

Substituto – A Synchronous Educational Language Game for Simultaneous Teaching and Crowdsourcing

Marianne Grace Araneta
University of Trento

Gülşen Eryiğit
AI & Data Engineering Dep.
Istanbul Technical University

Alexander König
CLARIN ERIC

Ji-Ung Lee
UKP Lab
Technical University of Darmstadt

Ana R. Luís
University of Coimbra
CELGA/ILTEC

Verena Lyding
Institute for Applied Linguistics
Eurac Research Bolzano/Bozen

Lionel Nicolas
Institute for Applied Linguistics
Eurac Research Bolzano/Bozen

Christos Rodosthenous
Open University of Cyprus

Federico Sangati
University of Naples, Italy
OIST University, Japan

Abstract

This paper investigates a general framework for synchronous educational language games that simultaneously allows researchers to crowdsource learner answers in a controlled environment. Our prototype Substituto¹ allows teachers and students to interact in real-time while undergoing language learning exercises; ensuring that the learner’s progress is not hurt by the introduction of crowdsourcing elements. We evaluate Substituto with a small-scale user study that focuses on training the use of English verb-particle constructions (VPCs), such as *break down* or *take over*, and test their use with second language learners of English of different proficiency levels over five pilot sessions. With the study we aim to ensure that our prototypical implementation behaves as expected and to identify any major design flaws that should be addressed. The preliminary results we achieved in order to evaluate the educational value, the user experience and the crowdsourcing capacity of Substituto confirm that it has the potential to become a valuable asset for language learning, a pleasant learning instrument and a crowdsourcing tool for collecting linguistic knowledge.

1 Introduction

In the last few years there has been a substantial growth in the number of language learning educational tools, and recent works have shown the importance of gamification and more specifically how

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>.

¹Code, data, and survey form are available at <https://gitlab.com/substituto/nlp4call2020>
All our study participants gave their consent.

game-based student response systems (SRS) help foster student motivation, engagement and learning (Turan and Meral, 2018; Göksün and Gürsoy, 2019). The main problem behind such systems is that teachers – usually with only little control over such tools – no longer play a central part in the educational process, and consequently, are not able to provide students with appropriate feedback and assist them in their progress. On the other hand, handing the teachers complete control over learning tools may place too much burden on them in constructing learning materials and would essentially resolve into a traditional, no-technology approach to education.

With Substituto as our prototype, we present an innovative language game framework that strikes a balance between these two challenges of game-based learning. Driven by well-established NLP technologies it addresses this issue as a game which proposes automatically generated learning content but at the same time involves teachers as moderators and allows them to guide their students through each round. The implemented system enables a teacher to interact in real-time with a group of students either in a virtual setting or a physical classroom scenario.

From an NLP perspective, engaging in the development of educational applications is not restricted to providing tools to real-life use cases, as illustrated by Litman (2016) and Settles et al. (2018). It is also an opportunity to collect students’ and teachers’ input, which can be used to construct robust NLP resources (e.g., annotated corpora) and educational datasets. While this is a long-term objective of this work, its success is strongly determined by demonstrating its value to education and foster adoption by users. Accordingly, the aim of

the preliminary evaluation was to test the game and confirm its educational value and user experience while observing its potential for crowdsourcing in real-world language learning settings.

In this work, we focus on a single type of exercises for learning English verb-particle constructions (VPCs) in context (such as *give in*, *check out* and *come along*). Students are presented with sentences containing one VPC for which they have to simultaneously find an appropriate synonym (e.g., one or multiple words expressing an identical meaning that is also syntactically appropriate within the context). After submitting their answer, students are shown all the responses given in class and are asked to choose one they approve of (excluding their own response).² The teacher then validates the responses by highlighting all the correct answers and points are awarded to students whose answers or votes match the teacher's choice.

The paper is structured as follows. After discussing related work in section 2, we describe the language learning exercise as well as our selection process in section 3. In section 4, we provide details about the application itself. Section 5 introduces the setup of our investigative experiments – carried out in real classroom settings – and provides analysis from a teacher's, data practitioner's, and user's perspective. Our paper concludes in section 6 with a discussion of open challenges and potential future use cases of Substituto.

2 Related Work

Recent years have witnessed a noticeable increase in the use of student response systems (SRS) which allow instructors to pose questions and gather students' responses during a lecture automatically. Studies investigating SRS reported enhanced student engagement (Wang et al., 2008; Mula and Kavanagh, 2009; Patterson et al., 2010; Wang, 2015; Licorish et al., 2017). It has been also shown by Turan and Meral (2018) and Göksün and Gürsoy (2019) that game-based SRS increase the achievement and engagement and decrease the test anxiety levels when compared to non-game-based SRS. It is worth pointing out that SRS in the literature are mostly click-based systems. This makes the investigation of SRS with open answers or no pre-defined correct answer a new area to explore in teaching. Truong (2016) reviews 51 studies related

²Currently, only one response can be chosen, even though multiple responses may be valid.

to adaptive e-learning systems with regard to learning styles and concludes that educational games are still at their early stages.

