Linguistic Constructs Represent the Domain Model in Intelligent Language Tutoring

Anisia Katinskaia[1,2], Jue Hou[1,2], Anh-Duc Vu[1,2], Roman Yangarber[2]
[1] Department of Computer Science, University of Helsinki
[2] Department of Digital Humanities, University of Helsinki
first.last@helsinki.fi

Abstract

This paper presents the development of the AI-based language-learning platform, Revita. It is an intelligent online tutor, developed to support learners of multiple languages, from lower-intermediate toward advanced levels. It has been in pilot use with hundreds of students at several universities, whose feedback and needs shape the development. One of the main emerging features of Revita is the system of *linguistic constructs* to represent the domain knowledge. The system of constructs is developed in collaboration with experts in language pedagogy. Constructs define the types of exercises, the content of the feedback, and enable detailed modeling and evaluation of learner progress.

1 Introduction

The focus of this paper is the novel Domain Model in Revita,[1] an Intelligent Tutoring System (ITS) for language learning (Katinskaia et al., 2018, 2017). Revita follows the classic design of ITS—with a Domain model, Student model, and Instruction model. The *Domain Model* describes what must be mastered by the learner: concepts, rules, etc.—known as *skills* in ITS literature—and *relationships* among them (Wenger, 2014; Polson et al., 1988). We represent the Domain Model as a system of *linguistic constructs*—a wide range of linguistic phenomena, including inflexion of various paradigms, government relations, collocations, complex syntactic constructions, etc. The system of constructs is developed in collaboration with experts in language teaching. It impacts all apcects and components of Revita—the variety of exercises that it generates automatically, the intelligent feedback given to the learner, modeling of learner knowledge, and evaluation of learner progress.

The *Student model* represents the learner's proficiency. It is based on the history of answers given by the learner to many exercises, and tries to build a detailed picture of what the user knows vs. does not know. The *Instruction model* embodies the pedagogical principles that lie behind the decisions: which exercises the learner is best prepared for next, and which feedback should be provided to guide the learner toward the right answer. These models are interconnected in ITS.

Revita is currently undergoing pilot studies with real-world learners and teachers at several universities (Stoyanova et al., 2021). Revita's main target group are learners who have passed beyond the beginner level — above A2 on the CEFR scale.[2] Revita is developed as a tool for learners and teachers of several languages: Finnish and Russian are currently the most developed languages. Several "beta" languages, including Italian, German, Swedish, and others, are in earlier stages of development. The user interface also works in several languages (English, Finnish, Russian, Chinese, Italian). Revita is not meant to replace the teacher. For students, it provides 24/7 access to an unlimited amount of personalized exercises for practice matching the learner's current level, with immediate feedback and progress estimation. For teachers, it provides time-saving benefits by allowing them to delegate the mundane work of creating hundreds of exercises for each topic for students at different levels. Revita allows the teachers to share learning materials, create their own exercises, work with groups, and monitor progress and evaluation.

The paper is organized as follows: Section 2 briefly reviews work on intelligent computer-assisted language learning (ICALL). The principles and ideas behind Revita are described in Section 3. It also describes its main components: linguistic constructs, automatic generation of exercises and feedback, and modeling of learner knowledge. Section 4 describes tools for learners and teachers. Section presents the conclusions and future work.

[1] revita.cs.helsinki.fi — Link to a short demo here.

[2] The Common European Framework of Reference for Languages: Learning, Teaching, Assessment

Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics System Demonstrations, pages 136–144

2 Prior Work

Interest in computer-assisted language learning (CALL) is growing with the rapid development of language technology. CALL is seen as the "study of applications of the computer in language teaching and learning" (Levy, 1997). Applying ITS to language learning and supporting CALL by intelligent and/or adaptive methodologies, such as expert systems (ES), natural language processing (NLP), automatic speech recognition (ASR), etc.—is the domain of intelligent CALL, or ICALL. The goal of ICALL is building advanced applications for language learning using NLP and linguistic resources—corpora, lexicons, etc. (Volodina et al., 2014).

The number of academic and commercial tools for language learning is growing drastically, e.g., popular commercial systems like Duolingo, Rosetta Stone, Babbel, Busuu, iTalki, etc. Some CALL systems aim to give learners access to *authentic* materials (White and Reinders, 2010), the opportunity to interact with teachers and native speakers (e.g., the app *Lingoda* is a platform for live video classes), and provide text or sound feedback based on learner needs and knowledge (Bodnar et al., 2017). Modern CALL systems are also mobile, which increases their accessibility (Derakhshan and Khodabakhshzadeh, 2011; Rosell-Aguilar, 2018; Kacetl and Klímová, 2019).

