

Correspondence Seminar: Bringing Linguistics to High Schools

Matěj Korvas

Charles University in Prague
Faculty of Mathematics and Physics
Institute of Applied and Formal Linguistics
korvas@ufal.mff.cuni.cz

Vojtěch Diatka

Charles University in Prague
Faculty of Arts
Department of Linguistics
Vojta.Diatka@gmail.com

Abstract

We present the concept of a correspondence seminar as a way to complement and support one-time contests, especially olympiads. We evaluate specific payoffs of this way of teaching linguistics, and compare its nature to that of a linguistics olympiad. We believe that the correspondence seminar is a great way to introduce talented high school students to linguistics.

1 Introduction

At high schools in the Czech Republic, linguistics is taught only marginally or not at all. Students talented in linguistics thus tend to focus their talent on other areas, not even knowing what linguistics is like. We have struggled to change the state of affairs, and provide an alternative to the state system in delivering linguistic education to high school students.

Up until recently, we exposed the high school students to linguistics only through a correspondence seminar. By the term *correspondence seminar*, we mean a form of voluntary education where students and teachers exchange assignments and their solutions by postal correspondence (or more recently, via electronic communication in a similar way). This concept is described in more detail in Section 2. However, as IOL¹ (International Linguistics Olympiad; Radev et al. (2008)) came to our attention, we learned that despite the strong Czech linguistic tradition (Vachek and Dušková, 1983), there was no contest organised to select the Czech team for IOL. Hence we started the Czech Linguistic Olympiad (ČLO)² last year, and we have since observed some notable differences

in the nature of the two formats, which we shall summarise in this paper.

We will start by giving a brief overview of the history of correspondence seminars, including Pralinka, the seminar in linguistics. In fact, correspondence seminar is not a new concept. The oldest contest based on postal correspondence, to our knowledge, is the Hungarian High School Mathematics and Physics Journal.³ It dates back to as early as 1894, and with two interruptions, it survived up to the present. Thanks to the fact that it is translated into English, it is open to international audiences.

To the best of our knowledge, most correspondence seminars are organised in the area of former Czechoslovakia. The Slovak seminars include KMS⁴ (mathematics), FKS⁵ (physics), and STROM⁶ (mathematics). The last mentioned one claims to have the longest tradition in the area of former Czechoslovakia, having been established in 1976. Correspondence seminars organised in the Czech Republic include MKS⁷ (mathematics; founded 1981), FYKOS⁸ (physics; 1986), and KSICHT⁹ (chemistry; 2002; cf. Řezanka et al. (2012)). The seminars mentioned above have grown very popular – they commonly have several hundred participants each year.

Our correspondence seminar in linguistics is called Pralinka.¹⁰ It was founded in 2008 and has about 14 participants each year, this low number being one significant difference to olympiads. During the five years, over 100 linguistic problems

¹<http://www.ioling.org>

²<http://lingol.cz>

³<http://www.komal.hu/info/bemutakozas.e.shtml>

⁴<http://www.kms.sk>

⁵<http://fks.sk/english/english.php>

⁶<http://seminar.strom.sk/>

⁷<http://mks.mff.cuni.cz>

⁸<http://fykos.org/>

⁹<http://ksicht.natur.cuni.cz/o-ksichtu>

¹⁰<http://ufal.mff.cuni.cz/pralinka/english.php>

have been created and published in Pralinka.

The contents and workings of the correspondence seminar are explained in detail in the following section. In Section 3, we point out differences between the correspondence seminar and an olympiad. In Section 4, we show sample problems created for Pralinka, and conclude with plans for the future in Section 5.

2 The Concept of a Correspondence Seminar

We will present the concept of a correspondence seminar on the concrete example of Pralinka. In Pralinka, we publish four issues featuring various linguistics problems each year. Students are supposed to solve as many problems as they can, and they have about six weeks to send us their solutions. We mark and comment on the solutions before sending them back to the participants. In a few weeks' time, a new issue is put together from new problem specifications and authors' solutions to previous problems. The specifications of the individual problems are linked together with a story, partly to provide motivation for the problems, partly to make the booklet more attractive for the reader.

The first issue is sent with other faculty propagation materials, including other correspondence seminars, in the printed form to high school teachers countrywide. This is the only occasion when Pralinka uses traditional post. Every issue is published on our web page as a PDF, and participants' solutions and their corrections are submitted to an integrated system again as document files. An appropriately formatted PDF of each issue is provided so that whoever is interested, can easily print out the booklet at their site. Apart from the system for collecting problem submissions and their corrections, another online interface of Pralinka is its Facebook page where every new issue is announced and participants can discuss with the organisers.

