Fictionary-Based Games for Language Resource Creation

Steinunn Rut Friðriksdóttir, Hafsteinn Einarsson

University of Iceland

Reykjavík, Iceland

{srf2, hafsteinne}@hi.is

Abstract

In this paper, we present a novel approach to data collection for natural language processing (NLP), linguistic research and lexicographic work. Using the parlor game Fictionary as a framework, data can be crowd-sourced in a gamified manner, which carries the potential of faster, cheaper and better data when compared to traditional methods due to the engaging and competitive nature of the game. To improve data quality, the game includes a built-in review process where players review each other's data and evaluate its quality. The paper proposes several games that can be used within this framework, and explains the value of the data generated by their use. These proposals include games that collect named entities along with their corresponding type tags, question-answer pairs, translation pairs and neologism, to name only a few. We are currently working on a digital platform that will host these games in Icelandic but wish to open the discussion around this topic and encourage other researchers to explore their own versions of the proposed games, all of which are language-independent.

Keywords: Games With A Purpose, Language Games, Deception Games, Crowd-Sourcing, Data Collection, Corpus Construction

1. Introduction

Language resources (LRs) are an essential part of natural language processing (NLP), linguistic research and lexicographic work. Recent years have seen a tidal wave of data-driven approaches, increasing the demand for large quantities of annotated data. However, creating these resources is a time-consuming and expensive process which often requires a considerable amount of manual labor. In this paper, we propose a novel method for crowd-sourcing LRs using a Game With A Purpose (GWAP) inspired by a parlor game known as Fictionary[1].

Fictionary is a deception game in which players guess the definition of an obscure word. In each round, one player selects and announces a word from the dictionary, and other players individually compose definitions for it. The made-up definitions, as well as the correct one, are collected blindly by the selector and read aloud, and the players vote on which definition they believe to be correct. Points are awarded for correct guesses, correct definitions, and for having a fake definition guessed by another player. If a player votes for their own guess they do not receive any points. However, they might still do that to deceive other players into voting for that guess as well.

Many games can be formulated within this framework that could be used to create or expand LRs via crowd-sourcing. For example, in a title generation game, players are given the first few lines of a news article and are then asked to guess its title. Subsequently, the players vote for the best title, receiving points when another player votes for their guess. This is where deception comes in as a player might vote for their own answer in order to get others to vote for it as well. The data generated by this game can be used to train a model that generates titles in an extreme summarization fashion or evaluates candidate titles for news articles.

The advantage of our method is that it is relatively quick and inexpensive to generate new LRs using this method if sufficiently many players participate. In addition, our method is potentially more engaging and fun for participants than other methods of data collection. Since the games are structured as competitions, the players are incentivized to create high-quality data as long as the incentives of the competition align with creating quality data. The voting phase of the game, explained in Section 3.3, can help identify good data as the number of votes can be considered a quality indicator. The games can also be customized to target specific languages or domains, making them very versatile.

However, it is important to note that the games must be designed carefully in order to ensure that the players are actually incentivized to create high-quality data. In some cases, players may be more interested in winning the game than in creating high-quality data, which could lead to lower-quality LRs. Therefore, it is important to carefully consider the game design in order to ensure that the players' incentives are aligned with the researcher's goals. Additionally, the data preparation costs (as input to the games) can be significant and would continue to be so if games were ported to new languages and domains.

We are working on a digital game for these types of games in Icelandic but we would like to start a discussion by pointing out this opportunity to other researchers who might be interested in studying this gamified framework of data collection. In this paper, we propose several games that fit within this framework and discuss further aspects of this framework to collect labeled data.

---

[1] Also known as "The Dictionary Game" along with a boardgame version called Balderdash.

2. Literature Review

Data-driven approaches have rapidly gained popularity in the field of natural language processing and with that comes the need for massive LRs. While certain types of data can be mined from various sources such as websites, newspapers and books, manual work is still needed in many cases where specifically annotated data is required. As manual labor can quickly become expensive, crowd-sourcing methods have been used to reduce costs and spread the workload. This can, however, lead to the problem of less engaged workers that quickly tire of their tasks, potentially rendering the data half-finished and thus inusable (Ogawa et al., 2020). Using motivation techniques through gamification, such as earning points or badges and climbing up leaderboards, can significantly increase user engagement and gratification when performing crowd-sourcing tasks.