In developing CALL, pedagogical goals—rather than technological means—should be the primary focus (Gray, 2008). It has been shown that using ICALL systems for education improves learner motivation and attitudes, increases options for self-study (Golonka et al., 2014), improves retention of various learning concepts, communication between students and teachers, academic self-efficacy (Bandura, 1977; Rachels and Rockinson-Szapkiw, 2018), and overall language skills (Yeh and Lai, 2019; Zhang and Zou, 2022).

Despite having existed for decades, ICALL still has a number of serious limitations to overcome. Apart from platforms where teachers directly interact with students in video classes, most existing ICALL systems are based on the so-called "canned" approach—the available sets of exercises are pre-made, and therefore limited. This limits the level of personalization: pre-made exercises can be arranged into different individual programs, but no personalized exercises can be provided.

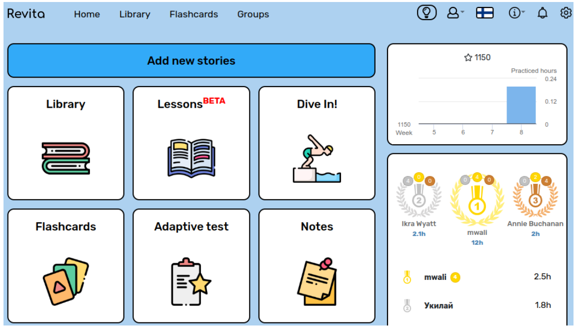

Figure 1: Revita's home page, with the main activities.

3 Core Components of Revita

3.1 Main Principles

Revita's approach is founded on the following primary principles:

1. *Practice should be based on authentic content*. By authentic we mean a text which is not artificial and written for learning purposes. The learner (or teacher) can upload any text from the Internet using a URL, upload any file, etc., to use it directly as learning content.
2. *Exercises are automatically generated* based on any authentic text chosen by the user, including any texts uploaded to the system.
3. *Exercises are personalized* to match the learner's current skill levels, so that each new exercise is selected to be a challenge that the learner is able to meet.
4. *Immediate feedback*: rather than saying only "right/wrong", the tutor *gradually guides* the learner toward finding the correct answer by providing *hints*, which begin as general hints about the context and then give more and more specific information about the answer.
5. *Continual assessment* of skills allows Revita to select exercises optimally personalized for each learner based on past performance.

The first principle is the bedrock of Revita's philosophy—in the story-based approach, all learning activities are based on authentic texts, which should be *inherently interesting* for the learner to read, which motivates her to practice longer. A few sample texts are available in the system's "public" library for each language; also, several new stories are recommended daily as "Stories of the day"—crawled daily from several selected websites. But the main idea is that texts be chosen and uploaded by the learners themselves (or teachers). The button "Add new stories" on the home page (see Figure 1)

Constructs	Examples
Finnish	
(1) Necessive construction: Present passive participle, with *-ttava* ending	*Energiakriisin lähestyessä kaikki keinot on* ***otettava*** *käyntiin.* (With the energy crisis approaching, all means must **be taken** into action.)
(2) Transitive vs. intransitive verbs	*Voisitko* ***sammuttaa*** *valon?* (Could you **turn off** the light?)
(3) Verb government: translative case	*Kaupungit eivät ole muuttuneet* ***energiatehokkaammiksi****.* (Cities have not become **more energy efficient**.)
(4) Substitute clause: participle substitutes for "that"-relative clause	*Maija kertoi* ***vanhempien asuvan*** *kaupungissa.* (Maija said that **her parents live** in the city.)
Russian	
(5) Verb: II conjugation	*Мы скоро* ***увидим*** *восход.* (We **will see** the sunrise soon.)
(6) Complex pronoun:	*Нам нужно* ***кое о чем*** *поговорить.* (We need to talk **about something**)
(7) Perfective vs. imperfective aspect	*Страны* ***согласовали*** *проект о будущих отношениях.* (The countries **agreed on** a draft on future relations.)
(8) Dative subject & impersonal verb	***Мне*** *необходимо поговорить с врачом.* (**I** need to talk to a doctor.)
German	
(9) Past perfect tense	*Ich* ***wäre*** *mit ihm* ***gekommen****, aber er wurde krank.* (I **would have come** with him, but he got sick.)
(10) Weak masculine nouns	*Ich möchte* ***den Jungen*** *kennenlernen.* (I want to meet **the boy**.)
(11) Prepositions governing dative case	*Wir sind* ***aus dem*** *Haus gelaufen.* (We ran **out of the** house.)

Table 1: Examples of *grammatical constructs* found in sentences (underlined). ***Candidates*** are words that will be chosen for exercises about the constructs (marked in bold).

allows the user to upload new text material into her private library.

