EURALEX XIX

Congress of the European Association for Lexicography

Lexicography for inclusion

7-11 September 2021

Ramada Plaza Thraki

Alexandroupolis, Greece

www.euralex2020.gr

Proceedings Book
Volume 1

Edited by Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras

EURALEX Proceedings

ISSN 2521-7100

ISBN 978-618-85138-1-5

Edited by: Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras

English Language Proofreading: Lydia Mitits and Spyridon Kiosses

Technical Editor: Kyriakos Zagliveris

2020 Edition

"Game of Words": Play the Game, Clean the Database

Arhar Holdt Š., Logar N., Pori E., Kosem I.

University of Ljubljana

Abstract

The paper presents the "Game of Words" (in Slovene: Igra besed), a mobile application purposed for a gamified improvement of two automatically compiled dictionaries for Slovene: the Collocations Dictionary of Modern Slovene and the Thesaurus of Modern Slovene. We provide a brief history of the game, and introduce its two modules that utilize collocation and synonym data respectively. A significant part of the paper is dedicated to the presentation of all the steps of the preparation of both datasets; this included addressing challenges brought by automatically extracted data from the corpus, and filtering out sensitive content considering the potential users. Crowdsourcing aspects of the game are discussed, especially in terms of the lessons learned in the development process, and how one needs to strike a good balance between the lexicographic intentions, numerous possibilities of using language information, and the enjoyment and motivation of playing the game. The paper concludes by outlining future plans, including further developments of the game both on the level of game modules and languages offered, in the framework of European projects and initiatives.

Keywords: Game of Words; GWAP; collocations; synonyms; crowdsourcing; gamification; responsive dictionary

1 Introduction

The recently published Collocations Dictionary of Modern Slovene and the Thesaurus of Modern Slovene[1] are innovative from the perspective of the dictionary-making process, introducing a concept of a "responsive dictionary" – a dictionary that is compiled entirely through automatic extraction methods, as soon as possible made available to the community both as a lexical database and as an online language resource, and after that continuously and transparently lexicographically improved, also with the help of user-provided feedback. For example, through the dictionary interface, users can vote positively or negatively on the automatically extracted data, and in the case of the Thesaurus also suggest additional synonyms to be included in the dictionary database. The Collocations Dictionary and the Thesaurus were already presented in literature (Krek et al. 2017; Arhar Holdt et al. 2018; Kosem et al. 2018a). Together with the methodology for user involvement, the need for getting users motivated to participate in resource improvement and enhancement was highlighted as one of the most crucial elements of the newly proposed workflows.

An important aspect of user involvement in the responsive dictionary development is that the involvement is direct and the task is explicit, i.e. the users are aware of the purpose of the task and the aim of their participation. However, we have to keep in mind that the involvement/feedback via the dictionary interface is secondary to dictionary consultation. This can significantly affect user motivation and the time they are willing to dedicate to providing feedback or suggestions. Consequently, we have decided to look into the possibilities offered by gamification, specifically games with a purpose, where the main purpose for the users is enjoyment, while the task remains in the background so that the users are often unaware they are providing information useful for linguistic/lexicographic purposes. As a result, we have developed a language game called Igra besed (Game of Words) that challenges players on their knowledge of Slovene collocations and synonyms, while supporting the improvement of the previously mentioned lexical resources: the Collocations Dictionary of Modern Slovene and the Thesaurus of Modern Slovene.

In the paper, we present the idea behind the mechanics of the game and the implementation of all its modules, together with the description of the data-preparation process. Namely, the data to be included in the game had to be filtered in order to avoid sensitive issues, such as derogatory and potentially offensive lexica. Then, the collocation and synonym module are presented in more detail, followed by the discussion of the crowdsourcing perspectives and shortcomings of the game. We conclude by offering final remarks and presenting future plans, including the development of next versions of the game in a wider European context, i.e. collaboration with current European projects and actions.