2.1. Games with a purpose

Deterding et al. refer to gamification as "the use of video game elements in non-gaming systems to improve user experience (UX) and user engagement". Using game heuristics when designing interfaces in non-game services increases participant enjoyment which in turn can raise interest and public participation in a given task (Deterding et al., 2011). While not prominent, the GWAP methodology has been used to collect NLP data for over a decade. In 2008, Chamberlain et al. developed the game Phrase Detectives where players collect anaphoric information in a gamified environment. The game Zombilingo, proposed by Fort et al. in 2014, uses several motivation techniques in order to incentivize players to create dependecy syntax data for French. In the same year, Jurgens and Navigli proposed an annotation paradigm that asks users to create a mapping from WordNet senses to images and perform word sense disambiguation while playing graphical video games.

In 2020, Araneta et al. introduced Substituto, a language learning game designed for English L2 learners that simultaneously crowd-sources NLP data. In 2021, Arhar Holdt et al. presented Game of Words, a gamified mobile application where users were encouraged to improve and enhance two automatically compiled Slovene dictionaries. In the same year, Eryiğit et al introduced a gamified approach to compiling an idiom corpora in Turkish and Italian. They designed a Telegram messaging bot that serves as a multiplayer game for native speakers that compete with each other while creating ideomatic and non-ideomatic sentences and rating each other's propositions. Users were additionally incentivized using gift cards (Eryiğit et al., 2021).

2.2. Crowd-sourcing projects in Iceland

The Common Voice project is a multilingual crowd-sourcing initiative where participants are asked to record their voice by reading sentences that they are presented with on the screen, and other participants are subsequently asked to verify the recordings using a simple voting system (Ardila et al., 2020). In July 2020, it was reported that the corpus had reached over 7,000 hours of voice data in over 50 languages. The Icelandic version of the project has used gamification in their marketing to great success. In 2022, 118 elementary schools competed for a prize where the goal was to read as many sentences as possible for the project. This has been an annual event since 2019 and has resulted in 1.5 million voice samples being collected for the project[2]. Additionally, over 360,000 voice samples were collected in a similar contest between Icelandic organizations and companies.

In 2021, Jasonarson used gamification and crowd-sourcing in order to collect LRs in Icelandic. His website, Málfróði (e. *linguistically knowledgeable* but in the form of a masculine name), incentivizes players to rate data according to their formality and inappropriateness on the one hand, and evaluate their linguistic correctness (spelling and grammar) according to their own conviction on the other hand. The players receive points for each submission they make. They receive more points if their submission is marked by the other players as having good quality, and they receive maximum points if their submission gets points from the majority of other players, indicating that their submission is reflective of public consensus (Jasonarson, 2021).

In 2021, Snæbjarnarson et al. published a resource where they present their extractive question answering (QA) dataset for Icelandic (Snæbjarnarson et al., 2021). Following the lead of Clark et al. (2020), they asked human annotators to write questions inspired by a 100-character-long prompt from Icelandic Wikipedia articles, but to make sure that the prompt did not answer their questions. In a second phase, the participants were asked to answer each other's questions. Based on that approach a mobile game was developed to build a larger crowd-sourced dataset for Icelandic[3]. The task was presented as a mobile game where users collect points and can receive prices based on their scores.

3. General game framework

In this section, we define the game and emphasize variations of it. The game is played over a predetermined number of rounds and the goal of each player is to maximize their points. We show an example of a game round in a title generation game in Figure 1.

3.1. Preparation phase

The round starts with the players receiving the same task. The task can come with a side-objective. For example, in a title generation game the side objective

---

[2]Scoreboard for elementary schools in Iceland: `https://samromur.is/grunnskolakeppni2022`.

[3]Available at `http://www.spurningar.is`

Preparation

Players receive a common task.

What is the title of the following news article?

After a safety meeting concluded at the Kermit Community Center recently, officers and an EMT were called to the scene of an accident that involved two men from the safety meeting…

Submissions

Players solve the task and submit their entry in private.