3.2 Linguistic Constructs

At the core of Revita's approach is the system of linguistic constructs that are represented in the Domain model. *Constructs* are linguistic phenomena or rules, that vary in specificity: e.g., a construct (in Finnish) may be *verb government*: the verb *tutustua* ("to become acquainted") requires its argument to be in the illative case (*"into something"*), while *tykätä* ("to like") requires its argument in the elative case (*"from something"*), etc. Constructs also include all *constructions*, as conceived in Construction Grammar (CG). CG treats many phenomena—grammatical constructs, multi-word expressions, collocations, idioms, etc.—within a unified formalism. Examples of constructs for several languages are shown in Table 1.

When customizing the system for a new language, we engage experts in language teaching in creating the inventory of constructs, which need to be mastered by the learners. Currently, Finnish and Russian have the most developed system of constructs, each with over 200 constructs. Potentially, the number of constructs can be much larger. The Russian constructs evolved from the extensive grammatical inventory covered in tests for second language (L2) learners developed at the University of Helsinki (Kopotev, 2012). The Finnish constructs are based on inventories of grammatical topics developed by experts in teaching Finnish as L2.

As shown in the examples in Table 1, each construct needs to be identified in the text, when the text is uploaded to Revita. For this purpose, we use finite-state morphological analyzers (e.g., HFST[3]), neural dependency parsers,[4] and rule-based pattern detection. Each morphological analyzer is wrapped into a "Revita" analyzer which modifies the output analyses into a uniform standard set of features used in the system. Considering that none of these tools can provide perfect performance alone, we rely on the agreement between morphological analyzers, parsers, and rules.

In Example (1), for construct "Present passive participle with *-ttava* ending," the rule matches the participle *"otettava"* by morphological features: participle, present tense, passive voice. This form is then recognized as the head of the "necessive" construction *"on otettava"* ("must be taken"), detected by a rule that matches: the singular 3rd person present form of modal verb *"olla"* ("to be") and the present passive participle, in the nominative case. The matching rule has to agree with the output of a dependency parser. In Example (2), the construct "Transitive vs. intransitive verbs" is detected

