Revita: a Language-learning Platform at the Intersection of ITS and CALL

Anisia Katinskaia, Javad Nouri, Roman Yangarber

University of Helsinki, Finland

first.last@cs.helsinki.fi

Abstract

This paper presents Revita, a Web-based platform for language learning—*beyond the beginner level*. We anchor the presentation in a survey, where we review the literature about recent advances in the fields of computer-aided language learning (CALL) and intelligent tutoring systems (ITS). We outline the established *desiderata* of CALL and ITS and discuss how Revita addresses (the majority of) the theoretical requirements of CALL and ITS. Finally, we claim that, to the best of our knowledge, Revita is currently the only platform for learning/tutoring beyond the beginner level, that is functional, freely-available and supports multiple languages.

Keywords: Computer-assisted language learning, CALL, intelligent tutoring systems, ITS, second language acquisition, SLA, foreign language acquisition, FLA, learner corpus.

1. Introduction

Over the recent years we observe a clear emerging trend toward flexible online language-learning tools that are accessible anywhere on demand. Despite the growing need for and popularity of such tools, most of the existing systems do not address the fundamental requirements of language learners and teachers.

In this paper, we demonstrate Revita, a freely available online platform, which has been designed to support language learning/tutoring *beyond the beginner level*. A variety of resources exist on the Web—various free and commercial applications—which support beginners, some with millions of users. However, once the learner has passed the beginner's level, and reached low-intermediate to advanced (LI-A) level (i.e., above A1 on the CEFR scale), resources available to her become drastically limited. As far as our research has shown, no systems today provide substantial support for LI-A learners in multiple languages.

Revita currently works with several languages—in various stages of development, ranging from initial, "beta" versions to fairly well developed ones. The languages include "big" languages—Finnish, Russian, Swedish[(β)], German[(β)], and Kazakh[(β)]—and several endangered minority languages, currently Komi-Zyrian, Udmurt, Meadow Mari[(β)], Erzya[(β)], North Saami[(β)], Komi-Permiak[(β)], and Sakha (Yakut)[(β)].[1] The system automatically generates a wide variety of randomized exercises targeted for the learner, based on arbitrary, *real* texts, which can be chosen and uploaded to the platform by the learner herself (or by her teacher). The system aims to adapt the level of exercises to every user depending on her level of competence, which it tries to estimate based on her answers to previously completed exercises.

Revita lies at the intersection of two established areas of research: intelligent tutoring systems (ITS) and computer-assisted language learning (CALL)—the project seeks intelligent solutions for language learning. On the other hand, Revita has the potential for enriching the language teaching process as well, because the platform can be used for collecting, mining and analyzing educational data.

The paper is organized as follows: Section 2. presents an overview of previous work in computational approaches to language learning. Section 3. describes the Revita system, a language-learning tool developed at the intersection of ICALL and ITS; Section 3.2. describes the main features of the system. Section 4. positions Revita in the field of Educational Data Mining (EDM), and Section 5. concludes with pointers for future work.

2. Prior work

The idea of using computers for language learning was introduced over 50 years ago. Computer-assisted language learning (CALL) is an active research area, which originated as a sub-area of computer-assisted instruction (CAI). CALL is broadly defined as "the search for and study of applications of the computer in language teaching and learning," (Levy, 1997). CALL includes a broad variety of technologies applied to language learning and teaching, in which the computer is accepted by researchers and by teachers as a *support* tool; it is not considered to be a *replacement* for the teacher. One of the early CALL systems, PLATO, (Hart, 1981) was created in the early 1970s.

Most of the major, popular CALL systems today are commercial, and are intended for learning/tutoring at the elementary/beginner level, viz., Duolingo, LinguaLeo, Babbel, and others. Those systems that claim to cater to advanced learners, typically offer a collection of learning materials and exercises, e.g., `https://www.deutsch-lernen.com/`. Crucially, most of these systems offer materials in the form of a closed, *fixed* set of exercises.

Exercises may be of several types. A type of exercise, known as a "*cloze*" (deletion) test,[2] is where a portion of the text has some of the words removed, and the learner is asked to recover the missing words. Clozes

---

[1]The system is online at `revita.cs.helsinki.fi`

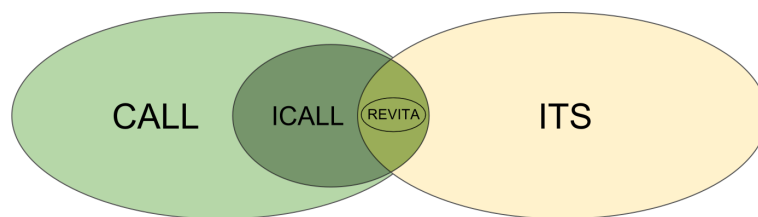[2]In the literature, first described in (Taylor, 1953).

