Integration of computer-aided language learning into formal university-level L2 instruction

Stoyanova, Nataliya

2021

Stoyanova , N , Hou , J , Kopotev , M & Yangarber , R 2021 , ' Integration of computer-aided language learning into formal university-level L2 instruction ' , L'analisi linguistica e letteraria , vol. 29 , no. 3 , pp. 117-126 . < https://www.analisilinguisticaeletteraria.eu/index.php/ojs/article/view/453/389 >

http://hdl.handle.net/10138/341350

unspecified
publishedVersion

ISSN 1122 - 1917

L'ANALISI
LINGUISTICA E LETTERARIA

FACOLTÀ DI SCIENZE LINGUISTICHE E LETTERATURE STRANIERE
UNIVERSITÀ CATTOLICA DEL SACRO CUORE

3

ANNO XXIX 2021

EDUCATT - UNIVERSITÀ CATTOLICA DEL SACRO CUORE

L'ANALISI
LINGUISTICA E LETTERARIA

FACOLTÀ DI SCIENZE LINGUISTICHE
E LETTERATURE STRANIERE

UNIVERSITÀ CATTOLICA DEL SACRO CUORE

3

ANNO XXIX 2021

PUBBLICAZIONE QUADRIMESTRALE

L'ANALISI LINGUISTICA E LETTERARIA
Facoltà di Scienze Linguistiche e Letterature straniere
Università Cattolica del Sacro Cuore
Anno XXIX - 3/2021
ISSN 1122-1917
ISBN 978-88-9335-906-1

Questo volume è stato stampato nel mese di dicembre 2021
presso la Litografia Solari - Peschiera Borromeo (Milano)

Indice

L'analisi linguistica e letteraria xxix (2021) 117-126

Integration of computer-aided language learning into formal university-level L2 instruction

* Nataliya Stoyanova, ** Jue Hou, *** Mikhail Kopotev,
** Roman Yangarber
* Università Cattolica del Sacro Cuore, ** University of Helsinki,
*** Higher School of Economics, St. Petersburg
nataliya.stoyanova@unicatt.it, jue.hou@helsinki.fi,
mkopotev@hse.ru, roman.yangarber@helsinki.fi

This paper presents our experience from pilot studies on integration of intelligent learning and tutoring tools into official curricula for foreign/second-language (L2) learning. We report specifically on initial studies with learners of Russian as a second language at major universities in Italy and in Finland. An important challenge in both of these educational situations is the heterogeneous nature of the student contingent, including the presence of a sizable proportion of 'heritage' learners. Furthermore, the groups are often very large, which motivates the integration of an ICALL system. We describe the first integration attempt, an analysis of the emerging aspects and problems, and the design of a new experiment, which is on-going and takes into account the lessons learned. To the best of our knowledge, this is the first report on large-scale ICALL studies involving substantial numbers of 'high-stakes' learners of Russian at the intermediate-to-advanced levels – i.e., learners beyond the elementary level.

Keywords: intelligent computer-aided language learning (ICALL), Russian language, adaptive testing, learning analytics, distance learning

1. *Introduction*[1]

We report on initial results of our efforts to integrate the ICALL system Revita into university-level L2 education. ICALL – intelligent computer-aided language learning – moves beyond 'standard' CALL, in which computers are used to support and enhance the learning process in any capacity, e.g., as a vehicle for delivery of learning content, storing results, etc. In ICALL, the system uses various AI methodologies to go much further – specifically to incorporate continual 'assessment', adaptivity and 'personalization', as described below.

The key factors that distinguish our experiments from those previously reported[2] are:

[1] The authors worked in close collaboration and made equal contributions; in the paper, N. Stoyanova contributed to 2.1, 3.1 and 5, M. Kopotev to 2.2 and 3.2, J. Hou to 4, R. Yangarber to 1.

