

Rosetta Stone Linguistic Problems

Bozhidar Bozhanov

Problem Committee

International Linguistics Olympiad

bozhidar.bozhanov@gmail.com

Ivan Derzhanski

Problem Committee

International Linguistics Olympiad

iad58g@gmail.com

Abstract

This paper describes the process of composing problems that are suitable for competitions in linguistics. The type of problems described is “Rosetta Stone”—a bilingual problem where typically one of the languages is unknown, and the other is the native language of the person solving the problem. The process includes selecting phenomena, composing and arranging the data and assignments in order to illustrate the phenomena, and verifying the solvability and complexity of the problem.

1 Introduction

1.1 What is a linguistic problem?

Linguistic problems are a genre of composition that presents linguistic facts and phenomena in enigmatic form (Derzhanski, Payne 2009). As an entertaining way of learning about language(s) and linguistics, they are suitable for a general audience (witness their occasional appearance in popular science journals, e.g., *Nauka i zhizn'* 1980.10, 2012.6) and can also be useful in the classroom or in linguistic textbooks as illustrations or exercises for the reader (cf., e.g., Testelet 2001), but at present their most common purpose is to be assigned to (usually) secondary-school students at contests such as the Moscow Traditional Olympiad in Linguistics, the North American Computational Linguistics Olympiad or the International Linguistics Olympiad (IOL).

Each problem may present phenomena from one or several subfields of the study of language—phonology, morphology, syntax, semantics, historical and comparative linguistics, writing systems, pragmatics, discourse analysis, etc.

There are two important requirements of the genre:

- The problem must be self-sufficient: it should contain all the necessary information for its solving, not expecting from the solver any prior knowledge (of languages, linguistics, mathematics, etc.) beyond what is commonly included into the secondary school curriculum.
- The problem must be unambiguous: it should not allow more than one plausible explanation of the data.

1.2 What is a “Rosetta Stone” linguistic problem?

In a “Rosetta Stone” linguistic problem¹ the material has the form of ordered matching expressions of two languages or language-like symbolic systems, so chosen as to enable deducing the regularities behind the correspondences, which is the essence of the problem.

In the most common subtype of Rosetta Stone the solver is given expressions (words, phrases, sentences) in an unfamiliar language and their translations into “Solverese” (a familiar working language, usually the solver’s native language)² and, in most cases, asked to translate more expressions in both directions (from the unfamiliar language to Solverese and *vice versa*). Less often the assignments require one to choose translations from a list, produce alternative transla-

¹ This term was introduced by Ivan Derzhanski in 2004 and gained currency within IOL’s Problem Committee. The idea is that the way to solve such a problem (by comparing matching structures in different languages) resembles Jean-François Champollion’s method of deciphering Ancient Egyptian with the aid of a parallel Egyptian and Greek text inscribed on a granodiorite stele that had been discovered near the town of Rashid (Rosetta) in the Nile Delta. Another term for this type of problem is *bilingua*, used by the team of the Moscow Linguistic Olympiad.

² This term was also invented by Ivan Derzhanski in 2008 and became part of the jargon of the Problem Committee of IOL. It is modelled upon language names such as *Chinese*, but is also inspired by *Motherese* ‘speech used by adults when talking to infants’ and similar coinages.

| IOL1 | IOL2 | IOL3 | IOL4 | IOL5 |
|-----------------|-----------------|-----------------|-----------------|----------------|
| Ⓔ4 15.24 | Ⓔ1 15.26 | Ⓔ1 12.90 | Ⓔ1 12.63 | #5 14.62 |
| #1 14.85 | #4 15.17 | #2 11.98 | #2 9.17 | #2 14.17 |
| #5 14.06 | #2 11.78 | Ⓔ4 11.56 | Ⓔ5 8.81 | #1 11.80 |
| #3 11.56 | #5 8.87 | Ⓔ3 10.66 | Ⓔ4 8.77 | #4 3.80 |
| #2 6.88 | #3 3.85 | #5 4.84 | #3 6.79 | #3 3.43 |
| IOL6 | IOL7 | IOL8 | IOL9 | IOL10 |
| Ⓔ4 13.00 | Ⓔ1 14.77 | #1 15.49 | Ⓔ3 13.62 | Ⓔ5 9.60 |
| #3 12.96 | #2 11.29 | #3 14.29 | #2 9.13 | Ⓔ4 8.92 |
| #1 12.94 | Ⓔ5 9.28 | #4 9.55 | Ⓔ1 6.38 | Ⓔ2 7.69 |
| Ⓔ5 9.78 | #3 4.38 | #5 9.43 | #4 4.75 | Ⓔ1 6.41 |
| #2 5.75 | #4 1.33 | #2 7.38 | #5 4.64 | #3 6.29 |