Figure 1: Landscape of tools for language learning.

require an understanding of the context, semantics and syntax in order to fill in the blanks with the correct missing words. Alternatively, exercises may offer selection from a multiple-choice menu.

Thus, most of systems in the literature that we have reviewed, are based on the so-called *"canned"* approach: i.e., exercises that the learner receives are drawn from a pre-made, *fixed*, and therefore limited set—even if the set may be large and varied. The exercises are created in advance of the study session by human experts, and are based on fixed textual material, pre-determined by the designers of the exercises , rarely expand on a context bigger than one sentence. Overall, however, the existing systems are mainly targeted toward beginners. Since hand-crafting a substantial volume of exercises for more advanced users requires much greater resources, it is more difficult to provide sufficient coverage to support the intermediate/advanced learner, due to the growing cost-to-benefit ratio: the demands on the systems grow, while the market shrinks—many more learners/consumers of language learning services are at the beginner, elementary level, than at the LI-A levels. This ratio is a key factor that has suppressed wider emergence of CALL/ITS beyond the beginner levels.

On the other hand, offering a fixed, limited set of exercises is in conflict with the principles of *adaptability* of the learning process to the profile of the particular user. Applying a "one-size-fits-all" paradigm to students with different levels of competence goes against established pedagogical practices.

Some systems do track personal progress and the user's state of knowledge; however, they do not *adapt* future exercises to the user depending on her prior errors; the set of exercises is the same for everyone.[3]

Further, most systems provide poor feedback about about the answers given by the learners. In the extreme, some systems provide to the learner only a summary of how many of her answers were correct, without any more specific information to help correct errors and improve in the future.[4]

Many of the CALL systems that do attempt to address the needs of intermediate-to-advanced learners employ complex linguistic or grammatical terminology[5]. This may be undesirable in certain settings, e.g.,

in cases where the learners may have quite substantial competency in the language, yet lack proficiency in linguistic concepts and terminology. This is especially true, e.g., of *"heritage"* learners; these are learners who acquired passive competency in a language from hearing parents/relatives speaking in diasporic settings, or in the settings of minority languages having no official state status—but who have no formal training in it.

As CAI developed, the field of intelligent tutoring systems (ITS) emerged with somewhat different goals. Within ITS another role for the computer was accepted—specifically, the role of computer as *tutor*.

ITSs have been applied in various knowledge domains—mathematics, the sciences, business management, etc.—and focus on the ability to generalize and apply knowledge to specific tasks, and on dynamic adaptation depending on the user and the performed tasks. Initially, ITSs were envisioned as aiming to replicate the function of human teachers (Self, 1990; Shute, 1991)—not only to support learning, but also to attend to the learner's progress, (Self, 1998).

Decades of research have shown the effectiveness of ITS technology for learning. Students who used ITSs showed improvements in performance, compared to students who were exposed only to traditional classroom settings, (Kulik and Fletcher, 2016). Several systems, which were used by tens or hundreds of thousands of students annually, significantly improved learning performance: e.g., Cognitive Tutor (Oxman et al., 2014; Ostrow and Heffernan, 2014), ALEKS (Craig et al., 2013), ASSISTments (Koedinger et al., 2010).

Most ITS approaches found in practice are simple; e.g., they employ simple heuristics for assessment of user progress based on the number of correct answers given in succession (Heffernan and Heffernan, 2014), and they do not attempt to model the underlying conceptual domain. Several seminal approaches have been developed to provide highly refined assessment, based on theoretical psychometric principles—e.g., the Knowledge Space Theory (KST), (Doignon and Falmagne, 2012; Falmagne and Doignon, 2010; Falmagne et al., 1990)—and incorporated into large-scale, commercial science-tutoring systems, such as ALEKS (Craig et al., 2013).

In KST the learner's competency is modeled not as a scalar—e.g., as the A1-C2 "levels" on the CEFR scale—but rather as a position in a complex *knowledge space*, which is modeled as a graph containing many

---

[3]For example, https://learning.lengalia.com

[4]For example, http://icelandiconline.is/

[5]For example: learning.lengalia.com , icelandiconline.is , www.easypersian.com

(possibly millions of) nodes. Each node in the graph represents a possible knowledge state of a learner—encoding the set of concepts mastered in that state—and the possible paths in the graph indicate the possible paths toward acquiring full competency in the domain. More precisely, the learner's competency is modeled as a probability distribution over the nodes in the knowledge graph: the learner is most likely to be positioned at those nodes where probability mass is concentrated.

ITS, when applied specifically for language learning—and supported by other intelligent and/or adaptive methodologies, such as adaptive systems (AS), expert systems (ES), natural language processing (NLP), automatic speech recognition (ASR), etc.—defines the domain of intelligent CALL, or ICALL. The goal of ICALL is broadly defined as building advanced applications for language learning using NLP and language resources—corpora, lexicons, etc., (Volodina et al., 2014).

