

# Katana and Grand Guru: a Game of the Lost Words (DEMO)

Alice Millour<sup>1</sup>, Marianne Grace Araneta<sup>2</sup>, Ivana Lazić Konjik<sup>3</sup>,  
Annalisa Raffone<sup>4</sup>, Yann-Alan Pilatte<sup>1</sup>, Karèn Fort<sup>1</sup>

(1) Sorbonne Université, STIH - EA 4509, Paris, France

alice.millour@etu.sorbonne-universite.fr, yann-alan.pilatte@etu.sorbonne-universite.fr, karen.fort@sorbonne-universite.fr

(2) University of Trento, Trento, Italy

marianne.araneta@studenti.unitn.it

(3) Institute for the Serbian Language of SASA, Belgrade, Serbia

ivana.konjik@isj.sanu.ac.rs

(4) L'Orientale University of Naples, NLP Unior Research Group, Naples, Italy

araffone@unior.it

## Abstract

We present here a prototype of a role playing game which allows to both i) crowdsource lexical units (including idioms) for a language and ii) help the player improve their knowledge of the language. Our implementation of the game is focused on non-standardized languages, for which the intergenerational transmission is not as efficient as it used to be. In order to address this, we incentivize the participation of a "Grand Guru", from whom the player needs help to fulfill their mission.

## 1. Motivations and Context

Developing natural language processing (NLP) or language learning (LL) tools for endangered languages suffering from language shift is a challenge due to the scarcity of linguistic resources of any kind. For languages that do not benefit from a standardized spelling system, two difficulties might add up: (i) the dialectal and spelling variants affect the exhaustiveness of digital resources as well as the performance of potential NLP tools, (ii) the material one might develop to support language learning can be unadapted to the habits of the learner or may not allow accurate evaluation because of the coexistence of equally correct, yet potentially unknown spelling variants.

For those of these languages which benefit from an on-line community of speakers, crowdsourcing represents a potentially fruitful way of collecting representative linguistic material (Millour and Fort, 2018). Nevertheless, motivation and retention of participants is a well known issue, that can be addressed by compensating the participants with various kinds of rewards such as financial compensation, social and personal incentives like entertainment (Chamberlain et al., 2013). Gamification has indeed shown to be an efficient way of bringing together a community of players to produce linguistic data as a side effect of their participation and enjoyment. Yet, this can only be achieved if the playful features are carefully intertwined with the task(s) to be solved.

To fulfill our objectives of (i) fostering intergenerational transmission for languages suffering from language shift, and (ii) collecting representative linguistic resources for these languages, we have designed a quest game for children between 8 and 10 years old, who share a language with, for instance, their grandparents (in the game, the "Grand Guru": GG) who speak another language which was not transmitted (the unknown language: UL).

We thus take advantage of the intrinsic motivation of

the player to learn the UL, and of the GG to transmit it (Tuite, 2014). The knowledge of the GG is collected through its transmission to the player interacting with the game.

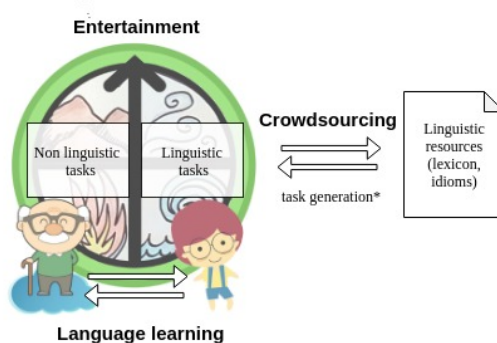


Figure 1: Combining three objectives: linguistic data collection, intergenerational transmission and entertainment.

\*The task generation has not been investigated yet.

## 2. A Game of the Lost Words

The Game of the Lost Words includes steps that can be achieved only if the players unlock a linguistic task (or step) in UL, with the help of the GG accompanying them through their quest. These steps are designed so that they should not represent a difficulty for the native speaker. So far, the variety of the designed linguistic tasks enables the collection of vocabulary items and idioms. This data can further be used for additional linguistic tasks or more generally as pieces of language documentation, resources for developing NLP tools, or language learning resources.

The prototype of the game we present here was designed and developed during a two-day hackathon (Lyding et al., 2018) (see the Acknowledgements section). It was developed with the open source software `rpgboss`<sup>1</sup>,

<sup>1</sup>See <https://rpgboss.com/>

a point-and-click Role Playing Game (RPG) engine.

