

# WordBricks: Mobile Technology and Visual Grammar Formalism for Gamification of Natural Language Grammar Acquisition

Running head: WordBricks Mobile Technology and Visual Grammar Formalism

Marina Purgina, Maxim Mozgovoy, and John Blake  
Email: {d8172102, mozgovoy, jblake}@u-aizu.ac.jp  
Tel.: +81-242-37-2664, Fax: +81-242-37-2528

The University of Aizu, Tsuruga, Ikki-machi, Aizu-Wakamatsu, Fukushima, 965-8580 Japan

Keywords: mobile-assisted language learning, learning environments, interface design, human-computer interaction

Biographical information

**MARINA PURGINA** received her M.E. in Computer Science from St. Petersburg State Polytechnical University. She is currently a Ph.D. student at Software Engineering Lab of the University of Aizu. Her research interests include natural language processing, artificial intelligence, and the principles of user interface design. She is a skilled mobile developer, and the principal author of WordBricks language learning system. Marina is a great lover of painting and photography, and tries to bring aesthetical value into visual components of software she designs. In addition to WordBricks, she is working on reinforcement learning algorithms for a machine learning-based game AI technology.

**MAXIM MOZGOVOY** received his Ph.D. degree in Applied Mathematics from St. Petersburg State University, and his Ph.D. in Computer Science from University of Joensuu. He is currently an associate professor at the University of Aizu, where he studies natural language processing and artificial intelligence. In particular, he is interested in machine learning, game AI technologies, educational software, and natural language processing. He has a solid academic record of over 50 published papers and industrial-level software development experience. His primary goal is to apply cutting-edge AI methods in practical projects. As an avid language learner, game AI specialist, and practical software developer, he is striving to design innovative and highly efficient language learning software.

**JOHN BLAKE** started his career teaching English as a foreign language in schools, colleges and universities in East Asia and the United Kingdom. John have lived and worked in Hong Kong, Thailand and Japan for extended periods. Over the years his professional interests have cycled through various roles, such as teacher trainer, assessor, and materials developer. He is currently an Associate Professor in the Center for Language Research at the University of Aizu. John uses corpus linguistics and discourse analysis to develop research-based materials to deliver learner-centered EAP or ESP classes that are enjoyable, challenging and relevant. He is an avid grammarian, and enjoys discussing any aspect of any grammar from anaphora to zeugma.

## Abstract

Gamification of language learning is a clear trend of recent years. Widespread use of smartphones and the rise of mobile gaming as a popular leisure activity contribute to the popularity of gamification, as application developers can rely on an unprecedented reach of their products and expect acceptance of game-like elements by the users. In terms of content, however, most mobile apps implement traditional language learning activities, such as reading, listening, translating, and solving quizzes. This paper discusses gamification of learning natural language grammar with a mobile app WordBricks, based on a concept of more user-centric lab-style experimental activities. WordBricks challenges users to create syntactically accurate sentences by arranging jigsaw-like colored blocks. Users receive instantaneous feedback on the syntactic compatibility each time any two blocks are placed together. This Scratch-inspired virtual language lab harnesses grammar models used in computational linguistics, and allows users to discover underlying grammatical constructions through experimentation. The system was evaluated in a

number of diverse settings, and shows how the principles of gamification can be applied to second language acquisition. We discuss general features that enable the users to engage in game-playing behavior, and analyze open challenges, relevant for a variety of language learning systems.

# 1 Towards Gamification in Second Language Acquisition

## 1.1 Motivation for Gamification

The practice of second language acquisition has been relying on modern concepts and the latest technologies for decades. In particular, language labs equipped with audio recording and playback facilities were in active use in 1970s-1980s (Khampusaen, 2013). Thus, the general idea of using modern technologies for language learning evokes little debate. The main discussions are related to the development of efficient ways of using the available instruments, and addressing their limitations. Technologically, language labs of the past century often suffered from unreliable tape-based systems and insufficient means of teacher control (Khampusaen, 2013). Methodologically, the prevalent audio-lingual method of teaching was considered inefficient and fell out of favor in the 1970s (Decoo, 2001). However, the technical issues were eventually resolved with the rise of computing machinery, and audio-lingualism was driven out by newer, presumably more efficient methods. In general, we believe that the modern practice of computer-assisted language learning (CALL) is evolving within according to the same principles. Language teachers are seeking to use the present technology (with its capabilities and limitations) in the most effective ways, while computer science specialists are trying to advance the technologies, providing more options for their practical use.

One of the most salient trends in modern computer-assisted education research is the rising interest to *gamification* of learning. While the idea to introduce certain game-like elements into learning is definitely not new, the word “gamification” came into wide use only in the 2010s (Morford, Witts, Killingsworth, & Alavosius, 2014), together with the surge of related research efforts. It is very likely that this process is connected to the rising popularity of smartphones and mobile games that turned a large number of phone owners into casual gamers. According to AdMob statistics, 59% of smartphone users install games within a week of getting their devices (AdMob, 2014), so “it is difficult to find a person now who hasn’t played at least one video game, making games more of an accepted and integrated part of our society” (Lavandier, 2013). In other words, certain exposure to computer or mobile game experiences can be expected now from a typical learner, so the developers of educational software can assume that the users perceive game-like elements as something familiar.

Following the work of Deterding et al. (2011), most authors draw a clear distinction between “gamification” (defined as “the use of game design elements in non-game contexts”) (p. 10) and related concepts of “(serious) games”, “toys” and “playful design”. While games and toys can be definitely used in educational context, the less restrictive concept of using “game design elements” can be arguably applied to a wider range of scenarios. Pure educational games often hide dull and repetitive tasks behind colorful graphics and animation, thus perceived as “chocolate-covered broccoli” by the users (Chen, 2016). Recent research efforts identified more subtle “fun factors”, such as *concentration*, *challenge*, or *immersion* that contribute to the enjoyability of the game experience (Sweetser & Wyeth, 2005). However, it is difficult to design a game that would combine engaging mechanics with high educational value. One relevant example is DragonBox Algebra game that lets the users practice solving linear equations. Experiments show that its visual formalism is hard to connect with the standard mathematical notation, so the students using DragonBox Algebra do not improve their math tests scores (Dolonen & Kluge, 2015; Long & Alevan, 2014).

Still, the motivation to combine game-like experiences with education is strongly supported with a simple observation: “since video games... can demonstrably motivate users to engage with them with unparalleled intensity and duration, game elements should be able to make other, non-game products and services more enjoyable and engaging as well” (Deterding et al., 2011, p. 10). Thus, gamification suggests a somewhat lightweight alternative to engaging in full-fledged educational game projects, evoking another question: what kind of “game elements” are able to motivate the users? Some authors express very bitter views on gamification, arguing that in practice it became a collective term for a number of exploitative techniques for increasing user spending, and has no relation to core game process (Bogost, 2011; Robertson, 2010).

Morford et al. (2014) introduces six basic traits of game-playing behavior:

- 1) direct impact on the game outcome and results;

- 2) clear goals and/or end conditions;
- 3) the presence of rules and barriers;
- 4) probabilistic outcome;
- 5) development of strategies and heuristics;
- 6) non-coerced initiation.

Strikingly, this list does not include elements or activities explicitly labeled as “fun”. One possible interpretation of the results obtained by Morford et al. is to conclude that many processes, possessing the stated traits (1-6) are *perceived as fun* or, at least, *engaging* by the users, and thus can be considered “gamified” even if they lack some other elements, typical for computer games (such as graphics, competitive gameplay, arcade controls or engaging narrative).

## 1.2 Technology for Conscious Learners

Wide adoption of technologies in education is primarily driven with advancements that make their use attractive for teachers and learners. One example of a technological success story is ubiquitous adoption of flight simulators in pilot training programs (Farmer, van Rooij, Riemersma, & Jorna, 2017). Computer-assisted language learning is also widely used in practice, if we follow its common definition as “*the search for and study of applications of the computer in language teaching and learning*” (Levy, 1997, p. 1). However, a closer look reveals that computer technologies are most commonly used in traditional learning activities, such as face-to-face chats, watching video clips and listening to audio, participating in group discussions, and reading texts. Consequently, the main emphasis is usually made on general-purpose computer instruments, such as Skype, YouTube, Facebook, and various online forums and resources (Chun, Kern, & Smith, 2016). On the other hand, dedicated CALL systems and specialized technologies are rarely mentioned.

The situation with specialized CALL instruments is somewhat complicated. To begin with, there is no general agreement about their effectiveness among experts, as shown in Hubbard’s survey, conducted in 2002. Hubbard notes: “...it is interesting that questions of effectiveness still tend to dominate. In fact, the basic questions of “*Is CALL effective?*” and “*Is it more effective than alternatives?*” remain popular even among those who have been centrally involved in the field for an extended period of time.” (Hubbard, 2002). We believe that, at least, in part this situation can be explained with relative immaturity of language processing technologies that could potentially be of great benefit for the learners. Among these technologies are machine translation, automated speech recognition, grammar checking, and feedback generation, to name a few. Automated speech analysis is used to certain extent (e.g., in Rosetta Stone software), but its quality is often criticized (Lewis, 2011; Santos, 2011). In addition, CALL applications are seemingly of limited interest for natural language processing community. As noted in (Amaral & Meurers, 2011, p. 7), “*the development of systems using NLP technology is not on the agenda of most CALL experts, and interdisciplinary research projects integrating computational linguists and foreign language teachers remain very rare*”. In addition, it is probably not easy to find the best use cases even for the existing language technologies in a way that would provide benefits for the learners despite technological limitations. Therefore, the progress in this field is hindered by the necessity of coordinated efforts between the teachers and technology experts, who have different agendas and constraints.

