



Australian Computational and Linguistics Olympiad

Preparation Package

What this package is for.

1. To prepare students for the competition. All students who register for OzCLO should have attempted at least some linguistic problems of this type before Round 1.
2. One or more sample problems can be used as an optional In-school Qualification Round run by teachers to select which students will represent the School.
3. As a fun activity to introduce students to Linguistics even if they will not be taking part in the competition. The Kiswahili problem is relatively straightforward and is a good discovery activity to develop analytical skills.

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.

and

2. Participate in the optional Training Session (where available)
 - 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.

For all students

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

- Problems from previous OzCLO first round and National competitions.
- The other practice problems, including problems from the North American competition.

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

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 Preparation

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

<p>• Subjects I </p> <p> (one who does the action)</p> <p> you </p> <p> s/he </p> <p> we </p> <p> they </p>	<p>• Objects me </p> <p> (one that the action is done to)</p> <p> you </p> <p> him/her </p> <p> us </p> <p> them </p>
<p>• Tense future</p> <p> (when)</p> <p> present</p> <p> past </p>	<p>• Verb bases like </p> <p> (describes the action)</p> <p> beat </p> <p> annoy </p>

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. abbbcdcc

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?

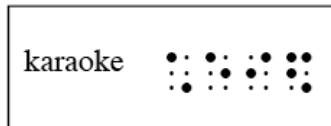
Problem 4: Tenji script for Japanese



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
------------	---------

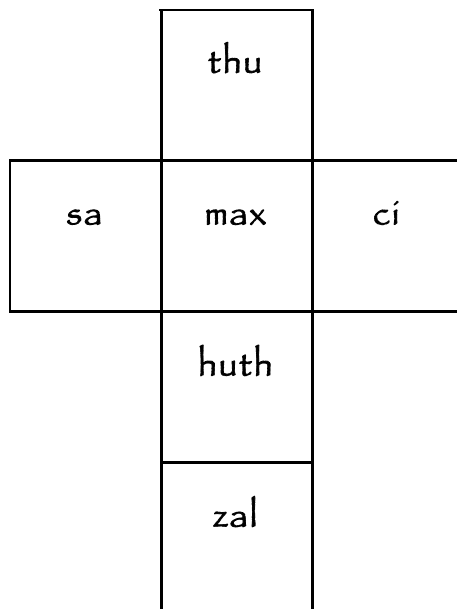
A4. Explain the logic behind the Tenji writing system.

Problem 5: Counting in Etruscan

Etruscans flourished as a separate people inhabiting parts of northern Italy centered on the region now known as Tuscany for several centuries until the 1st century B.C. when they were effectively absorbed into the expanding Roman Empire. They traded throughout the Mediterranean and acquired their alphabetic writing system from the Greeks with whom they traded extensively. They left many written texts which we can easily read, as the Greek alphabet was used. However, their spoken language became extinct and because Etruscan bears little resemblance to any Indo-European language, we cannot understand the meaning of many Etruscan words.

Generally, identification of Etruscan numbers remains difficult, but the first six numbers form a group apart. They are found in epitaphs, in which age of the deceased and the number of their children is given, and in the *Book of the Mummy* which specifies dates of the periodical religious ceremonies and the size of various offerings.

On a pair of Etruscan dice, known as the Tuscan dice, there are inscribed the following six words listed here in alphabetic order: *ci*, *huth*, *max*, *sa*, *thu*, *zal*. Each of these words corresponds to one of the numbers between 1 and 6 (compare English "one"-1; "two"-2; etc.). You can see how these number words are arranged on the two-dimensional representation of a die (cube) below:



OzCLO Preparation

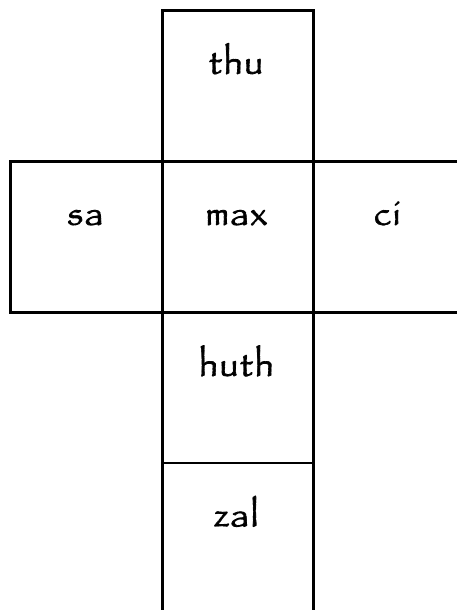
i. Which word corresponds to which number?

[6 points]

At the time of the decipherment, linguists had the following clues:

- 1) each die in a pair of dice has three pairs of opposite faces; the sum of the number on each pair equals 7;
- 2) *thu*, *ci* and *zal*, in a certain order, represent 1, 2, 3
- 3) *ci*, but not *thu* and *zal*, occurs very frequently in the Book of the Mummy;
- 4) the following pairs of words were found in epitaphs:
thu clan; *thu at*; *thu mezu*; *thu vinac*; *thu thuscu*;
ci clenar; *zal clenar*; *ci atr*; *zal atr*; *ci mesur*; *zal mesur*; *ci vinacr*; *zal vinacr*; *ci thuscur*; *zal thuscur*
- 5) in several ancient Mediterranean cultures the number '3' had special magic-like significance.

Write the correct number under its corresponding written version on the graphic of the die below.