[2] A. Katinskaia – J. Nouri – R. Yangarber, *Revita: a language-learning platform at the intersection of ITS and CALL*, LREC: 11th International Conference on Language Resources and Evaluation, Miyazaki, Japan, 2018; V. Slavuj – B. Kovačić – I. Jugo, *Intelligent tutoring systems for language learning*, MIPRO: 38th International

– focus on 'high-stakes' learners: e.g., incorporated into actual curricula at universities, high-schools, and other formal learning environments – as opposed to working only in experimental, research environments, or involving 'casual' learners;
– focus on learners at the intermediate-to-advanced levels: i.e., students at least at level A1 on the CEFR scale and beyond – as opposed to targeting learners at the elementary level,
– focus on L2 languages other than English;
– ICALL approaches that use intelligence: e.g., continual assessment, adaptive and personalized selection of material – beyond older CALL approaches, which might also include presentation of 'canned' (fixed, pre-selected) learning materials.

To the best of our knowledge, this is the first published report on experiments having all four of these characteristics.

This work is based on the language-learning system Revita (available at, revita.cs.helsinki.fi), which helps students learn languages by automatically creating exercises from arbitrary text materials, chosen by students or teachers[3]. The advantage of learning from authentic texts is that the texts can be chosen according to the learner's interest (rather than pre-selected, prepared texts and exercises), which stimulates the learner to spend longer time working with the system. The presence of the pragmatic component in the text moves the mode of practice to a new level; the student learns by operating within a coherent narrative, which requires comprehension on a much broader scale – compared to the kinds of exercises typically offered by language textbooks. Thus, learning 'in context' is qualitatively different from learning from isolated, artificially constructed exercises.

After a student uploads a text, the same text can be used for practice multiple times, because each learning session with this text will be different. The exercises offered by the system to the student are new each time, since the system generates them automatically and presents them in a randomized fashion. The exercises are also linked to a knowledge base, which describes the essential linguistic 'concepts', which the students must master to achieve high levels of proficiency. The set of concepts is language-dependent, and built into the system with the help of experts in the didactics of the language. The system offers extensive exercises on all aspects of grammar, vocabulary, and orthography. We are working on including additional modes of exercises into the system, and extending the power of the existing modes. For example, some exercises are offered on aural comprehension, but these will require further development. Creating the exercises in Revita relies on a wide range of computational tools – natural language processing (NLP) components[4]. Further, methods from artificial intelligence are employed to enable the system to moni-

Convention on Information and Communication Technology, Electronics and Microelectronics, Opatija, Croatia, 2015.

[3] A. Katinskaia – R. Yangarber, *Digital cultural heritage and revitalization of endangered Finno-Ugric languages*, DHN: 3rd Conference on Digital Humanities in the Nordic Countries, Helsinki, Finland, 2018; A. Katinskaia – J. Nouri – R. Yangarber, *Revita: a language-learning platform.*

[4] A. Katinskaia – J. Nouri – R. Yangarber, *Revita: a system for language learning and supporting endangered languages,* NODALIDA: 21st Nordic Conference on Computational Linguistics. Joint 6th NLP4CALL workshop on NLP for Computer-Assisted Language Learning and 2nd Workshop on NLP for Research on Language Acquisition, Göteborg, Sweden, 2017.

tor each student's progress and assess his/her proficiency, to optimize the selection of exercises based on these assessments[5].

It is important to note that the results reported in this paper could not be achieved without the complementary inter-disciplinary combination of the areas of expertise of the participants – applied linguistics, language teaching, artificial intelligence and language technology.

2. *Integration of Revita into curricula*

2.1 Italy

In Italy, the initial attempt at integrating Revita into university-level L2 education was undertaken at the State University of Milan. We list several problems with which we contend in the context of L2 education at this university: the high diversity of students' backgrounds, and exceedingly large class sizes. While the majority of students begin learning Russian from the elementary level, approximately 15% of the students are 'heritage' learners, coming from families where Russian is spoken, and who immigrated from the ex-Soviet states, in the first or second generation. The heritage students also form a highly heterogeneous group – some students have an excellent level of competency, whereas others understand standard Russian well, but speak in the dialect of the place of their family's origin.

The number of students is highly variable, and requires the university staff to adjust their teaching methodology to the specific situation of each year. Thus, the situation of teaching Russian language in the 2018-2019 academic year, during the initial integration of Revita into the official curriculum, was as follows: bachelor's 1st year: 115 students; bachelor's 2nd year: 140 students; bachelor's 3rd year: 120 students; master's 1st year: 25 students; master's 2nd year: 5 students.

This high and unpredictable enrollment for the bachelor program creates challenges from a didactic point of view, making it difficult to assure a good level of linguistic preparation independently of the number of students. For this reason, in 2018-2019 in the State University of Milan, Revita was integrated in the official curriculum of Russian language in all years of the Bachelor's and Master's programmes.