2 Gamification and Language Resources

Gamification is closely related with the notion of Games with a Purpose (GWAP), a crowdsourcing mechanism for (typically benevolent, i.e. voluntary, not paid) implicit crowdsourcing. In the categorisation of crowdsourcing approaches, "implicit" means that the purpose of the task is secondary or even partially hidden to the participants, as opposed to "explicit" crowdsourcing where the task is the primary purpose of participation (Lyding et al. 2018). In the case of

[1] Both dictionaries are freely available online, (also) through an English user interface: the Collocation dictionary at https://viri.cjvt.si/kolokacije/eng/ and the Thesaurus at https://viri.cjvt.si/sopomenke/eng.

GWAPs, participants' primary goal and the source of motivation is to play a game and by doing so, they perform a specific underlying, pre-designed task.
When it comes to creating lexical infrastructure, some successful GWAPs were designed to annotate language data, for example Phrase Detectives (Poesio et al. 2013), JeuxDeMots (Lafourcade 2007), and ZombiLingo (Guillaume 2016). Nonetheless, the use of gamification, and crowdsourcing in general, in lexicography is still very limited. While the benefits of crowdsourcing have been thoroughly established (Lew 2013; Abel & Meyer 2013; Benjamin et al. 2015; Fišer & Čibej 2017), the implementation lags behind. For example, one popular way of dictionaries promoting their activities as crowdsourcing (or citizen science) is enabling user feedback via online forms or emails. However, this is rarely crowdsourcing as it is based on individual rather than crowd contributions, plus the methodology of including suggestions in the lexicographic workflow is not necessarily transparent. Rather than turning to crowdsourcing for the sake of keeping up with the new trend, we propose a gradual inclusion of user-involving approaches, where new ideas and their implementation are thoroughly evaluated by the users and can be continuously improved. The evaluation of crowdsourcing techniques available through the dictionary interface of the Collocations Dictionary and the Thesaurus were presented by Pori et al. (2020) and Arhar Holdt (2020).
The beginnings of Igra besed (Game of Words) go back to 2014 when the first version was published as an online game.[2] The game was part of the project funded by the Slovenian Ministry of Culture, with the aim being to devise innovative ways to promote the Slovenian language, and its use. The first version of the game was not conceived as a Game with a Purpose, we simply wanted to make a game that would be fun and didactic at the same time. But due to the lack of suitable and free language resources we were forced into using automatically extracted data from the Sketch Engine, and relatedly, devising the game in a way where the potential noise in the data would not affect the playing experience. This was also the reason why the first version had only one playing mode: players had to type in three possible collocates of the word, which were then scored according to the ranking on the list. There were two formats of playing: practice and duel (participants were able to challenge another player to a duel, and the latter could accept or reject this challenge). As far as working with automatic data was concerned, the underlying assumption, which was later confirmed by the analysis of user data, was that the users will not intentionally type in wrong information (in this case collocates).
The transition of Game of Words to a GWAP was mainly driven by three developments. Firstly, the proposal for a new dictionary of Modern Slovene published at the time (Gorjanc et al. 2015), which was a response to the lack of lexicographic resources describing modern Slovene, described in detail how crowdsourcing methods could be implemented in lexicographic workflow to speed up the dictionary-making process. Moreover, also as the answer to the lack of resources on modern Slovene, was the introduction of responsive dictionaries - using the approach "publish good (automatically extracted) data now, clean later" -, of which the crowdsourcing component was a key part. Secondly, crowdsourcing experiments with collocations we have conducted have indicated that explicit crowdsourcing was not the most suitable method for dealing with this particular type of lexical information. And thirdly, a detailed analysis of Game of Words logs has pointed out a potential of the game for not only validation of collocational information, but also for other more complex tasks such as determining the definite or indefinite form of the adjective before noun.
In version 2, Game of Words was thus significantly upgraded, in terms of content, playing modes, and medium. Much more collocational data was included, not only in terms of number of lemmas and collocations but also in terms of syntactic structures. Also, synonym data and synonym playing module were added. Importantly, based on the feedback of the users of the first version, the game has moved from the online to the mobile medium, i.e. was developed as a mobile app, available both for Android and iOS devices. The development of the version 2 was made within two different projects funded by the Slovene Ministry of Culture; "The promotion of a language mobile app" funded the development of the mobile app and the upgrade of the collocation module, and "The promotion of the Thesaurus of Modern Slovene" funded the addition of the synonym module.