Gamification in learning games is an important and promising phenomenon (Sangati et al., 2015; Lafourcade et al., 2015). Creative approaches are especially welcome for language learning applications to trigger different learning styles.

Siyanova-Chanturia (2017) discusses the teaching and learning of MWEs as L2. Boers et al. (2017) examined the impact of textual enhancement (i.e., drawing the learner's attention to MWEs by physically manipulating certain aspects of the text to make them easily noticed) on the acquisition of MWEs for L2 learners and showed its positive impact. VPCs are a subtype of multiword expressions (MWEs) which pose interesting challenges for NLP, linguistics and language learning (Hernández, 2019).

Lastly, approaches that combine language learning and crowdsourcing have recently received increasing attention (Lyding et al., 2018; Agerri et al., 2018; Chinkina et al., 2020). For example, Rodosthenous et al. (2019), Lyding et al. (2019) and Rodosthenous et al. (2020), describe a vocabulary trainer with automatically generated content from a semantic network called ConceptNet (Speer et al., 2017) that crowdsources the answers to improve ConceptNet. Substituto partially follows the implicit crowdsourcing paradigm described by Nicolas et al. (2020) to improve NLP datasets that serve as a basis for language learning exercises by utilizing the respective learner answers. The difference of Substituto is that it also involves teachers to serve as supervisors in-between automatically generated exercises and students which allows us to crowdsource additional feedback from experts.

3 Exercise Generation

Although our proposed system can be used for any form of text completion or substitution exercise, we restrict our initial setup for demonstration and testing to the specific use-case of VPC replacement in sentences. VPCs, also known as phrasal verbs, are challenging for non-native speakers mainly because of their non-compositional semantics, which largely means that learners must memorize VPCs as holistic units in addition to learning their syntactic, semantic, and pragmatic use in sentence context. Through pedagogical tools like Substituto, students can familiarize themselves with the au-

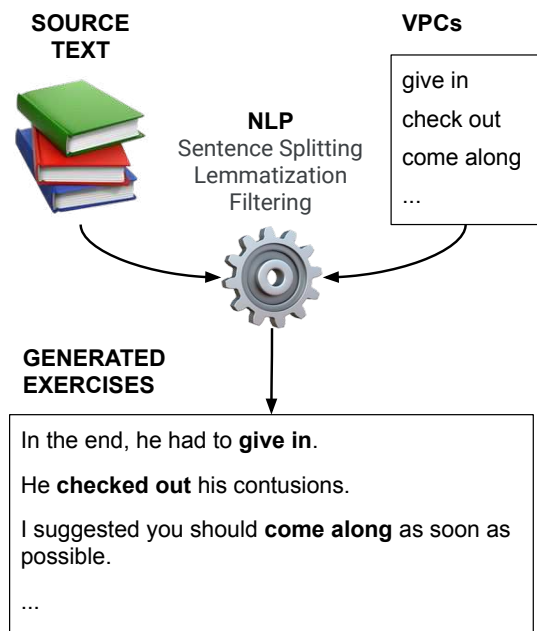


Figure 1: Pre-processing pipeline for exercise sentence selection.

thentic use of this complex language structure in both, receptive and productive vocabulary knowledge.

The goal of the exercises we created is to replace the highlighted VPC with an appropriate synonym which is then evaluated by a teacher. Similar to Boers et al. (2017) we also use textual enhancement for VPCs by bold-facing them. To make our application available to a wide range of users, it is essential to include an automated preprocessing pipeline that allows us to provide good exercise suggestions extracted from various sources (in our case books) to the teachers and allows to scale up the exercise generation at any moment as needed. As a starting point for such an automated procedure to select appropriate sentences for the exercises, we implemented a pre-processing pipeline shown in Figure 1. As our source texts, we selected books from contemporary English literature with Lexile levels of 700 to 1200L and recommendations from Oxford Readers Collections (2015). This reflects B1 difficulty based on the Common European Framework of References for Languages (CEFR). A VPC list of appropriate difficulty is extracted using the English Vocabulary Profile.³ To extract the sentences which include our target VPCs, we first generate a lemmatized version of a sentence using spaCy (Honnibal and Montani,

³<https://www.englishprofile.org>

2017) for all sentences which do not exceed a maximum number of 25 tokens including punctuation to ensure a good readability on a smartphone. We then filter for sentences which contain a VPC from our predefined list. This allows teachers to select only VPCs which fit their current syllabus to match their classes' proficiency. Afterwards, we apply a set of filter functions to remove sentences which are incomplete or are badly formatted.