An essential motivation for the participants are points we give them for their solutions. Seminars with a higher number of participants use the ranking of participants to select the ones eligible for the seminar summer school. Pralinka also has a short summer school each year, but we have no need to cut the number of its participants. Still, we reward the best ranking ones for their efforts with a prize.

Problems we publish in Pralinka can be divided

into three classes: single problems, thematic problems and running tasks. *Single problems* are one-off tasks that typically include little or no theory. They usually require the students to discover a pattern in the provided linguistic data, or expand on a given topic.

Thematic problems form a completely different genre. Their constitutive feature is that they go deep into the topic. Every year, a different topic is chosen to be investigated by the students under the supervision of the organisers. Theory for the topic is extended in each issue of the seminar, and a very open formulation of a problem to solve is given. The problem specification in the next issue is largely determined by the students' contributions. Students are thus introduced to the selected branch of linguistics step by step, both theoretically and practically. Thematic problems in the history of Pralinka examined topics like meaning of words in context or verbal aspects and their use.

Lastly, *running tasks* or *series* are similar to thematic problems in many respects. They are on a selected topic each year, and gradually build up the theory. In contrast to thematic problems, assignments in each instance of the running task are precisely specified and solvers' answers do not influence the future direction of the series. Some running tasks explore phenomena from different layers of linguistic description using an artificial language as the subject, others have the form of a textbook text split into chapters, providing exercises for each chapter. The latter kind covered topics such as language universals, Arabic (an introduction to the language) or semantics.

An important feature of the correspondence seminar is the individual attitude to students and their solutions. Correcting solutions does not consist only of assigning the appropriate number of points. More important is the feedback in the form of advice and questions related to the contestant's own text.

3 Comparison to Olympiads

We now turn to the comparison of a correspondence seminar and an olympiad as two alternative ways of promoting linguistics among gifted high school students. We will discuss the following aspects: time required for solving the problems, nature of the problems, attitude to linguistic knowledge, use of external information sources, and at-

type	Plka 10/11	Plka 12/13	ČLO 12/13
total	22 (100%)	18 (100%)	14 (100%)
seg+al	5 (23%)	7 (39%)	11 (79%)
theory	15 (68%)	7 (39%)	0 (0%)
open	16 (73%)	10 (56%)	4 (29%)

Table 1: Types of tasks in a sample from ČLO and Pralinka

type	Plka 10/11	Plka 12/13	ČLO 12/13
total	114 (100%)	119 (100%)	2393 (100%)
seg+al	23 (20%)	55 (46%)	2268 (95%)
theory	58 (51%)	43 (36%)	0 (0%)
open	86 (75%)	79 (66%)	665 (28%)

Table 2: Number of students that attempted solving different types of problems in a sample from ČLO and Pralinka

tractiveness to students.

Linguistic olympiads are generally one-time events whose participants are given just a few hours to solve a number of problems. They are thus motivated to quickly discover just as many features of the problem as needed to find answers for the questions posed. It is likely that solving such a problem involves just the short term memory, and the related ideas are much easier to forget. In contrast, participants of a correspondence seminar have lots of time to think each problem over, therefore, firstly, the problems need to provide enough food for thought, and secondly, this leads the solvers to internalise the ideas behind the problem much better.

For typically large numbers of participants of olympiads, especially in school rounds, olympiads need to have a clear grading scheme, hence closed-ended questions are the best suited. In contrast, problems in a correspondence seminar can be (and, in Pralinka, they often are) open-ended. This again supports deeper thinking about the problem.

The different composition of Pralinka versus ČLO in terms of problem types is quantified in Table 1. We counted problems from the 2010/2011 (the last school year before we launched ČLO) and 2012/2013 (this year; one issue could not yet be included) editions of Pralinka and this year’s ČLO. We assessed for each problem whether:

1. it is solved by applying the common pattern of establishing a segmentation of the linguistic data (e.g. words into morphemes, Chinese characters into two parts) and then aligning

the corresponding segments (row “seg+al” in the table);

2. theory is explained as part of the problem (“theory”);
3. an open-ended question is posed (“open”).

Table 2 follows the same layout but lists counts of students that attempted to solve each problem.¹¹ The numbers justify our claim that problems in Pralinka are more often open-ended than those in the olympiad. They also show that there is much more space for presenting theory as part of the problems in Pralinka, which we expand on in the following paragraph. Another fact illustrated by the numbers is that we adjusted Pralinka problems to be more similar to olympiad problems when ČLO was founded, in order to prepare Pralinka solvers for the olympiad. Lastly, the difference in the number of attempted solutions between Pralinka and ČLO is huge, as evident from Table 2. We try to explain this fact in the paragraphs below.