3 Data Preparation

3.1 Collocational Data

The basis for the collocation module of the game was the Collocations Dictionary of Modern Slovene, comprising 35,989 headwords and 7,338,801 collocations. Collocations were automatically extracted from the 1.2-billion-word Gigafida corpus of Slovene (Logar Berginc et al. 2012), using the Sketch Engine API, and also additionally filtered at the post-processing stage (for more see Kosem et al. 2018a). The Collocations Dictionary offered a rich resource of potential data for the game, allowing us to expand on the number of syntactic structures offered, something that was also requested by numerous players of the first version of the game. Based on the findings of the evaluation of automatically extracted collocational data, which was conducted within the KOLOS project[3] (Kosem et al. 2018b), we selected five syntactic structures that exhibited the highest percentage of good collocation candidates: adjective + noun (*osnovna šola* 'primary school'), noun + noun in genitive (*rezervacija sobe* 'room reservation'), verb + noun in accusative *(prevesti tekst* 'to translate text'), adverb + verb (*ironično komentirati* 'comment ironically'), and adverb + adjective (*zelo lep* 'very beautiful'). Since headwords (nouns, verbs, adjectives, adverbs) occupied different positions in the structures, this meant nine different syntactic structures altogether (for "noun + noun in genitive", only the version with the headword in the first position was taken). The total number of collocations in these nine syntactic structures was 2,723,551.

[2] The game is still available online at https://www.igra-besed.si/, however, only through a Slovene interface.
[3] KOLOS is the acronym for the national research project "Collocations as a Basis for Language Description: Semantic and Temporal Perspectives", funded by the Slovenian Research Agency (J6-8255).

The second step involved reducing the number of headwords and collocations according to statistical, morphosyntactic, and semantic criteria. This was needed in order to address certain problems that could affect the playing experience. The statistical filter we added was a minimum of 10 collocates per syntactic structure; this was mainly needed because of the new playing mode Choose (see Section 4.1) which required at least nine collocates. It should be stressed that this filter was implemented after the database had been filtered according to morphosyntactic and semantic criteria.

Morphosyntactic filters were related to either known problems with corpus annotation, or problems in collocation form due to lack of suitable resources. Thus, we removed all collocations containing collocates that were not in the Slovene morphological lexicon Sloleks (Dobrovoljc et al. 2018), which was used for assigning the right form to the collocates according to the case required by the syntactic structure. This filter was not applied in structures adverb + verb and adverb + adjective as lemma forms were always used in them. Also, we removed all collocations containing collocates beginning with a capital letter (there were no such headwords) as the evaluation showed that a large proportion of them is noise or are in incorrect form. In fact, many of these collocates were already removed in an earlier step as they were not in Sloleks. Another problematic group that was removed were 73 homonymous headwords as they were found as one lemma, so the automatically extracted data contained collocations for all, e.g. noun headword *tema* contained data for *tèma* ('dark') and *téma* ('topic'). Finally, we removed 1,370 verbs as headwords in the relation verb + noun in accusative, as these verbs never or very rarely occurred with an object, meaning that the majority of their collocates were errors.

The semantic filter used was in the form of a stoplist that contained words (featuring as either a headword or a collocate) with a negative connotation in at least one of its meanings. The list was based on existing resources such as dictionaries, and privately compiled lists by researchers or journalists.[4] Words on the list included insults, pejorative expressions, vulgar words, etc. but also words that could cause discomfort like verbs *ubiti* ('to kill'), *uničevati* ('destroy'), *groziti* ('threaten'). While it could be argued that many of these words are not really problematic, we gave priority to the fact that the game could also be used in educational settings with young(er) users.

