Guidelines for Linguistic Problem Creation

(Adapted from "The Linguistics Olympiads: Academic competitions in linguistics for secondary school students" © 2007 by Ivan Derzhanski and Thomas Payne)

At the heart of any Linguistics Olympiad (LO) is the self-sufficient linguistics problem, a unique genre of composition that presents linguistic facts and phenomena in enigmatic form. A steady supply of original, thoughtfully created and intriguing problems is absolutely necessary for the success of any ongoing LO programme. In a typical "live" competition, students are given several hours to solve a set of problems, many of which involve data from languages the students have never heard of, whilst others may highlight little-known features of commonly known languages or formal representations of natural languages.

Good LO problems require the solver to apply a formal style of thought familiar from the hard sciences to the realm of linguistic data, including orthographies, sounds, words and sentences. The most successful problem solvers are able to use this exposure to unfamiliar languages to discover new ways of thinking and categorizing the universe, and in so doing ultimately to develop an appreciation for both the unity of language and the diversity of languages. They come to view a language as a system built upon complex but logical and consistent principles rather than as a frustrating collection of impenetrable facts, as lessons in "grammar" so often present. This fusion of formal logic and cross-cultural perceptivity makes the problems attractive and profitable for mathematicians and language enthusiasts alike.[1] They bridge the "techie"/"fuzzy" divide that characterizes increasingly specialized academic cultures. For everyone there is also the appeal of the ludic element—the challenge of the puzzle, the same motivation that accounts for the popularity of crosswords, cryptograms, Sūdoku, etc. Finally, LO problems offer a chance to communicate a little, as it were, with cultures located far away in space or time by getting to know something about their languages, and to gain new and sometimes startling perspectives on the languages that the solver is already familiar with.

Here is a good example of a problem that exposes students to different cultural worlds, from the 16th Moscow LO (1979):

[1] The word "mathematics" (or "mathematical") that often appears in the full names of LOs highlights the fact that the gist of this activity is finding structures, regularities and correspondences (which is what mathematics is all about), rather than knowing languages (which is what linguistics amounts to, according to a popular misconception).

Six dates are given in Swahili, along with their translations in random order:

1. ***tarehe tatu Disemba Jumamosi***	A. Monday, October 5th
2. ***tarehe tano Oktoba Jumapili***	B. Tuesday, April 2nd
3. ***tarehe pili Aprili Jumanne***	C. Wednesday, October 5th
4. ***tarehe tano Oktoba Jumatatu***	D. Tuesday, April 4th
5. ***tarehe nne Aprili Jumanne***	E. Sunday, October 5th
6. ***tarehe tano Oktoba Jumatano***	F. Saturday, December 3rd

Assignment 1. Match each Swahili date with its proper translation.

Assignment 2. Translate the following additional dates into Swahili:

Wednesday, April 3rd; Sunday, December 2nd.

We challenge the reader to attempt to solve this problem before reading further.

This is but one template for LO problems, in which words, phrases or sentences in an unfamiliar language must be paired with their proper translations. There are many other possible templates, with correspondingly many possible strategies for solution (a few more samples will be found in the following pages). The exemplary quality of this problem is that it requires the solver not simply to infer correspondences between unfamiliar and familiar words, as though the subject language were a kind of coded English, but also to discover something about the culture of Swahili speakers. In particular, one must realize that Saturday is considered the first day of the week, which reflects the Islamic influence on the communities along the East coast of Africa. At the same time the names of the months are English borrowings, which reflects the British colonial history of the region. Many such surprising (to most teenagers in the Western world) facts are revealed, not by instruction, but by discovery in the process of solving engaging logic puzzles. Other mind-expanding details revealed in some of the traditional problems include: that in Hawai'ian, the words for siblings are distinguished by birth order (elder sibling *vs.* younger sibling) and relation between the genders (same-sex sibling *vs.* opposite-sex sibling); that in some Samoan nouns, singularity is marked and plurality is unmarked, and so on.

In every LO competition conscious effort is taken to provide problems that involve a variety of interesting phenomena, such as switch reference systems, (split) ergativity, noun classifiers, complex number systems, flexible word orders, etc. Problems are also chosen so as to highlight different areas of linguistic analysis: writing systems,

phonetics, morphology, syntax, semantics, historical change and formal modelling of natural language. Prominent researchers in all of these areas contribute problems, or at least data sets, so that the phenomena represented in the competitions reflect the issues addressed and the concepts and methods used in contemporary professional linguistics.

