

School-Tagging: Interactive Language Exercises in Classrooms

Federico Sangati¹, Stefano Merlo, Giovanni Moretti¹

¹Fondazione Bruno Kessler
Digital Humanities
Trento, Italy

federico.sangati@gmail.com, trecision@google.com, moretti@fbk.eu

Abstract

We present a prototype of a novel online platform for promoting playful learning exercises in classrooms, allowing teachers to engage with students in an interactive way. Differently from typical e-learning environments, it is the teacher, not the machine, who leads the learning activity, i.e., she is able to monitor students' individual and aggregated answers and provide them real-time feedback. Although the platform is subject-independent, we are currently focusing on language exercises based on standard school curriculum. Answers provided by students and validated by the teacher (or automatic assessment) could constitute, in a large scale scenario, a valuable source of linguistic annotation. The platform can therefore be seen also as a crowdsourcing system to create novel linguistic resources to be used by the scientific community in the Natural Language Processing field.

Index Terms: Web-based Learning Technologies, Online Educational Games, Crowdsourcing, Linguistic Annotation

1. Introduction

School-Tagging (ST) is a web-based platform for classrooms, which aims at engaging students in game-like interactive language exercises.

Language exercises are abundant in all education levels. Grammar exercises, writing tests and learning a foreign language are all intrinsically about language. However, also subjects like literature and history make extensive use of language comprehension tasks. Language is also a big research topic within the academic fields of Natural Language Processing (NLP), (Computational) Linguistics and Digital Humanities (DH).

The idea behind School-Tagging is to recast traditional pencil & paper language exercises into online game-oriented learning activities, to 1) stimulate students' interest and motivation, 2) promote cooperation among classmates with the help of teachers and 3) support scientific researchers related to language analysis and technologies.

Today's young generations are already fully immersed in technology, and teachers face a challenge of keeping their full attention on taught material. This often leads to classroom technologies being received with suspicion. Our platform would encourage teachers to incorporate a *controlled* level of technology in the class to keep students more engaged by means of game-like interactive activities, without abandoning traditional training methods and social dynamics.

A typical ST activity requires students in the same classroom to solve individually (but at the same pace) the same set of language exercises using digital devices (e.g., laptops, tablets).

Students' answers are compared and checked in real-time by the teacher, who can optionally encourage open discussions. Anonymized answers are collected and made openly available for scientific purposes, thereby creating a virtuous cycle between education and academic research.

2. Background and motivation

Figure 1 illustrates the data annotation process and the 3 main stakeholders benefitting from the platform.

1. Students The ST platform improves student engagement during exercise sessions in the classroom and promotes cooperation among the classmates with the teacher's assistance. It enables them to receive an immediate feedback from the teacher right after submitting each answer, and ask questions in case of difficulty.

2. Teachers The platform allows the teacher to control the classroom activity as she follows them step by step in real-time. It also provides a way to monitor overall class progress over longer time, as well as particular students, and guide those who have particular needs. Additionally, teachers can tailor the exercise sets to the class level and interest.

3. Language Researchers The collected annotated data can be made openly available to NLP and DH research communities. Most state-of-the-art systems in Human Language Technology research are based on supervised approaches, requiring large amounts of annotated data to be trained. The collection of such data is a resource- and time-consuming task. The ST platform enables the creation of high-quality training data to happen at the same time as education goals are being achieved.

2.1. Digital Learning Solution in Social Dynamics

School-Tagging is a social and innovative web platform which aims at using the full capabilities of today's Internet technologies, enabling a type of real-time multiuser interaction which was inconceivable only few years ago. However, it preserves the offline modality, essential in social dynamics: the classroom remains the central physical space where the learning process takes place, and it is only partially replaced by a virtual environment.

The platform aims at setting up a collaboration between schools, scholars involved in scientific research, and other possibly partners developing language technologies applications. Student's responses function as one collective intelligence, which pro-

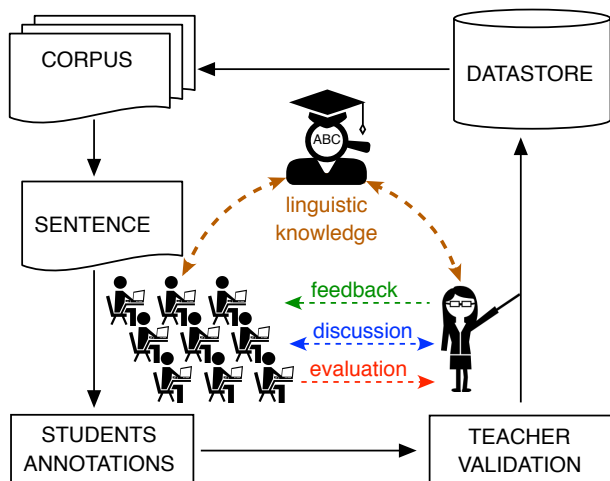


Figure 1: Diagram representing the data annotation process and the main relations between the 3 actors involved: students, teachers, and language researchers.