## 2.1. Building a Consistent Universe

To ensure that the linguistic steps fit in nicely within the storyline, we put our efforts on designing a game that makes sense from the player’s point of view and that is consistent with the crowdsourced data. The backstory is a metaphor of the linguistic diversity loss: the player progresses in a universe divided in 4 villages (one for each of the four elements) which souls have been stolen by a dragon, resulting in the disappearance of items of the everyday life (e.g. the food, the books etc.). The player has a special power: when they name something, it comes back to life in the universe. They use their power to help the villagers who, in return, provide them with tips and rewards needed to complete their quest—to find and defeat the dragon and its goblin minions.

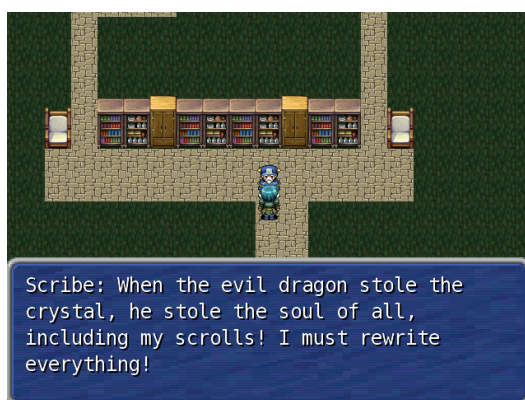


Figure 2: Example of a map included in the prototype.

## 2.2. The Crowdsourced Resources

During a two-day hackathon, we were able to create 10 maps in which 3 linguistic tasks are implemented, two of them being designed to foster interaction with the GG. The linguistic tasks include collection of lexicon items (see, for instance, Figure 3) and idioms.

### 2.2.1. Incentivizing original content

Collecting lexicon items from a variety of players from general instruction like “Name three fruits you know” can lead to the repetition of the most common items. To incentivize the production of original and diverse content, we encourage the players to relate to their own experience while answering with formulation such as: “what is your favorite fruit?” or “what is the last fruit your GG had?”.

### 2.2.2. Quality evaluation

We do not a priori possess much resources for the UL beforehand, and because the non standardized language can present spelling variants, the game is not aimed at providing feedback to the player over the correctness of the resources they provide. Yet, the tasks must be designed to avoid malicious behavior such as entering random strings instead of providing valid linguistic resources. This control can be performed thanks to various mechanisms, for instance with the use of a small size lexicon that serves as



Figure 3: Example of data collection.

reference, or with an additional task in which the player is asked to link the answers they gave to the original question. These steps have not been implemented yet, due to the limitation of the engine we used to prototype the game.

The prototype we designed is therefore being transferred to an engine that allows more control over the collected data<sup>2</sup>, and enables us to address the question of the spelling varieties: for example, when using a small lexicon to control, answers that approximately match the expected answer should be accepted. A more powerful engine will also allow us to deploy the game on mobile devices.

## Acknowledgements

This prototype has been developed during a hackathon, organized by the enetCollect (European Network for Combining Language Learning with Crowdsourcing Techniques) COST action. We wish to thank the organizer, R. Agerri, as well as enetCollect for providing us with such a great opportunity.

## 3. References

- Chamberlain, Jon, Karën Fort, Udo Kruschwitz, Mathieu Lafourcade, and Massimo Poesio, 2013. Using games to create language resources: Successes and limitations of the approach. In Iryna Gurevych and Jungi Kim (eds.), *The People’s Web Meets NLP*, Theory and Applications of Natural Language Processing. Springer Berlin Heidelberg, pages 3–44.
- Lyding, Verena, Lionel Nicolas, Branislav Bédi, and Karën Fort, 2018. Introducing the European NETWORK for Combining Language LEarning and Crowdsourcing Techniques (enetCollect). In *EuroCALL*. Jyväskylä, Finland.
- Millour, Alice and Karën Fort, 2018. Toward a lightweight solution for less-resourced languages: Creating a pos tagger for alsatian using voluntary crowdsourcing. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan.
- Tuite, Kathleen, 2014. Gwaps: Games with a problem. *Foundations of Digital Games 2014*.

<sup>2</sup>See for instance, RPG Maker <http://www.rpgmakerweb.com/>.