Another feature of LO problems makes them particularly attractive for talented students in schools in a wide range of socio-economic contexts. This is the fact that they do not depend on any specific previous educational experience. In contrast to other academic competitions and standardized tests such as the SAT, Stanford–Binet Intelligence tests, A-levels, etc., solving LO problems is mostly a matter of innate ability to understand and analyse a problem, envision a strategy for solution, and carry out that strategy in real time. No knowledge of linguistics or of the languages represented in the problems is expected.[2] Memorization of facts is useless. And while there are some frequently used techniques and methods that can be taught or learned through practice which help a student solve certain types of problems, no procedure guarantees success. Each problem must be viewed as a unique challenge, requiring its own strategies and approaches.

Over the years LO programmes evolve strict requirements that they impose on problem selection. There is some variation, but most tend to agree with the following:

a) All problems must deal with real languages that are (or once were) vehicles of human communication.[3]

b) Everything in the problem must be accurate. Simplification is very rarely acceptable.[4] Made-up data is never allowed.

c) Care must be taken to be respectful of the language and its speakers. The problems must not treat the languages as curiosities, but rather should present each language as a unique and valuable expression of a particular culture.

d) Problems must be self-sufficient: all the information needed to solve a problem must be present in the data or be common knowledge. Advanced

[2] In fact, familiarity with the language featured in the problem (and presumed unknown to the solver) would sooner be an obstacle than an advantage, because writing down an answer obtained from a source external to the problem does not normally count as a solution.

[3] More precisely, there is a hierarchy of acceptability, with natural languages at the top, followed by famous auxiliary and fantasy languages such as Esperanto or JRR Tolkien's languages, then by fragments of fictional languages or notational systems created expressly for the problem but featuring interesting phenomena found in natural languages (and perhaps difficult to illustrate using authentic material). Few LOs ever use problems on fictional linguistic phenomena, on ways in which languages don't actually work (though perhaps they might). Problems on sign systems other than human language, such as bar codes, computational automata, and scholarly notations used in various domains, constitute a category of their own.

[4] As a rule, simplification is allowed only in the orthography or the standard transcription when it contains much distracting detail and the problem highlights something else. In such cases, the solver should be informed that simplifications have been made.

linguistic terminology is avoided (if necessary, terms may be introduced in the statement of the problem).[5]

e) Simple analogy cannot be the only strategy needed to solve a problem. "Multi-stage logic" is required (see below for an example).

f) Sheer insight should not be the only way to solve a problem either. If no progress at all can be made unless the solver happens to stumble upon a particular idea, this is considered a serious shortcoming of the problem.

g) There can be one and only one reasonable solution, and that solution must be consistent with the facts of the language.

Here is an example of a problem that would *not* be acceptable, even in a much more complicated version, for any LO because it violates provision **e)** above (data from Turkish):

1. ***deniz*** 'sea'	3. ***denizde*** 'in the sea'	5. _____ 'pictures'
2. ***denizler*** 'seas'	4. ***resim*** 'picture'	6. _____ 'in the picture'

In this data set, the plural and locative forms of the word *resim* 'picture' are directly inferred from the forms of *deniz* 'sea'. An acceptable LO problem would require solvers to make at least one intermediate logical step; for example, consider the following:

1. ***deniz*** 'sea'	4. ***okul*** 'school'	7. ***okulda*** '_____'
2. ***denize*** 'to the sea'	5. ***okullar*** 'schools'	8. _____ 'seas'
3. ***denizde*** 'in the sea'	6. ***okullara*** '_____'	9. _____ 'in the seas'

In this example, in addition to recognizing the several noun suffixes, the solver must notice that they have different forms for the different roots and that the correspondences cut across the various suffixal categories (plural, locative and dative). Furthermore, there is no direct analogical model for the last translation—'in the ‹PLURAL NOUN›'. The solver must notice that plurality comes before case, and then infer the order *-ler+de* for the noun meaning 'sea', based on the order *-lar+a* in example (5). Crucially, no knowledge of suffixation, plurality or case, much less vowel harmony, is necessary to solve this problem. Rather, these features are "discovered" by the solver.

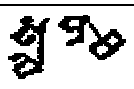
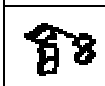
[5] What concepts count as too advanced depends on the school curriculum: in some places even "plural" or "suffix" are disallowed, in others they are expected to be known to all students from grammar classes. Terms such as "palatoalveolar" or "suppletion", however, are unlikely to be commonly known anywhere.