In this situation, gamification is an interesting direction of research, since it often deals with active *conscious learners* rather than participants of predominantly teacher-guided courses. Let us recall that one of the characteristics of game-playing behavior is *non-coerced initiation*, meaning that “*a player plays the game because he wants to, not because he has to*” (Morford et al., 2014, p. 30). Therefore, while gamification is possible both inside and outside the classroom, we believe that the best results can be achieved in voluntary user-initiated learning sessions, more closely resembling typical game-playing scenarios. Even if a certain application supports only basic traditional learning activities (such as reading and listening), it can reinforce user motivation and make the process of learning a language less burdensome (or more enjoyable, depending on one’s perspective). Regardless of a student’s attitude towards language learning, we must concede that this process involves numerous repetitive tasks and memory drills. For example, to master Japanese, one has to learn around 2000-3000 Chinese characters (Heisig, 2012). It is difficult to imagine learning strategies that would make this activity inherently fun and enjoyable. Indeed, most learners in practice rely on different variations of drills (Gamage, 2003) (mnemonics and other techniques still cannot liberate students from drilling sessions), so any technological tricks that make this undertaking less daunting should be appreciated. At the same time, it is not easy to estimate how many users would be willing to participate in such non-coercive game-like educational activities, and thus benefit most from

gamification. However, various studies conclude that at least people engaged in daily learning activities (such as university students) are willing to use their mobile phones for out-of-classroom learning as well (Foti & Mendez, 2014; Gikas & Grant, 2013). Thus, the ubiquity of mobile devices, wide spread of mobile gaming, and users' willingness to use mobile devices as learning tools constitutes a perfect combination for the success of gamification techniques at the present time.

### 1.3 Duolingo and Anki as Different Cases of Gamification

To illustrate some of the principles outlined in the previous section, let us quickly consider two successful language-learning applications, Duolingo and Anki, and discuss how these completely different kinds of tools agree with general principles of gamification.

Duolingo is often considered as one of the most successful language learning apps on the market with around 200 mln subscribers worldwide (Draycott, 2017). It received numerous positive responses by the users (Bogdan, 2016; Kumar, 2016), and its efficiency in keeping user attention and increasing language proficiency is reported in research literature (Castro, Hora Macedo, & Bastos, 2016; Huynh & Iida, 2017). The developers of Duolingo attribute their success directly to gamification. In particular, they mention four basic pillars of their approach (Draycott, 2017):

- 1) dissection of the large goal (learning a language) into a set of small daily user-chosen goals;
- 2) visual clues to track user progress;
- 3) emails and notifications for motivating inactive users to return to their studies;
- 4) rewards and achievement badges for continuous daily use (known as “the streak”).

Technically, Duolingo implements a number of traditional exercises, such as “translate a sentence”, “match words with their translations”, “type the pronounced phrase” or “pronounce the given phrase” (Karch, 2015). It is important to note that the developers do not try to cover ordinary tasks with a “chocolate layer” of a game. The exercises are presented in the same way as in conventional textbooks (see Figure 1). The application is clearly aimed at conscious learners who fully acknowledge that they are involved in a laborious and not always fun activity of learning a language rather than playing a game. Thus, Duolingo relies on more subtle principles of gamification, aimed to introduce game-playing behavior into learning. Indeed, typical learning sessions in Duolingo possess most of the traits of game-playing behavior listed in (Morford et al., 2014).

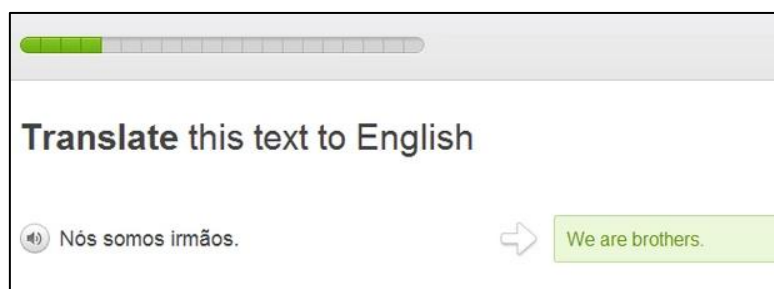


Figure 1. A fragment of Duolingo user interface.

Duolingo, however, is a showcase of success story that is hard to reproduce. The app implements vast functionality, so it is difficult to recommend following the same approach in smaller-scale projects. Therefore, it is interesting to consider the case of much smaller (in terms of functionality) project Anki that aims to create an intelligent flashcard organizer for desktop and mobile platforms.

At a glance, Anki is a plain and simple-looking application that implements only the required functions, necessary for its purpose, and does not adhere to any gamification principles. However, Anki is well known among language learners. The Android version of the app is installed on over one million devices, and is rated by over than 32,000 users (as of April 2018). It is also a subject of several research articles (Bailey & Davey, 2011; Librenjak, Vučković, & Dovedan, 2012), and often praised and recommended by the users (Kidd, 2014; Walker, 2015). Anki implements a space repetition procedure (Teninbaum, 2016) that constantly rearranges flashcards in such a way that new and poorly memorized cards are shown more frequently. This way, there is no need to review the whole deck of cards during each learning session: the system selects the cards for the next review automatically. In a sample session shown in Figure 2, the user has to recall the correct translation of the word “пациент”, and after revealing the

answer (“story, tale”) press the corresponding button. If the card is forgotten, the user should press “Again” and try this card again in a few minutes. Similarly, the button “Good” will schedule this card for review in 3 days.

While gamification was apparently not in the agenda of Anki developers, we should note some casual similarities between certain Anki features and deliberate game-like elements of Duolingo:

- 1) the large goal of memorizing the whole deck of cards is split into daily reviewing sessions;
- 2) users can track their progress by checking statistical data in a special window;
- 3) users are strongly encouraged to keep their “streak” and adhere to daily reviewing sessions to avoid a flood of unreviewed cards.

Since the users can freely create and share cards, there is even a certain social element in this activity. We also believe that reviewing sessions in Anki can be considered game-playing behavior according to (Morford et al., 2014).

The case of Anki shows that gamification does not necessarily have to be a well-thought strategy. Game-likeness can be an inherent property of a certain study process, so the software developers just have to recognize game-like elements and support them properly in their product.



Figure 2. Reviewing session with Anki. Source: Wikipedia.

## 2 The Game of Grammar: WordBricks Approach

Duolingo, Anki and numerous other related apps provide great examples of modern technology-driven way of supporting learning activities and learner motivation. However, in terms of content they represent traditional learning materials and exercises. Their true power comes mostly from mobility, multimedia capabilities, and game-like features, while the use of dedicated language processing technologies is still very limited (which probably can be explained with their relative immaturity, as we discussed above). In practice it means that some important elements of language learning remain outside the scope of most computer-assisted language learning tools. In particular, virtually no instrument can evaluate and correct user-constructed phrases, though such corrective feedback is usually considered an integral part of learning by many researchers (Bitchener, 2008) (some others, such as Truscott (1996), argue against it). In any case, feedback of some kind exists in most learning activities.

The effect of technological maturity can be seen in some educational tools, available for natural sciences, such as physics or chemistry. The foundations of these sciences are more precisely defined in mathematical terms, which opens new possibilities for educational software developers. For instance, Open Source Physics (Christian, Esquembre, & Barbato, 2011) and ChemCollective (Yaron, Ashe, Karabinos, Williams, & Ju, 2013) projects collect a vast amount of interactive simulations in physics and chemistry (see Figure 3). These instruments can be treated as “virtual labs” that enable the students recreate textbook experiments on their computers and even run their own experimental setups and analyze the outcomes. The equivalent of ChemCollective in language learning would be a virtual character (chatbot), able to discuss a range of predefined topics or engage in a free dialog with the user, and provide different kinds of feedback.

While the currently available technology cannot support such functionality, we argue that certain elements of user feedback can be automated. One example is automated speech analysis and recognition, mentioned previously. Another possible direction is the analysis of the structure of user-supplied text, implemented, e.g., in a Japanese language tutoring system Robo-Sensei (Nagata, 2009). In the subsequent sections, we will introduce our own system WordBricks that tries to gamify the process of grammar acquisition, using virtual lab approach, found in the systems like Open Source Physics and ChemCollective.

## 2.1 Exploring Grammar with Interactive Exercises

Natural language grammar is an essential topic of most common language courses. It is often integrated into general textbooks or covered in dedicated literature, such as well-known English grammar reference books by Azar & Hagen (2005) or Murphy (2012). While some educators, such as Stephen Krashen, argue against explicit teaching of grammar (Krashen, 2003), we are not taking sides in this debate and merely state that grammar as a subject is widely taught, and thus grammar acquisition can be considered a legitimate target for a CALL system.

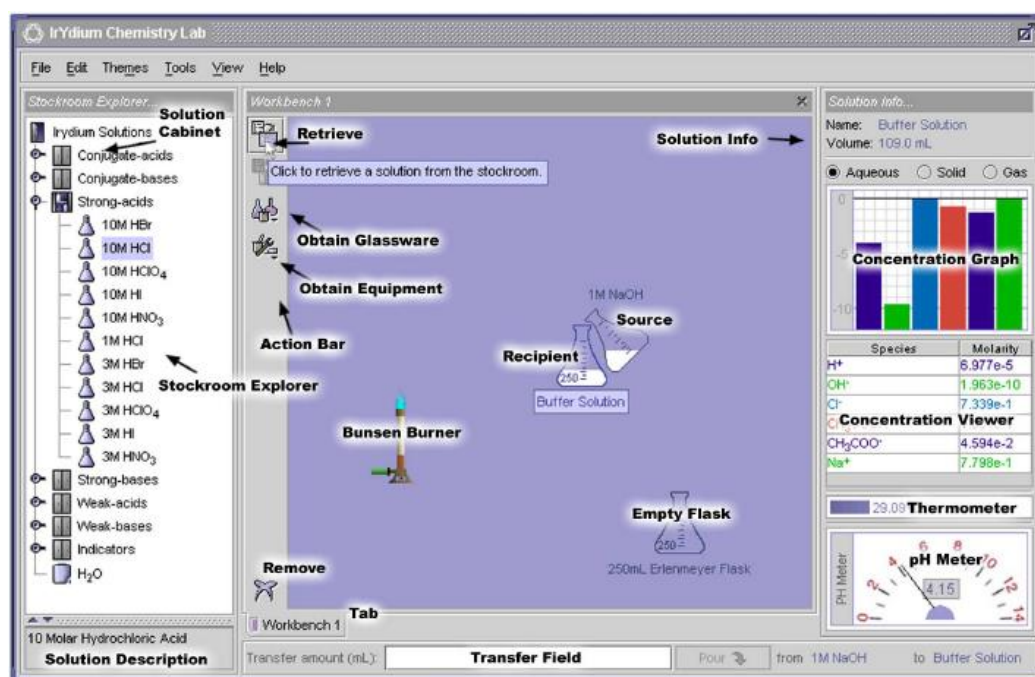


Figure 3. ChemCollective Virtual Lab. Source: chemcollective.org.