[3]GiellaLT
[4]Turku NLP, DeepPavlov

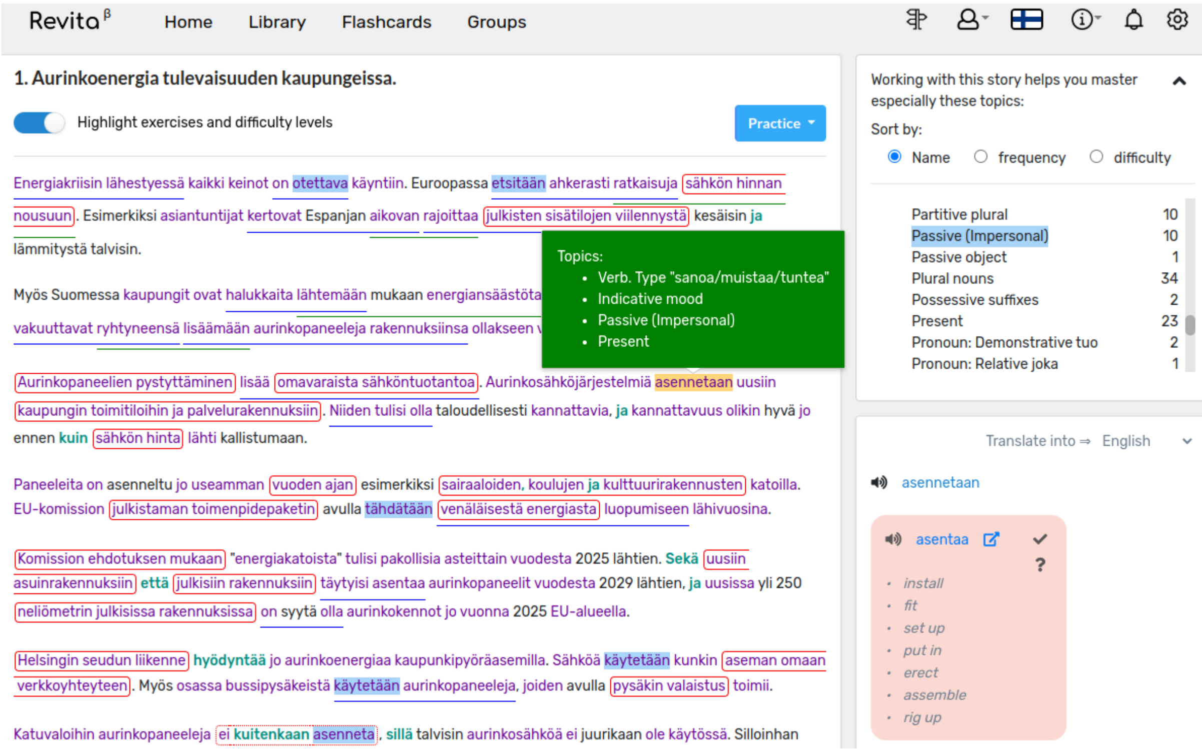

Figure 2: Preview mode for a story (before practice). All purple words can appear in an exercise. Noun phrases and prepositional phrases are circled in red. Government relations and constructions are underlined. Top-right corner—the list of constructs found in the story. Bottom-right corner—translation of the clicked word: "asennetaan", into English (target language can be selected). The green box over the clicked word lists all constructs related to it.

by using dictionaries of verb lemmas or by rules that detect regular transitive/intransitive ending patterns in verb lemmas (e.g., *sammuttaa* vs. *sammua*, "turn something off" vs. "turn itself off"). Dictionaries contain hundreds of the most frequent Finnish lemmas and are continually updated.

Verb government is detected by several components: large sets of government patterns (2000-3000 per language); pattern matching for noun phrases, prepositional phrases, and analytic verb forms; dependency relations detected by parsers. Each rule for government pattern tests the dependency roles of the arguments as conditions. In Example (3), a government pattern for the intransitive verb *"muuttua"* ("to change") requires an argument in the translative case—here, the comparative adjective *energiatehokkaammiksi* ("more energy-efficient"). The government detector will find an argument of *"muuttua"* regardless of its position in the sentence, and for any form of the verb, including complex analytic forms, e.g., the negative perfect tense *"eivät ole muuttuneet."*

Detecting longer and more complex syntactic constructions relies on all of the components mentioned above. In (4), to match the complex construction *"kertoi vanhempien asuvan"*, we use a rule states that the verb *"kertoi"* ("said") must govern a subordinate clause starting with *"että"* ("that"); the *substitute* clause contains a noun phrase in the genitive case, which acts as the subject (*"vanhempien"*) and a genitive active participle (*"asuvan"*).

The user can preview all constructs identified in a story in the Preview Mode prior to practice, see Figure 2. All noun and preposition phrases are circled, government relations and syntactic constructions are underlined. A list of all constructs found in the story is in the top-right window: clicking on a construct highlights all instances of the construct in the story. Figure 2 shows all impersonal passive forms highlighted in blue. Clicking on any word in the story will also show all constructs linked to it (green box above the clicked word, the see word *"asennetaan"*). The translation of the clicked word into the learner's chosen language (English here) is in the bottom-right. This lets the learner (or teacher) see what can be exercised in the given text.

3.3 Exercise Generation Based on Constructs

Revita offers several practice modes; the main activity for the learner is the Grammar Practice Mode

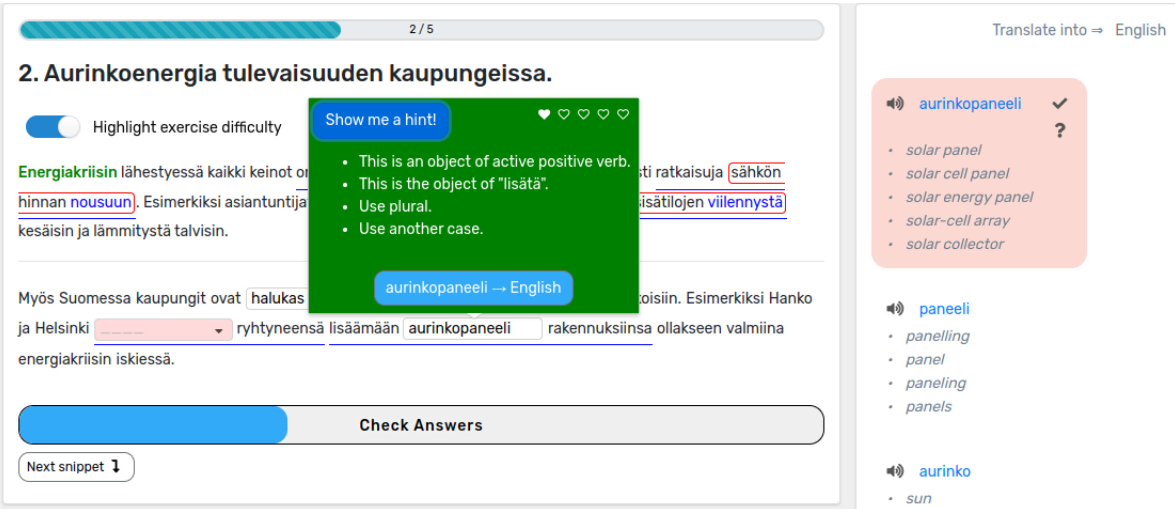

Figure 3: Practice mode with a story. Figure shows the second paragraph of a story with three exercises: two clozes ("halukas" and "aurinkopaneeli") and one MC. Previous answers are marked green and blue—correct and incorrect. The green box shows the hints requested so far for the cloze exercise.