Although ITSs were seen as generally useful in ICALL, they were not in the focus of ITS research over the last two decades, because: they were not considered to be intelligent enough, were felt to be too "behavioral" in nature, and were too bound to "drill-and-kill" types of exercises, (Hubbard and Siskin, 2004); another important reason was the lack of mature technology (Bush, 2008; Mozgovoy, 2012).

Currently, ITSs are acknowledged as useful for language learning, and "able to address more than just simple grammar and vocabulary teaching/learning, but they need to be designed keeping in mind learners' and instructors' needs" (Bush, 2008). We found that most of the existing ITSs for language learning remain in the *prototype* stage, having been developed purely for research purposes. Our literature and online search has revealed only three systems which claim to be functional, and to have been in use in classroom settings:

- E-Tutor, for learning university-level German (Heift, 2008; Heift, 2010)
- TAGARELA, for learning university-level Portuguese (Amaral and Meurers, 2007; Amaral and Meurers, 2011)
- Robo-Sensei, for learning Japanese, mostly focusing on translation (Nagata, 2009)

The first two systems build adaptive learner models, and tailor feedback messages depending on individual performance. The third system has a single sequence of activities distributed over 24 lessons, which is fixed and identical for all users. All of these systems are monolingual—designed for a single language. Most importantly, at the time of this writing, a search on the Web shows that *none* of these systems is currently available for use.

At the three most recent International Conferences on ITS—over the last 6 years—only four papers about language learning have been presented, and none of them relate to developing any new ITSs. Thus, overall we observe a lack of papers about ITSs for language learning, a lack of assessment, and a lack of functioning systems.

Golonka et al. (2014) published a review of computational methodologies in language learning; the review mentions three evaluations of ITS for language learning (Nagata, 1993; Nagata, 1997; Dodigovic, 2007). Although all three showed that ITS are more effective than traditional tutoring, the author was critical of the *evaluation* of the new methodologies.

3. Revita system description

The Revita system, situated at the intersection of ICALL and ITS—see Figure 1—attempts to address the requirements of both, and aims to move beyond existing solutions, (Katinskaia et al., 2017).

3.1. Addressing current problems in CALL and ITS

We briefly review the desiderata of language-learning systems and the key problems in developing ITSs—the main observed pitfalls. Many of these have been brought to light in prior surveys, e.g., in (Slavuj et al., 2015). For each desideratum or problem we briefly mention how Revita satisfies the requirement, or how we intend to address it in the future.[6]

1. *Over-restricting the learning domain horizontally*: Horizontal restriction refers to the tendency to focus on a single linguistic skill. Revita offers a variety of practice modes for exercising reading, grammar and vocabulary skills, to cover a broad range of linguistic concepts and phenomena.[7]

2. *Over-restricting the learning domain vertically*: lack of support for learners at different levels of proficiency (especially CEFR levels A2–C2), typically focusing on the beginner levels. Revita targets the *low-intermediate to advanced* levels (LI-A).

3. *Active vs. passive learning*: i.e., focus on the user's skills in active production of language vs. passive absorption of material. In Revita, the focus is on active skills, achieved by eliciting unrestricted user input; the learner produces word forms in the context of the story.

4. *Learning materials should reflect real-world communicative situations*: We allow the learners/teachers to use arbitrary, *real-world* texts, rather than allowing only artificial texts pre-fabricated for instructional purposes.

5. *Providing learning materials that keep learners motivated* during the learning process. In large part related to item 4, above: if the exercises are detached from real-world, useful contexts, they become uninteresting, and lead to loss of motivation. Revita provides various types of exercises,

---

[6]See the operational platform at `revita.cs.helsinki.fi`

[7]In the future, we will add exercises in the aural modality, by using text-to-speech tools, which are quite well-developed for many of the target languages.

and entertaining, engaging modes of practice, including competition against an opponent.

6. *Reference to assessment scales*: Most ITSs do not link to a generalized framework of reference, such as the CEFR or ILR scale. Detailed internal assessment is a crucial part of Revita. At present, the system keeps track of the learner's performance on a large set of grammatical and lexical "concepts,"—which serves as an indicator of competency in the corresponding linguistic skills (e.g., "mastered the inflection of nouns of class X"). In future work, we plan to incorporate these concepts fully into the KST framework, and model the knowledge space in accordance with the theory (Doignon and Falmagne, 2012). The model will be learned automatically from observed learner responses (Schrepp, 1999), and the inferred (statistical) precedence relations among the linguistic concepts.