Predictably, the great educational potential of LO problems has been appreciated by university lecturers and school teachers, and in several countries problems from past contests have been used in classrooms. Very many people outside the educational system also enjoy them as a tool for personal enrichment and entertainment.

We conclude this description with three examples of other formats for problem creation. The first is derived from one that the orientalist Svetlana Burlak wrote for the 26th Moscow LO (1995). The second was authored by Pilar Valenzuela for the a LO in Eugene, Oregon, in 1998, and the third is a computationally oriented problem used in the first NACLO competition in 2007.

Tocharian

Below are some words in Tocharian A, an ancient language that was spoken in the Chinese province of Xinjiang in the 6th–8th centuries C.E.[6] Each word is presented in the traditional script and in phonetic transcription and accompanied by an English translation.

	***lap***	"head"		***pkäl***	"bring!"
	***spaltäk***	"intention"		***täm***	"that"
	***maskäs***	"lives"		***sasak***	"only, alone"
	***pal***	"law"		***slamas***	"fires"
	***tmäk***	"therefore"		***pkal***	"he must be baked"
	säksäk	"sixty"		***tamät***	"he was born"
	pat	"or"			"jumping"
					"sit down!"

Fill in the blanks with the correct Tocharian word or transcription.

Note: the sound ***ä*** is a special vowel of Tocharian, different from the vowel written as ***a***.

[6] Most LO problems present such background information on the languages featured (genetic affiliation, region where spoken, number of speakers if living, historical significance otherwise), usually for the sake of interest only, although sometimes such notes contain important clues for the solving of the problem.

Quechua

Quechua was the official language of the *Tawantinsuyu* or Inca Empire before the Spanish invasion of 1532. For hundreds of years Cuzco, in what is now Peru, was the capital of the Empire. The sentences below represent the variety of Quechua currently spoken by thousands of people in Cuzco and in the area around Lake Titicaca.

The following are some sentences in Quechua, with their translations in random order.

1. ***Antukaq chakranpiqa t'ikashanmi papa.***

2. ***Siskuq chakranpiqa wiñashanmi sara.***

3. ***Siskuq chakranpiqa rurushansi kiwña.***

4. ***Antukaq chakranpiqa t'ikashanchá kiwña.***

5. ***Siskuq chakranpiqa wiñashansi sara.***

6. ***Antukaq chakranpiqa wiñashanchá papa.***

A. Potatoes may be growing in Antuka's field.
B. Barley may be flowering in Antuka's field.
C. Corn is growing in Sisku's field.
D. I've heard corn is growing in Sisku's field.
E. I've heard barley is yielding fruit in Sisku's field.
F. Potatoes are flowering in Antuka's field.

Problem 1. Indicate which translation goes with each Quechua sentence.

Problem 2. Provide English translations for the following Quechua sentences:

7. ***Istuchaq chakranpiqa t'ikashansi sara.***
8. ***Sawinaq chakranpiqa wiñashanchá kiwña.***
9. ***Tumasaq chakranpiqa rurushanmi papa.***
10. ***Kusiq chakranpiqa t'ikashanchá papa.***
11. ***Inashuq chakranpiqa rurushansi kiwña.***

Pooh's Encyclopedia

Once upon a time, a very long time ago, Winnie-the-Pooh and his friends bought an electronic encyclopedia, and tried to find answers to several important questions:

Winnie-the-Pooh:
Where should a bear stock his jars of honey?
How much honey should a bear store for the winter?

Eeyore:
Where should I look for my lost tail?
Which animals sleep during the winter?

Christopher Robin:
What is the shortest way from my place to the house of Winnie-the-Pooh?
Who wrote the books about Pooh Bear?

The encyclopedia's search engine identified a number of articles related to their questions; for example, it returned the following matches:

Winter food storage (for Winnie-the-Pooh)
Sleep patterns in mammals and other animals (for Eeyore)
Short stories and movies about Winnie-the-Pooh (for Christopher Robin)
Writers of children's books (for Christopher Robin)

On the other hand, the search engine missed several other relevant articles; in particular, it did *not* retrieve the following articles:

Planning of food supplies
Lost-and-found agencies
Finding shortest paths on a map
Biography of A.A. Milne, the author of Winnie-the-Pooh

Your task is to determine who received each of the following matches; two of these matches were for Winnie-the-Pooh, two for Eeyore, and two for Christopher Robin. Explain why!

Books about care and feeding of bears
Effects of honey on the sleep quality of humans and animals
Lost tales of "Bulls vs. Bears" stock trading
Ways to look for lost things
Ways to store food in the house
Winter hibernation of bears and rodents