Manual Filtering. Although the post-filtering process removes around 40% of the initially extracted sentences, some problematic cases were difficult to address automatically and had to be filtered-out manually (to be addressed in future work). The most relevant encountered issues are:

- Presence of inappropriate or vulgar content.
- Presence of words that require higher-level language proficiency.
- Insufficient context to disambiguate the VPC.
- The sentence contains a word sequence which is homographic to a VPC but it is not a VPC.⁴

For our preliminary study, the participating teachers hand-picked 15 sentences with 15 corresponding VPCs (3 exercise sets of 5 sentences each) that would fit well into their students' curriculum.⁵

4 System Description

The idea behind Substituto is inspired by an existing, turn-based game called PLAGIO⁶ where players take turns in choosing the beginning of a sentence from a book and the others complete it. Next, all collected completions (including the original) are displayed and participants try to guess the original continuation. Each player will gain a point if they correctly guessed the original completion, otherwise they will give the point to the player who wrote that completion. PLAGIO finally distributes points for correct guesses of the original sentence and substitutions voted by others. Players earn points by guessing the original sentence and each vote for their provided substitution.

⁴For instance: "What I see under the microscopes are cells sluggishly trying to reconstruct breaks in their walls" with the underlined word being a noun.

⁵The complete list of VPCs is: break down, break in, break up, check out, come along, come out, end up, fall down, fall over, get in, get on, get out, give in, go away, go down.

⁶<https://www.studiogiochi.com/en/publications/plagio-en-2>

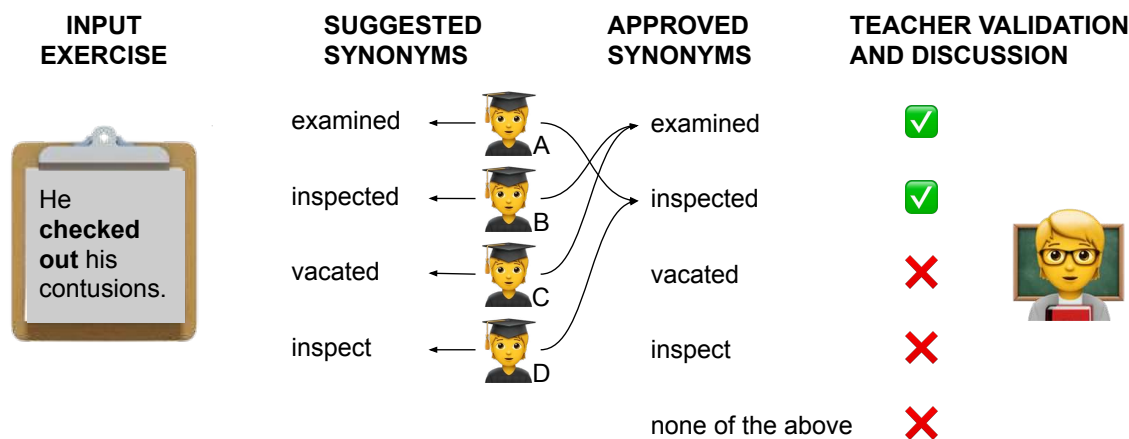


Figure 2: The four phases of each round of Substituto. Here, student A and B receive three points each (for their correct synonym and correct vote), while student C and D receive one point each (for their correct vote).

The interactive game has been implemented as a ChatBot using the Telegram Bot API.⁷ This enabled us to speed up development by focusing primarily on the game logic while using the built-in interface available in the messaging application, with keyboard buttons and commands.

Before starting a game, both the teacher and the students have to login to Telegram and search for the bot.⁸ Then the teacher has to create a game session by choosing a game name and communicate it to the students so that they can access the same session. Finally, the teacher chooses the exercise set to use in the game.

Each game is divided into five rounds (one per VPC in the exercise set) with each round being composed of four phases (cf. Figure 2):

- (1) A specific sentence from the exercise set is automatically selected and displayed to all participants, with the VPC in boldface.
- (2) Students type in a synonym of the VPC as a replacement in the given context, preserving the meaning.
- (3) Students then vote for a synonym suggested by another student or indicate that no other suggested synonym is correct. They cannot approve their own answers and identical answers are displayed only once.
- (4) Finally, the teacher validates which answers are correct (possibly none, or more than one).