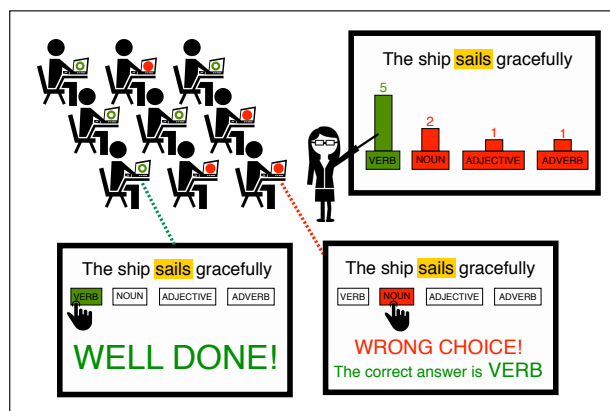


Figure 2: A schematic representation of the CHOOSE-CATEGORY game where 5 students out of 9 have answered correctly.

duces annotated resources. Such language resources remain essential for the development of more advanced linguistic knowledge, and for building state-of-the-art NLP systems by training language technologies that are widely used in today's society (e.g., Machine Translation, Automatic Question Answering). Besides, they contribute to the progress of Digital Humanities research, which exploits language annotation for developing better tools for humanities scholars (e.g., Historical Content Analysis). Finally, students are more engaged when using game-like interactive tools, and feel empowered by solving language tasks that contribute to scientific research.

3. Implementation

The general paradigm behind a School-Tagging platform is illustrated in the diagram of figure 3. There are two user entities: teachers and students. After a teacher logs in the platform, she can start a new session by choosing a unique string identifier. This token is communicated to the students in the classroom,

who have to use it to access the session. As they do so, their names become visible on the teacher's interface. After all students have accessed the platform, the teacher can choose a single exercise instance (e.g., a sentence to annotate) which is sent to all the students. As the students answer, the teacher is able to monitor in real-time each single answer and all aggregated answers via an interactive chart. After all students have answered (or the teacher decides that the time is up), the teacher can optionally promote some discussion in the class (e.g., asking specific students to justify the chosen answer). Finally, she is asked to select the correct answer. This validation generates an immediate feedback to the students. All answers are permanently written in a centralised datastore, which allows to keep track of students and classrooms progress as well as makes the data available for research purposes.

3.1. Grammar exercises

In the current prototype we have implemented 2 grammar game-exercises.

1. In the first one, the **CHOOSE-WORD**, students are presented with a sentence and they have to select all entities¹ (morphemes, single words or word sequences) that belong to a specific grammatical category.
2. In the second exercise, the **CHOOSE-CATEGORY**, students have to provide syntactic information of a linguistic entity which is highlighted in a given sentence. Figure 2 illustrates an instance of this exercise.

The two exercises are complementary to each other with respect to the type of annotation the students input into the system. The first exercise identifies how each sentence is segmented into linguistic units (entities), each being mapped to a coarse grammatical category. The second exercise, enables to ask students more fine grained syntactic information about highlighted entities (e.g. number for nouns, mode and tense for verbs). Currently, the set of sentences from which exercises are generated is a pre-built corpus of few hundred sentences. In future versions of the platform, we are planning to allow teachers to input their own sentences or extract them automatically from the Internet.

3.2. Technical Specifications

The current project's implementation² is based on the Google Application Engine (GAE) framework³. GAE enabled us to quickly develop a web-based prototype, which was easily deployable on the Internet, while eliminating all efforts in maintaining a local server, and providing immediate integration of a cloud system for data hosting (the datastore). This technology allows full scalability up to an unrestricted number of users: with no extra effort, the platform could be potentially extended to an unrestricted number of schools across different countries, producing a significantly large volume of multilingual and cross-cultural annotated data. This could foster the creation of a network of students and teachers who could exchange not only their experience, but could provide suggestions for improving the system, while building an international community of expert users.