OzCLO Preparation

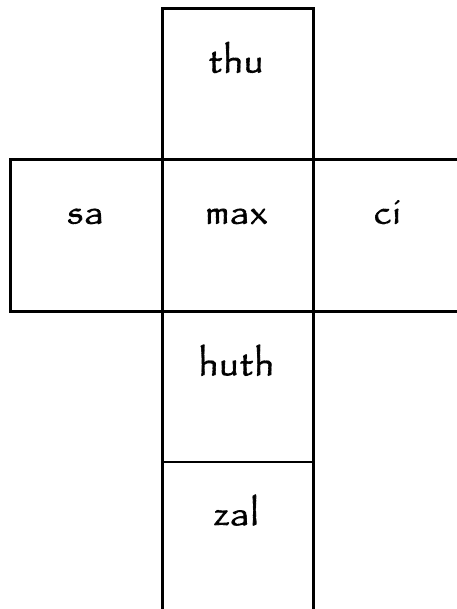
Now here's another twist.

It seems that Etruscans enjoyed gambling as many pairs of dice have been found. On another pair there are inscribed the following six words which we give here in their alphabetic order: *caius*, *est*, *i*, *va*, *volote*, *urti*. These were inscribed on the dice rather than the number words found on the "Tuscan dice".

Moreover, this choice of words is not random. It is claimed that they make up a sentence expressing a popular Etruscan proverb: *volote i va est, caius urti* meaning 'to a docile horse, the ford is pleasant'.

ii. Supposing that these words were arranged on these dice to symbolize the numbers written on the Tuscan dice, inscribe each word of the proverb below its corresponding number word on this two-dimensional figure of a Tuscan die.

[6 points]



iii. Briefly explain your reasoning for the answer you gave to Question 2. [2 points]

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

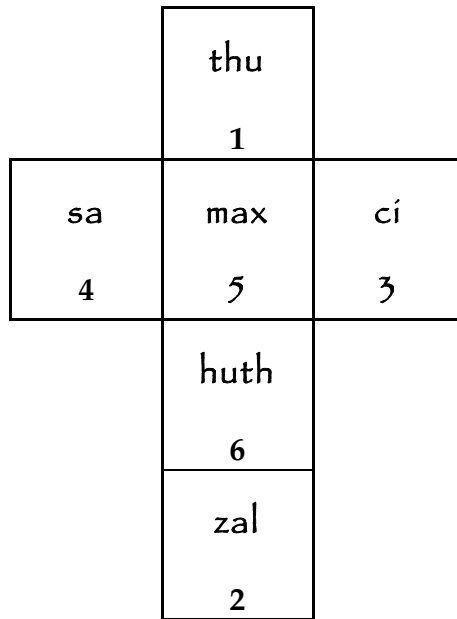


A4: Tenji is an alphabetical writing system in which each vowel and each consonant has a specific sign. A consonant and a following vowel sound that together make up a syllable can be combined in a single Braille symbol. Some syllables only have a vowel sound.

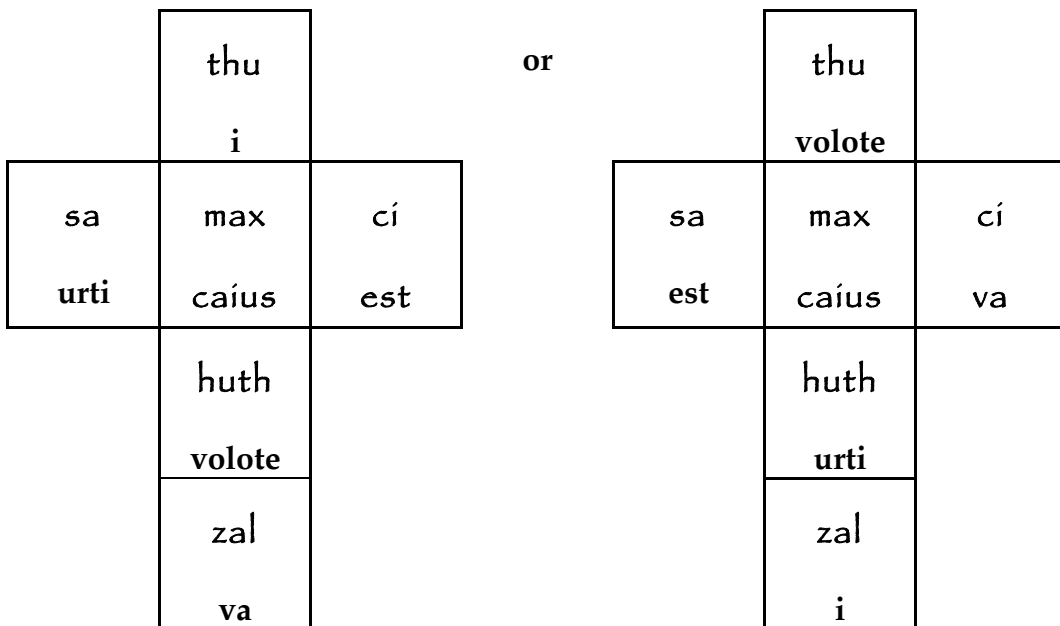
The vowel signs use the cells in the top row and the middle row left cell. The consonant signs use the remaining cells: bottom row cells and the middle row right cell.

5. Etruscan

i.



ii. There are two possible solutions. Either



iii. For the first solution above: The number of letters in each word symbolizes the number.

For the second solution above: The order of the words in the proverb corresponds to the number, e.g., first word = 1, last word = 6.