2.2 Finland

University of Helsinki offers two subprograms in the Russian language that are intended for: a) L2 speakers with basic proficiency in Russian, and b) 'heritage' speakers of Russian.

[5] A. Huhta, *Diagnostic and formative assessment*, in *Handbook of Educational Linguistics*, B. Spolsky – F. Hult ed., Blackwell, Oxford 2008, pp. 469-482; J. Hou – M. Koppatz – J. Hoya Quecedo – M. Kopotev – N. Stoyanova – R. Yangarber, *Modeling language learning using specialized Elo ratings*, ACL: 56th Annual Meeting of the Association for Computational Linguistics. BEA: 14th Workshop on Innovative Use of NLP for Building Educational Applications, Florence, Italy, 2019. For a more detailed description of the system and the ideas that underlie its approach to didactics please see, A. Katinskaia – J. Nouri – R. Yangarber, *Revita: a language-learning platform*.

The latter group is by definition highly diverse in terms of their proficiency: some of the heritage students speak Russian at home, but do not demonstrate advanced language skills due to a lack of systematic schooling in Russian. Others graduated from or studied for a period in schools in Russia, which results in a very high level of proficiency[6]. Each year, about ten heritage students attend the course in Russian and use Revita. The attrition of students across the years is quite small, only one student has dropped out of the studies.

Since 2005, the Department has been running a progress test, based on a large bank of questions, called *Karttu*[7], which enables students studying Russian to monitor their own progress in the language. The teachers can also use the test results for planning instruction and for updating their teaching materials. Since 2015, Karttu tests have been integrated into Revita. In the current curriculum, all heritage students attending the course *Written and Oral Skills in Russian* are required to pass an initial test in Revita, which functions as a placement test. The results obtained at this stage become a starting point for creating student-centered individual curricula. This includes both reading a textbook and practicing inside Revita, which offers exercises personalized for each particular student. Revita is designed to monitor the student's progress and to try to adjust to the individual progress of learning. Furthermore, the data collected from the students allows the teacher to carefully monitor, to individually consult, and to easily carry out the final test for the students.

3. *Initial results*

3.1 Aspects emerging as important for application in Italy

The aspect of Revita, which the students reported as most important and stimulating, and which the teachers consider to be critically needed for very large groups of students, is the individualized approach. Each student can liberally choose the texts and stories, which become the basis for his/her personal grammar training. This choice valorizes their personality and involves the emotional aspect in the language study. This helps the students to create their personal idiolect in Russian L2, with the lexical strata and expressions of their interest – based on informal interviews with the learners conducted by the teachers, we find that the students are enthusiastic about this aspect of the approach. Moreover, they can track their progress with precise references to the various grammatical topics defined by the system.

The system currently covers 140 grammatical concepts, and this set is being continually extended by the system developers in accordance with the teachers' feedback. This set does

[6] M. Polinsky – O. Kagan, *Heritage languages: In the 'wild' and in the classroom*, "Language and Linguistics Compass", 1, 2007, 5, pp. 368–395.

[7] M. Kopotev, *Тест прогресса языковых навыков KARTTU: структура заданий*, in *Изучение и преподавание русского языка в Финляндии*, Златоуст, Санкт-Петербург 2010, pp. 331-340; M. Kopotev – A. Mustajoki, *Progress test on Russian language KARTTU*, LINDAT/CLARIN digital library, Praha, Czech Republic, 2014, http://hdl.handle.net/11372/LRT-755 (last accessed November 29, 2021).

not cover all aspects of grammar, and not all concepts are covered in equal depth, but we believe that it does provide a sufficient and growing foundation to support the students' out-of-classroom off-line development.

Another important practical advantage is the possibility of offering an efficient means of individualized support to the heritage learners/speakers: usually they have to start from the very beginning along with other students (foreign language learners), whereby they lose their initial 'privileged' status. With Revita, rather than 'forcing' their pace to match that of the entire group, they can practice and progress in accordance with their own, more advanced level.