Table 1. IOL1–10: the average scores for the problems (ordered from highest to lowest).

tions, judge the well-formedness of phrases or sentences in the unknown language, locate errors, or explain the meanings of words or phrases that don't translate readily into Solverese.

If the material of the unfamiliar language consists of number names, their meanings can be given in figures instead of Solverese expressions. Problems on number names are often thought to form a separate type, but arguably (Zhurinsky 1993; Derzhanski 2007) they are ordinary problems on a somewhat peculiar discrete domain of semantics.

As such, Rosetta Stones contrast with “Chaos and Order” problems, in which the expressions in the two languages are not ordered and matching them is part of the solution, or problems on inferring the laws of a single system (a fragment of a language's grammar, a poetic genre, a mnemonic system) without comparing it to another.

In all cases solving the problem involves discovering and analysing the regular correspondences and deriving a mini-grammar and vocabulary of the unfamiliar language from the data before proceeding to the assignments.

The genre described above is the bread-and-butter type of problem at linguistic contests. Although the classification of a linguistic problem is often a fuzzy issue, among the 50 problems that have been assigned at the individual contests of the first ten instalments of IOL, 18 (36%) can be counted as classical Rosetta Stones, as are eight (40%) of the 20 problems in (Derzhanski 2009). Not surprisingly, experienced solvers are better prepared to handle these than problems of other types: at all eight IOLs where such problems were assigned, the best-solved problem was always one of them, and the worst-solved problem never was one. This can be seen in Table 1, which presents the contestants' average scores for the problems of IOL1–10, ordered from high-

est to lowest within each year, with the classical Rosetta Stones marked by “Ⓔ” and boldface. (The maximal possible score for each problem was 20. There were two exceptions at IOL1, but in the table the scores for those problems have been normalised to enable comparison with the others.) Or it can be observed that none of the ten worst-solved problems at IOL1–10 have been classical Rosetta Stones, whilst among the 40 others they are evenly distributed, meaning that they are relatively well received, but not trivially easy.

Some increasingly non-prototypical subtypes of Rosetta Stones include problems in which:

- the unfamiliar language is not a speakable human language but a symbolic system such as a pasigraphy (e.g., Linzbach's “Transcendental Algebra”³);
- or the two matching sets of data are not expressions in an unfamiliar language and in Solverese but expressions in two unfamiliar languages;
- or they are words or sentences of a single language written in two scripts, or in orthography and a transcription, and one has to derive the rules of spelling and pronunciation;
- or they are cognate words (or loanwords and their sources) of two languages or dialects, and the rules to derive are phonetic correspondences;
- or both sets are non-language data which, however, share some important characteristics with human language and thus can be said to be of linguistic interest

³ IOL1, problem 1 (Ksenia Gilyarova).

(e.g., messenger RNA sequences and the corresponding polypeptide chains⁴).

In some problems there are more than two languages (or language-like systems) involved: the data may consist of parallel sentences in two unfamiliar languages and in Solverese, or in an unfamiliar language in two scripts (or in orthography and transcription) as well as Solverese⁵, or of cognate words of several languages.

With the concept so extended, the ratio of Rosetta Stones at the first ten IOLs rises to more than a half (27 of 50). However, although all subtypes of Rosetta Stone share the same general method, all are not equal in linguistic content. What follows will concentrate on the classical subtype.

2 Selecting phenomena

Like most other types of problems, a Rosetta Stone problem is normally built around an interesting linguistic phenomenon that is not present in Solverese.⁶ In order to illustrate that phenomenon, some side phenomena must be included which allow forming actual sentences, phrases, or word forms.