Irony: Safety meeting causes an accident.

Accident at a safety meeting proves its importance.

Safety meeting goes awry when an accident occurs.

Voting

Players vote publicly one by one on the submissions. Some options might represent correct answers or old guesses from other players.

Safety meeting ends in an accident.

Irony: Safety meeting causes an accident.

Accident at a safety meeting proves its importance.

Safety meeting goes awry when an accident occurs.

Results and point distribution

Players see the identity behind each submission and receive points for voting for the correct answer or if another player votes for their suggestion.

1 point (for correct vote)

0 points

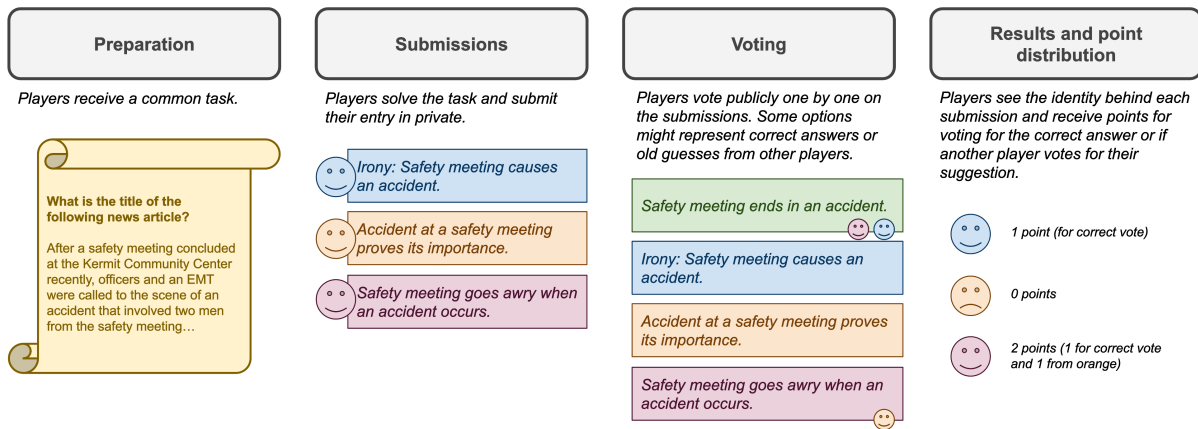2 points (1 for correct vote and 1 from orange)

Figure 1: Example of a round in a title generation game. The players are presented with a task and everyone writes their solution that they submit in a private manner. After everyone has submitted their entry they proceed to a voting round. In a voting round, additional options might be available such as a correct title (green) in this case, or possibly old submissions from external players for the same task (not shown). The players vote for what option they think is correct in this case (or the one they like the most with respect to a given objective in case a correct option is not available). Finally, players receive points based on the voting result and the identity behind each submitted item is revealed.

could be to write a short or witty title. Such side objectives can serve as additional labels for data generated by the players in that round.

3.2. Submission phase

Each player writes a submission in private for the given task. This phase can be played with a timer if the players are playing in real-time or without a timer if the players are ready to spend time on their submissions and play asynchronously.

An asynchronous approach can be implemented in a manner similar to a popular game called Wordle where each day the participants play a single round and need to vote before tha day ends.

3.3. Voting phase

After all players have submitted their entry they proceed to a voting phase. The players can either vote publicly one by one or they can all vote simultaneously. There is a qualitative difference between these two approaches because if players vote one by one then their vote can influence the decision of the next players in line. This presents an opportunity for deception where a player might vote for their own submission in order to deceive the other players into voting for it as well.

The options available in the voting phase do not necessarily need to consist only of the submissions of the players. They can also include a correct answer (if one is available) or old submissions made by external players for the same task. When the number of options is greater than the number of players it can be sensible to give players more than a single vote to increase the chances of them receiving points in the round and even allow them to vote for the same item more than once.

This could further affect the point calculation, for example, by doubling the number of points a player assigned to a correct item.