based on a story from the public or private library, see Figure 3. Practice mode offers "cloze" (fill-in-the-blank) and multiple-choice (MC) exercises. A cloze exercise is shown as a text box, with the lemma of the expected answer given as a hint to the learner. In Figure 3, the lemma in the box is *aurinkopaneeli* ("sun panel"). The learner is expected to insert the correct form of this word that suits the context; here, it is the plural partitive case (*"aurinkopaneeleja"*)—the expected answer is the original word form from the story, which was replaced with the exercise. Each word picked to be exercised must be disambiguated—we must know the correct lemma to show to the learner. Disambiguation is performed by agreement rules and by dependency parsers. For *analytic* verb forms, such as *"on otettava"* ("should be taken"), the cloze box will show the lemma of the *head* verb: *ottaa* ("to take").

All *candidates*—potential exercises in practice mode—are based on the constructs detected in the story. In Example (3) for Finnish in Table 1, an exercise on the construct "Verb government" is in bold: the learner will see the lemma *energiatehokas* ("energy-efficient"). To insert the correct form in the translative case, the learner needs to know which case is required by the governing verb. MC exercises are more targeted: the options to choose from—known as "distractors"—are generated based on the exercised construct. Therefore, the same word or construction may have more than one set of distractors, since more than one construct may be linked to the candidate.

Distractors are created by rules and morphological generators. In Example (6), for the construct "Complex pronoun", tests the knowledge of joint vs. hyphenated spelling—a rule generates distractors like: *"кое о чем"*, *"кое-о-чем"*, *"о кое-чем"* ("about something"). For transitive vs. intransitive verbs, we use dictionaries of lemma pairs. However, the distractors must be *inflected* forms that fit the context, not lemmas. We use morphological generators to produce the required inflected forms.

MC distractors are often an effective way of learning a particular construct, and choosing good distractors is a task that requires pedagogical expertise. In Example (4), e.g., the construction requires the subject to be in genitive case. It is useful to offer the lemma *"vanhemmat"* ("parents") in other cases which can mark the subject in other constructions, e.g., nominative, partitive, etc. These forms, which differ only by case, are produced by the morphological generator.

In addition to the mentioned exercise types, Revita generates MC exercises for stress in Russian.[5] Distractors are generated using the finite-state morphological analyser UDAR[6] (Reynolds, 2016). Another kind of exercises is based on Text-to-Speech technology[7]—the learner needs to to listen to a spoken fragment and insert the missing forms. These exercises are not generated based on constructs, and are therefore outside the scope of the paper.

[5]Stress is a very complex topic in learning Russian.
[6]https://github.com/reynoldsnlp/udar
[7]Text-to-Speech

3.4 Feedback

Feedback is a second essential feature of Revita. The learner gets *multiple attempts* for every exercise. Feedback is designed to gradually guide the learner toward the correct answer by providing a sequence of hints that (a) depend on the context, (b) on the constructs linked to the exercise, and (c) on the answer that was given by the learner. Feedback hints are ordered so they become more specific on subsequent attempts: starting from referring to syntactic construction or word paradigms and then to grammatical features required in the answer. For example, for an object of a verb governs the partitive case, the feedback sequence may be: *"The is the object of the verb 'xyz'."* →*"Use another case."* →*"Use partitive case."* The learner can also request hints *before* giving an answer: as seen in the green box in Figure 3, four of the available hints are already "used up" (one heart remaining). Requesting hints indicates that the learner has *not* mastered the construct, and affects the learner's scores.

Feedback that depends on the context gives information on whether a word in question is part of some construction or relies on a governing word (verb, noun, or adjective), etc. Hints also appear as *underlining* of syntactically related elements in the context. These hints are generated based on detected syntactic constructions.

Some feedback hints are generated in the stage when the construct is linked to the text. For example, a hint *"Use past perfect tense"* will be attached to *"wäre gekommen"* (see example (9) in Table 1). A more complicated example is for an exercise with the participle *"asuvan"* in *substitute that-clause* construction (see example (4) in Table 1). We generate the feedback for it: *This is equivalent to "...kertoi että vanhemmat asuvat..."* ("...said that parents live...")—by generating the actual clause which is substituted by the participle. To produce this feedback message, Revita uses information about the syntactic roles of each word in the original construction *"kentoi vanhempien asuvan"*, and the required grammatical features of the forms in the feedback—to produce these forms, we use the morphological generator.