In the final step, rather than filtering the results, we conducted an additional step of post-processing. Namely, we added reflexive pronouns "si" or "se" to 1,358 reflexive verbs, or when the verb was reflexive in one of its meanings, indicated the possible use of reflexive pronoun in brackets, e.g. *umivati (se)* ('to wash (oneself)'). This was needed because listing a verb without the reflexive pronoun could elicit incorrect collocates from the players.

The final dataset for the collocation module contained 23,303 headwords (9,132 nouns, 8,423 adjectives, 3,953 verbs, and 1,795 adverbs) and 2,448,994 collocations. For comparison, the first version of the game contained 10,578 headwords (5,237 nouns and 5,341 adjectives) and 2,928,177 collocations.[5]

3.2 Synonym Data

The basis for the synonym module of the game was the Thesaurus of Modern Slovene. In its current version, the Thesaurus comprises 105,473 (single- or multi-word) headwords. Synonyms for these headwords were obtained automatically from The Oxford®-DZS Comprehensive English-Slovenian Dictionary (Šorli et al. 2006) and the Gigafida reference corpus of written Slovene (Logar Berginc et al. 2012). The synonyms -- or more precisely, 'synonym candidates', as the data has not yet been lexicographically checked -- are separated into two groups: "core" and "near". According to their assigned score of relatedness to the headword, "core" synonyms are believed to be the most relevant, and "near" synonyms only optionally useful (see Krek et al. 2017 for a more detailed methodology on the scoring and ranking of synonym candidates). From the moment the thesaurus was published, the user community also had the chance to provide additional suggestions for synonyms of any given headword (for more information on user involvement techniques see Arhar Holdt et al. 2018).

For the game, we wanted to use headwords that: (a) have enough synonyms for enjoyable gameplay; (b) are likely to be reliable considering the automated procedures that were used for the preparation of the Thesaurus; (3) are non-problematic for pedagogical use. We also wanted the game to progress in difficulty, as explained in Section 4.2. In the following paragraphs, we describe the decisions made to achieve the listed goals.

First, we arranged the headwords by the frequency of their corresponding synonym candidates. In the frequency count, we included only the automatically acquired core and near synonyms, not also the user-provided synonyms. The result was an ordered list ranging from the headword *hud* ('wild') with 110 synonym candidates to *gnati na vso moč* ('pushing with everything one has'), which is an example among 43,088 headwords with only one synonym candidate. Next, we filtered the list. To begin with, we filtered out 18,165 headwords consisting of three or more words, e.g. *ukvarjati se z; postaviti na glavo; po drugi strani* ('to attend to; to turn upside down; on the other hand'). We have furthermore eliminated from the frequency count all the synonim candidates with 3 or more words. This filtering step was conducted because it had been determined (Čibej & Arhar Holdt 2019) that due to the methodology features, multi-word synonym candidates include a higher portion of irrelevant material. Additionally, we filtered from the list 117 headwords with only one or two letters, e.g. *da*; *po; za* ('that; after; for'), as these comprise solely abbreviations and grammatical words, less suitable for the game. We have also filtered out any headwords that were on the previously described stoplist of derogatory and vulgar words (see Section 3.1), e.g. *pizda; peder* ('cunt; queer'). The final filtering condition was that among the remaining synonym candidates, at least five had to be in the core category. In this way, when a player entered three of these synonyms, the game could offer the remaining core synonyms as didactical suggestions (Section 4.2.3).

After filtering, we separated the list into single- and two-word headwords. For the first and main step of the manual selection, we focused on the 5,085 single-word headwords with at least 10 synonyms. In the manual check-up, we

[4] The list is continuously updated for future versions of the game, and other purposes.