⁷<https://core.telegram.org/bots/api>

⁸A running demo is provided in <https://gitlab.com/substituto/nlp4call2020>

Students receive points at the end of each round, rewarding both their productive and receptive language skills. They receive two points for suggesting a correct synonym. During voting, one point is awarded for selecting a synonym that has been approved by the teacher. In case no synonym is deemed correct, only voting “None” is awarded one point. All incorrect votes get a penalty point.

Before the next round starts, the game is put on hold; allowing the teacher to interact with the students and discuss the solved exercise. The game provides a built-in */chat* command that enables participants (students and teacher) to communicate with each other in a group chat that is visible for everyone; allowing teachers to openly engage with the students and provide appropriate hints if they notice that their students have difficulties with an exercise. This is particularly relevant when the game is held virtually without using other ways of interaction such as with a parallel virtual conference or a separate group chat.

5 User Study

In order to investigate the viability of Substituto in real-world teaching scenarios and make preliminary evaluations in terms of educational value, crowdsourcing capacity and user experience, we have conducted five pilot test sessions involving a total number of 26 participants (20 unique students), as reported in Table 1. The tests were conducted in a virtual classroom setting using a parallel video conference setup, with each participant using their own device. Figure 3 shows the teacher’s perspective of Substituto during one pilot session.

The first pilot test comprised five university stu-

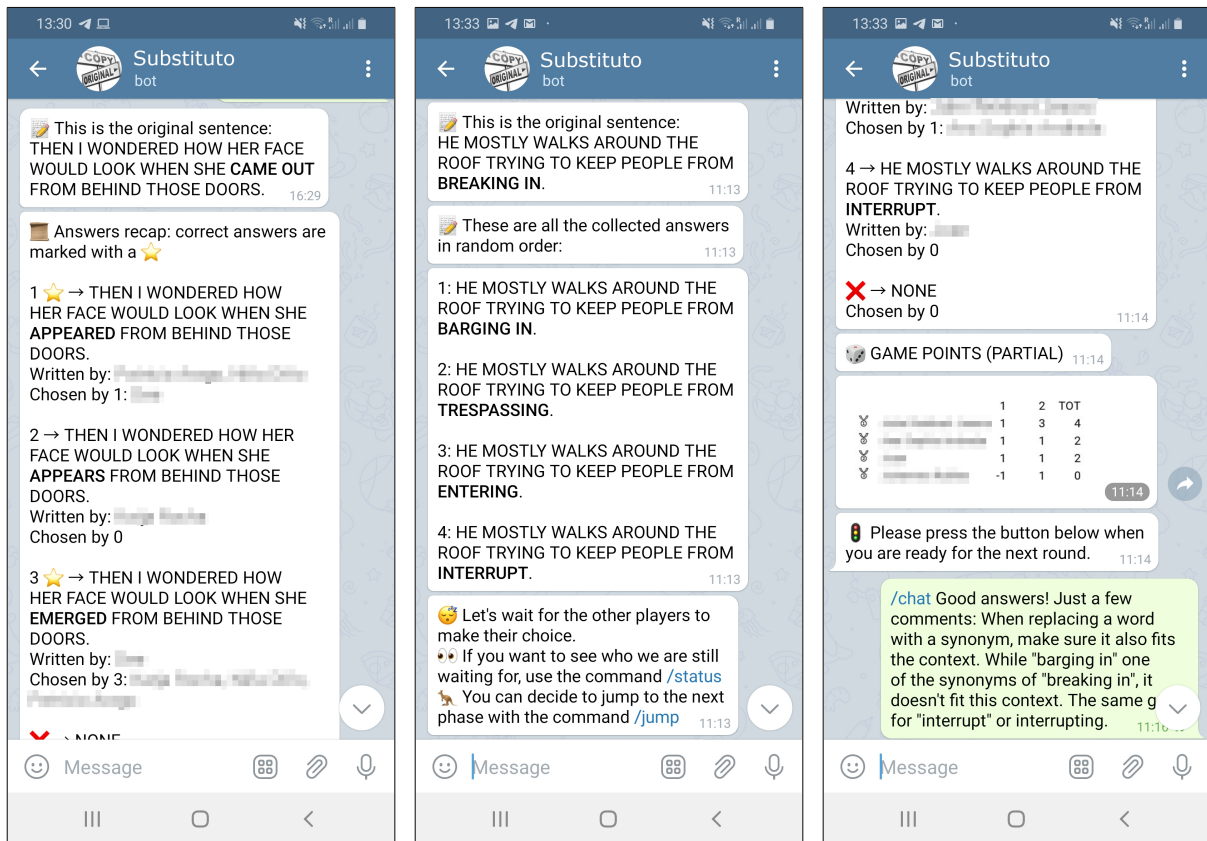


Figure 3: Screenshots from one of our pilot sessions showing the teacher’s perspective.