When the learner inserts an answer which does not match the expected one (i.e, the one in the original story), Revita analyzes the answer and checks which grammatical features are incorrect. To give feedback on these features in the order of increased specificity, Revita uses a language-specific hierarchy of features. For example, in Russian, the hierarchy specifies that the hint about an incorrect gender of an adjective is shown before hints about an incorrect number or case.

All mechanisms which define the order and the content of feedback hints and algorithms of sampling exercises for students are part of the Instruction Model of Revita.

3.5 Learner Modeling and Exercise Sampling

All learner answers and all requested hints to each exercise are recorded. A learner may attempt to answer each exercise multiple times. For each attempt, Revita analyzes the answers and the requested hints to calculate *credits and penalties* for the corresponding language constructs. Partial correctness of answers is taken into account, e.g., if the answer used the correct tense but wrong number, only number will be penalized, and tense will receive credit.

The collected information on performance with constructs is used to model learner skills and the difficulty of the constructs. To model learner skills and exercise difficulty, we use Item Response Theory (IRT) (Embretson and Reise, 2013; van der Linden and Hambleton, 2013). IRT comes from psychometrics and is widely used in education (Klinkenberg et al., 2011). The *Item* in IRT is a task that the learner should solve. Most IRT applications have a clear definition of an *item*, and a clear credit standard. The classic example of an item is a test question in mathematics: it is unambiguous and there is a clear judgment of the answer—correct or wrong. Our major challenge is that language constructs are not directly judged, as test items in other learning domains. It is challenging to determine the credit and penalty for each construct based on the student's answer, because the link from exercise to constructs is *one-to-many*.

We leave the details of modeling difficulty with IRT outside the scope of this paper. To date, we have collected 570K answers for Russian exercises. Experiments with this data show a strong correlation between students' proficiency estimated by IRT vs. by their teachers. This suggests that with IRT we are able to reliably model learner proficiency. Interestingly, the estimates of exercise difficulty do not correlate with teachers' judgments, which agrees with the findings of other researchers (Abbakumov and Lebedeva, 2016).

At present, we assume that the difficulty of an exercise depends on the *hardest* construct linked to it.

Thus, exercises are sampled for practice based on the difficulty of the hardest construct linked to each exercise. The difficulty of constructs is modeled by IRT. We aim to provide exercises that are best suited to each student's proficiency level. For each possible exercise, IRT first estimates the probability that the student will answer the exercise correctly—then the probability of picking this exercise for practice is sampled from a normal distribution around the mean of a 50% chance that the learner would answer correctly. Thus, on average, the exercises are not too difficult and not too easy.

For languages with insufficient learner data for training IRT, we ask teachers to assign manually CEFR difficulty levels to constructs. Earlier experiments using specialized Elo ratings[8] for assessing learner skills and evaluating the difficulty of linguistic constructs based on the learner data collected by Revita are presented in Hou et al. (2019).

4 Tools for Students and Teachers

At any time, the student can set her CEFR proficiency level manually or take an *adaptive* placement test to estimate proficiency (see the button "Adaptive Test" on Figure 1). The test draws on a bank of questions prepared by teachers; the sampling of questions is driven by an IRT model trained on learner data. After that, the estimate of the learner's proficiency levels is adjusted according to the correctness of answers to exercises.

The learner can upload a story from any website or a local file. To each uploaded text, Revita applies classification by semantic topic—culture, science, sport, politics—and difficulty classifiers. In case the learner does not want to choose a text for practice, there is a "Dive-in" option to practice with a randomly sampled story from a selected category (private vs. public or story tagged by a semantic topic). Another option is to choose a "story of the day" suggested by Revita.

The *Preview mode* (see Figure 2) allows the user to read a story, edit it (in case it was extracted from a web page inaccurately), and review the grammatical topics that can be learnt through practice with this story. Clicking on each word provides its translation into a number of languages. The learner can mark whether she knows a word or not. All unfamiliar words are added to the learner's personal set of flashcards, which are used for Vocabulary Practice.

[8]The Elo rating system is a method for calculating the relative skill levels of players in zero-sum games such as chess.

The *Practice mode* presents the grammar and listening exercises—the learner can hear a segment of text in context and is expected to type the missing words correctly in the empty box. The user can also practice with a story in the Competition Mode, against a bot: the difference from normal practice is that the learner needs to do the exercises faster than the bot—whose skill levels approximately match those of the learner. Another option is to practice with a Crossword built on the authentic text—the translations of words are used as hints. Practice mode also allows the user to create notes during practice, which can be attached to words in the story. All learner's notes are collected together (see button "Notes" in Figure 1) for easy review; each note has a reference to words in the story it was attached to.