[5] The reason why the first version had such a high number of collocations per headword lies in the fact that we included complete lists of collocates without any frequency threshold, so even collocates with a frequency of 1 made it to the list.

eliminated from the list 529 additional headwords that could be potentially problematic for pedagogical use. This step was entirely subjective, its primary goal was to ensure the safe use of the game in the classrooms. If the headword raised any doubt, it was marked for removal. It was interesting to notice that in the case of the synonym module, we were prone to eliminate not only headwords that were vulgar or sensitive, but also headwords that were probing the player for vulgar or sensitive synonyms. Typical examples of that were words alluding to sexual activities, such as *drgniti; položiti; poriniti* ('to rub; to lay; to push'). Another group that was module-specific were non-derogatory words describing unwanted human features, as these headwords might encourage students to enter (as "synonyms") names of their classmates or similar. Some examples: *lizunka; parazit; čudak* ('kiss-ass; parasyte; weirdo').
From the list of 453 two-word headwords with at least 10 synonyms, 200 were manually selected for the game, all of them verbs with a reflexive pronoun, e.g. *umakniti se; obrniti se* ('to remove oneself; to turn oneself'). Compared to other examples of two-word headwords, these demonstrated the most reliable synonym candidates. After this step, the list comprised 4,756 headwords. To reach the desired 5,000, we manually selected the remaining 244 headwords from the single-word headwords with 9 synonyms, following the same criteria for selection as described above. Some examples of headwords that made it to the dataset in this final step were e.g. *dragocen; nalepiti; čarobno* ('valuable; to stick; magically'). Finally, the 5,000 headwords were separated into 500 sets by 10 according to the number of their synonyms. The sets were manually checked and rearranged to ensure that words from the same word-families or/and with the same meaning were not included in the same set, e.g. *odločen; odločno* ('decisive; decisively') where the first headword remained in set 003, while the second was moved to set 005. Arranged headwords were provided to the developers together with core and near synonyms that are used for scoring the player-provided entries (Section 4.2.2).

4 Game Modules

In this section, we present both collocation and synonym modules in more detail. It is noteworthy that even though the main focus in developing the mobile app was on gamification and language data, the visualization part of the end product was almost equally important. We explicitly wanted a clear and non-confusing appearance of the application; a design that would not distract the players and would enable them to focus on the content, rather than colors, shapes, or movements. Some initial players' reactions to the visualization of the Game of Words suggest we succeeded in this attempt, yet further user evaluations will be needed to confirm (or discard) this - for now - satisfying response.

4.1 Collocation Module

4.1.1 Playing Modes

The collocation module, launched in September 2019, has introduced significant changes to the game compared to version 1. Namely, the old, online version of the game had only one grammatical structure to be completed with collocates (*adjective* + *noun*), only one game mode to be played (*typing*), and was also quite basic and straightforward regarding scoring of the results. In the mobile version, three different modes of playing are available: TYPE, CHOOSE, and DRAG.
In the Type mode, the format of which was not changed from the online version, players have to complete collocations by typing in three collocates of the given headword, e.g. three adjectives that typically precede a given noun (as shown in Figure 1). One game room of this mode included three headwords.
In the Choose mode, players are presented with three groups of three collocates and have to choose the most typical collocate in the group. Then, for bonus points, they need to arrange their selection according to the (perceived) typicality. Collocations used in the game are selected from three ranges in the list of collocates, one from each range – top 30%, 30-55%, 55-100%. In that way we ensure that for example three collocates next to each other in the list are not selected, and that it is easier to detect the most typical ones. One game room of the Choose mode included three headwords.
In the Drag mode, the players need to drag the collocates to one of the three options: headword A, headword B, and Bin. They are provided with nine randomly ordered words, consisting of three collocates from the list of one headword, three from the list of the other headword, and three distractors (at the moment, taken from completely different headwords and grammatical relations). The headwords compared are picked at random, but belong to the same word class and the same position in a given syntactic structure. Only one headword pair per game room was offered.