It is characteristic of the classical subtype of Rosetta Stone that the Solverese expressions are given for the sake of their meaning only, so the bilingual analysis involves matching components of the *structure* of each expression in the unknown language with components of the *meaning* of its translation. Overestimating the importance of the structure of the Solverese (and treating the unfamiliar language as a code for Solverese) is an error frequently made even by experts in the field. Thus solving a problem of this type always involves some amount of semantic analysis of the Solverese expressions.

2.1 The main phenomenon

Having a single main phenomenon in a problem is not a rule—there may be two of them—but having more usually makes the problem too hard to solve in a limited timeframe. The main phenomenon is usually something interesting and

intriguing that the author of the problem has stumbled upon while researching (or sometimes authoring, typically on the basis of fieldwork) a description of the unfamiliar language.

Some phenomena pertain to the ways in which the unfamiliar language expresses information that is also present in the translations (though likely not in the same form), others do not. Here are some examples of the former:

- semantically determined noun classes;
- ergativity (and split ergativity);
- direct-inverse argument marking;
- obviative (fourth person);
- overcounting in numerals.

The latter include phonological processes such as distant assimilation or dissimilation and sandhi, as well as complex allophony and allomorphy.

Sometimes a problem illustrates variation, i.e., is built on the fact that a class of Solverese expressions can be translated in more than one way into the unfamiliar language (which may or may not reflect an ambiguity of Solverese that is resolved there), or *vice versa* (in such case the data are often introduced as Solverese expressions with their translations into the other language).

2.2 The side phenomena

Side phenomena are included in order to properly construct examples that demonstrate the main phenomenon and to achieve the desired level of complexity. As a rule they are of lesser interest than the main phenomenon, but still require deducing from the data. They often concern such things as:

- word order;
- agreement;
- number or case marking;
- person and number marking;
- marking of verb tense and mood;
- relatively straightforward allomorphy.

The side phenomena can vary in difficulty. Some may be trivial (e.g., a plural affix) and some may be harder to discover (e.g., assimilation). There is no strict distinction between main and side phenomena. For example, assimilation may be a main phenomenon in a simpler problem and a side phenomenon in a more complex one.

The author should balance the number of the side phenomena: too few may make the main phenomenon too conspicuous and the problem too easy; too many may obscure the main phenomenon.

⁴ IOL8, problem 4 (Alexander Berdichevsky).

⁵ As in Champollion's original Rosetta Stone, which was in fact a *trilingua*, featuring Ancient Egyptian in hieroglyphic and demotic script as well as Greek.

⁶ In the case of a problem intended for a multilingual contest such as IOL, this means that the main phenomenon must be absent from all working languages which will be used at the current instalment. With respect to the side phenomena this requirement is relaxed.

3 Constructing the data

The data in a problem is the language material (word forms, phrases or sentences) that is fully given to the solver—in the unfamiliar language and in Solverese. This material must represent all the chosen phenomena without any unaccountable exceptions.

For a phenomenon to be unambiguously discoverable, it must be illustrated by several examples in the data. The bare minimum, sufficient for simpler side phenomena, is two; the main phenomenon takes more. A statistical measure of the sufficiency of the material is developed in (Testelet 1994), but to the best of our knowledge neither this theoretical method nor any other is applied in practice to evaluate the quality of new problems.

3.1 Techniques for constructing the data

Here is a non-exhaustive list of frequently used techniques and approaches to constructing the data:

- Preselect a number of words that are usable in the problem—i.e., meet the requirements for representing the phenomena. For instance, if the phenomenon involves direct vs. indirect objects, transitive verbs will be needed.
- Make a table (or tables) of all possible forms of the chosen words that can appear in the problem, and choose some for the data, leaving some for the assignments.
- Group the words you have preselected according to their properties. For example, put stems ending in vowels and in consonants in separate groups if the suffixes depend on the final sound.
- If working with phrases or sentences, don't focus on the meaning. In general, it is sufficient if they aren't so absurd as to confuse the solver.
- Consider using assignments on translation from the unfamiliar language to Solverese to complement the data in showing that certain forms are possible.

This process often requires extensive search in dictionaries and work with reference grammars or informants (the latter is very desirable, but seldom done, for practical reasons).

Simplifying the grammatical patterns of the language or changing them in any other way is considered impermissible, but some parts may be

concealed in order to make certain regular portions stand out.