#	L1	CEFR	ppl.	dur. (min)
1	PH	B1-B2	5	44
2	PH	A2-B1	4	29
3	FA, POR, RO	C1	7	30
4	PH	A2-B1	4	25
5	FA, POR	C1	6	60

Table 1: Summary of pilot sessions.

dents whose first language is Filipino but have high English proficiency, ranging from B1 to B2. They completed all rounds of two VPC exercise sets within 40 minutes. The second pilot test was conducted with four Filipino students with A2 to B1 English proficiency and only one exercise set was completed. The third group of students comprised seven students with different L1 backgrounds (Portuguese, Farsi and Romanian) and C1 proficiency. They finished three rounds of one exercise set in 30 minutes. The fourth pilot test was conducted with four Filipino students with varying English proficiency levels, from A2 to B2. The last test comprised six of the seven students in the third pilot test. They completed all five rounds of one

exercise set in one hour.

5.1 Educational Analysis

Our preliminary evaluation suggests that Substituto could indeed be useful in a real-world scenario with regards to educational value. We discuss hereafter the teachers’ observations in terms of user engagements, quality of student responses, their capacity to evaluate different options and overall learning opportunities. Later in Section 6, we discuss how we intend to further explore the educational value in a quantitative manner.

With respect to user engagement, students were generally engaged during the game and those who were already familiar with each other showed higher levels of engagement – cheering on those who gave good answers or expressing their frustration for not getting a point. There was also much interaction, both between the teacher and students and among students themselves. The teacher would give feedback after each round and students would clarify and discuss the responses. The time spent on feedback and discussion determined the game’s pace.

In terms of the quality of student responses,

groups with higher average proficiency, as was the case with the students from the first, third, and fifth session, demonstrated a greater variety of correct answers. In such cases, the teacher-student interaction served the purpose of exploring the sentence context to clarify why some answers may have been more correct than others. For those groups with lower average proficiency, there was less variety and some instances where the literal synonym would be given rather than the contextualized one, as if students had consulted a thesaurus without understanding the word’s correct use. For example, given the sentence “*He mostly walks around the roof trying to keep people from **breaking in***”, with “*breaking in*” as the target VPC, one response was “*barging in*”, while another was “*interrupt*”. While these two are synonyms of “*breaking in*”, they do not fit the context of the sentence. This became an opportunity for the teacher to discuss the importance of context when looking for synonyms.

Regarding the ability of students to evaluate different options, teachers observed that giving students the chance to see correct responses of their fellow students constitutes a valuable learning opportunity that serves to increase their creative use of language. The point system encourages students to give the best possible response but it also triggers discussion among students who feel the need to understand why some answers are better than others. Teachers made use of the chat option to give feedback on student responses; but it was also used to cheer on those who were leading and encourage those who were trailing behind. Being able to trace responses to those who answered them was helpful in monitoring their progress, especially for students who are struggling.

More problematic aspects however have also been identified. With respect to the evaluation of multiple answers, the game allows the selection of more than one correct response, but they all receive the same number of points. This may not reflect “degrees of correctness”, where some answers, aside from being correct, reveal a higher level of proficiency that might merit higher points. With respect to the examples, common to all sessions was the difficulty in replacing certain VPCs due to the sentence structure, particularly those VPCs that were followed by a preposition. For example, in the third group, given the sentence: “*All at once the clear voice of Reepicheep **broke in** upon the silence*”, students had difficulty replacing “*broke in*”

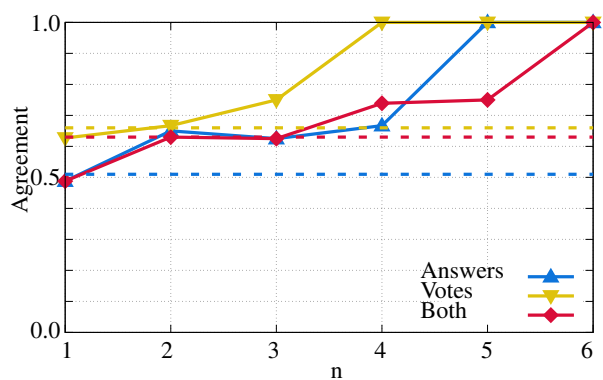


Figure 4: Agreement for different student feedback and thresholds n . Dashed lines show the averaged agreement resulting from a majority vote for each exercise.

with other synonyms due to the prepositional clause “*upon the silence*”, which has a more restricted use. Another observation noted the sentence complexity of the literary sources. Despite the VPCs and reading material being on-level for B1 learners, they were still too challenging for some students.