We note that a digital experience also presents more opportunities for labeling. In the voting phase, players could also be presented with the option to assign additional labels to the submissions that do not award points. For example, they could tag submissions with emojis or some fixed reactions that are there to drive engagement in the game but could serve as interesting labels as well.

3.4. Results phase

The identity behind each suggestion is revealed and players receive points based on the votes in the results phase. A player receives points when another player voted for their suggestion. When a correct option is available the players also receive points for voting for it.

After this phase, the game proceeds to the next round.

3.5. Single player variant

In a single player variant of the game the player skips the submission phase and proceeds directly to the voting phase, where they are presented with several items. In case of a correct item, the aim of the player is to spot it. When a correct item is not available the aim of the player is to spot the most popular item where the popularity of an item is determined by its past success. This approach also allows for a more passive participation as this type of voting could be done at any given moment, serving more as a validation of previously generated data.

3.6. Target users

We note that the game could be implemented, for example, in the form of an app for mobile phones and tablets or as a website. On an accessible digital platform, the game can be played by a broad group of users. However, some of the games might be more relevant to a classroom environment where the aim is to have students learn about e.g. domain-specific vocabulary in a game that closely resembles the original Fictionary.

With a sufficiently general platform the user could define tasks themselves. This can be beneficial for teachers who want to use this method in their classroom to engage students in a novel manner.

3.7. Data logging

For a game built using this framework, it is necessary to log the configuration of each game session along with the data generated in that session. Such logging helps researchers filter out data that was not generated in a particular manner. For example, if players are given a choice between voting simultaneously or one by one when a game session is started then that choice should be recorded such that one could select data from sessions where everyone voted simultaneously.

4. Game suggestions

In this section, we present several ideas for games that could be used for data collection, particularly for NLP. We further suggest the usability of the data collected for each game.

4.1. Summarization

In a title generation game, players receive the first few paragraphs of a news article and are asked to generate a title for it. The players are then asked to vote for the best title. Points can be awarded based on a majority vote or such that each player receives points when another player votes for their submission. Creating a good source dataset for a game of this type does not require significant work given the amount of publicly available news articles.

The resulting dataset could be used to train a model to evaluate candidate titles for a news article. A model could also be used to generate titles in an extreme summarization fashion. We believe that this game has the potential to be engaging for players since they have the opportunity to come up with witty and clever title suggestions. Additional objectives can also be provided to the players if the aim is to create serious titles, funny titles, short titles, long titles et cetera.

4.2. Word sense disambiguation

In this game the aim is to create a new data set for word sense disambiguation. Players would be given a sentence with a word highlighted, and would need to write a definition of that word. Points could be given in a similar manner to the original game where the player who provides the correct definition gets a point

as well as the player whose definition is voted for by the other players, regardless of truth value. This game can be considered a generalization of Fictionary since the word is given with the addition of surrounding context. If the examples are hard, then the players will generally be wrong and the value of the data might appear less clear than for other game suggestions. However, we note that wrong suggestions that are good in deceiving players could be used as negative examples when training a model for word sense disambiguation or as a test set to get a better measure of the performance of word sense disambiguation models.

4.3. Question answering and generation

In this game, the goal is to generate new data for question answering systems. Players would be given a question and would need to write an answer for that question. Points could be given as in the previous game where players that provide the correct answer get points as well as the players who manage to deceive other players into voting for their answers. The game can also be reversed where players are given an answer and have to write an appropriate question based on some additional requirements such as the question needing to be serious, witty or sounding like a riddle.

4.4. Paraphrasing

In this game, the aim is to generate paraphrases for a given sentence in a particular style. Players would be given a sentence and an objective and would need to write a paraphrase for that sentence while trying to satisfy the objective. The objective could be stylistic, e.g., to make the sentence more serious or more funny. The objective could also be to make the sentence shorter, longer or simpler.

Data from a paraphrasing game could be useful for training paraphrasing models that can change the style of a given text. It could also aid in training models with the aim of making text simpler to read and more accessible, e.g. for people with disabilities or L2 learners.