Today's actual grammar teaching practice is primarily focused on traditional exercises aimed at acquisition of proper grammatical forms and rules. Numerous studies indicate that most research on "innovative" grammar teaching methods have little impact on textbook content and classroom activities (Larsen-Freeman, 2015). Jean and Simard (2011, p. 467) note that "grammar instruction is perceived by both students and teachers as necessary and effective", and thus most educators are reluctant to abandon it, especially in the absence of universal agreement on possible alternatives.

There is, however, an ongoing discussion of particular ways to implement grammar instruction in practice. For example, common advice is to focus on student communication, and to draw attention to grammar forms arising naturally in the process rather than following a predefined list of grammatical structures (Long, 1991). Still, this approach can be implemented differently by different teachers, and there are no universally preferred ways to explain grammatical phenomena. For example, Larsen-Freeman (2000) suggests to focus on *reasons* rather than rules (e.g., while considering a sentence "There is a snowstorm coming", the teacher should explain that *there* introduces new information, and new information is marked with indefinite determiners such as *a*, rather than quote the corresponding formal grammar rule).

Certain attention is paid to the problem of balance between input processing and production activities (Shintani, Li, & Ellis, 2013) and to the creation of "focused tasks" designed to practice specific grammatical structures (Pica, 2005). In general, most conventional activities are not marked as "inherently (in)efficient" in research literature. Effectiveness depends primarily on their appropriate implementation.

Judging from typical grammar book contents, most common types of exercises require the language learner to form grammatically correct sentences. These exercises come in numerous variations, such as:

- 1) jumbled sentence: put the words in the correct order (possibly, with altering their forms);
- 2) fill the gap: fill the gap in a phrase using the appropriate word from the given list;
- 3) find errors: decide which phrases from the given list are grammatically correct;
- 4) rephrase: rewrite the given phrases using the specified grammatical construction.

The exercises are usually designed to have a single correct answer, provided in the “Answers” section.

We decided to elaborate this scheme by providing the user more interactivity and more visual clues, fostering better understanding of grammatical constructions. We believe that the lack of interactivity is one of the most salient shortcomings of traditional grammar book exercises. A learner can confirm own understanding of how to use certain words in certain combinations using the rules described in the given book section, but has no way to experiment with these words and rules. For instance, the learner might want to try substituting one word with another, using a word in another context, or combining two rules to formulate a more complex sentence. Our roadmap included the following scenarios (partially implemented at the present time):

- 1) The user sees on the screen a number of movable words, related to an individual exercise in a particular grammar book section. The task is to combine the words into a single sentence (so, this is a variation of a “jumbled sentence” exercise type). The user is also able to substitute certain words with their word forms.
- 2) In addition to the first scenario, the user is able to add new words related to the same grammar book section, and freely experiment with them (i.e., change their word forms and connect them into sentences).
- 3) The user can select any words from the available word bank and freely combine them.
- 4) The user can add new words to the word bank and analyze the structure of arbitrary sentences.

The viability of this plan (both in terms of technical feasibility and in terms of pedagogical value) strongly depends on a particular approach to visualization. In our case, graphics reflect a certain “visual grammar language” that directly influences learner perceptions and system capabilities.

## 2.2 Grammar Visualization Principles

Our approach to designing such a visual language is influenced by Scratch (Resnick et al., 2009), which is a system for learning the basics of programming. Programming languages have a grammar (albeit much simpler than human languages do), so it is essential to understand how individual instructions can be combined into complex structures. Scratch expresses grammar rules by representing instructions as blocks of different shapes, so that only matching blocks can be connected into a single structure (see Figure 4).

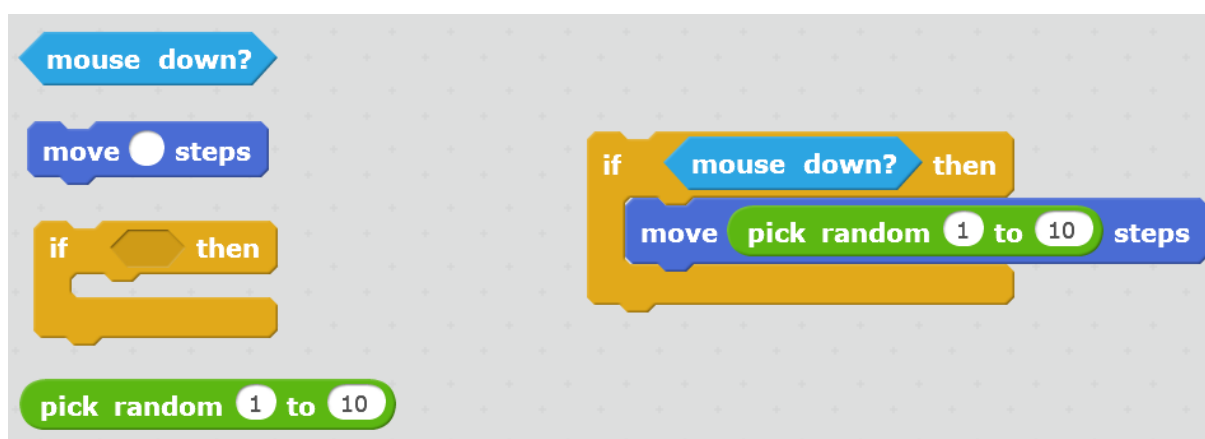


Figure 4. Combining blocks in Scratch.

Scratch’s graphical editor is not just a simpler way to write computer programs, but is helpful for the beginners. We believe that it can be treated as a construal (Gooding, 1990) that forms a model of a programming language in the learner’s mind. This way, the learner understands both the rules of grammar and the reasons why they work in

a certain way (because one cannot fit a rectangular block into a round hole). A similar idea is used to some extent in natural language learning as well (Ebbels, 2007) (see Figure 5).

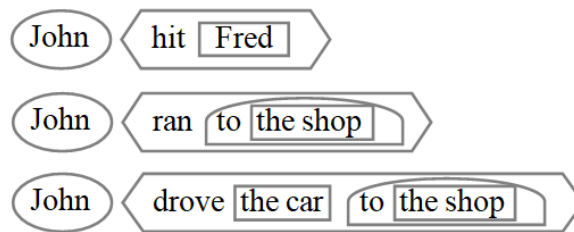


Figure 5. Shaped blocks in language learning materials (adopted from (Ebbels, 2007)).

Obviously, it is much harder to design a consistent set of blocks for a natural language than for a simplified programming environment. Words in natural language have many grammatical attributes (such as part of speech, gender, person and number), and the rules of grammar are often complex and contain numerous exceptions. Therefore, we do not strive for a perfect system (in fact, even Scratch blocks do not always adhere to the principle of shape matching), but aim to illustrate at least the basic phenomena of natural language grammar.

It is probably even more difficult to decide the logic of block arrangement inside an individual sentence. As shown in the Figure 4, a Scratch program resembles a two-dimensional jigsaw puzzle. Certain blocks in Scratch, such as “if...then” construction, have several “connectors”. Other blocks can be connected to the top or the bottom of an “if” block, placed inside it, or between the words “if” and “then”. Simpler blocks, such as “mouse down?”, can only be placed into the connectors inside other blocks. Therefore, it is necessary to decide what kind of connectors individual blocks should have, and how to arrange them on the screen so that they adequately represent syntactic structure of natural language sentences.

Existing linguistic theories approach the problem of sentence structure from different perspectives. We base our project on dependency grammar theory (Debusmann, 2000) that suggest connecting individual words in a sentence with direct links, reflecting “head/dependent” relationships. Dependency grammar formalism is widely used in natural language processing, and practical principles of dependency-based sentence markup are well documented (Marneffe & Manning, 2008). Our main motivation for relying on dependency grammar formalism was its resemblance to the structures of Scratch and to the Shape Coding system introduced by Ebbels (2007). Furthermore, dependency relations require no additional visual blocks (all blocks represent sentence words), which reduces the number of onscreen objects.

As a result of converting a sentence into a set of head/dependent pairs, we obtain a tree-like structure that has to be visualized. Unfortunately, common types of visualizations can be difficult to understand for non-specialists (see Figure 6). Therefore, we had to design our own scheme, somewhat similar to the Ebbels’s Shape Coding system.