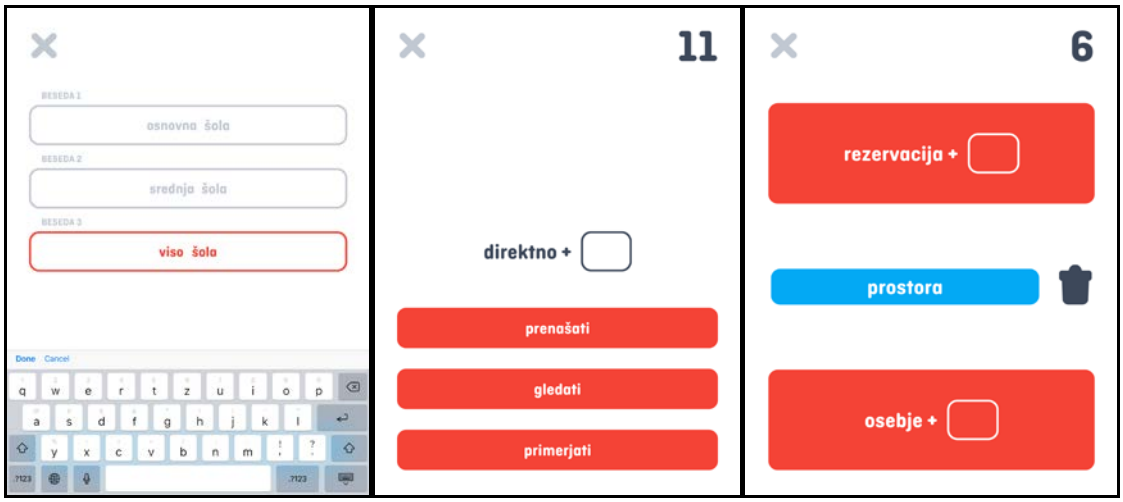

Figure 1: Game modes in the collocation module: *Type; Choose;* and *Drag*.

All modes are available in the Competitive format, which automatically creates game rooms at regular time intervals. Thus, the players, after picking the mode they want to play, enter the game room running at the time (or wait until the next one starts), play the same words/collocations and compete against each other. The second format used was called Thematic and gave us more control in picking group headwords on a certain common topic (e.g. winter holidays, Christmas, 50-year anniversary of Moon landing). The Thematic format, which was primarily devised to facilitate the promotional activities of the game, was open in a specified time span, each player could play the topic batch only once, and the top players received practical prizes.

4.1.2 Scoring

The numerical score is one of two key feedback pieces of information that language games provide. The non-numerical feedback can be of different nature: the correct answer (if a wrong one is given), a suggestion of another possible response, etc. Due to the project financial restrictions, the collocation module of the Game of Words only provides the numerical score.

When discussing the scoring options, we first and foremost wanted to simplify the scoring system used in the first version as it was too detailed and difficult to comprehend. Furthermore, we paid attention to two issues: *when* in the gaming cycle should the player be presented with the score, and *what would be the best way* to distribute the points. As to *when*, the score is now shown at the end of each headword (pair) - while this extends the total time in the Type and Choose modes, it was thought important that the players have time to inspect their results on individual headwords rather than having a long scrollable list of results at the end.

In terms of point distribution, the scoring system was founded on the collocation salience data; however, the number of collocates per headword and the rating of each collocate were taken into account as well. For the Choose mode, three-point groups were formed based on the collocate ranges. For the Type mode, five scoring groups are used. For the Drag mode, we also used three groups, adding the small "reward bonus" (when a collocate from the list of a particular headword was thrown in the bin) to the correct and incorrect ones. The scoring was devised in a way that enabled comparability across different modes, given that, in addition to having leaderboards for each game mode, we also had the common leaderboard. In order to make the scoring system easy to comprehend, points are also translated into a five-star scale.

4.2 Synonym Module

4.2.1 Playing Modes

The synonym module of the game was launched in January 2020. The module introduces a solo playing format in which players have to enter three synonyms for a given word. The module is oriented towards didactical purposes and thus offers some learning-oriented features (described below). The didactical angle was included on the request of language teachers who were participating in the project as user evaluators of the Thesaurus of Modern Slovene. In their feedback, it was highlighted that dictionary-based gamification for vocabulary acquisition in Slovene would be extremely valuable, as existing teaching resources for his topic were very scarce and limited in scope.