¹An entity can be a discontinuous span such as the underlined verb in: *She has always wanted to visit China.*

²The source code of the project is publicly available at [link provided at camera ready]

³<https://cloud.google.com/appengine>

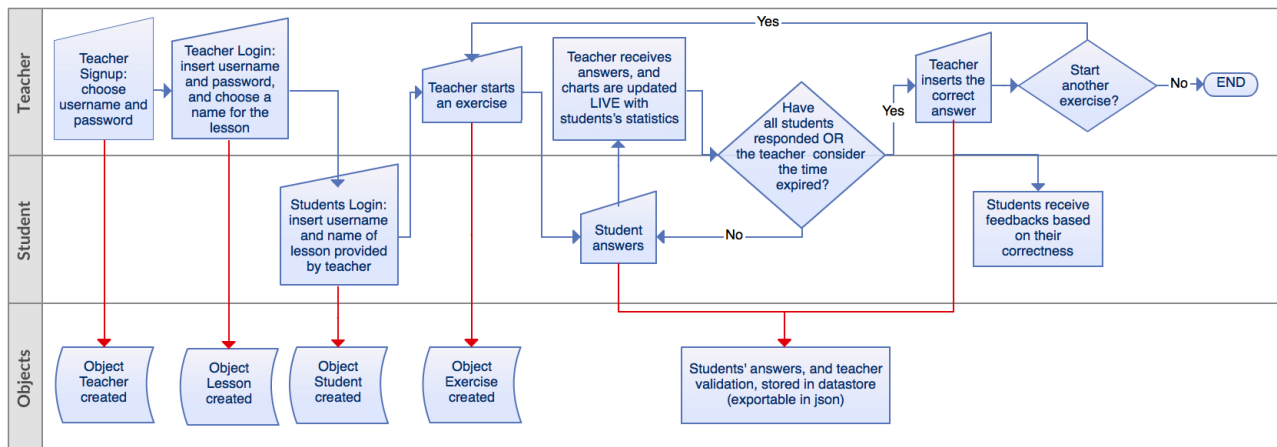


Figure 3: Diagram representing the flowchart of the School-Tagging platform.

The backend is implemented in Python, which makes use of two main libraries: webapp2⁴ (lightweight Python web framework compatible with Google App Engine) and jinja2⁵ (full featured template engine for Python).

The frontend development is based on Javascript, and makes use of the HighCharts⁶ graphical library for displaying interactive charts on the teacher interface.

Two main interface templates are available, one for the teacher and one for the students. The Google Channel API is used as interaction framework to enable real-time communication between the students and the teacher. The interactions between the server and the javascript clients works by opening a unique channel using a token identifier.

Regarding the Data Storage, we are using the Google App Engine Datastore, which is built on top of the Bigtable distributed storage system [1] with implementation of memcache to improve performance and avoid transactions latency issues and accessible via the NDB Python API⁷, through which the datastore can be exported in JSON format.

3.3. Testing

We have conducted a pilot study of the current platform in the computer-science lab of an Italian⁸ junior-high school equipped with 16 PCs. Two different classrooms (grade 6 and 9) participated in the test with a total number of 48 students, divided in 4 different groups. They have accessed the online platform using the Chrome browser⁹ installed in the local machines.

The students responded very positively to the novel methodology: in an anonymous questionnaire the overall majority has answered that they liked the system (88.5%), that the teacher's feedback was useful (77.1%) and that they would like to use the system regularly in class (87.5%).

⁴<https://webapp-improved.appspot.com>

⁵<http://jinja.pocoo.org>

⁶<http://www.highcharts.com>

⁷<http://cloud.google.com/appengine/docs/python/ndb>

⁸The current prototype has an Italian and an English version.

⁹Other browsers (i.e., Firefox and Safari) have been successful tested in our lab.

4. Demo Planning

In the demo session we would like to present the latest working version of the School-Tagging platform to the public, by simulating a classroom game-learning activity. We would use a large screen to show the teacher interface, and make a few devices available that are connected to the same session as students. Members of the audience can test the system in both roles of student or teacher using their own or provided devices.

The current version of the prototype can be tested¹⁰ at school-tagging.appspot.com. A video demonstrating the platform is embedded in the same webpage.

5. Related work

The type of prototype we have implemented sits in the boundary between existing e-learning technology and citizen-science solutions.

Many e-learning technologies are being introduced in European school systems, the great majority of which is commercially oriented. A big portion of these systems are Learning Management Systems (e.g., blackboard.com, moodle.org) which provide the infrastructure to help the administration, tracking, reporting of e-learning resources, but typically do not include actual content. Other training-oriented tools typically consist of non-collaborative solutions, i.e., the computer acts as a teacher offering a finite set of exercises, while already knowing the correct answers (e.g., khanacademy.org, softschools.com, and visl.com¹¹).

Solutions which are more similar to our project, are those which are collaborative and classroom oriented (e.g., [socrative](http://socrative.com), infuse-learning.com, getkahoot.com, clickerschool.com). However, they typically offer very rigid exercise templates (e.g., multiple answers questions, true/false), which either contain predefined answers or need to be filled in by the teacher at home. Most importantly, the answers to these tests are not used for scientific purposes.