In terms of learning opportunities, teachers observed during the pilot sessions that the game is more than a vocabulary expansion tool for L2 learners, offering various other learning opportunities. It develops teaching comprehension skills, in so far as students need to fully understand the written examples so as to find an adequate synonym. Because the replacement must both be meaningful and adjusted to the given sentence structure, the game also exercises the students’ grammatically correct use of verb forms within context. During the teacher-student discussion time, issues concerning the adequacy of the students’ answers (e.g., style, collocation, denotation vs. connotation) were addressed and clarified, promoting language awareness raising opportunities. By and large, the teacher plays a salient role in the learning process, by encouraging class-appropriate discussion and offering real-time and individualized feedback to students.

5.2 Crowdsourcing Analysis

To evaluate the effectiveness of our prototype Substituto in a peer-reviewed learning setup without the teacher’s assessment (i.e., an expert opinion), we investigated different levels of student feedback (their answers and votes) to identify correct answers. Our claim is that the collected answers can be used to obtain expert quality labels for a learner-crowdsourced corpus. Due to the restricted exercise context, the students’ answers may overlap; resulting in ≈ 0.75 diverse answers per student and

exercise. This allowed us to not only consider the number of votes for each answer but also to take into account how often an answer was provided by several students.

We identified correct answers using different thresholds n and computed by how much they agree with the teacher’s assessment, e.g., an agreement of 0.66 at $n=2$ votes shows that 66% of the answers which got 2 or more votes were correct.⁹ The yellow curve in figure 4 shows that using the students’ votes leads to substantially more correct answers than only using the answer quantity (blue) or the sum of answer quantity and vote (red).

We observed a similar outcome when conducting a simple majority vote (tied votes were treated as multiple correct labels) for each exercise as the averaged agreement shows (dashed lines). Collecting the students’ votes has an additional advantage; we observed that 73% of the answers with zero votes were also identified as incorrect by the teachers.

Our preliminary results seem to confirm the possibility that collecting more answers and votes, and aggregating them allows us to either build a dataset to provide automatic feedback or to use the aggregated knowledge for upgrading semantic-oriented datasets describing, among other things, synonymy relation between words (e.g., semantic networks such as ConceptNet) or datasets focusing on using synonyms in context (e.g., datasets for automated paraphrasing). We will further discuss the possibility of compiling a gold standard, the required amount of answers to achieve a certain quality, and how to efficiently combine answers of teachers and learners in section 6.

5.3 User Experience Analysis

After concluding the pilot tests all 20 students provided feedback through a structured online survey. The survey addressed (1) the overall experience of using the game, (2) the interaction with the application, and (3) the learning experience. The questionnaire combined multiple-choice answers and open questions. Overall, the evaluation of the students was very positive. 90% of the respondents enjoyed working with Substituto indicating that drawing inspiration from an existing game was a good choice. They particularly liked the gaming aspect and appreciated very much reading the answers of other learners and the feedback provided

⁹Note that in our pilot sessions, the maximum possible number of votes is $n = 6$.

after each exercise. 75% of the users were positive that the game helped them to improve their verb usage skills. All users confirmed that they were able to follow the instructions and operate the Substituto, while some users commented that the point system was not clear. Also, most respondents appreciated the appearance of the game, and some highlighted positively the scoreboard and status updates.

6 Conclusion and Future Work

This paper presents Substituto, an online language game that promotes synchronous interaction between teachers and students in a virtual or physical setting while providing the possibility to perform crowdsourcing. The system prototype is in a stable version that supports parallel game sessions and an unlimited number of students for peer-reviewed learning scenarios. We tested the system in five pilot sessions carried out in actual educational environments. Our preliminary study and the results we obtained in terms of educational value, crowdsourcing potential and user experience indicate that Substituto has the potential to become a valuable asset for language learning as a pleasant learning instrument and a crowdsourcing tool that can be used in order to collect linguistic knowledge.

We further plan to address several improvements of Substituto. Whereas we manually selected a set of sentences in our user study, we plan to improve the NLP module for automated sentence retrieval (1) to generate the list of relevant exercises with higher accuracy, and (2) to assess the sentence complexity, which would allow us to tailor exercises to students from a wide range of language proficiency levels. Other mechanisms to control the exercise difficulty include inverting the current task and asking students to find VPC synonyms for non-VPC words – which we expect to result in more difficult exercises – and including other linguistic resources like ConceptNet (Speer et al., 2017) during generation. In addition, we also foresee to use annotated corpora for the exercise generation, such as for example the PARSEME corpus (Ramisch et al., 2018), which is manually annotated for verbal MWEs. Furthermore, as automated methods cannot fully cover the selection of unsuitable sentences, we will include the possibility for teachers to skip any exercise at the beginning of each round. Following suggestions of teachers from our study, we plan to allow them to distribute bonus points

to reward answers that reveal higher proficiency or creativity.

