / NP)
    4. ((S / NP) \ NP) \ ((S \ NP) / NP)

The first of these is probably the most appropriate. Some possible reasons:
- It is by far the simplest. (After all, all our other categories are relatively simple.)
- It keeps the existing structure of the sentence (where "enjoy" combines with what follows it and then with what precedes it).
- "Long" describes "books" and not "enjoy", so it might make sense to keep them together.
- The first would be the only one to work if "long books" were in any other position.

O3. Possible answers: "I enjoy sleep", topicalized object sentences like "Books I enjoy" and "Sleep I enjoy".
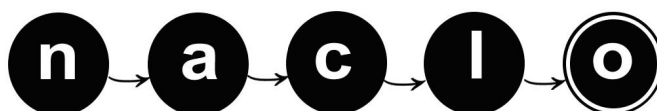
(P) Combining Categories in Tok Pisin (1/2)

P1.

1.	*Brata bilong em i stap rit.*	E
2.	*Ol i stap dringim wara.*	H
3.	*Ol i ken ritim buk bilong mi.*	C
4.	*Em i ritim buk pinis.*	A
5.	*Em i laik rit.*	G
6.	*Susa bilong em i ken rait.*	D
7.	*Susa bilong mi i boilim wara.*	B
8.	*Wara i boil pinis.*	F

A.	He has read the book.
B.	My sister boils the water.
C.	They can read my book.
D.	His sister can write.
E.	His brother is reading.
F.	The water has boiled.
G.	He wants to read.
H.	They are drinking water.

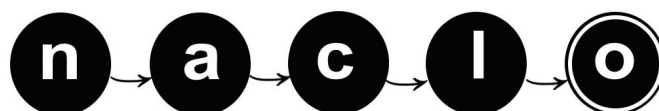P2.   My brother is reading my sister's book.

P3.   Susa bilong ol i laik raitim buk.

P4.

1.	bilong	B
2.	brata	A
3.	boil	D
4.	boilim	E
5.	buk	A
6.	dringim	E
7.	em	A
8.	i	C
9.	ken	G
10.	laik	G

11.	mi	A
12.	ol	A
13.	pinis	F
14.	stap	G
15.	raitim	E
16.	rit	D
17.	ritim	E
18.	susa	A
19.	wara	A

A.	NP
B.	(NP \ NP) / NP
C.	(S \ NP) / (S_b \ NP)
D.	(S_b \ NP)
E.	(S_b \ NP) / NP
F.	(S_b \ NP) \ (S_b \ NP)
G.	(S_b \ NP) / (S_b \ NP)

n → a → c → l → o

P5.   A. Any noun or pronoun is category A (NP) because they can be used as a noun.

B. The word "bilong" shows possession of the preceding NP by the following NP; therefore, it is (NP\NP)/NP. Also, the phrase [NP bilong NP] yields a noun phrase (NP).

C. The word "i" is necessary for a grammatical sentence, so it is $(S\backslash NP)/(S_b\backslash NP)$. It wants a following verb phrase (indicated by $(S_b\backslash NP)$) and a preceding noun phrase (NP). $NP+i+(S_b\backslash NP)$ forms a sentence.

D. Each intransitive verb (boil and rit) can stand on its own as $S_b\backslash NP$, forming the verb phrase.

E. Transitive verbs (boilim, dringim, raitim, ritim; the ones ending in -im), need a following NP, so they are categorized as $(S_b\backslash NP)/NP$, a verb phrase followed by a noun phrase.

F. The verbs "stap," "ken," and "laik" precede the primary verb phrase and need another verb phrase to create an $S_b\backslash NP$, so they are the category $(S_b\backslash NP)/(S_b\backslash NP)$.

G. The verb "pinis" comes after the main verb, so it is of the category $(S_b\backslash NP)\backslash(S_b\backslash NP)$ which requires a $(S_b\backslash NP)$ to precede it.
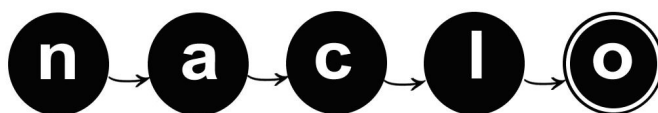
Sentence	Incorrect Word	Corrected Word
A.	nyuniny	nyundu
B.	banggaaldu	banggaalda
C.	bama	bamaal
D.	ngayu	nganda
E.	bamaal	bama
F.	guluguluugu	gulugulu
G.	bama	bamaal
H.	nganjiiny	nganjiinda
I.	biwuudu	biwuuda
J.	nyundu	nyuniny
K.	dunguu	dungu
L.	bunyaang	bunya

A. "Nyundu" is attested in sentence 10 as the subject of a transitive verb.
B. "Banggalda" is attested in sentences 8 and 9 meaning "with an axe".
C. "Bamaal" is attested in sentences 2, 12, and 16 as the subject of a transitive verb or part of the subject of a transitive verb in combination with "bunyaang" (woman).
D. Following the pattern of sentence 3: when the sentence is about possession, the possessor is in its "nda" form, meaning "to" or "with". Sentence 3 might be paraphrased as "How much food is with/at/to you?"
E. Following the pattern of sentence 17: The subject of an intransitive verb is in its shorter form. For example, "woman" in 17 is "bunya", not "bunyaang".
F. Combining information from sentences 1, 2, 3, and H: "Mayiigu" means "for food", "mayi" means "food". "Guluguluugu" means "for black bream", "gulugulu" means "black bream" used as the object of a verb.
G. "Bamaal" is attested in sentences 2, 12, and 16 as the subject of a transitive verb.
H. Following the pattern of sentences 10 and 18: When the verb means "give" or "show", the person who is receiving/being shown something is in its "nda" form. You could think of "nda" as meaning "to". Sentence 18 shows "nganjiinda" meaning "to us".
I. "Biwuuda" is attested in sentence 5 meaning "with a fishing spear".
J. Follow the pattern of sentence 7 where "nganyany" is used instead of "ngayu" when the verb ends in "jinyu".
K. Follow the pattern of "jinaa" in sentences 19 and 20 corresponding to "jina" in sentence 7, or "mandii" in sentence 21 corresponding to "mandi" in sentence 1.
L. "Bunya" is attested in sentence 15 as the object of a verb.