

# Gamification for Question Processing Research – the *QuestGen* Game

**Paweł Łupkowski, Patrycja Wietrzycka**

Adam Mickiewicz University, Poznań  
pawel.lupkowski@amu.edu.pl | patrycjakarolinaw@gmail.com

---

**Abstract:** We present the *QuestGen* game which is aimed at facilitating data gathering for research on question processing. *QuestGen* is an online game in which players solve detective puzzles by asking yes/no questions to the presented stories. We present the prototype of the game and its user testing.

**Keywords:** gamification, scientific data processing, games with a purpose (GWAP), question processing

---



## Introduction

In recent years we have been witnessing a growing popularity of gamification in various domains (see e.g. Deterding et al. 2011). What is interesting, game-like elements are successfully used in the scientific domain. One may point at such examples as *Foldit* (Cooper et al., 2010), *Galaxy Zoo* (Darg et al., 2010), or *Wordrobe* (Venhuizen, Basile, Evong, Bos, 2013). In all these cases games are designed and used to facilitate processing of scientific data (classifying already existing data, or creating new data). In *Galaxy Zoo* <<http://www.galaxyzoo.org/>> users classify pictures of galaxies obtained from *Sloan Digital Sky Survey*. *Foldit* <<http://fold.it/portal/info/science>> users try to establish possible proteins' structures, while *Wordrobe* <<http://wordrobe.org>> players help to produce semantic annotations of natural language data from the *Groningen Meaning Bank* <<http://gmb.let.rug.nl/>>. What these games have in common is that when playing the game (and having fun) players solve a serious scientific problem (or rather well defined parts of it).<sup>1</sup> Games of this type are referred to as *games with a purpose* (GWAP – see von Ahn, 2006) or *scientific discovery games* (see Cooper et al., 2010).

### 1. The *QuestGen* game idea

The idea of the game was presented in Łupkowski, 2011, pp. 89–91. The aim of the *QuestGen* game is to engage players in generating a large collection of questions for a certain piece of story written in natural language. The collection (along with the stories) will be then used as input data for research on question processing.

#### 1.1. *Motivation*

In general, question processing takes place when a question is not followed by an answer but by a new question or a strategy of reducing it into auxiliary questions (for more details see <<https://intquestpro.wordpress.com/>>).

.....

1 For a more detailed description and discussion see Łupkowski, 2011 and Kleka, Łupkowski, 2014.

This issue is discussed within theoretical frameworks of Inferential Erotetic Logic (see Wiśniewski 1995, 2013), inquisitive semantics (see Groenendijk, Roelofsen, 2011), or KoS<sup>2</sup> (see Ginzburg, 2012). It also constitutes an attractive subject for empirical research. However, it is a demanding task to prepare adequate natural language input data for such a purpose. One possible solution is to use natural language corpora (like e.g. in Łupkowski, Ginzburg, 2013). This method allows for obtaining natural language data but is time-consuming and has some limitations (e.g. there is no possibility of obtaining additional explanations from dialogue participants). Yet another method is to ask experts to prepare the necessary data (like e.g. in Urbański, Paluszkiwicz, Urbańska, 2014). This solution allows for much more control but is also time-consuming. What is more, it bears the risk of introducing a certain bias to the output data (since the professional knowledge of experts would be involved), and often the results are far from ordinary language conversations. With the *QuestGen* game we aim at providing a flexible solution for this problem.

## 1.2. Game rules

In the game, two randomly chosen players are engaged in solving a detective puzzle. One of them plays as the Detective, while the other is called the Informer. The aim for the Detective is to solve the presented puzzle by questioning the Informer. Each story in the game has two formulations (one for the Detective and one for the Informer), containing all the additional data necessary to solve the puzzle. Figure 1 presents one of the stories used in the game.

**Figure 1.** An exemplary story used in the *QuestGen* game (Detective and Informer parts). The story is based on Wiśniewski, 2003, p. 391.

### Detective part:

Imagine that you are a detective who is following the well-known international villain Arsen L. You are trying to establish if Arsen L. went to Paris, London, Kiev, or Moscow. You look through your notes and this is the information you have managed to gather so far:

.....

2 “KoS” is a toponym – the name of an island in the Dodecanese archipelago – bearing a loose connection to conversation oriented semantics (Ginzburg, 2012, p. 2).

1. Arsen L. left for Paris or London if and only if he departed in the morning.
2. Arsen L. left for Kiev or Moscow if and only if he departed in the evening.
3. If Arsen L. took a train, then he did not leave for London or Moscow.
4. If Arsen L. left for Paris or Kiev, then he took a train.

So, where did Arsen L. go?

Before you answer this question you may ask several auxiliary questions to the railway station's employee.

Remember: Your time is limited. Ask only yes/no questions. It is pointless to ask the employee directly about where Arsen L. went, because he does not have a clue.

.....