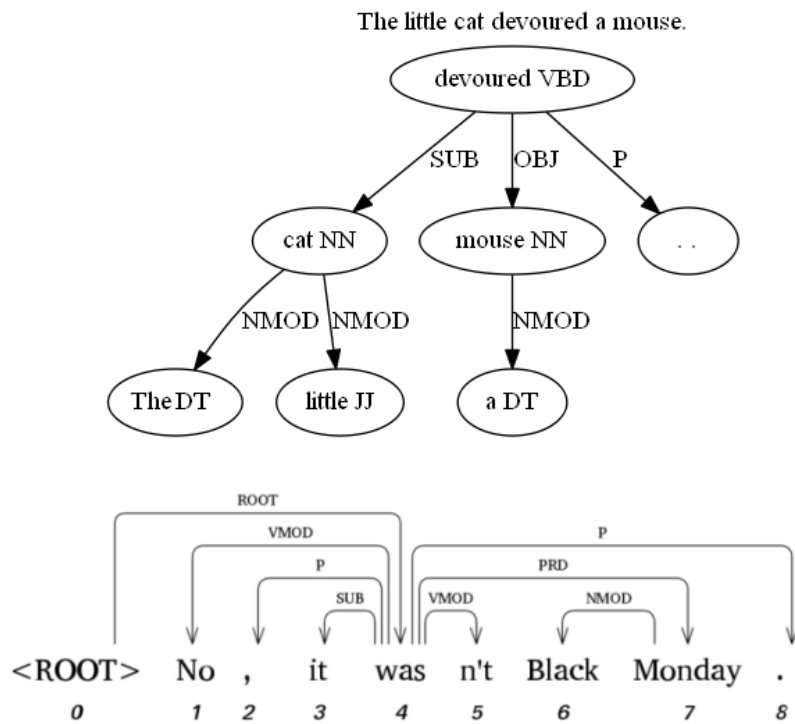


Figure 6. Visualizations obtained with AT&T GraphViz (above) and ZPAR parser (below).

### 3 WordBricks: User Interface and Capabilities

#### 3.1 General Design of the System

Initially, WordBricks was conceived as a desktop application (Mozgovoy & Efimov, 2013). It followed the traditional methods of visualizing word links and implemented a rich system of graphical elements represented various grammatical phenomena. Recently we redesigned WordBricks as a mobile app based on simpler and cleaner principles of visualization that greatly improved user experience.

In WordBricks, all syntactic elements of a sentence, such as words or the whole phrases are represented with blocks. The shapes and colors of such blocks depend on a set of their language-dependent grammatical attributes, such as part of speech, person, gender, and so on. Some blocks also have one or more same-colored connectors. Each connector is shaped according to the set of grammatical attributes associated with it. Connectors are “placeholders” for dependent syntactic elements, such as words or phrases. For example, most verbs have a connector for a subject in the left-hand side of the block, and for an object in the right-hand side of the block. If a shape of a connector matches a shape of a certain block, and the set of grammatical attributes of a block forms a subset of grammatical attributes of a connector, the user can insert the block into the connector.

In the current version of the system, the user has to select a particular exercise in the main menu, and the corresponding predefined blocks will appear on the screen. In addition, some optional blocks will be made available via the “word bank” menu of the application.

This way, the user sees on the screen the blocks of different colors and shapes, representing words and phrases, and can connect them to get a completed sentence. In many cases, the user only needs to make sure that the shapes of the block and the connector match, to join them together. If the shapes do match, but the attributes do not, the system will display a hint, explaining which mismatching attributes prevent the elements to be connected. In most of our experiments, the shape of a block is defined by its part of speech, but this configuration is flexible. Unfortunately, in practice it is difficult to show all grammatical attributes visually on the block, so we have to rely on the system of hints to provide additional error feedback to the user.

This method of displaying word links can be seen as a way to visualize dependency relationships, similar to the ones shown in Figure 6. Our approach enforces a certain word order in accordance to the order of connectors, and let us display the resulting sentence in a natural linear way (see Figure 7). However, it cannot handle *non-projective*

*dependencies* that rarely appear in English, but may constitute up to 25-27% of constructions in some languages such as Czech and German (Havelka, 2007).

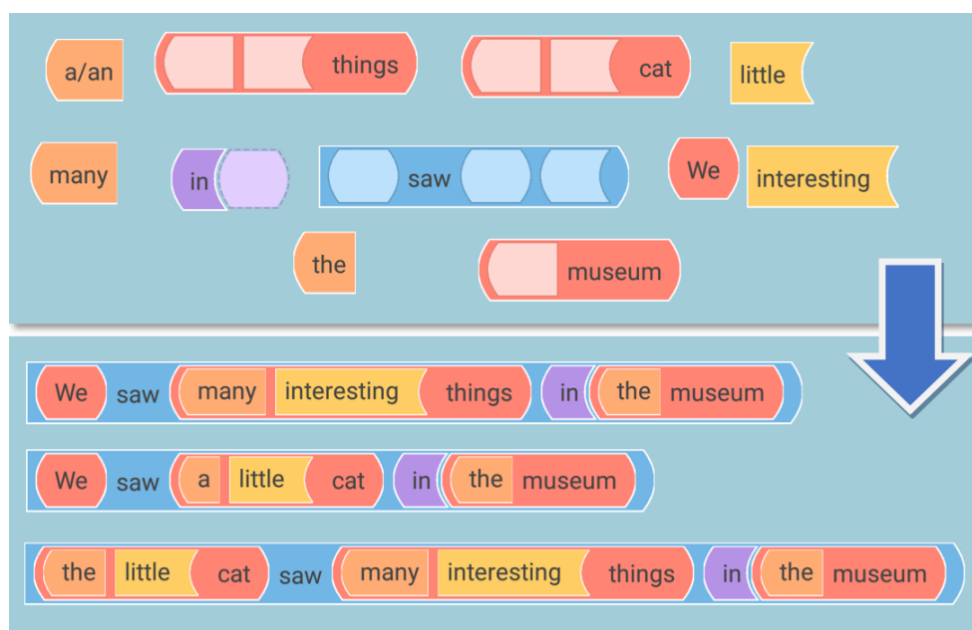


Figure 7. Combining bricks into sentences in WordBricks

Since WordBricks is a mobile application, it follows conventional touchscreen interface conventions. The user can move blocks and fragments of sentences in any direction on the screen using drag-and-drop (see Figure 8). The whole screen area except the menu bar at the top and the status bar at the bottom is used for brick arrangement. Double tap on a block opens its settings. Currently, the main functionality of the settings dialog is the selection of the desired word form. For example, if an exercise contains the word “cat”, the only way to obtain “cats” on the screen is via this dialog.

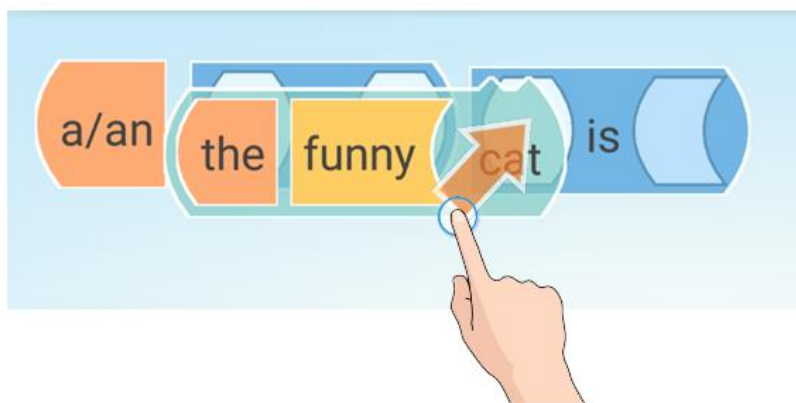


Figure 8. Combining blocks with drag-and-drop interface.

### 3.2 Technical Details

The current version of WordBricks is available for Android platform. It relies on standard functionality of the Android framework, and uses the capabilities of Android SDK classes to implement application logic (Purgina, Mozgovoy, & Klyuev, 2016). Previously, each block was represented with a widget based on a customized `View` class of the standard Android library (Purgina & Mozgovoy, 2017). However, nested blocks caused unacceptable performance drops, so we had to rewrite the entire block rendering functionality ourselves. The application draws all the blocks on the main view. Shapes, colors and the content of the blocks are rendered according to their XML definitions.

The configuration of each block is described in XML as follows:

Listing 1. XML description of the word *devoured*

```

<brick id="devoured_1" lemma="devoured" type="Verb phrase">
  <item type="brickConnector" connector="1" value="Noun phrase">
    <attrs case="common" person = "third" number="singular"/>
    <attrs case="nominative" person = "third" number="singular"/>
  </item>
  <item type="word">devoured</item>
  <item type="brickConnector" connector="2" value="Noun phrase">
    <attrs case="common" person = "third" number="singular"/>
    <attrs case="oblique" person = "third" number="singular"/>
    <attrs case="common" person = "second" number="singular"/>
    <attrs case="oblique" person = "second" number="singular"/>
    <attrs case="common" person = "third" number="plural"/>
    <attrs case="oblique" person = "third" number="plural"/>
    <attrs case="common" person = "second" number="plural"/>
    <attrs case="oblique" person = "second" number="plural"/>
  </item>
</brick>

```

Such description defines the content (words, grammatical attributes, and connectors) for all the blocks available in the given exercise. The format is designed to be simple and easy to use. The basic assumption is that any word form and any alternative set of attributes of the given word is described as a separate block. The final section of the XML file describes the exercises, and their expected solutions. Thus, the teacher needs to create an XML document with the description of words, syntactic forms and attributes with connectors to create a new exercise or a subset of language grammar for student experiments. The shapes of the blocks should be developed considering the most frequent combinations of syntactic phrases to further emphasize the correct order of the words with smooth transitions (see Figure 9). However, the teacher can change the shapes to better adapt WordBricks to another language or lesson. We are also planning to create a graphical tool to design XML rules without actually having to write XML.



Figure 9. Block shapes reflect the natural flow of words in a sentence.