Revita offers various statistics and info-graphics to track progress on grammar constructs and vocabulary. These analytics are available to the learner and to the teacher. Revita allows teachers to build groups of students, share texts with them, and create tailored exercises that can be shared with the group. Revita allows teachers to track how their students practice and how well they perform on various tasks.

5 Conclusions and Future Work

This paper presents an in-depth discussion of the novel core component of the Revita language learning system—the Domain model embodied in a system of linguistic constructs. This system of constructs underlies and supports all aspects of the learning experience in Revita, it supports the generation of the quality exercises and feedback. It also supports the modeling of learner skills more accurately to provide informative progress analytics, and to offer exercises most appropriate for the learner's current level.

We have results from pilot studies with Finnish and Russian L2 learners using the new Domain Model, but the discussion of the results is beyond the scope of this paper. In the future, we plan to improve the Domain Model by adding more information about the *interactions* and dependencies among the constructs—which will enable the creation of more intelligent learning paths. We also plan to add new types of activities, e.g., speech exercises.

Limitations

Revita works with many languages, however, at present, only Finnish and Russian have a sufficiently developed inventory of constructs that can be actually used by students in real-world scenarios. Most other languages have a limited set of constructs, which affects the quality and variety of the exercises, as well as limited feedback. Developing a substantial inventory of constructs is a complex task, that requires expertise in computational linguistics, as well as in language pedagogy. As mentioned above, Finnish and Russian have the order of 200 constructs. Meanwhile, "the Great Finnish Grammar" has over 1500 articles (Vilkuna et al., 2004), each of which introduces at least one construct, which, in principle, constitutes an aspect of the linguistic competency of a native speaker. A fascinating research challenge is determining the "essential" core inventory of constructs, which can support effective learning. Our experience so far with the rather modest inventories suggests that they already bring enormous value to learners and teachers (Stoyanova et al., 2021).

Revita's approach relies on arbitrary authentic texts being uploaded from the web; sometimes these texts cannot be extracted from the website without some inaccuracies. Also, the original texts may contain typos, mistakes, etc. These problems should be fixed manually by editing the text. Of course, learners with a low proficiency level cannot do that independently. To avoid having these mistakes negatively affect learning, the stories can be checked by a human teacher/tutor. We also plan to employ strong language models for grammatical error detection to identify such potential problems and highlight them to alert the user that additional checking may be needed.

Revita relies on the text when checking the learner's answers. Currently, only one correct answer is allowed—the one that is present in the story. Sometimes the word form entered by the learner may also be valid in the given story context—"alternative correct" answers. In such cases, Revita may still tell the learner that the answer is not correct. This is one of the important problems that we are researching at present, using neural models for the detection of grammatical errors (Katinskaia et al., 2019; Katinskaia and Yangarber, 2021).

Revita also has certain limitations related to the use of external tools and services: dependency parsers, morphological analyzers, and external dictionaries—all may contain inaccuracies and errors. All of these factors can be a source of mistakes in the intelligent tutor: wrong analyses, incorrectly disambiguated lemmas, missing translations, etc. The system tries to collect *multiple sources of evidence* for its predictions to raise the confidence in—and precision of—the predictions. When the confidence is low—e.g., in the presence of conflicting evidence—the exercise, feedback, etc., is *not* offered to the learner.

Ethics Statement

Revita is designed to carefully guard the privacy of its users—learners and teachers. It does not share any personal information collected during the learner's practice with any third parties. The teacher can track the learner's performance only if the learner has explicitly accepted the invitation to join the teacher's group.

Any authentic text material uploaded into the system is visible only in the user's personal *private* library. If the teacher shares a story with a group of students, it is visible only inside the group library, never to anyone outside the group. Texts pre-loaded into Revita's public library come either from sources that have given us explicit permission to use their content or from the public domain.

Acknowledgements

References

Dmitry Abbakumov and Maria Lebedeva. 2016. The RFL grammar and vocabulary placement test: design, pilot test and psychometric analysis. *Russian Language Abroad*, (5).

Albert Bandura. 1977. Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, 84(2):191.

Stephen Bodnar, Catia Cucchiarini, Bart Penning de Vries, Helmer Strik, and Roeland van Hout. 2017. Learner affect in computerised l2 oral grammar practice with corrective feedback. *Computer Assisted Language Learning*, 30(3-4):223–246.

Ali Derakhshan and Hossein Khodabakhshzadeh. 2011. Why call why not mall: An in-depth review of text-message vocabulary learning. *Theory and Practice in Language Studies*, 1(9):1150–1159.