Another important difference regards the amount of linguistic knowledge presented *together with* the problems and required to solve them. The olympiad is primarily concerned with testing the contestant’s skills in analysing an unknown language, often their analytical thinking in general. On the other hand, the correspondence seminar puts stress on teaching not only skills, but also knowledge, in order to widen contestants’ horizons. To this end, problems published in Pralinka often comprise two parts: a theoretical one and a practical one, the latter part helping the solver practise immediately what was expounded in the former part.

The two formats differ also with respect to the approach to various *external sources* of knowledge. Olympiads strictly forbid using them, whereas in the correspondence seminar, their use is welcome. O. Šteffl, a prominent Czech education specialist, claims that “... the accessibility of information has dramatically changed. If I type ‘coelenterate’ into Google, what appears in a few seconds are BBC documentaries, pictures, explanations, curiosities, and I can search for context and links in this topic.”¹² Situation is the same for linguistic knowledge. There is a vast amount of

¹¹Entries in Table 2 are measured in student-problems.

¹²source (original in Czech): respekt.ihned.cz/c1-55775590-jsme-posedli-selekci-det

linguistic information accessible on the Internet. In Pralinka, even though we usually aim to create problems that cannot be simply solved by consulting Wikipedia or other sources, we are pleased to hear when our participants bother to use on-line resources, or even borrow a grammar book of a language in order to understand the topic more deeply, as provoking students to study on their own is one of our goals.

Finally, it is remarkable how few students get involved in the correspondence seminar, compared to the olympiad. The olympiad started only last year and it already has an order of magnitude more participants. This can be attributed to a simple fact that general awareness of linguistics as an interesting discipline among Czech high school students is very poor compared to mathematics or physics. Students thus do not show active interest in linguistics, although they get involved once their teachers give them linguistic problems at school during the olympiad school round.

We believe that participants of Pralinka are generally more interested in linguistics than participants of ČLO, and hence are more likely to enroll in a linguistics study programme at the university and be successful in it. Unfortunately, data we could use to test this hypothesis is not collected yet, although currently there is an effort of the Faculty of Mathematics and Physics to quantify the effectivity of its propagation activities including Pralinka.

4 Sample Problems

In this section, we present two examples of problems that appeared in Pralinka, which we think are particularly suitable for a correspondence seminar.

4.1 Labovian cups

In this problem, we motivate the participants to replicate Labov's famous experiment with tableware (Labov, 1973). General introduction to the topic of categories which do not have clear-cut boundaries is given through a simple dialogue led by two characters, rather than a technical exposition. Participants are then presented with tasks connected to two pictures of tableware (see Figure 1).

Tasks were as follows:

1. Do you think that the content of a container would influence whether it is called *cup*,

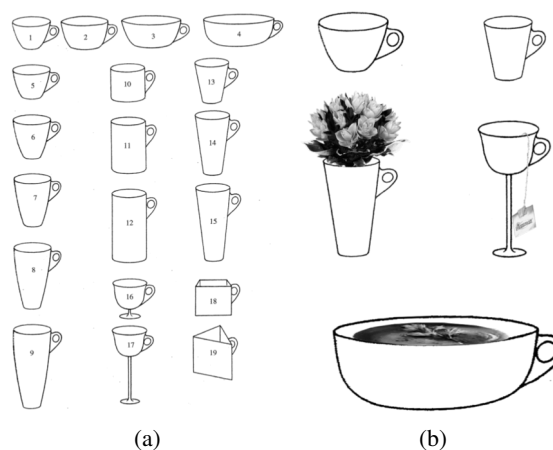


Figure 1: Labov's tableware experiment in Pralinka

bowl, or *vase*? We could fill it with flowers, tea or mashed potatoes.

2. Do your own experiment. Print out the pictures (those shown in Figure 1) and ask at least three people to name the objects. Show them first the container without contents, and then the same container with contents. Make sure that you show only one container at a time. Instruction could be that simple: "Tell me what you see." Write down the answers and expand on them.

4.2 Word alignment

In this problem, we present the concept of word alignment as used in machine translation (Koehn, 2007, pp. 113–124), and ask the students to elaborate on possible configurations in word alignment tables.

The task is motivated by the main character of Pralinka constructing a dictionary, trying to capture all translation options for every phrase.

This problem is clearly not suitable for olympiads, whereas it fits nicely the format of a correspondence seminar. Let us now comment on how well some students can cope with such problems, using two quite different kinds of analysis.