## 4 Classroom Experiments with WordBricks

In this section, we will discuss several experiments conducted to prove feasibility of WordBricks. We wanted to evaluate its pedagogical merits in diverse environments and scenarios. To the present day, three independent studies were completed. The first study was aimed to prove that WordBricks can be helpful as a learning aid in a conventional English as a second language classroom, taught for the students of computer science at a Japanese university. The second study explored the capability of WordBricks to serve as a demonstration tool used by the teacher to illustrate certain grammatical phenomena. It was conducted with a different group of students at the same university. The goal of the third study was to test WordBricks in a course of some other language rather than English. WordBricks is, in principle, language-independent, but the difficulty of expressing grammatical constructions with specific types of visual blocks may vary. This study was performed at an Irish language class of a junior school in Ireland. As a part of each experiment, we also asked the users to provide their feedback on the interface of WordBricks and their suggestions for its subsequent development. Some of these suggestions were implemented in later versions of the software.

The diversity of experimental setups and different approaches to evaluation of the system is driven by the needs of teachers and students participating in our studies. As mentioned above, WordBricks was initially designed as a tool for “conscious learners” who would download the app and use it for their own language learning needs (like Duolingo). However, teacher interest to the system motivated us to do a series of pilot studies in classrooms, which would give us some perspective on the possibility to use WordBricks in schools. Experimental settings reflect the difference in educational goals. The teacher in the first study was motivated to improve his students’ test scores. Many of these students were intended to re-take a TOEIC exam after the course, and wanted to see how their

results improve after the course. The teacher in the second study conducts a dedicated English grammar course, based on traditional rules-and-exercises textbook. He was looking for a way to provide better visualizations of grammatical phenomena he had to explain (mostly in a non-interactive style). The teacher in the third study deals with young learners of primary and junior high school, having low motivation to study Irish language, which is widely regarded as a compulsory subject with little practical utility. Thus, her primary interest was to introduce interactive, game-like experiences that would increase learners' motivation to engage in educational activities. Therefore, our evaluation concerns three loosely related categories of merits of the app: a) capability to serve as a learning aid that facilitates better understanding of language grammar (that results in higher test scores); b) capability to serve as a visualization tool for illustrating particular grammatical points; c) capability to introduce game-like elements that make educational process more enjoyable even if it does not immediately translate to language proficiency.

#### 4.1 WordBricks as a Learning Aid

Let us begin with a brief discussion of our first attempt to use a relatively early version of WordBricks in a real classroom environment. Our goal was to gather initial responses from the teacher and the students, and assess the feasibility of the chosen approach (the study is described in more detail in (Park, Purgina, & Mozgovoy, 2016)). The participants of the experiment were two randomly chosen groups of predominantly male sophomore computer science students at a public Japanese university. The students were enrolled in an elective advanced English grammar course, and ranged in age from 19 to 21 years. First, we conducted a diagnostic English grammar test to ensure that the pre-course English level of both groups is very similar (see Table 1). Approximately one half of these students have passed a TOEIC test recently, their average score is 350 points.

Table 1. Descriptive Statistics for Two Groups in Diagnostic Test

<b>Group</b>	<b>Number of participants (N)</b>	<b>Mean Test Score (M)</b>	<b>Standard deviation (SD)</b>
Control Group (G1)	11	65.25	7.50
Experimental Group (G2)	10	65.90	7.99

To investigate whether WordBricks had any observable effect on students' grammar learning, we adopted a pre-test/post-test design with a control group and experimental group. In this setup, all 21 participants studied units 69 and 70 from the same *English Grammar in Use* textbook (Murphy, 2012) with the same teacher. Both units are about countable and uncountable nouns. However, Unit 70 seems to be more demanding than introductory Unit 69, since it is dedicated to more advanced grammatical points.

Though each group covered the same content and underwent the same English grammar assessment procedures, the control group was taught with an English grammar textbook in a traditional way (teacher-centered, grammar focused), but the experimental group autonomously interacted with WordBricks using Android-based tablets, given to each participant. Before the experiment, control group members could familiarize themselves with WordBricks by solving predesigned exercises, based on the first paragraphs of the Azar and Hagen's grammar book (Azar & Hagen, 2005).

Based on the course textbook, we developed two sets of paper-based English grammar tests to measure participants' English grammar performance over two course units. For Unit 69 pre-/post-test, participants were asked to correct given sentences focusing on the nouns of the sentences. For Unit 70 pre-/post-test, they were asked to complete sentences using correct noun form. According to the test results, the experimental group showed greater improvement for both topics. The average score of the experimental group (G2) increased from 15.90 to 24.20 points (out of 30 possible) for the first grammar topic, while the average score of the control group (G1) increased from 15.18 to 21.00. Subsequently we conducted a similar experiment with a group of 16 students who studied material of both units 69-70 as a single block, where the average score improved from 17.13 to 20.69 for the control group, and from 17.94 to 20.31 for the WordBricks group (see Table 2). To compare groups' achievements, we performed a paired samples t-test using aggregate pre-test/post-test data of G1 and G2. The resulting values (0.00005 for G1, 0.00026 for G2) indicate that both groups achieved statistically significant progress, and

WordBricks (G1) group performed better. These results show that our application can be as efficient as a textbook, at least, in certain contexts.

Table 2. Results of the Quantitative Experiments

Exp.	Test type	Group	Group size	Mean score	Std. dev.
#1 Unit 69	pre-test	G1	11	15.18	5.04
		G2	10	15.90	4.43
	post-test	G1	11	21.00	5.80
		G2	10	24.20	4.02
#2 Unit 70	pre-test	G1	11	6.00	2.72
		G2	10	4.20	2.57
	post-test	G1	11	9.18	4.17
		G2	10	11.60	2.84
#3 Units 69-70	pre-test	G1	8	17.13	3.80
		G2	8	17.94	4.64
	post-test	G1	8	20.69	2.91
		G2	8	20.31	2.83

Figure 10 illustrates progress achieved by individual students of both groups. Test scores indicate the overall percentage of correctly accomplished assignments. The diagrams show that in both groups teaching was effective, and students in general were able to improve their test scores. Most progress was made by the participants having lower initial scores, which is not surprising. It is also noticeable that the difference in pre-test and post-test scores is larger in the WordBricks group.



Figure 10. Progress of individual students in WordBricks and control groups.

## 4.2 WordBricks as a Demonstration Tool

Early classroom experiments with WordBricks made us believe that the system can also be used by the teachers as a demonstration tool within the framework of traditional grammar instruction. Currently, teachers often rely on PowerPoint or similar presentation software to present grammatical concepts or analyze the grammar of sentences (Hu, Clark, & Ma, 2003). However, PowerPoint presentations are comprised of linearly organized display units (slides) (Farkas, 2005), which might not be the best way to present non-linear or hierarchical concepts to the audience. When using presentation software, language teachers tend to use different fonts, colors and shapes to visualize the grammar of a sentence. WordBricks alleviates this need by providing ready-made building blocks for

typical grammatical structures. More importantly, as discussed above, their selection was based on established theories of grammar (Debusmann, 2000) and pedagogically sound approaches (Ebbels, 2007). Therefore, there is no need for language teachers to create customized elements, saving their time and providing advantage over more generic demonstration tools.

The possibility of using WordBricks as teacher’s aid was evaluated in a small group of seven 4-year computer science undergraduate students (22-25 years of age), enrolled in an Advanced English course at a Japanese public university. This course is primarily focused on helping the students write a graduation thesis, which is structurally similar to short research articles in computer science. As part of this thesis writing course, useful sentence structures are discussed using sentences extracted from an example research article (Washio & Watanabe, 2014). We selected ten sentences for division into WordBricks blocks to explain particular target structures (see Table 3). Since the course was focused on larger structures than individual words, we had to design custom WordBricks blocks that represent these structures (see Figure 11).

Table 3. Model Sentences and Target Structures

#	Sentence	Target structure
1	Secret sharing is a method of encrypting a secret into multiple pieces called shares so that only qualified sets of shares can be employed to reconstruct the secret.	A is a method of B so that C can be D.
2	Audio secret sharing (ASS) is an example of secret sharing whose decryption can be performed by human ears.	A is an example of B whose C can be D.
3	This paper examines the security of an audio secret sharing scheme encrypting audio secrets with bounded shares, and optimizes the security with respect to the probability distribution used in its encryption.	This A examines B and optimizes C.
4	Figure 1 illustrates an example of two shares and their superposition of a (2; 2)-threshold VSS scheme.	Figure # illustrates X.
5	Desmedt <i>et al.</i> [4] proposed information-theoretically secure schemes that encrypt a binary string secret.	X [#] proposed A that B.
6	It is conventional to use the mutual information to measure the statistical independence between random variables.	It is A to B.
7	Let P be a finite set, and let $A_Q$ and $A_F$ be subsets of $2^P$ .	Let A be B.
8	Table 1 summarizes the existing works on ASS and VSS schemes as well as this work.	Table # summarizes A.
9	First, we provide a formal definition of ASS schemes and a construction of the simplest ASS scheme, namely a (2; 2)-threshold ASS scheme.	We provide A and B, namely C.
10	The result indicates that the security is optimized when the variance of the normal distribution approaches infinity.	A indicates that B when C.

Since WordBricks is mobile software, we had to setup a virtual Android machine using Oracle VirtualBox on a teacher’s Windows 10 laptop to display presentations. In total, five presentations of 15-20 minutes were delivered. Each presentation analyzed two new model sentences, and reviewed previous sentences. Presentations consisted of the teacher constructing sentences using the blocks while eliciting and explaining the reasons from the placement of each block. Both suitable and unsuitable selections of blocks were made to provide students with opportunity to contribute ideas. After each presentation, students worked in pairs or threes to discuss the two new target structures. This was followed up with a writing task in which students created a sentence related to their research using the target structures.