In 2018-2019 both aspects – practice with authentic texts and progress tracking – were part of the official curriculum of the Department of Foreign Languages and Literatures, at the State University of Milan. Moreover, we introduced a separation between the practical and theoretical tracks: those who had studied Russian previously and were able to demonstrate a level superior to the minimum required by the end of the year – as measured by Revita – could join the following year's group for the practical part of lessons. The order of the theoretical exams remained the same:

– bachelor's 1st year: phonetics and basic morphology;
– bachelor's 2nd year: advanced morphology;
– bachelor's 3rd year: syntax;
– master's 1st year: pragmatics;
– master's 2nd year: stylistics.

This innovation resulted in more homogeneous groups, and encouraged and rewarded the personalized progress of the best students.

To offer a more consistent approach to language learning, especially with respect to practicing grammar within a coherent narrative – a novel possibility, which Revita gives to learners – the entire structure of practical lessons was changed. The 'pragmatic' aspect of language learning was emphasized and the 'traditional' phrase-translation approach was abandoned. Language teachers were trained to focus on the correspondence between Russian and Italian expressions on the pragmatic level during their practical classes. This active oral learning from real-world situations[8] triggers the mechanisms of 'implicit learning', and it matches well with the individual grammar training component of Revita. At the same time, the setting of the lessons, which compels the students to speak and react fluently in a natural conversational rhythm, is complemented by the timed features in Revita, where, for example, progress tests place a time limit (e.g., 15 seconds) for every question. Moreover, to be admitted to exams students had to do their individual reading practice, working with a large volume of authentic, non-adapted texts, which enhances their passive language proficiency: recognition of grammatical patterns and enrichment of vocabulary. Some of the students preferred using Revita also for this reading task with non-adapted texts.

[8] For the first 4 years we based our program on the *Assimil* courses: the first volume leads up to CEFR level B2 and the second up to C1, which we enrich with extensive additional materials.

In this scheme, there is no need for training on single phrases taken out of context and unrelated to anything. On the contrary, at all learning stages the students deal with materials which express the intent of the speaker, i.e., with pragmatically coherent materials:

i. oral practice with the texts chosen by the teacher;
ii. grammar practice with texts chosen according to individual interest in Revita;
iii. individual reading practice with a larger amount of text; and
iv. practicing to write essays.

This last point produced an excessive amount of work for the teachers, so one of our future goals is to integrate additional components into Revita, which will assess the proficiency of students in active text production, such as composition, or written story retelling tasks. However, these are more challenging tasks from the perspective of the state of the art in ICALL.

3.2 Aspects emerging as important for application in Finland

Heritage students demonstrate a high level of spoken proficiency, typically C1-C2. However, they are known to have stark gaps in their written skills and in formal communication. The typical heritage speaker is able to carry on a fluent conversation on everyday topics, but has many difficulties in giving a presentation, let alone creating an academic essay. In such situations, Revita helps to develop the writing skills based on more formal written texts, e.g., academic texts in the fields relevant to the students. These exercises improve both the vocabulary and grammatical patterns characteristic of academic texts.

A further issue that Revita addresses is the gaps in competency in specific grammatical topics. Such competency is typically acquired in school (by native speakers) or in advanced classes (by L2 students). Heritage speakers often miss either of these possibilities. As a result, they experience serious difficulties in those topics, which are less used in oral communication, e.g., numeral declension, constructions with numerals, participial and gerund formation, formal collocations, to mention a few.

Finally, the latest version of Revita offers exercises in orthography, which are motivated by a practical need: heritage learners typically demonstrate lower levels of literacy, as compared to both native and L2 learners; some are even unable to use the Cyrillic alphabet.

We are faced with balancing two complementary challenges: on the one hand, heavy study loads on the students, in terms of the amount of material. On the other hand, teachers must contend with heavy workloads due to large class sizes. The Revita approach will foster innovation in the ways of adapting the academic curriculum to meet both of these challenges.

4. *Emerging problems and solutions*

We are working to refine the learning and assessment features, based on the feedback from user studies. More accurate calculation of the students' proficiency level is needed, by considering the likelihood of (accidentally) guessing answers correctly, even when the student

does not know the correct answer[9]. In cloze exercises, the probability of a correct guess may be negligible, but in a multiple-choice exercise it is substantially higher. For example, if there are 4 choices, then the probability is .25. This can be taken into account in the calculation of the score.