4.5. Generating NER data

In this game the aim is to generate new Named Entity Recognition data. Players would be given a sentence with some words replaced by blanks. The players would need to fill in the blanks with named entities that satisfy a given tag, e.g. person, location or organization. The task can also come with an objective, such as finding entities that make the sentence funny while still satisfying the objective. The task can also be flipped, i.e., the players receive some fixed entities and their task is to write a sentence involving said entities, possibly with some side objective as before.

The resulting data of entities, in context, labeled with their NER tags can be used to train NER models. The task for fixed entities could be especially useful for generating training data for entities that occur rarely in text.

4.6. Poem games

4.6.1. Finishing poems

In a poem finishing game, players receive the first few lines of a poem and are asked to finish it with respect to a given rhyme scheme. This is a variation of a game that was (and still is, on special occasions) commonly played in Iceland, usually between two players that took turns finishing each other's poems. The objective of the original game is usually to be witty or pointed towards the opponent, an element that could easily be adapted into the voting system. A point-scale could even be added (using emojis, for instance) where the players rank the proposed lines based on their wittiness. The resulting dataset could be used to train a poem generator. Such a generator could be a good source of inspiration for song and story writers as well as being interesting on its own: what type of poems does an AI write?

4.6.2. Writing poems for a given subject

In this game, players receive a given context (for example, a news article) and are required to write a poem with respect to a given rhyme scheme that reflects its contents. The objective could be similar to that of the previous game, that is, to make the poem particularly witty, sarcastic or pointed with ranked scores.

The resulting dataset could be used to train an abstractive summarization model whose output is in the form of a poem. We are not aware of models that perform this type of summarization although we speculate that large generative language models might have such capabilities at some point.

4.7. Story writing

4.7.1. Story by a committee

In a story game, players start with a blank prompt or some general objective, and everyone writes the beginning of a story. The players then vote on the suggestions and the winner becomes the prompt for the next round where the process is repeated until the story ends. This way, the players collectively write a story about a given subject. Alternatively, the game could be played in turns where each player has a specific amount of time to write their prompts, skipping the voting until the end where a player could be voted as being the most creative or the funniest contributor. The advantage of this approach is that it does not require any data to get started. As an objective the players could be given a list of characters, settings, and objects, and then have to come up with a story that includes all of those elements.

The disadvantage of this approach is that the players might take a lot of time to write and each game round is not independent of the rounds that came before. This might lead to lower engagement than with the other games but it could still be used in a proper setting, for example, as an exercise in a class on writing short-stories. Additionally, as a single-player mode, a com-

puter player such as GPT-3 could be used as an opponent, giving the player an instant response.

4.7.2. Natural language dialogues

In this game the aim is to generate natural dialogue. Players would take on the role of characters in a dialogue and would need to continue a conversation. In any given round, everyone responds on behalf of the same character and the players vote which response will be chosen to continue the dialogue. This type of game could be used in a teaching setting, particularly with L2 learners which in turn would collect a language variant that is often underrepresented in textual data.

The data generated from a game like this one could be used to generate training data for a chat bot. If the users are given additional instructions then that information can be used for finer-grained dialogue tasks. For example, if users are instructed to be rude, then the data could be used to train a rudeness detection model.

A game of this type could also be an interesting exercise for students writing scenes in a play in a democratic manner. In this process, everyone can collectively decide on how to move a dialogue forward without the risk of a single individual taking over the process.

4.8. Machine translation

4.8.1. Translations of technical terms

A game could be designed to generate suggestions for translations of technical terms and domain-specific words. Players would be given a foreign word along with its definition and would need to suggest a translation, which could in principle be a neologism. The suggestions could then be voted on by other players, scoring the suggestor of the winning candidate points on the leaderboard.

Creating suggestions in this manner could help committees and professional translators settle faster on good translations for new technical terms.

4.8.2. Translating sentences

In this game, the players receive a sentence they need to translate into a given language. This can be played as a language learning game where a group helps each other learn a new language, similar to an online tandem partner. But for native speakers of a given language, the game could lead to high-quality paired training examples. Since players would suggest many possible translations and one might not obviously be the best one, it might be better to let the players rate each translation in this game than to vote for a single one.