The game includes 5,000 words, separated into 500 levels with a set of 10 words. The levels progress in difficulty established by the number of existing synonym candidates in the Thesaurus of Modern Slovene. There are two playing modes available, the Game mode (the players progress from one level to another), and the Practice mode (the players can choose the level). For example, level 1 includes headwords like *odličen* ('excellent'), *divji* ('wild') and *uničiti* ('to destroy'), which all have more than 50 synonyms in the Thesaurus. A more difficult level 300 includes headwords like *posušen* ('dried'), *poskus* ('a trial') and *osvoboditi* ('to liberate') that have around 15 synonyms in the Thesaurus. The adjective *odločen* ('decisive') presented in Figure 2 appears at level 3 (or 003) as indicated on the top of the screen.

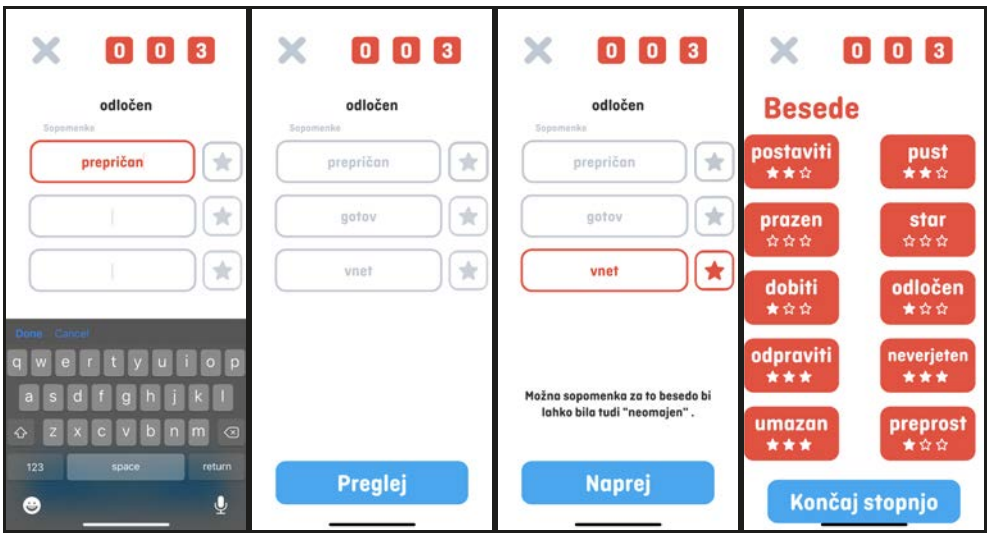

Figure 2: Synonym module: entering synonyms for *odločen* ('decisive') as part of Level 003.

4.2.2 Scoring

The scoring system for the module is very straightforward. After every game, each player-provided synonym that is also found in the Thesaurus database is attributed a star. In this way, players can collect between zero and three stars per game. After a set of 10 games, the stars are transformed into game points, where each star is worth 100 points. The points are then added to the overall score and the player progresses in the Hall of Fame accordingly. For example, player-provided synonyms for *odločen* ('decisive') in Figure 1 are *prepričan* ('positive'), *gotov* ('certain') and *vnet* ('ardent'). For each of them, the player received one star. It is obvious that as long as the Thesaurus database is under development, such scoring is not entirely precise and can cause some frustration among the players. For this reason, we have included a disclaimer explaining that with every upgrade more user-suggested synonyms will be included in the scoring.