**Informer part:**

You are a railway station's employee. You may expect a detective's visit in a while. The detective is in pursuit of Arsen L., a well-known international villain. He is trying to establish if Arsen L. went to Paris, London, Kiev, or Moscow. The detective already knows that:

1. Arsen L. left for Paris or London if and only if he departed in the morning.
2. Arsen L. left for Kiev or Moscow if and only if he departed in the evening.
3. If Arsen L. took a train, then he did not leave for London or Moscow.
4. If Arsen L. left for Paris or Kiev, then he took a train.

You would like to help the detective. You know that:

- A. Arsen L. departed in the morning: YES
- B. Arsen L. departed in the evening: NO
- B. Arsen L. took a train: NO
- C. Arsen L. left for London: YES

The Detective is allowed to use only yes/no questions and cannot ask straightforwardly for the solution (this is enforced by the wording of the plot, see Figure 1, Detective part). The Detective may ask as many questions as he/she wants (needs). The Informer is obliged to answer the Detective's questions in accordance with the information presented in his/her part of the story. Each story is played within a time limit. There are two scenarios that may be employed here: a cooperative and a competitive one. In the cooperative scenario the Detective and the Informer play together against the time limit and obtain points for each (correctly) solved puzzle. In the competitive scenario the Detec-

tive will gain points for each correctly solved puzzle, while the Informer will gain points when the Detective fails (i.e. the solution given is wrong or the allowed time has passed). For both possible scenarios score boards are used to present the ranking of players' individual scores.

The game rules are inspired mainly by the *ESP game* (see von Ahn, 2006). An unintended similarity in the gameplay might be also drawn to the *Mind Maze* by Igrology <<http://www.igrology.ru/>>.

## 2. The *QuestGen* game prototype

### 2.1. Implementation

The game was implemented (in Polish) by the second author of this paper. The *QuestGen* prototype uses the following rules of time limit and scores. The time limit for a single story is 3 minutes. If the Detective manages to solve the puzzle correctly within the time limit, he/she earns points. When the time limit passes or the solution is wrong, the points go to the Informer. One may observe that such a rule employs a competitive scenario for game participants described in Section 1.2.

For the prototype, 9 stories were prepared. Table 1. presents these stories' basic characteristics in terms of premises and facts used and the number of words in the Detective's and Informer's parts. These data give some insights about the complexity of the stories used. The complete stories (in Polish) are available at <<https://plupkowski.wordpress.com/projects/questgen-game/>>.

### 2.2. User testing

The test group consisted of 30 testers grouped in 15 pairs (13 males, 17 females, aged 18 to 51, average age: 24 years). Each pair played all 9 stories in supervised sessions. The supervision was aimed at identifying problems that might occur during the play. At the end of the session, the testers were asked to fill out a short questionnaire addressing the *QuestGen* playing experience and their Internet and gaming habits. A short characteristic of the group is the following. All testers use the Internet on a daily basis, mainly for searching for information, social networking, and email.

They play games less than once in a month (12 testers declared that they did not play games). The most frequent game categories pointed out were arcade and logical games. Only one tester declared that he/she played a GWAP game before (i.e. *Galaxy Zoo*).

**Table 1.** Stories used in the *QuestGen* prototype – basic characteristics

Title	Number of premises used (the Detective part)	Number of facts used (the Informer part)	Number of words (the Detective part)	Number of words (the Informer part)
Bomb	6	5	166	173
Cookies	6	7	194	213
Character C.	5	5	243	263
Washing machine	5	4	111	137
Murder	8	4	127	198
Warehouseman	9	8	181	185
Goal	4	7	117	113
Coffee	4	5	111	107
Arsen L.	4	3	159	129

### 2.3. Results and discussion

15 pairs of testers played 9 stories each, which gives a total of 135 times the game was played. The testers asked 304 questions in total (which took less than 7 hours). The average number of questions needed for solving one puzzle was 2, while the average number of questions per pair of testers was 20. This suggest that the *QuestGen* may be effectively used to collect language data for the intended purpose.

When it comes to data validity, 67% of puzzles were solved correctly. In 19% of cases the puzzle was not solved because of the time limit. The most problematic story in this respect was *The Warehouseman*, since it ended without a solution 10 times. This might have been caused by the complexity of this story (9 premises used for the Detective part and 8 facts available for the Informer make this story hard to process for players – see Table 1). These data suggest that the established time limit was too short. This is in line with the testers’ reports expressed in the

questionnaire (53% of them declared that the time to solve one puzzle was too short).

The summary of the tests is presented in Table 2.

**Table 2.** The summary of user tests (15 pairs of players, 9 stories played by each pair)

Title	Average no. of asked questions	Correct solutions	Incorrect solutions	No solution (time ended)
Bomb	1.86	13	1	1
Cookies	2.20	8	2	5
Character C.	2.13	10	1	4
Washing machine	1.73	13	0	2
Murder	1.66	9	2	4
Warehouseman	1.60	5	0	10
Goal	3.20	14	1	0
Coffee	3.10	7	8	0
Arsen L.	2.86	14	1	0

A very important conclusion drawn from the tests is that the players do not respect the rule that forbids them to ask directly for the solution (e.g. the question “Did Arsen L. leave for Paris?” in the story presented in Figure 1). Such questions constitute 25% of all the questions gathered during the test. It seems that the rule needs an additional motivation (e.g. in the scoring system) to be respected.