3.2 The size of the dataset

There is no strict requirement for the number of examples presented to the solver. There have been problems with as few as 4 given sentences and as many as 25 word forms. The main factor to consider is phenomenon density. If the author can properly illustrate all the selected phenomena in a couple of sentences, then the number of examples is low, but the phenomenon density is high. Contrariwise, if an example may contain only a single instance of one of the phenomena, then the density is low, and many examples are required.

3.3 Parasitic solutions

A parasitic solution is one that correctly and plausibly accounts for the data in the problem but differs from the fact of the language. Although the plausibility of a parasitic solution can sometimes be a matter of debate, in general it indicates a flaw of the problem. If discovered by test-solving the problem, it can be blocked, usually by adding a specific example which it does not account for.

For example, a problem on split ergativity in Inuktitut⁷ had an early version which allowed one to think that the ergative construction was used whenever the verb begins with a vowel (whereas in fact its use is triggered by semantic properties of the object). That would have been an unlikely explanation, but as pointed out in Section 1 the solver is not expected to possess linguistic proficiency and may not be able to tell a plausible hypothesis from an implausible one. Therefore an example was added where an ergative construction was used with a verb starting with a consonant. In that way the parasitic solution was no longer accounted for the data.

3.4 Scrambling the data

If the complexity of a Rosetta Stone problem is deemed insufficient, it may be increased by scrambling the data, that is, presenting the material without indicating which Solverese expression corresponds to each expression of the unfamiliar language. In this way the Rosetta Stone problem is turned into a problem of another type, Chaos and Order. Chaos and Order problems are beyond the scope of this article, so suffice it to say here that the prevailing expert opinion is that

⁷ IOL6, problem 5 (Bozhidar Bozhanov).

this technique should be reserved for occasions where the possibility of presenting a phenomenon as a problem depends on it, and not used for adding mere technical complexity to an easy Rosetta Stone (section 6.1).

4 Constructing the assignments

The assignments are exercises on using the rules that the solver is expected to have deduced from the data. Their purpose is to verify that this has been done correctly, which is why they must not be doable by using simple analogy with the data; they must ask the solver to construct forms (or combinations of forms) that have not been given previously. Sometimes the assignments include more explicit material that will be used in them but could not have been given in the data. For example, if the data has the form of sentences and one of the phenomena is that the semantics of a noun determines its class which in turn determines the choice of an obligatory article, the assignments may include a short list of nouns in citation form for the solver to classify and use in translations (including them in the data would have revealed the articles).

As noted before, an assignment on translating from the unknown language to Solverese may also be used as a way of showing more examples of some phenomenon, thus reducing the number of examples required in the data. Any sentences (or phrases) assigned for translation from the unknown language should, however, be reasonably intuitive; an improbable translation may cause the solver to unduly question the rules.⁸

As also noted, apart from the most common “translate from X to Y” assignments, there can be assignments of other, less common types. This usually happens when the understanding of the main phenomenon is hard to assess only on the basis of translations. Questions of the form “Can you translate the following? If so, how? If not, why not?” may be asked, or an additional “story” may be told, with new pairs of matching explanations presented, especially if the phenomenon is a complex one and should be deduced in parts. Of course, formulations should be kept as simple as possible.

⁸ A sentence meaning “The dog shot itself” was nearly assigned for translation from Inuktitut into Solverese in IOL6, problem 5 (Bozhidar Bozhanov), but was eliminated in the final version.

5 Auxiliary information

Apart from data and assignments, a problem nearly always contains an introductory text, and frequently notes as well. Both may contain valuable information about the data or hints to the solver.

5.1 Introductory text

In most cases this is a mere cliché such as “Here are sentences in Such-and-Such language and their translations”, also indicating whether the translations are ordered or scrambled and sometimes making other useful statements, e.g., that certain parts of the data are there for completeness but can be ignored when solving the problem, that the transcription has been simplified, etc. Such statements are usually made directly; it is not common for information to be expressed in this text in oblique ways.

5.2 Notes

In most cases notes (footnotes or endnotes) contain information on the language and descriptions of unknown sounds, and sometimes also explanations of unfamiliar concepts.