4.8.3. Sentences from fixed words

In this game, players are given a list of words and need to write sentences that include those words. The sentences could be evaluated by the other players based on different factors, such as grammar, fluency, and appropriateness. This is particularly suitable in an L2 learning setting where the list of words can even be given

in the players' native language but the players must use them in their target language.

4.9. Neologism

In this game, players are asked to come up with synonyms or definitions for already existing words. A particular objective could be to create words in a certain style or register (e.g. for academia, slang, or for a specific, potentially made-up, dialect). Players could also be asked to invent new words that convey a certain proposed meaning. Players would then vote for the suggestions using a scale and receive points based on their rank. In this game, the focus would be on creativity rather than accuracy.

As with the game proposed in Section 4.8.1, the data generated by this game could be used when coming up with neologisms and translations for new vocabulary entering the language.

4.10. Recipe generator

In this game, players are given a list of ingredients and have to come up with a dish that contains those ingredients. Certain criteria could be introduced as variants of the game, e.g. to create the best vegan recipe or make the highest-calorie or most frugal meal possible from the list of ingredients. Players then vote for the recipe that they like best, scoring the author points on the leaderboard. This data could be used to train a model whose objective is to automatically retrieve recipes from a list of proposed ingredients, which people could then use to get new ideas based on what they currently have in their kitchen. Since a game like this could be challenging for novices it is crucial to record the cooking skill level of each user beforehand.

5. Competition to improve labels

The data acquisition approach we have presented has several interesting qualities when compared to other approaches. First of all, players can be incentivized to create high-quality data since their examples are reviewed by other players. Second, we note that the voting phase of the game can provide interesting information on the quality of the players' entries. Such information could be helpful to train models to rank examples with respect to a given task description. For a title generation game, the model would receive as input the task description as well as the players' entries. The output of the model would then be a score for each entry that can be used to rank which candidate fits best. To get a better estimate of the quality of an entry, it can be used in another round with different players. Players could then be voting not only on their own entries but also on entries submitted by external players. Under such conditions, it could make sense to give the players more than just a single vote since otherwise it might be more challenging for them to get any points at all.

Additionally, each game could start with the instigator configuring which game type they want to play first, what type of voting system they want to use, whether or not they only wish to participate in the voting etc. This metadata would be logged, making it possible to filter out language resources that are created in some specific way.

6. Testing the idea

As a proof of concept, and a qualitative evaluation, we played some of the games proposed in Section 4 with a few colleges. As our platform is not ready, the games were played on paper but in essence they were the same as they would be in a computerized form. None of our colleges had played Fictionary before but they agreed that the framework had the potential to work well. They compared the idea to Kahoot (Dellos, 2015) or Jackbox [4] and mentioned that well designed graphics and music could do a lot for making the game more appealing to users. They agreed that some games were more interesting than others and could be played for entertainment purposes but others resembled a traditional crowd-sourcing task that would quickly get boring. They mentioned that all of the games that involved a side task such as making the answer funny would work well and compared those games in particular to Jackbox. They additionally mentioned that games that involve a single correct answer could be played for entertainment if presented as a trivia game that allows users to level up, collect badges or climb up leaderboards.

When asked whether they would be more likely to play the games if they would be preceded with an explaination regarding their importance for data collection for Icelandic NLP tasks, one of our colleges pointed out that the platform could be presented in two seperate ways. If the idea was to appeal to the masses and get the average user to play, the entertainment value would always be the selling point and the idea of unpayed labour might even put some users off. On the other hand, the platform could be presented in schools as a learning instrument as well as having the higher purpose of helping advance Icelandic to the digital age. Our college had played Kahoot in school before and mentioned that the diversion from traditional teaching methods was highly appreciated by the students. They added that if the tasks were presented as a multiple choice, the students' input could provide additional information to train language models. Wrong answers that receive a lot of votes from students would be labelled as particularly hard and could be used as challenging negative examples for language models.

7. Discussion

We have presented a new framework for building LRs in a gamified manner. We have demonstrated several tasks that fit within this framework and which could potentially lead to voluntary participation or participation as an exercise in a classroom environment.