What is more, an additional answer should be available for the Informer, like e.g. “this is not relevant” or “I do not know”. The reason for this is twofold. First of all, there were cases when the question asked by the Detective was not a yes/no question. Secondly, the creativity of players is sometimes larger than the creativity of the stories’ authors and there are cases where the Informer is not capable of providing an answer to a question on the basis of provided facts.

There were also several cases where the Informer wrongly answered the Detective’s question. The supervision revealed that these were simple mistakes and were not motivated with the competitive scenario used in the discussed implementation.



### 3. Summary and future work

User testing of the prototype of the *QuestGen* game suggests that the presented method can be an effective way of collecting data needed for research on question processing. The test also revealed a series of necessary corrections and possible modifications to the game design.

Future work will mainly include:

- creating new stories to be used in the game,
- modifications of game rules (e.g. adding additional answers to the Informer's repertoire),
- exploring competitive vs. cooperative game scenarios,
- testing the game in an online environment without any supervision.

### Acknowledgements

This work was supported by funding from the National Science Centre, Poland (DEC-2012/04/A/HS1/00715). The Authors would like to give their thanks to Michał Łupkowski, Katarzyna Paluszkiewicz, and Mariusz Urbański for helpful feedback and comments on a draft of this article. We also thank anonymous reviewers for their helpful remarks.

### References

- Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., Popović, Z., Foldit Players (2010). Predicting protein structures with a multiplayer online game. *Nature*, 466(7307), 756–760.
- Darg, D. W., Kaviraj, S., Lintott, C. J., Schawinski, K., Sarzi, M., Bamford, S., Vondenberg, J. (2010). Galaxy Zoo: the fraction of merging galaxies in the SDSS and their morphologies. *Monthly Notices of the Royal Astronomical Society*, 401, 1043–1056.
- Deterding, S., Sicart, M., Nacke, L., O'Hara, K., Dixon, D. (2011). Gamification: Using Game Design Elements in Non-Gaming Contexts. In: *Proceedings of the International Conference on Human Factors in Computing Systems*, CHI 2011, Vancouver, BC, Canada, May 7–12, 2011.

- Ginzburg, J. (2012). *The Interactive Stance: Meaning for Conversation*. Oxford University Press, Oxford.
- Groenendijk, J., Roelofsen, F. (2011). Compliance. In: A. Lecomte, S. Tronçon (eds.), *Ludics, Dialogue and Interaction* (pp. 161-173). Berlin – Heidelberg: Springer Verlag.
- Kleka, P., Łupkowski, P. (2014). Gamifying science – the issue of data validation, *Homo Ludens*, 1(6), 45-55.
- Łupkowski, P. (2011), Human computation – how people solve difficult AI problems (having fun doing it). *Homo Ludens*, 1(3), 81-94.
- Łupkowski, P., Ginzburg, J. (2013). A corpus-based taxonomy of question responses. In: *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)*; pp. 354-361). Potsdam, Germany, March 2013. Association for Computational Linguistics.
- Urbański, M., Paluszkievicz, K., Urbańska, J. (2014). Deductive Reasoning and Learning: A Cross Curricular Study. Research Report no 2(4)/2014. DEC-2012/04/A/HS1/00715, Adam Mickiewicz University. Online: <<https://intquestpro.wordpress.com/>>.
- Venhuizen, N. J., Basile, V., Evong, K., Bos, J. (2013). Gamification for word sense labeling. In: *Proceeding of the 10th International Conference on Computational Semantics (IWCS-2013)*; pp. 397-403).
- Ahn, L. von (2006). Games with a Purpose. *Computer*, 39(6), 92-94.
- Wiśniewski, A. (1995). *The Posing of Questions: Logical Foundations of Erotetic Inferences*. Dordrecht – Boston – London: Kluwer AP.
- Wiśniewski, A. (2003). Erotetic search scenarios, *Synthese* 134(3), 389-427.
- Wiśniewski, A. (2013). *Questions, Inferences and Scenarios*. London: College Publications.

**Paweł Łupkowski PhD** – Department of Logic and Cognitive Science, Institute of Psychology, Adam Mickiewicz University, Poznań.

**Patrycja Wietrzycka M.A.** – Institute of Psychology, Adam Mickiewicz University, Poznań.

---

## **Grywalizacja badań nad przetwarzaniem pytań – gra *QuestGen***

**Abstrakt:** W artykule omawiamy grę *QuestGen*. Jej celem jest ułatwienie zbierania danych językowych na potrzeby badań nad zjawiskiem przetwarzania pytań. Gra *QuestGen* jest grą online, w której gracze rozwiązują zagadki detektywistyczne, zadając pytania do przedstawionych historii. Omawiamy implementację prototypu tej gry oraz rezultaty testów z użytkownikami.

**Słowa kluczowe:** grywalizacja, przetwarzanie danych naukowych, gry skierowane na cel (GWAP), przetwarzanie pytań

---