As a result of this study, the teacher identified several strong features of WordBricks, helpful in the context of demonstrating grammatical constructions. In particular, he noted that automatic handling of colors, shapes, and grammatical correctness of the resulting phrase helps to avoid errors during presentation, when attention is focused on the audience rather than manipulating elements using a mouse. The interactive nature of WordBricks that allows to build sentences gradually, block by block, was also mentioned. Finally, it was suggested that preparing

demonstrations in WordBricks can be faster than using other demonstrational software, though at the present moment it requires manual XML file editing.

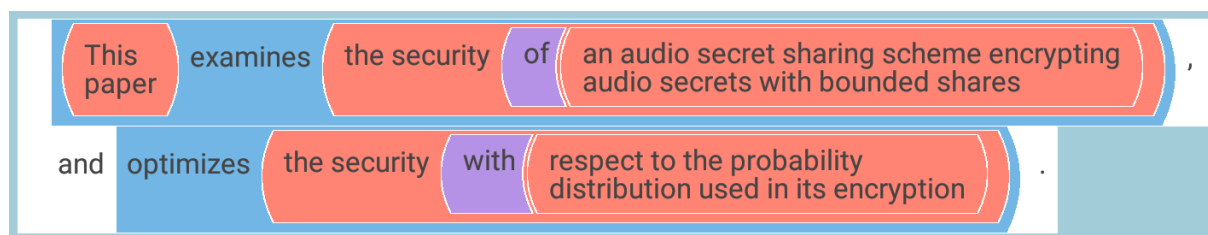


Figure 11. Model sentence #3 chunked into grammatical units.

### 4.3 WordBricks at an Irish Language Class

WordBricks was initially designed to be language-independent in a sense that it is not based on presumptions, specific for certain particular language. However, we wanted to have practical evidence that WordBricks is flexible enough to handle non-English grammatical structures. Our current experiments are focused on integrating WordBricks into Irish language classes at a junior school in Ireland (Purgina, Mozgovoy, & Ward, 2017). We also hope that gamification of Irish classes through WordBricks would help to support learners' interest in the subject. While the Irish language is a compulsory subject in Irish schools, only a very small minority (3%) of country's population use Irish as a community and household language (Government of Ireland, 2006). These Irish speakers are bilingual (Irish/English) and there is no communicative need to learn Irish (Watson, 2008). The overwhelming majority of people in Ireland (82%) believe that Irish should be taught in schools (Darmody & Daly, 2015), but in practice many learners tend to struggle both with the language and with lack of motivation.

The current version of Irish WordBricks application deals with some of the basic constructs of Irish that learners must master, yet find difficult due to their structural difference from English. Most school teaching of Irish follows the traditional model of books, workbooks and teacher-led activities. Irish WordBricks introduces some game-like aspects by enabling learners to construct their own grammatically correct sentences in Irish, and reinforce Irish word order. For example, the phrase "I have a hat" is "Tá hata agam" in Irish (literally, "Is a hat with me"). Learners can find this structure difficult, as they may try to map the Irish words onto the English word order, and WordBricks can help them to see the real word connections in this phrase (see Figure 12).

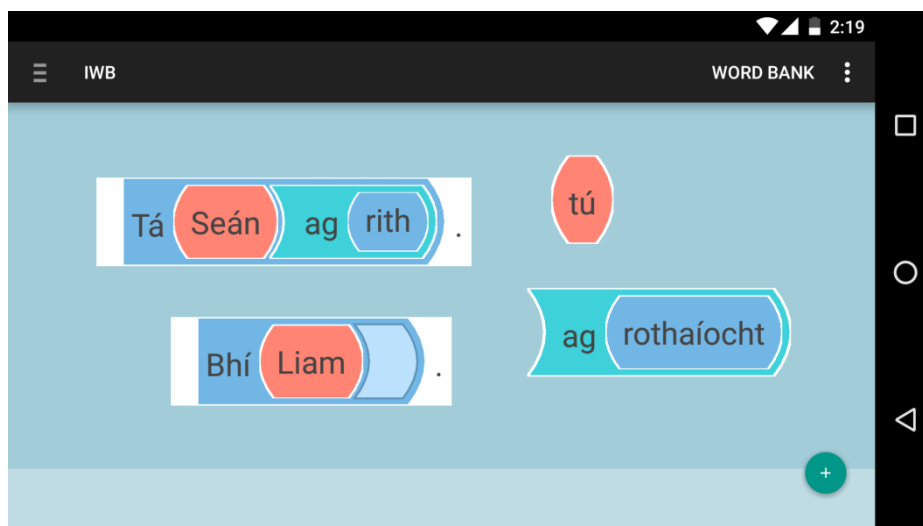


Figure 12. Irish "have" construction in WordBricks.

We conducted several pilot studies to find out how enjoyable is the system for users, and the initial responses of the teachers, parents, and school pupils were positive (Ward, Mozgovoy, & Purgina, 2018). The learners in general reported that they enjoyed working with the application, found it easy to use, and would like to use it as a part of their homework. The parents also enjoyed using the application and thought it was a very good idea to have such an application for Irish. Several parents reported that they struggle to help their children with their Irish homework and have tried in vain to find something useful for them as parents to either revise their knowledge of Irish or learn

it from scratch in the case of immigrant parents. Several primary school teachers also reviewed Irish WordBricks. They were positive about the application, and found its interactive elements appealing for their students. Even though Irish WordBricks was initially designed for a single user in an independent learning situation, the teachers plan to use the application in their classrooms. The idea is to ask the students to form sentences using the classroom computer so that all students can see and become familiar with the grammatical structure being studied.

The first study involved a mixed group of 46 school students, 8-12 years of age. Our goal at this stage was to gather their initial impressions about the app and understand possible directions of subsequent development. We asked the students to play freely with the app, perform basic assignments, and analyze the structure of several suggested sentences.

The purpose of the second study was to test the applicability of WordBricks in real classroom setting, i.e., when it is used to illustrate a particular language phenomenon according to the plan of a lesson, and the app is primarily used by the teacher rather than students. WordBricks was ran with Android emulator installed on a desktop machine. Each grammar topic was illustrated with two or more example sentences. The participants in this case represented two cohorts:

- a) two groups of 5<sup>th</sup>-year school students (10-12 years of age), 44 participants in total. Had 7 years of Irish language education, including 5 years of written Irish. Worked with five different grammatical constructs over a five-week period.
- b) three groups of 3<sup>rd</sup>-year school students (8-9 years of age), 75 participants in total. Had 5 years of Irish language education, including 3 years of written Irish. Worked with three grammatical constructs over a four-week period.

The results of student surveys are summarized in Tables 4 and 5.

Table 4. Summary of student responses of the first pilot study (46 students)

Question	Yes (%)	No (%)	A bit (%)
Did you enjoy the Irish WordBricks app?	95	0	5
Did you find the Irish WordBricks app easy to use?	78	0	78
Do you think IWB helped you to learn Irish?	78	2	20
Would you like to use the Irish WordBricks app at home?	56	2	42

Table 5. Summary of student responses in the second pilot study (119 students)

Question	Yes (%)	No (%)	A bit (%)
Did you find WB easy to use?	82	2	16
Did you think it helped you to learn Irish?	73	7	20
Would you like your teacher to use WB in class?	81	5	14
Did you enjoy WB?	84	1	15

We recognize that the actual impact of WordBricks on language education has to be reconfirmed with both quantitative and qualitative evaluations involving larger groups. We must note, however, that such experiments are hard to perform in the context of minority-language primary school classes, where there are many variables outside our control (including differences in teaching approaches, student abilities, textbooks used, and class size), and the total number of learners is limited. Still, we consider the obtained results encouraging, as they demonstrate the feasibility of our approach, serving as primary motivation for subsequent work.

#### 4.4 User Suggestions

As one might expect, we also received numerous reports on shortcomings of the present version of our software. Some were related to particular bugs or user interface inconsistencies that were later improved. Some users suggested to turn WordBricks into a universal application, able to work on both mobile and desktop operating systems, which is especially important for its use as a demonstration tool. There were some specific requests such as “let the user save the current block arrangement to return to the sentence later”. Other comments reveal common needs of language learning software users in general, so we believe that they deserve additional discussion.



It seems that certain fraction of students in any given group has a natural proclivity for teacher- and book-centered learning. They perceive mobile apps as “not serious” types of learning aids, and often ask to make WordBricks more “book-like”, for example, by following a traditional structure of sections, containing explanations, examples, and exercises. Some of these students are also sufficiently proficient, so they found WordBricks content irrelevant for their knowledge. They also feel more comfortable when the application is designed as a “textbook companion”, containing exercises strongly following the textbook structure and vocabulary.

Likewise, certain users like WordBricks just for the sake of being a mobile app, as they find appealing more “technological” way of learning a language. Such people explained their positive attitude with responses such as “*I like fiddling with a tablet*”, “*WordBricks is like puzzle games, and I enjoy to study and play games*”.

Probably, the largest number of suggestions were related to the current limited set of supported words and constructions. Users found the system too restrictive, as it only implements predefined constructions taken from the textbook units used in our experiments. Some users explicitly requested the capability to add own words and rules for independent studies.

Another large portion of suggestions was directly related to gamification. During the above described experiments, users had to deal with a “plain” version of WordBricks implementing only the basic functionality of building up phrases from blocks, and containing no explicit game-like features. However, they immediately recognized potential for further gamification, and requested to implement simple additions, such as victory fanfare sounds, scoring system, and explicit user progression through goals and subgoals. This observation supports our earlier note that mobile gaming is such a common leisure activity nowadays so that the users often suggest moving the project further into this direction themselves (we must note though that most of our testers are young people, more likely to be engaged in gaming). Here we should also mention common requests to implement a system of hints and other feedback mechanisms.

## 4.5 Open Challenges

It would be unbalanced to focus on WordBricks advantages without discussing principal shortcomings of the present system. User interface limitations, limited vocabulary or inadequate user feedback can be addressed, but there are also harder challenges that should be discussed, as they highlight inherent difficulty of natural language, and problems relevant to a wider range of CALL systems.