---

[4] http://jackboxgames.com

The key factor which determines the success of this LR generation strategy is how engaging the game turns out to be. An engaging game has the potential to be entertaining for users while simultaneously creating high-quality LRs. Given the success of games such as Balderdash, which have been sold in millions of copies, we believe that this approach has great promise.

We must acknowledge that some games might not be as engaging as others and it is likely not possible to fit every dataset creation task into this format. As an example, a task more challenging than extreme summarization would be to write a summary of the news article that is longer than a single sentence. That is a more tedious task than title generation and possibly less engaging for that reason. One approach to make tasks like this more engaging for the user could be to mix the tasks up so that they are randomly sampled from the set of available games. In case users get bored of a particular task, each round could start with a majority vote where users can vote on whether to cancel or continue with the currently proposed game. Having a good variety of tasks can potentially increase the sense of novelty, which can further drive engagement. The framework could, in principle, also potentially be used for traditional crowd-sourcing tasks where the objective is simply to generate data, without regards to the entertainment value or even scoring points.

Finally, we want to acknowledge that this LR creation process can introduce new biases into a dataset. Dynamics that arise due to the competitive nature of this approach might lead to submissions that are not representative of data acquired through other means. Studying the extent of such a bias remains an open problem and can further help to understand the value of this approach for creating LRs.

8. Acknowledgements

9. Bibliographical References

Ardila, R., Branson, M., Davis, K., Kohler, M., Meyer, J., Henretty, M., Morais, R., Saunders, L., Tyers, F., and Weber, G. (2020). Common voice: A massively-multilingual speech corpus. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4218–4222, Marseille, France, May. European Language Resources Association.

Arhar Holdt, Š., Logar, N., Pori, E., and Kosem, I. (2021). "game of words": Play the game, clean the database. *EURALEX XIX*.

Chamberlain, J., Poesio, M., Kruschwitz, U., et al. (2008). Phrase detectives: A web-based collaborative annotation game. In *Proceedings of the International Conference on Semantic Systems (I-Semantics' 08)*, pages 42–49.

Clark, J. H., Choi, E., Collins, M., Garrette, D., Kwiatkowski, T., Nikolaev, V., and Palomaki, J. (2020). Tydi qa: A benchmark for information-seeking question answering in typologically diverse languages. *Transactions of the Association for Computational Linguistics*, 8:454–470.

Dellos, R. (2015). Kahoot! a digital game resource for learning. *International Journal of Instructional technology and distance learning*, 12(4):49–52.

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., and Dixon, D. (2011). Gamification. using game-design elements in non-gaming contexts. In *CHI'11 extended abstracts on human factors in computing systems*, pages 2425–2428.

Eryiğit, G., Şentaş, A., and Monti, J. (2021). Gamified crowdsourcing for idiom corpora construction. *Natural Language Engineering*, pages 1–33.

Fort, K., Guillaume, B., and Chastant, H. (2014). Creating zombilingo, a game with a purpose for dependency syntax annotation. In *Proceedings of the First International Workshop on Gamification for Information Retrieval*, pages 2–6.

Grace Araneta, M., Eryigit, G., König, A., Lee, J.-U., Luís, A., Lyding, V., Nicolas, L., Rodosthenous, C., and Sangati, F. (2020). Substituto-a synchronous educational language game for simultaneous teaching and crowdsourcing.

Jasonarson, A. (2021). Málfróði: Að nota lýðvirkjun og leikjavæðingu til að afla gagna fyrir íslenska tungu.

Jurgens, D. and Navigli, R. (2014). It's all fun and games until someone annotates: Video games with a purpose for linguistic annotation. *Transactions of the Association for Computational Linguistics*, 2:449–464.

Ogawa, H., Nishikawa, H., Tokunaga, T., and Yokono, H. (2020). Gamification platform for collecting task-oriented dialogue data. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 7084–7093.

10. Language Resource References

Snæbjarnarson, Vésteinn and Einarsson, Bergur Tareq Tamimi and Auðunardóttir, Ingibjörg Iða and Sæmundsson, Unnar Ingi and Bjarnadóttir, Hildur and Gunnarsson, Helgi Valur and Einarsson, Hafsteinn. (2021). *NQiI - Natural Questions In Icelandic - v1.0.*