The timing of test sessions poses challenges. The baseline, simplistic approach (as currently) may be to accept uniformly, e.g., 15 seconds per item as an appropriate timing condition, and to adjust the difficulty of all items accordingly. Alternatively, we may use more flexible conditions, where items are allowed longer times, but the student's time is taken into account in the computation of the student's score for each item[10]. Lastly, students may take the tests under different conditions. A student who takes frequent breaks (which are permitted by the system) works under less stressful conditions than one who does many exercises over an extensive period of time. All timing information is preserved in the database, and it can be used to refine the assessment of a student's competency.

Based on the data we collected from the original test scheme we observed several important problems: the difficulty of the overall test may not be suitable for every student, the length of the test (over 1.5 hours) may be exhausting, and the timing may be quite challenging for many students, especially those at lower levels of proficiency. These 'negative' lessons learned point to the need for a smarter testing scheme. Thus, we turn to adaptive testing.

We implemented adaptive testing based on Item Response Theory (IRT) – widely accepted in the education assessment community[11]. The idea of applying the IRT approach to testing in Revita can be summarized into the following key points:

1. Question difficulty: before testing begins, each question in the question pool is assigned a 'difficulty' score. This score is learned from a large amount of data collected from many students who answered many questions (previously, during 2018-2020).
2. Item selection: during the test, the IRT model defines the 'information' of each question with respect to a given user. This means that on each iteration the model will select the test question from the pool which will yield the greatest amount of information about the user.
3. Ability estimation: after the user answers the question, we re-estimate the user's ability based on his answers so far, and the difficulty of the questions.

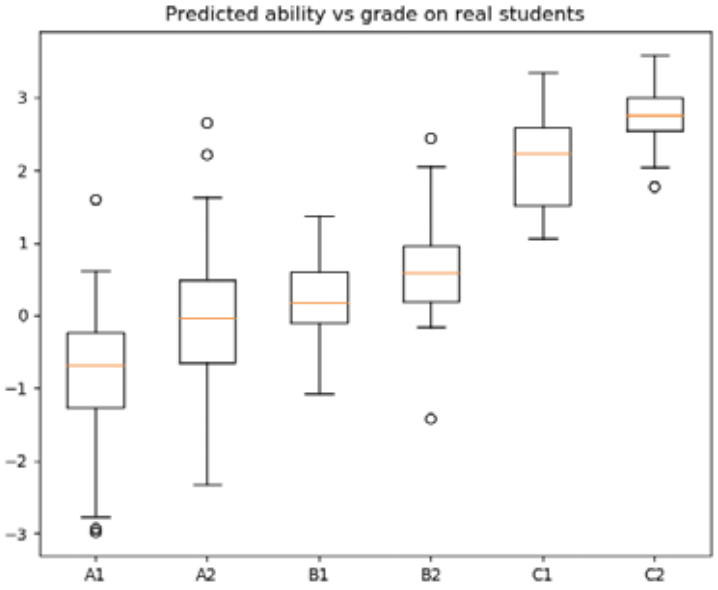
To train the IRT test model, we used data which was collected from our original testing scheme – the answers of many students to many questions over time. We ran several simulations to validate the resulting IRT model.

[9] R. Pelánek, *Applications of the Elo rating system in adaptive educational systems*, "Computers & Education", 98, 2016, pp. 169-179.

[10] S. Klinkenberg – M. Straatemeier – H.L. van der Maas, *Computer adaptive practice of maths ability using a new item response model for on the fly ability and difficulty estimation*, "Computers & Education", 57, 2011, 2, pp. 1813-1824.

[11] W.J. Van der Linden – R.K. Hambleton ed., *Handbook of modern item response theory*, Springer Science & Business Media, New York 2013.

Figure 1 - *Ability of students predicted by IRT-based model vs. CEFR level assigned to students by teachers*



The figure above shows a simulation of the IRT model, which was learned from actual student data collected from real students on the 'long' test. This is a special set of approximately 200 students – for whom the CEFR level was individually estimated by their teachers (subjectively, based on the teachers' knowledge of the students' overall skills). The X-axis shows the teachers' CEFR estimate, the Y-axis shows the student ability assigned by the adaptive test. Student abilities are first learned on a scale (arbitrarily) fixed to be between -5 and +5 (through calibration, this ability scale can be mapped to the CEFR scale).