7. *Feedback messages*: should be personalized, and should guide the learner toward finding the correct answer. Rather than responding to the user simply with "correct vs. incorrect", the system should point to the source of the problem and offer a chance to attempt the exercise again using new clues returned by the system. Revita returns immediate feedback to the learner—about grammatical errors, incorrect choices in multiple-choice exercises, or wrong translations in flash-card vocabulary exercises. We plan to extend the feedback mechanisms to be more personalized by analyzing the history of the learner's responses.

8. *Mimicking a language teacher*: ITS is envisioned as being capable to mimicking (and complement) successfully the functions of a language teacher. In order to do so, the ITS must be designed specifically for the purpose of language learning—i.e., based on sound pedagogy and language-teaching methodology, supported by foreign/second language acquisition theory (FLA/SLA). This theory must form the basis for designing intelligent CALL systems. In Revita we work with SLA/FLA experts, to assure adherence to this requirement.

Lastly (also related to problems 5, and 7) when the feedback that the system provides is limited to "right/wrong," the system gives the impression of not caring about the learner's development, and failing to fulfil the requirements of a good tutor.

In summary, CALL systems are seen in the literature as widely available, but not sufficiently "intelligent," nor adaptive. ITSs in the language-learning domain are intelligent, but not available to end-users, or not free of charge. Further, all ITSs that we found are monolingual; ICALL systems are multilingual, but the more languages, the simpler the system, and the more basic the level of the exercises.

The principal characteristics that distinguish an intelligent tutor are: the ability to diagnose the knowledge structure and the skills of the learner; personal adaptation of instruction to the learner; provision of personalized feedback. To meet these requirements, the ITS needs to implement a *(domain) knowledge model*, a *student model* and an *instruction model*.

1. *Domain knowledge model:* the (linguistic) domain knowledge model in Revita is embodied in language-specific rules, which drive the creation of exercises, described below in Section 3.2.. The rule component is specialized for each language,[8] and requires some input from language experts.

2. *Student model:* the system records all of the student's answers (correct and incorrect), which affect the choice of future exercises. These responses will be used to implement KST functionality, to determine learner's current state of knowledge. Both the Knowledge model and the Student model are based on data collected from the students' practice sessions.

3. *Instruction Model:* embodies the pedagogical principles: in which order concepts and exercises should be presented, what feedback to provide, etc. Currently, Revita uses rules as Instruction model to determine which new exercises should be presented to the user based on her Student model. The system then attempts to emphasize practicing those concepts: A. in which the user is not proficient, and B. which the user is best prepared to absorb next. In this process, the exercises turn less frequently—i.e., with lower probability—to concepts: 1. on which the learner previously made very few mistakes (to avoid boredom), and 2. on which she (almost) always makes mistakes (to avoid frustration).

Future work includes incorporating state-of-the-art theories for driving instruction by means of KST and Bayesian knowledge tracing (BKT), (Pardos and Heffernan, 2010).

3.2. Exercise modes

In a broad sense, active language learning is reflected in the functionality of Revita on several levels, including: support active learning by providing the ability to seek out materials to match one's own interests (and sharing interesting materials with friends), as well as selecting categories on which to concentrate in an exercise session, using progress feedback to direct the focus and emphasis of exercises.[9]

The bulk of Revita's exercises are based on the stories the user has selected. One of the bedrocks of Revita's philosophy is that language learning is stimulated and becomes more productive when the learner

---

[8]For example, in German and Swedish, Revita creates flash-card exercises to test the knowledge of noun gender via their articles—the article is determined by the noun's gender. This is not relevant for gender-free languages.

[9]This feature is helpful for preparation for standardized tests—by specifying a set of topics to focus on.

Figure 2: Practice mode. Current snippet of the story (in Finnish) at the bottom, over grey background, contains multiple choice (menu) and cloze (textbox) exercises. Previous snippets above; correctly answered questions in dark green, incorrectly answered questions in blue. User clicked the last word in the current snippet to see its meaning (shown in box on the left).

is working with materials which hold *personal interest for the learner*. The system has a small public library of texts for every language; however every learner is encouraged to upload materials into her own, private library.[10] Learning materials can be shared with other users of the platform ("friends").

Several modes are available for working with a story. The *reading mode* allows the learner to skim through a story and view translations of unfamiliar words— which also adds them to the personal vocabulary (stack of flashcards) for later practice.

In *practice mode* (see Figure 2), the system automatically generates exercises based on a story chosen by the user. The story is presented piece by piece; each piece ("snippet") is about 30-40 words in length, with several words chosen for exercises—these words are omitted and replaced by a multiple-choice quiz, or a "close" quiz. In the figure, the current snippet is at the bottom of the screen (over a grey background). Previously completed snippets remain above the current snippet (over a white background). In the prior snippets, correctly answered exercises are highlighted in green, and incorrectly answered ones are in blue; hovering over an incorrectly answered exercise provides more insights about the mistake.