The information about the language usually contains taxonomy information, locality and number of speakers, for the solver’s edification and for putting the problem in context. It is rarely useful for solving the problem.⁹

If the note mentions unfamiliar sounds (or spellings), it may do one of the following:

- simply state that these are sounds of the featured language, as a way of saying that they (or the letters used to write them) should be distinguished from others (and as a hint that their precise phonetic value is immaterial);
- give rough approximations to familiar sounds, usually with the only purpose of making it easier to read the problem

⁹ The popular notion that the solver should strive to use independently acquired knowledge about language families in order to guess what phenomena may be present in the problem is at variance with the principle of self-sufficiency: information that is neither part of the problem nor common knowledge is as likely to be harmful as to be beneficial. On the other hand, the author may expect the reader to apply, for example, some generally known fact of geography along with the information from the note on where the language is spoken in order to deduce something about its lexicon.

(many people find it easier to handle words if they have some idea as to what they sound like);

- explain the phonetic characteristics of the sounds, often (though not always) to indicate a phonological phenomenon.

6 Assessing the complexity of the problem

As was said above, there is no objective way of assessing the complexity of a problem; test-solving is the only procedure. However, one can try to estimate the complexity on the basis of the triviality or obscurity of the main phenomenon, the number of side phenomena, and the quantity of the assignments. This can put the problem into the broad categories of “easy”, “medium” and “hard”.

The assessment of complexity is needed, first, in order to choose an appropriate forum and audience for the problem, and second, to design a scoring system if it is to be used at a competition where the scoring must be devised *a priori*. As a rule, finding the main phenomenon is harder than finding any of the side phenomena, and translating from the unfamiliar language is easier than translating into it.

6.1 Types of complexity

There are two types of complexity of a problem—linguistic complexity and technical complexity.

The former is the complexity related to figuring out the linguistic phenomena and deducing the grammar—grouping the examples into categories, determining the structure of the sentences, segmenting the word forms into morphemes, identifying phonological processes, etc.

The latter is about the technical complexity of doing the above and involves mechanical or logical tasks rather than linguistic ones. For example, unscrambling translations given out of order (a stage of solving Chaos and Order problems) is a purely technical task, done on the basis of the number of occurrences of the instances of the phenomena. This type of complexity simply makes the problem harder without adding anything linguistically interesting to it.

In a problem linguistic complexity is favoured upon technical complexity. Rosetta Stone problems rarely exhibit undesirable amounts of technical complexity, and this is one of the reasons for which they are the dominant type of problems at contests.

6.2 Specifics of scoring

Designing a scoring scheme is a process separate from composing the problem (in most cases the author doesn’t even know at what contest the problem will be used and what the scoring system will be there, nor has any control over it; different contests seldom score a problem in the same way). In the case of a Rosetta Stone points may be allocated for finding the phenomena (as a rule, more for the main one and fewer for the side ones), as well as the assignments (reflecting an assessment of their relative importance and complexity). A study of the point counts won by the participants in the first ten instalments of IOL shows that, while on the average for each problem about ¼ of all contestants had scores in the middle third of the actual range and the rest were equally divided between high and low scorers, for classical Rosetta Stones the middle scorers outweighed the low ones. Table 2 presents the ratio of high to middle to low scorers for each problem, again with the classical Rosetta Stones marked by “@” and boldface.

| IOL1 | | IOL2 | | IOL3 | | IOL4 | | IOL5 | |
|------|-----------------|------|-----------------|------|-----------------|------|-----------------|-------|-----------------|
| #1 | 52:33:15 | @1 | 70:17:13 | @1 | 58:26:16 | @1 | 61:24:16 | #1 | 49:13:38 |
| #2 | 24:03:73 | #2 | 57:24:20 | #2 | 54:10:36 | #2 | 27:37:35 | #2 | 66:28:07 |
| #3 | 56:16:28 | #3 | 09:09:83 | @3 | 48:16:36 | #3 | 24:25:51 | #3 | 10:11:79 |
| @4 | 67:27:06 | #4 | 61:17:22 | @4 | 40:44:16 | @4 | 18:49:33 | #4 | 15:02:84 |
| #5 | 73:06:21 | #5 | 04:36:60 | #5 | 22:06:72 | @5 | 27:27:45 | #5 | 52:30:18 |
| IOL6 | | IOL7 | | IOL8 | | IOL9 | | IOL10 | |
| #1 | 55:34:10 | @1 | 67:23:09 | #1 | 76:18:06 | @1 | 18:38:43 | @1 | 13:31:56 |
| #2 | 19:16:64 | #2 | 45:27:28 | #2 | 32:08:60 | #2 | 23:38:38 | @2 | 36:13:51 |
| #3 | 60:27:13 | #3 | 20:05:76 | #3 | 69:23:08 | @3 | 61:22:17 | #3 | 17:24:60 |
| @4 | 51:33:16 | #4 | 03:05:92 | #4 | 45:07:47 | #4 | 13:13:74 | @4 | 17:47:36 |
| @5 | 30:39:31 | @5 | 30:42:28 | #5 | 27:35:37 | #5 | 06:22:72 | @5 | 37:47:16 |

