



Australian Computational and Linguistics Olympiad 2010 Preparation Package

For students who have not participated in OzCLO before

1. Try some of the linguistic problems in this package, and then check your answers against the solutions provided.

Teachers may choose to use one or more of these problems with their classes to select which students will represent the School. The Kiswahili problem is relatively straightforward and makes a good way to introduce a class to linguistic analysis.

and

2. Attend the optional Training Session
 - to get some practice with more sample problems, and tips for solving them, and
 - to find out what linguistics, computational linguistics and language technologies are all about.

then ...

For all students

Try out the problems that are available via the OzCLO website.

- Problems from the OzCLO 2009 State and National competitions.
- The other practice problems, including problems from the North American competition.

<http://www.ozclo.org.au/>

All students who register for OzCLO should have attempted at least some linguistic problems of this type before the State Round.

Problem 1: Kiswahili

Kiswahili is spoken by more than 50 million people (including first- and second-language speakers) in a number of countries in eastern Africa. Carefully study these Kiswahili words and their English translations, and answer the questions on the next page.

1.	atanipenda	s/he will like me
2.	atampenda	s/he will like him/her
3.	atatupenda	s/he will like us
4.	atawapenda	s/he will like them
5.	atakupenda	s/he will like you
<hr/>		
6.	nitakupenda	I will like you
7.	nitampenda	I will like him/her
8.	nitawapenda	I will like them
<hr/>		
9.	utanipenda	you will like me
10.	utampenda	you will like him/her
11.	tutampenda	we will like him/her
12.	watampenda	they will like him/her
<hr/>		
13.	atanipiga	s/he will beat me
14.	ananipiga	s/he is beating me
15.	alinipiga	s/he beat me
<hr/>		
16.	atakupiga	s/he will beat you
17.	anakupiga	s/he is beating you
18.	alikutiga	s/he beat you
<hr/>		
19.	atampiga	s/he will beat him/her
20.	anampiga	s/he is beating him/her
21.	alimpiga	s/he beat him/her
<hr/>		
22.	atakusumbua	s/he will annoy you
23.	unamsumbua	you are annoying him/her

OzCLO 2010 Preparation

i. Work out which *parts* of the Kiswahili words go with each of the following meanings:

<ul style="list-style-type: none"> • Subjects: (one who does the action) 	I	<ul style="list-style-type: none"> • Objects: (one that the action is done to) 	me
	you		you
	s/he		him/her
	we		us
	they		them
<ul style="list-style-type: none"> • Tense: (when) 	future	<ul style="list-style-type: none"> • Verb bases: (describes the action) 	like
	present		beat
	past		annoy

ii. In what order do the different parts (subject, tense etc.) occur in a Kiswahili word?

iii. What would the meanings be for these Kiswahili words?

- walikupenda
- utawapiga
- tulimsumbua

iv. What would the Kiswahili words be for these meanings?

- We beat them.
- You are annoying me.
- They liked him/her.

Problem 2: Kannada

by Mirjam Fried

Kannada is one of the major languages of India, spoken by more than 25 million people primarily in the South of the country, near Bangalore. It's a very old language and it uses its own writing system. For the purpose of this puzzle, the Kannada letters are transcribed using the Roman alphabet. The letters L, D, T, and N represent a special pronunciation with the tongue curled upward. Note that there is no translation for 'the' in Kannada.

mane 'house'	manege 'to (the) house'
peeTe 'market'	peeTege 'to (the) market'
tande 'dad'	tandege 'to dad'
roTTi 'flat bread'	roTTige 'to (the) flat bread'
chaTNi 'chutney (a relish)'	chaTNige 'to (the) chutney'
hakki 'bird'	hakkige 'to (the) bird'
taayi 'mother'	taayige 'to mother'
jooLa 'corn'	jooLakke 'to (the) corn'
pustaka 'book'	pustakakke 'to (the) book'
simha 'lion'	simhakke 'to (the) lion'
kalkatta 'Calcutta'	kalkattakke 'to Calcutta'
manushya 'man'	manushyanige 'to (the) man'
amma 'mom'	ammanige 'to mom'
huDuga 'boy'	huDuganige 'to (the) boy'
sneehita 'friend'	sneehitanige 'to (the) friend'

i. Carefully examine the pairs of Kannada words on each line above. Which parts of the words in the righthand column correspond to the English word 'to'?

ii. Now, here are some new words in Kannada:

hamsa 'swan'
akka 'older sister'
tangi 'younger sister'

Study the examples above to work out how would you say:

'to (the) swan'

'to (the) older sister'

'to (the) younger sister'

Hint: In understanding how a language works, we often look at the individual sounds that make up the words and the kinds of meanings that the words have.