For example, "Pohjanmaa $\boxed{on}$ täynnä $\boxed{jokilaaksoja}$."

is the first sentence (in Finnish) in the snippet in Figure 2, (meaning "Pohjanmaa *is* full *of river valleys*"). The system created cloze exercises by hiding the words in the boxes and providing their base forms (*lemmas*) to the learner as hints:

"Pohjanmaa $\boxed{olla}$ täynnä $\boxed{jokilaakso}$."

From the verb lemma "*olla*" ("to be") and noun lemma "*jokilaakso*" ("river valley") the learner should derive the original, correctly inflected forms. The task of the user is to guess which form of the skipped word was used in the context by the author.

Non-inflected words (prepositions, conjunctions, etc.) are used for multiple-choice quizzes. Inflected words (nouns, verbs, etc.) are used for "close" quizzes: the base form (*lemma*) of the word is shown as a hint, and the learner needs to guess the appropriate grammatical form in context.

This approach has its advantages as well as problems. On one hand, it is easy to validate the user's answer, because the system knows the original form used in the story. On the other hand, the learner may possibly insert a different form which is valid in the given context, yet different from the form found in the story. It is a challenging NLP problem to determine whether the user's answer is *also* acceptable—grammatically and semantically—for the given lemma in the given context. This requires a high-precision language model, which is one of the objects of our current research.

---

[10]This is done by loading a local file (.txt or .doc), by copy-pasting text, or by providing a remote URL of the website containing text to upload into the library.

4088

Figure 3: Crossword mode. The current story (in Finnish) on the right contains cloze exercises, in textboxes (exercises that have been resolved are in dark green, e.g., the first word in the story). Each cloze corresponds to entries in the crossword (middle); currently selected entry (15→) is highlighted. The clues to the entries are translations of the missing words (on the left).

After the learner responds to all quizzes in the current fragment of the story, the system provides her with immediate feedback, and then moves on to the following fragment with new exercises. All possible candidates for exercises are determined and saved at the time when the story is uploaded into the system and analyzed by several lower-level natural language processing (NLP) modules including morphological analysis, disambiguation, etc.

Candidates are created using language-specific rules. For example, for Russian, the system may include rule A: [word='в', POS=prep, case={loc, acc}], Another rule, B, in 3 parts, may be:

1. [POS=prep, case=**X**]
2. [POS=adj, case=**X**, number=**Y**, gender=**Z**]
3. [POS=noun, case=**X**, number=**Y**, gender=**Z**].

These (declarative) rules state general facts about the language, and refer to the part of speech tags and morphological tags of words that may be encountered in a story. Rule A states the preposition "в" (meaning "in") governs the locative or accusative case. Rule B stipulates the agreement[11] within a noun phrase that contains the parts of speech (POS) [prep adj noun]: case agreement between preposition, adjective and noun, as well as number and gender agreement between the adjective and noun. Using such rules, the system finds constructions, such as:

"… в тёмном лесу …"
in dark-Masc.Sg.Loc forest-Masc.Sg.Loc
"... in a dark forest ..."

where the case, number and gender categories agree. Then this sequence may be offered as an exercise, e.g.,

---

[11]By unification over the variables, which are in bold.

with a multiple choice for the preposition (with the remaining options filled by *distractor* prepositions), and "cloze" boxes for the noun and adjective. Various approaches to the problem of generation of reliable distractors are described in (Lee and Luo, 2016; Correia et al., 2010; Rakangor and Ghodasara, 2014; Sakaguchi et al., 2013; Hill and Simha, 2016; Liang et al., 2017).

The actual choice of candidates for exercises depends on the user model—the history of answers given previously by the user. The system computes weights (probabilities) for all potential candidates in the story snippet; words or constructions receive a lower weight if they had been answered mostly incorrectly (or mostly correctly) in previous sessions—since that implies that they were too difficult (or too easy) for the learner at present. Candidates with smaller weights appear with a lower probability in the next exercises. Weights are assigned not only to particular words, but also to their various grammatical categories; thus, if a user, e.g., sometimes makes mistakes in a certain nominal case, the system will provide more exercises for this case.

The *crossword mode* encourages practicing grammar and vocabulary simultaneously. A crossword is built automatically and randomly based on the text of a story (see Figure 3). Candidates for the crossword are selected according to the same principles as in practice mode (above); hidden words need to be inserted back into the story in the correct inflected form. As clues, the learner receives the translations of the missing words (rather than their lemmas, as in practice mode). All answers are saved and used for updating the weights of candidates on subsequent exercises, as is done for the practice mode.