Susan E Embretson and Steven P Reise. 2013. *Item response theory*. Psychology Press.

Ewa M Golonka, Anita R Bowles, Victor M Frank, Dorna L Richardson, and Suzanne Freynik. 2014. Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer assisted language learning*, 27(1):70–105.

Lisa Gray. 2008. Effective practice with e-portfolios. *Higher Education Funding Council for England, JISC, Bristol*.

Jue Hou, Koppatz Maximilian, José María Hoya Quecedo, Nataliya Stoyanova, and Roman Yangarber. 2019. Modeling language learning using specialized elo rating. In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 494–506, Florence, Italy. Association for Computational Linguistics.

Jaroslav Kacetl and Blanka Klímová. 2019. Use of smartphone applications in English language learning—a challenge for foreign language education. *Education Sciences*, 9(3):179.

Anisia Katinskaia, Sardana Ivanova, and Roman Yangarber. 2019. Multiple admissibility in language learning: judging grammaticality using unlabeled data. In *Proceedings of BSNLP: the 7th Workshop on Balto-Slavic Natural Language Processing, ACL: 56th Annual Meeting of the Association for Computational Linguistics.*

Anisia Katinskaia, Javad Nouri, and Roman Yangarber. 2017. Revita: a system for language learning and supporting endangered languages. In *6th Workshop on NLP for CALL and 2nd Workshop on NLP for Research on Language Acquisition, at NoDaLiDa*, Gothenburg, Sweden.

Anisia Katinskaia, Javad Nouri, and Roman Yangarber. 2018. Revita: a language-learning platform at the intersection of ITS and CALL. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Anisia Katinskaia and Roman Yangarber. 2021. Assessing grammatical correctness in language learning. In *Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications*. The Association for Computational Linguistics.

Sharon Klinkenberg, Marthe Straatemeier, and Han LJ van der Maas. 2011. Computer adaptive practice of maths ability using a new item response model for on-the-fly ability and difficulty estimation. *Computers & Education*, 57(2):1813–1824.

Mikhail Kopotev. 2012. Karttu: результаты языкового тестирования в школе и вузе (Karttu: results of language testing in schools and universities). *Формирование и оценка коммуникативной компетенции билингвов в процессе двуязычного образования (Formation and assessment of communicative competency of bilinguals in bilingual education)*, pages 312–339.

Michael Levy. 1997. *Computer-assisted language learning: Context and conceptualization*. Oxford University Press.

Martha C. Polson, J. Jeffrey Richardson, and Elliot Soloway. 1988. *Foundations of intelligent tutoring systems*. Hillsdale, N.J.: L. Erlbaum Associates.

Jason R Rachels and Amanda J Rockinson-Szapkiw. 2018. The effects of a mobile gamification app on elementary students' spanish achievement and self-efficacy. *Computer Assisted Language Learning*, 31(1-2):72–89.

Robert Reynolds. 2016. Russian natural language processing for computer-assisted language learning: capturing the benefits of deep morphological analysis in real-life applications.

Fernando Rosell-Aguilar. 2018. Autonomous language learning through a mobile application: a user evaluation of the busuu app. *Computer Assisted Language Learning*, 31(8):854–881.

Nataliya Stoyanova, Jue Hou, Mikhail Kopotev, and Roman Yangarber. 2021. Integration of computer-aided language learning into formal university-level l2 instruction. *L'analisi linguistica e letteraria*.

Wim J van der Linden and Ronald K Hambleton. 2013. Handbook of modern item response theory. Springer Science & Business Media.

Maria Vilkuna, Riitta Korhonen, V Kovisto, Tarja Riitta Heinonen, and Irja Alho. 2004. *Iso suomen kielioppi (The Great Finnish grammar)*. Suomalaisen Kirjallisuuden Seura.

Elena Volodina, Ildikó Pilán, Lars Borin, and Therese Lindström Tiedemann. 2014. A flexible language learning platform based on language resources and web services. In *Proceedings of LREC: the Ninth International Conference on Language Resources and Evaluation*, pages 3973–3978.

Etienne Wenger. 2014. *Artificial intelligence and tutoring systems: computational and cognitive approaches to the communication of knowledge*. Morgan Kaufmann.

Cynthia White and Hayo Reinders. 2010. The theory and practice of technology in materials development and task design. Cambridge University Press.

Hui-Chin Yeh and Wei-Yun Lai. 2019. Speaking progress and meaning negotiation processes in synchronous online tutoring. *System*, 81:179–191.

Ruofei Zhang and Di Zou. 2022. Types, purposes, and effectiveness of state-of-the-art technologies for second and foreign language learning. *Computer Assisted Language Learning*, 35(4):696–742.