Table 2. IOL1–10: ratio of high:middle:low scorers for each problem.

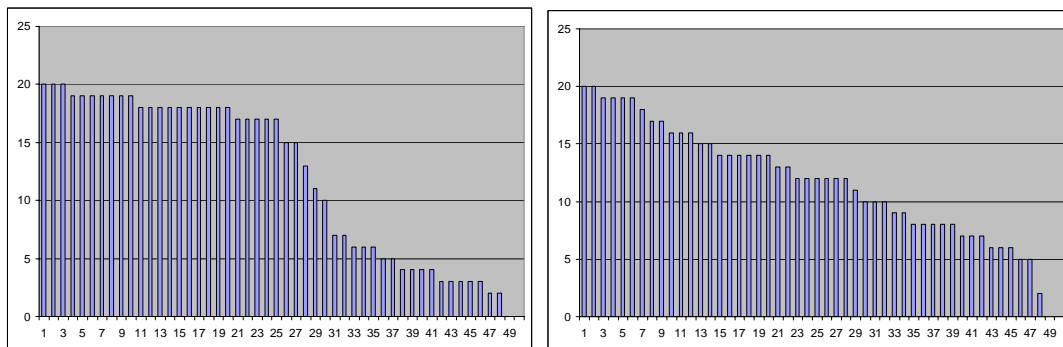


Figure 1. The distribution of scores for two IOL problems.

The reason for this is the inclusion of multiple phenomena. Finding a side phenomenon and using it in the assignments carries a portion of the points, even though the main phenomenon may not have been discovered; Rosetta Stones are almost never “all-or-none” problems.

This is an important factor when considering the complexity of a problem within a problem set. “All-or-none” problems create the danger of anomalies in the end results of a competition (contestants scoring lower than their abilities due to random factors) and of discouraging less experienced participants.

Figure 1 further illustrates the difference in the distribution of scores. The problem on the left was a Chaos and Order on Lango.¹⁰ The one on the right was a classical Rosetta Stone on Yoruba.¹¹ The contestants’ average scores for the two problems were extremely close (11.98 and 11.56, respectively), but the Rosetta Stone took less insight, though more work on the whole, and this made for a smoother ranking.

7 Co-authoring a problem

It is not uncommon for a problem to be authored by two people. This sometimes means that the authorship has been divided chronologically: one person wrote a problem that the other thoroughly revised (to an extent thought to amount to co-authorship). Or else they may have worked jointly on creating the problem from the idea and the original data, possibly dividing among themselves the tasks, which include (in the case of a Rosetta Stone problem) selecting side phenomena, selecting material, constructing the table of usable forms, and constructing assignments. This modularity of the authoring process greatly facilitates co-authoring.

¹⁰ IOL3, problem 2 (Ksenia Gilyarova).

¹¹ IOL3, problem 4 (Ivan Derzhanski).

The common-sense iterative procedure for collaborative work when the two authors are not physically present in one place and cannot hold discussions while constructing the problem (each author in turn making changes and sending them over to the other to review) has some specifics in this case—changes must be explicitly accounted for, so that one does not by accident remove an example that the other thought necessary for illustrating a phenomenon.