In the simulation, we follow the same steps as in the 'real-life' adaptive test. We use the student's answer history (from earlier test sessions that we collected) and follow the steps toward estimating the student's ability by offering questions that contain the most information from among those questions that the student has actually answered. In other words, in the simulation, the next question is the most informative within the student's answer history – rather than the most informative overall (as would be done in a real test setting with a live student).

As we can see from Figure 1, the estimated ability of the students is strongly correlated with their CEFR levels as actually judged by the teachers. At the same time, in this simulation, the majority of the tests are completed within 55 questions or less – the test is complete when the IRT model has converged on an estimate of ability (i.e., it does not change over some predetermined number of iterations). That means that the test is quite short on average – this is important because it does not stress the student. This simulation shows that despite the ostensibly 'negative' experience where we encountered many problems with the initial testing scheme, we nevertheless can use the results from learners collected

during those initial trials to train an IRT model, which is able to give reliable estimation of ability after a much smaller number of questions in the adaptive test.

5. *New experiment*

The initial attempts to integrate Revita into formal university-level L2 instruction described in Sections 2 and 3 helped us identify many critical problems and shortcomings in the Revita approach, which needed to be addressed. We used this opportunity to collect feedback and comments from students and teachers on a continual basis. This experience provided many new insights into how the approach can be developed to support the needs of the students and teachers. As a result of analysing the lessons learned, many new features were added to the system, besides the adaptive testing, described above. These new improvements include:

- a new user interface (UI) based on the latest in Web technology for UI design, including a mobile interface;
- a user interface localized for several new languages, including Italian;
- personalized feedback to the learner in case the learner gives erroneous answers to grammatical exercises;
- support for learning about Russian word derivation;
- support for teachers, including selection of topics to be used in the students' practice sessions, and management of student groups;
- a special mode for reviewing exercises done with a text previously, including mistakes made and feedback received.

This implementation encouraged a new extensive, *longitudinal* experiment phase of integrating Revita at the university level. The experiment, starting in the fall semester of academic year 2021-2022, involves the following universities:

- Catholic University of Sacred Heart in Milan, Italy
- Milan State University, Italy
- Pushkin Institute, Moscow, Russia
- University of Helsinki, Finland
- Tampere University, Finland
- Asfendiyarov Kazakh National Medical University, Almaty, Kazakhstan.

The experiment is designed as follows. At each level the students are free to decide to join the 'experimental group' which uses Revita or the 'control group' which does not use it. During the semester students are tested at least twice: at the beginning and at the end. Testing can follow one of these schemes: i) the usual form – commonly used in each university program, ii) Revita's adaptive test, or iii) Revita's exhaustive test, customizable by the teachers, according to the specific requirements for the tested group.

Throughout the semester, both groups of students follow the regular curriculum. Revita's team supports the teachers of the experimental group by: i) offering preliminary workshops to discuss platform features, possible exercise types, and exercise settings for the

students, ii) monitoring student activity during the semester; if experimental students drop off, inquire about reasons, iii) developing of additional functions, as required by teachers.

Students from the control group are free to use Revita during the semester, and vice versa, students from the experimental group are free to opt out; after the final testing, we can explicitly check who used Revita during the semester, which will allow us to obtain more reliable statistics.

During the semester and after the final testing, we analyze the data collected from the experimental groups and provide detailed reports to the teachers about all skills measured in Revita. The specific content of the reports and the parameters of the analysis are agreed with the teachers before the start of the experiment and are clarified during the semester.

Comparison of the results of the experimental and control groups, which follow identical curriculum except for the use of Revita, will allow us to measure the effectiveness of the integration of Revita into formal university-level L2 instruction.

FACOLTÀ DI SCIENZE LINGUISTICHE E LETTERATURE STRANIERE

L'ANALISI LINGUISTICA E LETTERARIA

ANNO XXIX - 3/2021

EDUCatt - Ente per il Diritto allo Studio Universitario dell'Università Cattolica
Largo Gemelli 1, 20123 Milano - tel. 02.72342235 - fax 02.80.53.215
e-mail: editoriale.dsu@educatt.it (produzione)
librario.dsu@educatt.it (distribuzione)
redazione.all@unicatt.it (Redazione della Rivista)
web: www.educatt.it/libri/all

ISSN 1122 - 1917