**Scalability issues.** The present version of WordBricks assumes that the user picks up blocks from a “tray” and moves them to the main application window. Alternatively, a predefined set of blocks is assigned to a specific exercise, so when the user opens the exercise, the corresponding blocks appear in the main window. This approach is hardly applicable for large word lists, and organizing words into classes according to their part of speech is not sufficient either. We are working on a combination of word tray and text input interface to facilitate easier search of words.

**Unintuitive structures.** Dependency grammars provide intuitive word-linking rules for simple types of dependency, such as “subject-verb” or “noun-property”. However, for certain structures these rules are often based on conventions rather than on rigorous linguistic theory. They include relations between words making up proper names (such as “Joe” and “Doe” in “John Doe”); relations between the main and subordinate clauses; relations between the words in phrasal verbs (such as “look up”); relations involving auxiliary words, such as “have been doing”, and so on. Arguably, understanding sentence structure is beneficial to the learner (explicit structural diagrams are used, e.g., in Richard Webb’s *80/20 Japanese* textbook (Webb, 2016)), but some of the present constructions can be more confusing than helpful.

This situation can be improved to some extent by designing blocks corresponding to separate *logical entities* rather than words. For example, we can consider the construction “will have been” as an atomic block, thus removing the need to examine the relations between words inside this entity. In fact, this approach is in line with the original concept of dependency grammars described by Lucien Tesnière, who distinguished *words* as syntactic elements from *nuclei* as compound elements carrying the same role as words (Kahane, 1996).

**Interface/visualization constraints.** Many blocks should have optional, variable or dynamic list-like connectors, while the current system assumes that blocks have predefined connectors, specified in the configuration. For example, nouns can have optional associated properties (“[large] book”), the verb *to be* can be used with a noun

or an adjective as an object (“I am a student / I am funny”), and many verbs can be linked with a number of indirect objects (“I bought a book [where / when / why]”).

One way to handle such flexibility is to let the users to add, remove or change block connectors while arranging sentences, if these changes do not violate grammatical rules. In other words, the system will provide certain “basic” blocks, and it will be a user’s responsibility to configure them properly. Such a method is adopted in Scratch. It provides a range of mathematical functions, such as  $\sin(x)$  or  $\log(x)$ , but the user sees only the  $\sqrt{x}$  block in the tray. Other functions are accessible via a drop-down list of the  $\sqrt{x}$  block. However, we must acknowledge that this approach will make user interface more complicated and will introduce new required actions into sentence building.

In addition, as noted above, our current visualization subsystem supports projective dependences only. However, so far we had no requirements to deal with non-projective dependencies in practice.

**Pedagogical considerations.** One may feel compelled to use WordBricks to encode a large number of specified rules of grammar. However, the flexibility of natural language grammar lets the system to interpret certain constructions as correct, while in practice they are most likely to be erroneous. Many “grammatical rules” described in textbooks are actually dictated by semantics rather than syntax. For example, *English Grammar in Use* (Murphy, 2012) clearly states: “we do not use *the* with names of people (‘Helen’, ‘Helen Taylor’, etc.)”. However, a book on advanced grammar provides a case where *the* is used to disambiguate the subject of speech: “that’s not *the* Stephen Fraser I went to school with” (Hewings, 2013). Similarly, rules related to the choice of past vs. present perfect tenses in English often mention that the words *already*, *yet*, and *just* are used with present perfect tense (Murphy, 2012). However, one may argue that they deal with semantics rather than syntax. Syntactically, adverbs (such as *already*) can be used with any verb forms. Finally, it is unreasonable to accept student-produced sentences that can be considered grammatical only with the help of counter-intuitive interpretations, such as in the classic garden-path sentence “*The old man the boat*” that relies on the meaning of “man” as “operate” (Guo, 2016).

The teachers designing the blocks have to decide which constructions are include and which are exclude, given the target level of learners. It is far more likely that the beginners will erroneously use “the” with a person’s name rather than do it correctly in few situations where it is acceptable. However, many cases require deeper involvement of semantics, and thus are beyond the scope of WordBricks.

## 5 Discussion

State-of-the-art technologies have been used in language education for a long time. One of the recent trends is the rise of gamified mobile apps for language learning, supported by widespread reach of smartphones, and by the rise of mobile gaming as a popular leisure activity. This allows application developers to presume that many of their potential users are ready for game-like activities, and even expect to experience them in non-game apps. Language learning requires long-time commitment, and often involves going through routine tasks that hardly can be considered entertaining, so reasonable attempts to exploit human propensity for games should be supported. However, it might be tempting to interpret this suggestion too literally and endeavor to develop a real “educational game”, which in practice often turns out to be a substandard educational tool, and a substandard game. Successful projects are typically targeted at conscious learners and do not try to disguise themselves as “games”. Instead, they implement certain game-inspired tricks that help the users to stay on track.

In terms of content, most projects are based on traditional learning materials, (such as texts for reading, audio- and videoclips, and textbook-style explanations), and traditional exercise activities (quizzes, jumbled sentences / fill the gap / translate phrases grammatical exercises). We believe that natural language processing technologies are potentially able to support a variety of innovative educational scenarios, not available with traditional learning materials, but in practice few technologies are mature enough to reliably address learners needs. For example, automated speech analysis is often criticized for providing misleading feedback.

Our primary motivation for creating WordBricks was to explore certain “technology-driven” educational scenarios that would make use of dedicated technologies, specifically designed for a purpose of language learning. At the same time, we tried to address the problem of technological limitations by restricting the users with activities that can be reliably supported. For example, it is nearly impossible to design a reliable grammar checker that would

evaluate any given sentence and find errors. However, it is possible to restrict the users with the set of grammar rules, and let them compose sentences that are considered correct according to the rules. Our current experiment show potential of this approach, and WordBricks is regarded highly both by teachers and learners. However, the flexibility of human language and the lack of formalized grammar rules presented in a textbook order makes the design of WordBricks exercises a very nontrivial and challenging task. Fortunately, in many scenarios it is sufficient to implement the structures that makes sense from a pedagogical point of view, which is only a subset of all grammatically correct constructions. These considerations give us the motivation to continue experiments.

## 6 Conclusion

In this paper, we have briefly discussed the rising gamification of language learning via mobile apps, and introduced our work-in-progress system WordBricks, targeted for natural language grammar acquisition. WordBricks allows the users to combine words into sentences using Scratch-inspired “blocks and connectors” approach that prevents them to form ungrammatical constructions. Currently, the system supports three primary use cases: 1) as an “open lab” for free experiments with language grammar structures; 2) as an exercises platform to be used in combination with a grammar textbook; 3) as a demonstration tool for a teacher.

We are evaluating WordBricks in diverse settings, involving different educational goals, student profiles, and different target languages. Our first experimental setup confirmed that the system was able to help students to improve their English grammar test scores within the context of a dedicated grammar course. The second study demonstrated the capability of WordBricks to serve as a handy visualization mechanism of particular grammatical constructions in a primarily non-interactive lecture-based course. The third study emphasized user enjoyment and game-like elements of the app, appealing to young learners with low motivation to learn a language, taught as a compulsory school subject.

Our evaluation shows that the chosen approach is regarded positively by all involved parties. Students feel game potential in the app, and request for more game-like features, such as the ones found in Duolingo. Implementing them is our primary goal. At the same time, we have to admit that even formal adherence to textbook grammar cannot hide the whole degree of complexity of natural language. Grammar rules often rely on vaguely defined categories, semantics, and general knowledge, and thus can be hard to implement in WordBricks. Furthermore, the system of blocks provides an impression that all constructions are “equal” in a sense that they are equally correct according to the rules of grammar. However, in practice from a didactical point of view it might be preferable to stick to fewer rules, and to introduce less commonly used constructions at later stages.

To extend current experiments, we are also working on an improved and simplified version of XML format, describing blocks and block linkage rules. Ultimately, we are planning to make this process accessible to a wider audience of educators and language learners. In general, we hope to see more works in technology-driven language education, and more apps implementing innovative approaches to facilitate second language acquisition.

## References

- AdMob. (2014). *Six Essential Tips for App Developers*. Retrieved from <https://www.thinkwithgoogle.com/advertising-channels/apps/six-essential-tips-for-app-developers>
- Amaral, L. A., & Meurers, D. (2011). On using intelligent computer-assisted language learning in real-life foreign language teaching and learning. *ReCALL*, 23(1), 4–24.
- Azar, B., & Hagen, S. (2005). *Basic English Grammar, 3rd Ed*: Pearson Longman.
- Bailey, R. C., & Davey, J. (2011). Internet-based spaced repetition learning in and out of the classroom: Implementation and student perception. *CELE Journal*, 20, 39–50.
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of second language writing*, 17(2), 102–118.
- Bogdan, D. R. (2016). Duolingo as an “Aid” to Second-language Learning. An Individual Case Study. *Bulletin of the Faculty of Education (Ehime University)*, 63, 199–212.