In the School-Tagging platform, for a given task there are a

¹⁰In order to simulate multiple users on the same machine we recommend to open the webapp on different browsers (Chrome, Firefox, or Safari). Two sessions can run on the same browser only if one is set to "private" mode.

¹¹http://beta.visl.sdu.dk/games_gym.html

potentially infinite number of language exercises (i.e., the sentences provided to the system), and their solution is not previously known. The teacher has a central role in the choice of the exercise type and the validation process, and students are made aware that their effort can help a bigger scientific quest. This last element brings our platform closer to existing citizen-science and game with a purpose (GWAP) projects (e.g., fold.it, zooniverse.org, wordrobe.org [2], [phrasedetectives](http://phrasedetectives.com)¹² [3]) and other GWAP and human computation systems (e.g., the ESP game¹³ [4], and duolingo.com [5]). However, none of these crowdsourcing systems offers a collaborative multi-user and real-time environment for classrooms with exercises which are tightly aligned with school programs and linguistic research. The concept of a crowdsourcing application for educational activities has been also suggested in [6]. Finally [7] presents a more complete overview of different crowdsourcing solutions for language annotation using different methodologies.

5.1. Conclusions and Further Work

In this paper we have presented the current prototype of the School-Tagging platform. There are several aspects of the system that still require improvement:

Authentication Currently only the teacher can authenticate in the system whereas students can access only with the session name provided by the teacher. This simplifies the process since a standard authentication system for students would create a big barrier (as students easily forget their passwords). However this prevents the system from keeping track of the progress of the same student across different sessions. A possible solution would be to create a classroom entity and request the teacher to insert once for all the names of the student in each classroom.

Exercise Set The current set of sentence in the system is rather limited. In future versions, we would like to introduce the possibility for the teacher to insert her own sentences or integrate an automatic system, which could crawl them from the Web.

Home mode Even though the current modality is envisioned for a classroom, we would like to consider a similar system for students who want to practice at home. Since the teacher cannot always be available for real-time validation, we would consider using some aggregator methods [8] for deciding the correct answer based on a number of answers to the same exercise provided by students at the same time.

Gamification The current prototype offers only to teachers the possibility to monitor the progress of the students. In the future, we plan to include more advance gamification strategies, e.g., badges and points to assign to students as they progress with their learning activities.

The long-term goal of this project is to build a free and open-source online collaborative environment to be used by students and teachers in classrooms. Within this platform, we are planning to develop, in collaboration with school teachers, a set of language games/exercises which lie in the intersection between school programs and the NLP/DH agenda.

¹²<https://anawiki.essex.ac.uk/phrasedetectives/>

¹³http://en.wikipedia.org/wiki/ESP_game

6. References

- [1] F. Chang, J. Dean, S. Ghemawat, W. C. Hsieh, D. A. Wallach, M. Burrows, T. Chandra, A. Fikes, and R. E. Gruber, “Bigtable: A distributed storage system for structured data,” *ACM Trans. Comput. Syst.*, vol. 26, no. 2, pp. 4:1–4:26, Jun. 2008.
- [2] V. Basile, J. Bos, K. Evang, and N. Venhuizen, “Developing a large semantically annotated corpus,” in *LREC*, Istanbul, 2012.
- [3] J. Chamberlain, M. Poesio, and U. Kruschwitz, “Phrase detectives - a web-based collaborative annotation game,” in *In Proceedings of I-Semantics*, 2008.
- [4] L. Von Ahn and L. Dabbish, “Labeling images with a computer game,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2004, pp. 319–326.
- [5] L. von Ahn, “Duolingo: Learn a language for free while helping to translate the web,” in *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, ser. IUI '13. New York, NY, USA: ACM, 2013, pp. 1–2.
- [6] M. Skaržauskaitė, “The application of crowd sourcing in educational activities,” *Social Technologies*, vol. 2, no. 1, pp. 67–76, 2012.
- [7] J. Chamberlain, K. Fort, U. Kruschwitz, M. Lafourcade, and M. Poesio, “Using games to create language resources: Successes and limitations of the approach,” in *The People’s Web Meets NLP*, ser. Theory and Applications of Natural Language Processing, I. Gurevych and J. Kim, Eds. Springer Berlin Heidelberg, 2013, pp. 3–44.
- [8] U. Endriss and R. Fernández, “Collective annotation of linguistic resources: Basic principles and a formal model,” in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Sofia, Bulgaria: Association for Computational Linguistics, August 2013, pp. 539–549.