Aside from improving Substituto itself, we intend to conduct more grounded evaluations to confirm our preliminary results. With respect to educational value, a classic strategy would be to work with control groups and compare the improvement of the learning capacity of Substituto against other baselines. Such an approach requires groups of students that are truly comparable to perform evaluations in equivalent conditions, which by themselves are two difficult challenges to tackle. We therefore intend to proceed differently and in a more indirect fashion by involving and formally interviewing a larger number of teachers. By doing so, we intend to obtain an indirect expert evaluation of the educational value.

Regarding the crowdsourcing potential, the involvements of more teachers will allow us to collect a larger amount of learner and teacher answers. We will use the answers of the teachers to create a gold standard that we will manually correct. This gold standard will then be used to evaluate how many aggregated answers from student are needed to decide on the synonymy of two VPCs and will allow us to explore the question of how to best combine the answers of the students and teachers.

Finally, with respect to the user experience, we will take advantage of the larger number of students to run a larger user survey with a new set of questions derived from the practical experience we obtained from this work.

Acknowledgments

We thank all anonymous reviewers for their helpful comments and suggestions. This article is based upon work from COST Action enetCollect (CA16105), supported by COST (European Cooperation in Science and Technology). The work presented in this paper was started during the WG2/WG4/WG5 meeting organized by the Action in November 2019 in Sliema, Malta.

References

- Rodrigo Agerri, Montse Maritxalar, Verena Lyding, and Lionel Nicolas. 2018. enetcollect: A new european network for combining language learning with crowdsourcing techniques. *Procesamiento del Lenguaje Natural*, 61:171–174.
- Frank Boers, Murielle Demecheleer, Lin He, Julie Deconinck, H el ene Stengers, and June Eyckmans. 2017. [Typographic enhancement of multiword units in second language text](#). *International Journal of Applied Linguistics*, 27(2):448–469.
- Maria Chinkina, Sim on Ruiz, and Detmar Meurers. 2020. [Crowdsourcing evaluation of the quality of automatically generated questions for supporting computer-assisted language teaching](#). *ReCALL*, 32(2):145–161.
- Derya Orhan G oks un and G ulden G ursoy. 2019. [Comparing success and engagement in gamified learning experiences via kahoot and quizzz](#). *Computers & Education*, 135:15–29.
- Sim on Eduardo Ruiz Hern andez. 2019. *Individual Differences and Instructed Second Language Acquisition: Insights from Intelligent Computer Assisted Language Learning*. Ph.D. thesis, University of T ubingen.
- Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear.
- M. Lafourcade, A. Joubert, and N.L. Brun. 2015. *Games with a Purpose (GWAPS)*. Focus Series in Cognitive Science and Knowledge Management. Wiley.
- Sherlock Licorish, Jade George, Helen Owen, and Ben Daniel. 2017. [“Go Kahoot!” Enriching Classroom Engagement, Motivation and Learning Experience with Games](#). In *25th International Conference on Computers in Education*, Christchurch, New Zealand.
- Diane Litman. 2016. [Natural language processing for enhancing teaching and learning](#). In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI’16, pages 4170–4176. AAAI Press.
- Verena Lyding, Lionel Nicolas, Branislav B edi, and Kar en Fort. 2018. Introducing the european network for combining language learning and crowdsourcing techniques (enetcollect). *Future-proof CALL: language learning as exploration and encounters—short papers from EUROCALL*, 2018:176.
- Verena Lyding, Christos Rodosthenous, Federico Sangati, Umair ul Hassan, Lionel Nicolas, Alexander K onig, Jolita Horbacauskiene, and Anisia Katin-skaia. 2019. v-trel: Vocabulary trainer for tracing word relations—an implicit crowdsourcing approach. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 674–683.
- Joseph M Mula and Marie Kavanagh. 2009. [Click go the students, click-click-click: The efficacy of a student response system for engaging students to improve feedback and performance](#). *E-Journal of Business Education and Scholarship of Teaching*, 3(1):1–17.