One participant performed a principled analysis of the possible configurations of alignment points, distinguishing cases with a single alignment point in the row and column, multiple points in a row or column, and multiple points in the same row *and* column. He illustrates his classification using examples including those shown in Figure 2. The first two examples in Figure 2 (the first one meaning "my country") are based on

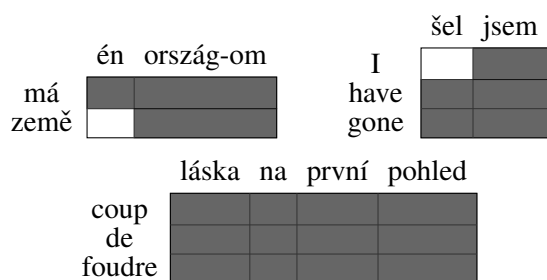


Figure 2: Examples of alignments from Solver 1

different ways grammatical categories are marked in different languages, including the first person and possessive markers in the first example, and first person and the tense in the second example. The third example (“love at first sight”) shows two phrases that are translation of each other but cannot be analysed into smaller units that would also translate one to the other.

Another participant focuses in her solution on the most interesting cases, showing typical properties of different languages, such as the tendency towards analytic or synthetic forms, and different word order. Even though she does not come up with a classification system, she gives a comprehensive overview of particular interesting examples. A few of her alignment examples are shown in Figure 3. The first example again demonstrates

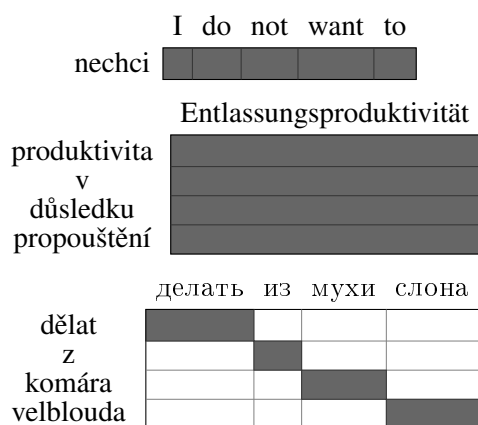


Figure 3: Examples of alignments from Solver 2

(a more extreme) difference in expressing grammatical categories in a fleective (Czech) and an isolating (English) language. The second example (“productivity resulting from layoffs”) illustrates the same meaning being expressed using multiple words in Czech, as compared to the synthetic German. The last example (“make a mountain out of a molehill”) is parallel to the example with *coup de foudre*, but the phrases decompose in this case,

even though they result in the *mosquito–mouse* and *camel–elephant* translation pairs.

5 Future Work

It is our ambition to make Pralinka international, either by organising the summer school jointly for Czech and foreign students, or by translating our problems into English and inviting foreign students to solve them. However, the latter could meet with a larger response than the existing organising team can handle, thus we greatly welcome any helpers before that transition is made, both to participate in the organisation, and help in promoting the seminar.

Acknowledgments

We thank Lěra Ivanova and Adam Pospíšil for the consent to cite their problem solutions, and Ondřej Dušek, Lukáš Žilka and the anonymous reviewers for their corrections and helpful comments. Pralinka has been supported by the Faculty of Mathematics and Physics of the Charles University, ČLO has been supported jointly by the Faculty of Mathematics and Physics and the Faculty of Arts of the Charles University.

References

- Philipp Koehn. 2007. *Statistical Machine Translation*. Univ. Pr., Cambridge.
- William Labov. 1973. The boundaries of words and their meanings. In Bailey and R. W. Shuy, editors, *New Ways of analyzing variation in English*, pages 340–373. Washington, D.C.: Georgetown University Press.
- Dragomir R. Radev, Lori S. Levin, and Thomas E. Payne. 2008. The North American Computational Linguistics Olympiad (NACLO). In *Proceedings of the Third Workshop on Issues in Teaching Computational Linguistics*, TeachCL ’08, pages 87–96, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Michal Řezanka, Pavel Řezanka, Luděk Míka, Pavla Perlíková, and Karel Berka. 2012. Korespondenční seminář inspirovaný chemickou tematikou (KSICHT) [Correspondence seminar inspired by chemistry topics (KSICHT)]. In *Chemické listy*, volume 106, page 319–322, Prague, Czech Republic, April.
- Josef Vachek and Libuše Dušková. 1983. *Praguiana: some basic and less known aspects of the Prague Linguistic School*. Number v. 12 in *Linguistic & literary studies in Eastern Europe (LLSEE)*. J.Benjamins, Amsterdam ; Philadelphia.