In all modes—reading, practice, crossword—the
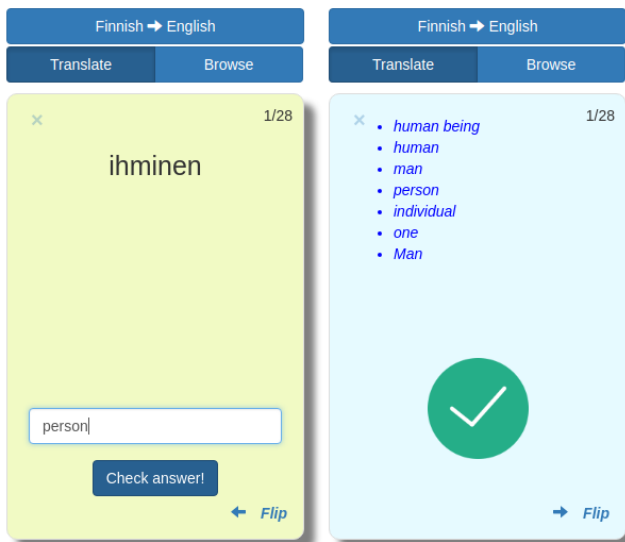
Figure 4: Flashcards for practicing vocabulary.

Figure 5: Progress visualization for numerals in Russian

learner can request a translation of any unfamiliar word. Translations are looked up in third-party on-line multi-lingual dictionaries. Words for which translations were requested are added into the user's stack of flashcards—used in the *flashcard* mode, see Figure 4, to practice vocabulary. In the figure, the user first receives the word to be translated (left), types a translation in the box below, flips the card and receives feedback (right) whether the translation is among those that are recognized by the system.

For all of the above exercise modes—reading, practice, crossword, flashcards—Revita provides a *Competition mode extension*.[12] For example, for the practice mode, the idea behind the competition is that the user must guess the correct grammatical word given the context, as before, but now she is simultaneously racing to answer the questions faster than an opponent, who is solving the same exercises at the same time. The learner needs to get more correct answers in a shorter period of time to win the competition. At present, the opponent is a *bot*, which aims to imitate as nearly as possible the learner's own rate of answering questions and state of knowledge (probability of answering the given question correctly). In this way, the learner is *competing with herself*—i.e., trying to improve on her own previous performance.[13]

These exercise modes—all building on the original story text selected by the user—provide a variety of interactions, which help the user practice on one's own, outside the classroom, and which are designed to stimulate and engage the user—more so than a traditional textbook might.[14]

---

[12]At present, competition is implemented only the practice mode, but the idea is straightforward and the extension will be implemented for all exercise modes.

[13]We plan to implement competition with a human friend as well—another user on the platform.

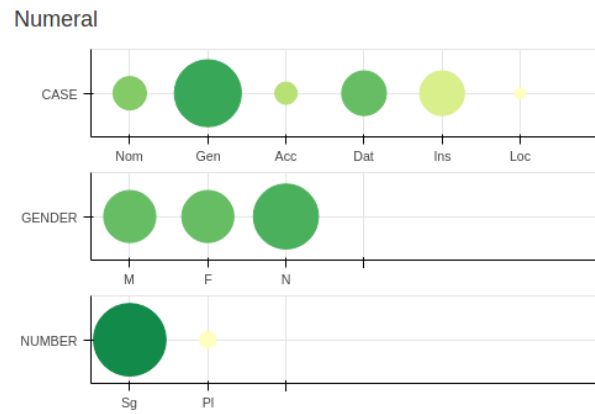[14]In the future, the goal is to incorporate exercise modes

3.3. Assessment

Currently the system implements a simple system for progress assessment: it checks all of the learner's answers, tracking how exercises involving various grammatical concepts—which are categories, such as tense and mood for verbs, number and case for nouns, etc.—were answered by the learner, and how many exercises with different concepts were practiced.

The learner may keep track of her progress via a visualization (see Figure 5). The size of the balls indicates the relative frequency with which this grammatical concept was encountered by the user in exercises so far; the color ranges from green to red, depending on the percentage of correct answers given so far; if the majority is correct, the color tends toward green, otherwise toward red. This visualization is one way to summarize the learner's knowledge state at the present time; we plan to extend this visualization with information about progress across time.

3.4. Evaluation by learners and teachers

To date, we have introduced Revita to professional teachers of 12 different languages, who provided extensive and useful feedback about their particular language. All of them have stated an interest in using the system in their pedagogical practice.

We are currently in the process of testing the system with students. These initial tests focus on LI-A learners of Finnish and Russian, with wider testing planned to follow.

4. Revita as a source of educational data

Educational data mining (EDM) is defined as the "research field concerned with the application of data mining, machine learning and statistics to information generated from educational settings (e.g., universities and intelligent tutoring systems)," (Baker and Yacef, 2009). The goal of EDM is to help predict the learner's behavior—which means deriving accurate

---

for reading and aural comprehension, and for speaking.