- Lionel Nicolas, Verena Lyding, Claudia Borg, Corina Forăscu, Karén Fort, Katerina Zdravkova, Iztok Kosem, Jaka Čibej, Špela Arhar Holdt, Alice Millour, et al. 2020. Creating expert knowledge by relying on language learners: a generic approach for mass-producing language resources by combining implicit crowdsourcing and language learning. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 268–278.
- Oxford Readers Collections. 2015. *Oxford Readers Collections B1 - B2*. Oxford University Press.
- Barbara Patterson, Judith Kilpatrick, and Eric Wobkenberg. 2010. Evidence for teaching practice: The impact of clickers in a large classroom environment. *Nurse education today*, 30(7):603–607.
- Carlos Ramisch, Silvio Ricardo Cordeiro, Agata Savary, Veronika Vincze, Verginica Barbu Mititelu, Archana Bhatia, Maja Buljan, Marie Candito, Polona Gantar, Voula Giouli, Tunga Güngör, Abdelati Hawwari, Uxoá Iñurrieta, Jolanta Kovaljevskaitė, Simon Krek, Timm Lichte, Chaya Liebeskind, Johanna Monti, Carla Parra Escartín, Behrang QasemiZadeh, Renata Ramisch, Nathan Schneider, Ivelina Stoyanova, Ashwini Vaidya, Abigail Walsh, Cristina Aceta, Itziar Aduriz, Jean-Yves Antoine, Špela Arhar Holdt, Gözde Berk, Agnė Bielinskienė, Goranka Blagus, Loic Boizou, Claire Bonial, Valeria Caruso, Jaka Čibej, Matthieu Constant, Paul Cook, Mona Diab, Tsvetana Dimitrova, Rafael Ehren, Mohamed Elbadrashiny, Hevi Elyovich, Berna Erden, Ainara Estarrona, Aggeliki Fotopoulou, Vassiliki Foufi, Kristina Geeraert, Maarten van Gompel, Itziar Gonzalez, Antton Gurrutxaga, Yaakov Ha-Cohen Kerner, Rehab Ibrahim, Mihaela Ionescu, Kanishka Jain, Ivo-Pavao Jazbec, Teja Kavčič, Natalia Klyueva, Kristina Kocijan, Viktória Kovács, Taja Kuzman, Svetlozara Leseva, Nikola Ljubešić, Ruth Malka, Stella Markantonatou, Héctor Martínez Alonso, Ivana Matas, John McCrae, Helena de Medeiros Caseli, Mihaela Onofrei, Emilia Palka-Binkiewicz, Stella Papadelli, Yannick Parmentier, Antonio Pascucci, Caroline Pasquer, Maria Pia di Buono, Vandana Puri, Annalisa Raffone, Shraddha Ratori, Anna Riccio, Federico Sangati, Vishakha Shukla, Katalin Simkó, Jan Šnajder, Clarissa Somers, Shubham Srivastava, Valentina Stefanova, Shiva Taslimipour, Natasa Theoxari, Maria Todorova, Ruben Urizar, Aline Villavicencio, and Leonardo Zilio. 2018. Annotated corpora and tools of the PARSEME shared task on automatic identification of verbal multiword expressions (edition 1.1). LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.
- Christos Rodosthenous, Verena Lyding, Federico Sangati, Alexander König, Umair ul Hassan, Lionel Nicolas, Jolita Horbacauskiene, Anisia Katinskaia, and Lavinia Aparaschivei. 2020. Using crowd-sourced exercises for vocabulary training to expand conceptnet. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 307–316.
- Christos T. Rodosthenous, Verena Lyding, Alexander König, Jolita Horbacauskiene, Anisia Katinskaia, Umair ul Hassan, Nicos Isaak, Federico Sangati, and Lionel Nicolas. 2019. Designing a prototype architecture for crowdsourcing language resources. In *Proceedings of the Poster Session of the 2nd Conference on Language, Data and Knowledge (LDK 2019)*, Leipzig, Germany, May 21, 2019, volume 2402 of *CEUR Workshop Proceedings*, pages 17–23. CEUR-WS.org.
- Federico Sangati, Stefano Merlo, and Giovanni Moretti. 2015. School-tagging: interactive language exercises in classrooms. In *Language Teaching, Learning and Technology (LTLT-2015)*, pages 16–19, Leipzig, Germany.
- Burr Settles, Chris Brust, Erin Gustafson, Masato Hagiwara, and Nitin Madnani. 2018. Second language acquisition modeling. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 56–65, New Orleans, Louisiana. Association for Computational Linguistics.
- Anna Siyanova-Chanturia. 2017. Researching the teaching and learning of multi-word expressions. *Language Teaching Research*, 21(3):289–297.
- Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI’17*, page 4444–4451. AAAI Press.
- Huong May Truong. 2016. Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities. *Computers in Human Behavior*, 55:1185 – 1193.
- Zeynep Turan and Elif Meral. 2018. Game-based versus to non-game-based: The impact of student response systems on students’ achievements, engagements and test anxieties. *Informatics in Education*, 17(1):105–116.
- A. I. Wang, T. Øfsdahl, and O. K. Mørch-Storstein. 2008. An evaluation of a mobile game concept for lectures. In *2008 21st Conference on Software Engineering Education and Training*, pages 197–204.
- Alf Inge Wang. 2015. The wear out effect of a game-based student response system. *Computers & Education*, 82:217–227.