8 Problem approval process

When the problem is finished it has to go through an approval process before being used in competitions. The formality of the process depends on the contests and the rules of problem committee. Some steps are:

- Idea validation: not every language phenomenon can be used for a linguistic problem at all.
- Pretesting/beta-testing: no amount of reasoning can substitute for people’s actual attempts to solve the problem, as a way of evaluating its difficulty and verifying its unambiguity.
- Assessing suitability for a given competition: a problem which is good in principle may be deemed unsuitable for a specific place or time. It may be too hard for an introductory round (or too easy for an advanced one), or it may contain phenomena that are very similar to ones used at a recent instalment of the same contest. If judged usable in principle but not momentarily, the problem may be put in a repository, where it is saved for future competitions.

This is a generic process that applies to all problem types, but some details are relevant specifically to Rosetta Stones.

8.1 Beta-testing

The process of beta-testing is the most important step before finalising a problem. A solution (or non-solution) usually leads to modifications in the problem and a further version is released, which should be beta-tested again. Modifications carried out as result of test solutions include:

- blocking parasitic solutions;
- showing more examples of an under-represented phenomenon;
- removing or adding side phenomena in order to reduce or augment the complexity of the problem;
- amending assignments in order to prevent them from being doable by analogy with the data;
- clarifying assignments that are hard to understand.

The process continues until no more changes are required.

8.2 Translating problems

Linguistic problems have always been translated for a variety of purposes, from using problems made in one country at contests (or in lectures) in another through accommodating overseas guest competitors to running international contests. This is not always easy. The new Solverese may make some things less or more evident, it may share the main phenomenon with the featured language (which means that the problem ceases being a genuine problem in translation), or the solution may depend on recalling some facts of the original Solverese that are lost in translation. Often the choice is between an awkward wording and a problem that is not functionally equivalent to the original.¹² Which is preferable may depend on the occasion: an old foreign problem is worth translating and reusing only if it sounds natural in translation; it may be acceptable for guest participants in another country's national contest to be at a slight disadvantage, but at an international competition equality is crucial.

Rosetta Stone problems often involve phrases or sentences, which means that in principle they contain more opportunities for untranslatability.

¹² This may happen, for instance, when translating glosses of sentences from Russian, which lacks articles, into a language that has them: if some nouns become definite and others indefinite, this will create a new opposition that the solver will have to consider.

In light of this it may seem a paradox that they are so frequent at IOL. Yet it appears that problems of other types, and especially unclassifiable problems, are harder to make work equally well in several languages than Rosetta Stones are. The type wins out thanks to its familiarity.

9 Conclusion

Composing linguistic problems is a challenging task, which involves many steps and considerations. A good problem is unambiguous, contains well-presented interesting phenomena, does not have parasitic solutions and has prevailing linguistic complexity.

A Rosetta Stone problem enables authors to illustrate the most interesting linguistic phenomena, allows for smoothly distributed and fine-grained results and, as described above, has a relatively straightforward and well-defined composition workflow. No surprise, then, that it has become an expected feature at every linguistic contest.

References

- Ivan A. Derzhanski. 2007. Mathematics in Linguistic Problems. In: L. Dimitrova and L. Pavlov (eds.), *Mathematical and Computational Linguistics. Jubilee International Conference, 6 July 2007*, Sofia, 49–52.
- Ivan A. Derzhanski. 2009. *Linguistic Magic and Mystery*. Union of Bulgarian Mathematicians, Sofia.
- Ivan A. Derzhanski and Thomas E. Payne. 2009. The Linguistics Olympiads: Academic competitions in linguistics for secondary school students. In: K. Denham and A. Lobeck (eds.), *Linguistics at School: Language Awareness in Primary and Secondary Education*, Cambridge University Press, Cambridge, UK, 213–226.
- Yakov G. Testeleets. 1994. Linguistic Problems and the “Presumption of the Author’s Mildness”. In: V.I. Belikov, E.V. Muravenko and N.V. Pertsov (resp. eds.), *Sign: A collection of papers on linguistics, semiotics and poetics in memoriam of A.N. Zhurinsky* (in Russian), Russian Educational Centre, Moscow, 213–224.
- Yakov G. Testeleets. 2001. *An Introduction to General Syntax* (in Russian). Russian State University for the Humanities, Moscow.
- Alfred N. Zhurinsky. 1993. *Word, Letter, Number: A discussion of self-sufficient linguistic problems with an analysis of a hundred samples of the genre* (in Russian). Nauka, Moscow.