Student models from user's feedback; the same feedback can be used to improve the Domain models inside the system. More generally, EDM also allows us to study the effects of the educational support provided by the system, and to advance scientific knowledge about the process of language learning and acquisition.

Thus, the final aspect of Revita to which we draw attention in this paper is that as the platform is used by the learners, over time it builds a rich and valuable resource: the detailed educational data in the form of learner responses—correct and incorrect,—the contexts of the responses within the stories, all words which were unfamiliar in their contexts (i.e., for which translations were requested)—as well as the timings of all of these interactions.

This is valuable data, no less valuable than learner corpora, which language educators have worked very hard to collect for decades. The larger a learner corpus grows, the better it allows us to build accurate models of the learners and the learning process, to identify the common mistakes, and the patterns of mistakes. This information allows the system to derive more accurate Domain, Student and Instruction models.

The Student model for each user is based directly on the totality of the data collected for that user. Important aspects of the Domain model are implicit in—and can be inferred from—generalizations over data collected from many learners. For example, if we consider two concepts, A and B, and over a large number of learners we can observe that those who answer B correctly also (almost always) answer A correctly, but not vice versa, then we can infer that, with high probability, concept A is a pre-requisite for B.

Further, looking for patterns of the *order*, in which concepts are mastered by students—over a large population of students—we can obtain statistical information about how the learner's native language affects the order in which concepts in the target language are likely to be acquired. It seems intuitively clear that a group of learners with a common native language will likely exhibit common traits—observable through their responses.[15]

It is also reasonable to assume that having the ability to model these commonalities explicitly, and exploit them in the learning process (in the Instruction model) will provide for a more intelligent tutoring system.

The Instruction model is currently computed by rules, which operate on the data stored in the learners' history. The same data will form the foundation for more complex Instruction models, which we plan to explore, viz., Knowledge Space Theory, and Bayesian knowledge tracing.

It is clear that the educational data collected through Revita over time will find a variety of purposes and important applications. It will impact various aspects of our overall enterprise—namely, gaining insight into the process of language learning, and improving the methodology of automated language instruction and tutoring.

High-quality, detailed and large-scale data is a primary requirement for the application of modern data-driven machine learning and statistical methods. The key point, however, is that *without* a powerful, automated platform such as Revita, data on a comparable scale would be impossible to collect.

5. Conclusion

An in-depth literature review and investigation of the currently available tools indicates that ICALL systems do not exhibit sufficiently intelligent characteristics, and ITS systems described in the literature appear to be laboratory experiments, not available to learners in practice.

We have presented Revita, an open platform developed at the intersection of ITS and CALL, aiming to address all of the main drawbacks of existing language-learning systems.

Revita aims to bring intelligent language technology into the freely available online language-learning space, which is becoming more popular, not only among the learners, but also among the teachers. It also serves as an instrument for collection of valuable educational data.

The functionality of the system is under continual development. Modifications to Revita are motivated by feedback from the users and SLA/FLA experts.

6. Acknowledgments

7. Bibliographical References

Amaral, L. and Meurers, D. (2007). Conceptualizing student models for ICALL. In *User modeling*, volume 2007, page 11th. Springer.

Amaral, L. A. and Meurers, D. (2011). On using intelligent computer-assisted language learning in real-life foreign language teaching and learning. *ReCALL*, 23(1):4–24.

Baker, R. S. and Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *JEDM-Journal of Educational Data Mining*, 1(1):3–17.

Bush, M. D. (2008). Computer-assisted language learning: From vision to reality? *CaLICO Journal*, 25(3):443–470.

---

[15]For example: native speakers of Turkic languages may find some aspects of learning Finnish (the target) easier than native speakers of Slavic languages—vowel harmony is one such aspect, because it is present in similar forms in both Finnish and Turkic. On the other hand, other aspects of Finnish may be much more intuitive to Slavic speakers, because historically Finnish has been in more immediate contact with Slavic languages, and the development of many of its linguistic features was affected through this contact.

Correia, R. P. d. S., Baptista, J., Mamede, N., Trancoso, I., and Eskenazi, M. (2010). Automatic generation of cloze question distractors. In *Second Language Studies: Acquisition, Learning, Education and Technology*.

Craig, S. D., Hu, X., Graesser, A. C., Bargagliotti, A. E., Sterbinsky, A., Cheney, K. R., and Okwumabua, T. (2013). The impact of a technology-based mathematics after-school program using ALEKS on student's knowledge and behaviors. *Computers & Education*, 68:495–504.

Dodigovic, M. (2007). Artificial intelligence and second language learning: An efficient approach to error remediation. *Language Awareness*, 16(2):99–113.

Doignon, J.-P. and Falmagne, J.-C. (2012). *Knowledge spaces*. Springer Science & Business Media.

Falmagne, J.-C. and Doignon, J.-P. (2010). *Learning spaces: Interdisciplinary applied mathematics*. Springer Science & Business Media.