Problem 3: Using rules to make strings.

In one sense we can think of a sentence as a string or sequence of words. But it's not a random string of course: there are rules. This problem is about a type of rule that builds up ('generates') strings, but we'll use characters (letters) here instead of words.

You start with a string of characters. If your string contains a character that appears on the left side of the arrow in a rule, you can turn that character into whatever is on the right side of the arrow in that rule. You can apply different rules to your string over and over again until no more moves are possible. You're not allowed to twiddle the order of the characters in your string.

Here are the rules:

$S \rightarrow AB$

$A \rightarrow ab$

$A \rightarrow aAb$

$B \rightarrow bcd$

$B \rightarrow bBc$

i. If you start with 'S', which of these strings is it possible to end up with using these rules? (Put a tick to the right of the possible strings, and a cross to the right of the impossible ones.)

1. abcd

2. abbcd

3. aabbbcd

4. aaabbbcd

5. abbbbcdcc

6. aabbccdcc

7. aabbbbcdc

8. aaabbbbcd

9. aaabbbbcdc

10. aabbbbbcdcc

11. aaabbbbbcdcc

ii. Here is a string that can *not* be generated by these rules: **bbbcdccc**

Can you add a rule to all the others so that this string *can* be generated?

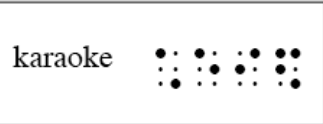
Problem 4: Tenji script for Japanese



Braille is a tactile writing system, based on a series of raised dots, that is widely used by the blind. It was invented in 1821 by Louis Braille to write French, but has since been adapted to many other languages.

English, which uses the Roman alphabet just as French does, required very little adaptation, but languages that do not use the Roman alphabet, such as Japanese, Korean, or Chinese, are often organized in a very different manner!

To the right is a Japanese word written in the *tenji* (“dot characters”) writing system. The large dots represent the raised bumps; the tiny dots represent empty positions.



A1. The following *tenji* words represent *atari*, *haiku*, *katana*, *kimono*, *koi*, and *sake*. Which is which? You don’t need to know either Japanese or Braille to figure it out; you’ll find that the system is highly logical.

a. _____	b. _____
c. _____	d. _____
e. _____	f. _____

A2. What are the following words?

g. _____	h. _____
----------	----------

A3. Write the following words in *tenji* characters:

i. samurai	j. miso
------------	---------

A4. Explain the logic behind the Tenji writing system.

Solutions

I. Kiswahili

• <i>Subjects:</i> (one who does the action)	I	ni	• <i>Objects:</i> (one that the action is done to)	me	ni
	you	u		you	ku
	s/he	a		him/her	m
	we	tu		us	tu
	they	wa		them	wa
• <i>Tense:</i> (when)	future	ta	• <i>Verb bases:</i> (describes the action)	like	penda
	present	na		beat	piga
	past	li		annoy	sumbua

ii. Subject – Tense – Object – Verb base

iii. walikupenda They liked you.
utawapiga You will beat them.
tulimsumbua We annoyed him/her.

iv. We beat them. Tuliwapiga.
You are annoying me. Unanisumbua.
They liked him/her. Walimpenda.

2. Kannada

i. Last parts: '**ge**', '**kke**' or '**nige**'

ii. 'to (the) swan' **hamsakke**
'to (the) older sister' **akkanige**
'to (the) younger sister' **tangige**

In this language, the way to express the meaning 'to' is to add one of three endings to a word: 'ge', 'kke' and 'nige'. Which ending is used is determined by two factors:

- what the final sound of the word is,
- and, if the word ends in /a/, by whether the word denotes a human or not.

So, if a word has:

Final 'e' or final 'i' → 'ge'
Final 'a' + non-human → 'kke'
Final 'a' + human → 'nige'

3. Using rules to make strings

i. Strings 2, 3, 5, 7, 8, 10, and 11 can be generated using the rules.

ii. There are a couple of possibilities for the extra rule required:

$$\mathbf{S} \rightarrow \mathbf{B}$$

or, with an empty right-side (meaning that the character is deleted):

$$\mathbf{A} \rightarrow \emptyset$$

4. Tenji

- A1 a. *haiku*
b. *sake*
c. *katana*
d. *kimono*
e. *koi*
f. *atari*

- A2 g. *karate*
h. *anime*

- A3 i. samurai



- j. miso



A4: Tenji is an alphabetical writing system in which each vowel and each consonant has a specific sign. A consonant and a following vowel sound that together make up a syllable can be combined in a single Braille symbol. Some syllables only have a vowel sound.

The vowel signs use the cells in the top row and the middle row left cell. The consonant signs use the remaining cells: bottom row cells and the middle row right cell.