- Bogost, I. (2011). *Persuasive Games: Exploitationware*. Retrieved from [https://www.gamasutra.com/view/feature/134735/persuasive\\_games\\_exploitationware.php](https://www.gamasutra.com/view/feature/134735/persuasive_games_exploitationware.php)
- Castro, A., Hora Macedo, S., & Bastos, H. (2016). Duolingo: An Experience In English Teaching. *Journal of Educational & Instructional Studies in the World*, 6(4), 59–63.
- Chen, S. (2016). *Facing Edutainment's Dark Legacy*. Retrieved from <http://www.gamesandlearning.org/2016/01/25/facing-edutainments-dark-legacy/>
- Christian, W., Esquembre, F., & Barbato, L. (2011). Open source physics. *Science*, 334(6059), 1077–1078.
- Chun, D., Kern, R., & Smith, B. (2016). Technology in language use, language teaching, and language learning. *The Modern Language Journal*, 100(S1), 64–80.
- Darmody, M., & Daly, T. (2015). *Attitudes towards the Irish Language on the Island of Ireland: The Economic and Social Research Institute*.
- Debusmann, R. (2000). An introduction to dependency grammar. *Hausarbeit fur das Hauptseminar Dependenzgrammatik SoSe*, 99, 1–16.
- Decoo, W. (2001). *On the mortality of language learning methods*. James L. Barker Lecture, Brigham Young University.
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments* (pp. 9–15).
- Dolonen, J., & Kluge, A. (2015). Algebra Learning through Digital Gaming in School. In *11th International Conference on Computer Supported Collaborative Learning* (Vol. 1, pp. 252–259).
- Draycott, R. (2017). *Gamification is the key to Duolingo success says product manager Gilani at Canvas conference*. Retrieved from <http://www.thedrum.com/news/2017/10/26/gamification-the-key-duolingo-success-says-product-manager-gilani-canvas-conference>
- Ebbels, S. (2007). Teaching grammar to school-aged children with specific language impairment using shape coding. *Child Language Teaching and Therapy*, 23(1), 67–93.
- Farkas, D. K. (2005). Explicit structure in print and on-screen documents. *Technical communication quarterly*, 14(1), 9–30.
- Farmer, E., van Rooij, J., Riemersma, J., & Jorna, P. (2017). *Handbook of simulator-based training*: Routledge.
- Foti, M. K., & Mendez, J. (2014). Mobile learning: how students use mobile devices to support learning. *Journal of Literacy and Technology*, 15(3), 58–78.
- Gamage, G. H. (2003). Perceptions of kanji learning strategies. *Australian Review of Applied Linguistics*, 26(2), 17–30.
- Gikas, J., & Grant, M. M. (2013). Mobile computing devices in higher education: Student perspectives on learning with cellphones, smartphones & social media. *The Internet and Higher Education*, 19, 18–26.
- Gooding, D. (1990). *Experiment and the making of meaning: Human agency in scientific observation and experiment*: Kluwer Academic Publishers Dordrecht.
- Government of Ireland. (2006). *Statement on the Irish language*.
- Guo, J. (2016). *Google's new artificial intelligence can't understand these sentences. Can you?* Retrieved from <https://www.washingtonpost.com/news/wonk/wp/2016/05/18/googles-new-artificial-intelligence-cant-understand-these-sentences-can-you>
- Havelka, J. (2007). Beyond projectivity: Multilingual evaluation of constraints and measures on non-projective structures. In *45th Annual Meeting of the Association of Computational Linguistics* (Vol. 45, pp. 608–615).

- Heisig, J. W. (2012). *Remembering the kanji (Book 3)*: University of Hawai'i Press.
- Hewings, M. (2013). *Advanced grammar in use: A self-study reference and practice book for advanced learners of English : with answers* (3rd ed). Cambridge, New York: Cambridge University Press.
- Hu, P. J.-H., Clark, T. H. K., & Ma, W. W. (2003). Examining technology acceptance by school teachers: a longitudinal study. *Information & management*, 41(2), 227–241.
- Hubbard, P. (2002). *Survey of unanswered questions in Computer Assisted Language Learning*. Retrieved from <http://www.stanford.edu/~efs/callsurvey/index.html>
- Huynh, D., & Iida, H. (2017). An Analysis of Winning Streak's Effects in Language Course of "Duolingo". *Asia-Pacific Journal of Information Technology and Multimedia*, 6(2).
- Jean, G., & Simard, D. (2011). Grammar learning in English and French L2: Students' and teachers' beliefs and perceptions. *Foreign Language Annals*, 44(4), 465–492.
- Kahane, S. (1996). If HPSG were a dependency grammar... *Actes de TALN*, 22–24.
- Karch, A. (2015). *Duolingo Review: The Quick, Easy and Free Way to Learn A Language*. Retrieved from <https://www.fluentin3months.com/duolingo/>
- Khampusaen, D. (2013). Past, Present and Future: From Traditional Language Laboratories to Digital Language Laboratories and Multimedia ICT Suites. In *Tenth International Conference on eLearning for Knowledge-Based Society* (pp. 12–13).
- Kidd, E. (2014). *Learning Ancient Egyptian in an Hour Per Week with Beeminder*. Retrieved from <https://blog.beeminder.com/hieroglyphs/>
- Krashen, S. D. (2003). *Explorations in language acquisition and use*: Heinemann Portsmouth, NH.
- Kumar, S. (2016). *My Gamified Language Learning Experience With Duolingo*. Retrieved from <https://elearningindustry.com/duolingo-gamified-language-learning>
- Larsen-Freeman, D. (2000). Grammar: Rules and Reasons Working Together. *ESL Magazine*, 3(1), 10–12.
- Larsen-Freeman, D. (2015). Research into practice: Grammar learning and teaching. *Language Teaching*, 48(2), 263–280.
- Lavandier, A.-M. (2013). *Mobile Gaming is Huge...and it's Staying: Rise of the Casual Gamer*. Retrieved from <https://medium.com/the-nerd-castle/mobile-gaming-is-huge-and-its-staying-rise-of-the-casual-gamer-12a07333df66>
- Levy, M. (1997). *Computer-assisted language learning: Context and conceptualization*. Oxford [u.a.]: Clarendon Press.
- Lewis, B. (2011). *Review of Rosetta Stone: Detailed and honest look at latest version (TOTALE)*. Retrieved from <https://www.fluentin3months.com/rosetta-stone-review/>
- Librenjak, S., Vučković, K., & Dovedan, Z. (2012). Multimedia assisted learning of Japanese kanji characters. In *MIPRO, 2012 Proceedings of the 35th International Convention* (pp. 1284–1289).
- Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. *Foreign language research in cross-cultural perspective*, 2(1), 39–52.
- Long, Y., & Alevin, V. (2014). Gamification of joint student/system control over problem selection in a linear equation tutor. In *International Conference on Intelligent Tutoring Systems* (pp. 378–387).
- Marneffe, M.-C. de, & Manning, C. D. (2008). *Stanford typed dependencies manual*: Stanford University.
- Morford, Z. H., Witts, B. N., Killingsworth, K. J., & Alavosius, M. P. (2014). Gamification: the intersection between behavior analysis and game design technologies. *The Behavior Analyst*, 37(1), 25–40.

- Mozgovoy, M., & Efimov, R. (2013). WordBricks: a virtual language lab inspired by Scratch environment and dependency grammars. *Human-centric Computing and Information Sciences*, 3(1), 1–9.
- Murphy, R. (2012). *English Grammar in Use, 4th Ed*: Cambridge University Press.
- Nagata, N. (2009). Robo-Sensei's NLP-based error detection and feedback generation. *Calico Journal*, 26(3), 562–579.
- Park, M., Purgina, M., & Mozgovoy, M. (2016). Learning English Grammar with WordBricks: Classroom Experience. In *2016 IEEE International Conference on Teaching and Learning in Education*.
- Pica, T. (2005). Classroom learning, teaching, and research: A task-based perspective. *The Modern Language Journal*, 89(3), 339–352.
- Purgina, M., & Mozgovoy, M. (2017). Visualizing Sentence Parse Trees with WordBricks. In *3rd IEEE International Conference on Cybernetics (CYBCONF)* (pp. 1–4).
- Purgina, M., Mozgovoy, M., & Klyuev, V. (2016). Developing a Mobile System for Natural Language Grammar Acquisition. In *14th IEEE International Conference on Dependable, Autonomic and Secure Computing* (pp. 322–325).
- Purgina, M., Mozgovoy, M., & Ward, M. (2017). Learning Language Grammar with Interactive Exercises in the Classroom and Beyond.
- Resnick, M., Silverman, B., Kafai, Y., Maloney, J., Monroy-Hernández, A., Rusk, N., . . . . (2009). Scratch: Programming for All. *Communications of the ACM*, 52(11), 60–67. doi:10.1145/1592761.1592779
- Robertson, M. (2010). *Can't Play, Won't Play*. Retrieved from <http://www.hideandseek.net/2010/10/06/cant-play-wont-play>
- Santos, V. (2011). Rosetta Stone Portuguese (Brazil) levels 1, 2, & 3 Personal Edition Version 4 (TOTALe). *Calico Journal*, 29(1), 177–194.
- Shintani, N., Li, S., & Ellis, R. (2013). Comprehension-based versus production-based grammar instruction: A meta-analysis of comparative studies. *Language learning*, 63(2), 296–329.
- Sweetser, P., & Wyeth, P. (2005). GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment (CIE)*, 3(3), 3.
- Teninbaum, G. H. (2016). Spaced Repetition: A Method for Learning More Law in less Time. *J. High Tech. L.*, 17, 273.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language learning*, 46(2), 327–369.
- Walker, N. (2015). *Hacking the Kanji: 2,200 Kanji in 97 Days*. Retrieved from <https://nihongoshark.com/learn-kanji/>
- Ward, M., Mozgovoy, M., & Purgina, M. (2018). Can Word Bricks Make Learning Irish More Engaging For Students? *International Journal of Game-Based Learning*, in press.
- Washio, S., & Watanabe, Y. (2014). Security of audio secret sharing scheme encrypting audio secrets with bounded shares. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on* (pp. 7396–7400).
- Watson, I. (2008). Irish language and identity. In C. Nic Pháidín & S. Ó Cearnaigh (Eds.), *A New View of the Irish Language* (pp. 66–75). Cois Life.
- Webb, R. (2016). *80/20 Japanese*. Retrieved from <https://8020japanese.com/>
- Yaron, D., Ashe, C., Karabinos, M., Williams, K., & Ju, L. (2013). ChemCollective. <http://www.chemcollective.org>.