Falmagne, J.-C., Koppen, M., Villano, M., Doignon, J.-P., and Johannesen, L. (1990). Introduction to knowledge spaces: How to build, test, and search them. *Psychological Review*, 97(2):201.

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., and Freynik, S. (2014). Technologies for foreign language learning: a review of technology types and their effectiveness. *Computer Assisted Language Learning*, 27(1):70–105.

Hart, R. (1981). Language study and the PLATO system. *Studies in Language Learning*, 3(1):1–24.

Heffernan, N. T. and Heffernan, C. L. (2014). The assistments ecosystem: building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, 24(4):470–497.

Heift, T. (2008). Modeling learner variability in CALL. *Computer Assisted Language Learning*, 21(4):305–321.

Heift, T. (2010). Developing an intelligent language tutor. *CALICO journal*, 27(3):443–459.

Hill, J. and Simha, R. (2016). Automatic generation of context-based fill-in-the-blank exercises using co-occurrence likelihoods and Google N-grams. In *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 23–30.

Hubbard, P. and Siskin, C. B. (2004). Another look at tutorial CALL. *ReCALL*, 16(2):448–461.

Katinskaia, A., Nouri, J., and Yangarber, R. (2017). Revita: a system for language learning and supporting endangered languages. In *Proceedings of the Joint 6th Workshop on NLP for Computer Assisted Language Learning and 2nd Workshop on NLP for Research on Language Acquisition at NoDaLiDa*. Linköping University Electronic Press.

Koedinger, K. R., McLaughlin, E. A., and Heffernan, N. T. (2010). A quasi-experimental evaluation of an on-line formative assessment and tutoring system.

Journal of Educational Computing Research, 43(4):489–510.

Kulik, J. A. and Fletcher, J. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. *Review of Educational Research*, 86(1):42–78.

Lee, J. and Luo, M. (2016). Personalized exercises for preposition learning. *Proceedings of ACL-2016 System Demonstrations*, pages 115–120.

Levy, M. (1997). *Computer-assisted language learning: Context and conceptualization*. Oxford University Press.

Liang, C., Yang, X., Wham, D., Pursel, B., Passonneaur, R., and Giles, C. L. (2017). Distractor generation with generative adversarial nets for automatically creating fill-in-the-blank questions. In *Proceedings of the Knowledge Capture Conference*, page 33. ACM.

Mozgovoy, M. (2012). Towards WordBricks—a virtual language lab for computer-assisted language learning. In *Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on*, pages 229–232. IEEE.

Nagata, N. (1993). Intelligent computer feedback for second language instruction. *The Modern Language Journal*, 77(3):330–339.

Nagata, N. (1997). An experimental comparison of deductive and inductive feedback generated by a simple parser. *System*, 25(4):515–534.

Nagata, N. (2009). Robo-Sensei's NLP-based error detection and feedback generation. *Calico Journal*, 26(3):562–579.

Ostrow, K. and Heffernan, N. (2014). Testing the multimedia principle in the real world: a comparison of video vs. text feedback in authentic middle school math assignments. In *Educational Data Mining 2014*.

Oxman, S., Wong, W., and Innovations, D. (2014). White paper: Adaptive learning systems. *Integrated Education Solutions*.

Pardos, Z. and Heffernan, N. (2010). Navigating the parameter space of Bayesian Knowledge Tracing models: Visualizations of the convergence of the expectation-maximization algorithm. In *Educational Data Mining 2010*.

Rakangor, S. and Ghodasara, Y. (2014). Automatic fill in the blanks with distractor generation from given corpus. *International Journal of Computer Applications*, 105(9).

Sakaguchi, K., Arase, Y., and Komachi, M. (2013). Discriminative approach to fill-in-the-blank quiz generation for language learners. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 238–242.

Schrepp, M. (1999). Extracting knowledge structures from observed data. *British Journal of Mathematical and Statistical Psychology*, 52(2):213–224.

Self, J. (1990). Theoretical foundations for intelligent tutoring systems. *Journal of Artificial Intelligence in Education*, 1(4):3–14.

Self, J. (1998). The defining characteristics of intelligent tutoring systems research: ITSs care, precisely. *International Journal of Artificial Intelligence in Education (IJAIED)*, 10:350–364.

Shute, V. J. (1991). Rose garden promises of intelligent tutoring systems: Blossom or thorn. *Nasa technical report*.

Slavuj, V., Kovačić, B., and Jugo, I. (2015). Intelligent tutoring systems for language learning. In *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015 38th International Convention on*, pages 814–819. IEEE.

Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 30.

Volodina, E., Pilán, I., Borin, L., and Tiedemann, T. L. (2014). A flexible language learning platform based on language resources and Web services. In *LREC*, pages 3973–3978.