4.2.3 Didactic Value

From the didactical point of view, the game facilitates and encourages the use of ICT in the classroom. Both modules of the game are aligned with the curriculum for Slovene as a school subject. The focus of the game is on working with empirical data relevant for specific thematic parts of school curriculum, which enables teachers and students easier transfer of knowledge into practice. In particular, under the guidance of the teacher, the game helps develop metalinguistic competence by teaching the students the concepts of synonyms, collocations, parts of speech (noun, adjective, verb etc.); develop linguistic competence through increased vocabulary and knowledge of syntax; learn to identify synonyms and their semantic/stylistic differences; learn to identify collocations as multiword units; learn to evaluate pros and cons of different types of language resources and to anticipate and identify errors in automatically prepared language resources; learn the importance of openly accessible language data in the digital era and the possibilities of including language community in the creation of openly available language resources
In comparison to the collocation module, the synonym module has developed special features that support the process of teaching Slovene. For this, the Practice mode, where the player can jump to any given level and play without being scored on the joint leaderboard, is particularly useful and was in fact developed with the pedagogical purpose in mind. Using this mode, teachers have the possibility to find the levels optimally suited for their specific teaching purposes and focus on those in the classroom. Another important feature of the module is that it provides learning material. After the player-provided entries are evaluated, the game shows a possible synonym from the database that was not entered, thus helping the player enrich their vocabulary in Slovene. For example, on the last screen in Figure 2, the game suggested: *Možna sopomenka za to besedo bi bila tudi "neomajen"*. (A synonym for this word might also be 'unwavering'.) At the moment, the suggestions are acquired automatically from the database. In the future, we plan to manually check the suggestions as well as complement them with selected corpus examples to demonstrate their use in context.

5 Crowdsourcing Perspectives of the Game of Words

In this section, we discuss crowdsourcing aspects of the Game of Words and its playing modules in more detail. The whole crowdsourcing workflow consists of three stages, namely data preparation, annotation (when the game is played), and data analysis or results implementation. We have already described in detail how both datasets were prepared, but it is important to add that each headword, collocate, and collocation, as well as each headword and its synonyms had to be indexed before being uploaded into the database of the game. The same IDs are then part of the exported data, as this is the only way to ensure valid and quick analysis of the results. One shortcoming in terms of gamifying collocational and synonym data, and any type of linguistic data for that matter, is that sensitive and vulgar content has to be left out, especially if the game also serves pedagogical purposes.
The quality and reliability of data annotation is largely dependent on the playing mode. For example, the Type mode,

which is the most mobile unfriendly and was initially questioned by our designers, is the most reliable mode for crowdsourcing as the players need to enter their answers, whereas in the Choose and Drag modes, they simply (have to) choose between three given options. This is particularly problematic in the Choose mode where there is a chance, albeit a small one, that all three collocates offered are bad ones. The mobile unfriendliness of Type did not seem to bother the players in the synonym module, but it is true there they were not presented with a choice. In the collocation module, however, it was the Drag mode that proved the most popular mode among the players.

The differences in reliability of different modes made us think of what that means for evaluating annotator agreement. For example, how many player decisions are necessary and what needs to be the level of player agreement for a collocation/synonym that we can consider it to be good? We have looked at the data from the first version in attempt to get an answer to this, and we agreed that the acceptable number of annotations would be around 20. Clearly, the number of accepted answers coming from the Type mode could be much lower than at Choose and Drag, which is why we started to devise a scoring system in which a Type "votes" would have a higher annotation value than the Choose and Drag ones.

Another issue closely related to game modes is the crowdsourcing tasks one can do with them. The Type and Choose mode, for example, are much more suitable for the validation of good collocations or synonyms than the identification of bad ones. One can for example consider the never entered or chosen collocates/synonyms, especially those at the top of the list, as potentially bad, but this could still mean a great deal of manual analysis. The Drag mode, however, with the Bin option is also suitable for cleaning the bad collocation candidates.

One crucial matter that is vital for gamification, and which we have perhaps neglected a little bit when preparing both modules, is keeping good and regular control over the data that is being annotated. With that we mean that you cannot import a large dataset into the game, leave it for several months and hope that as much data as possible gets annotated. Let us take collocations, for example. From the lexicographic perspective, everything revolves around headwords, so the best possible method would be to crowdsource all collocations in all the syntactic structures of one headword first, and then move to the next headword. But this approach is not game-friendly as the players would get bored easily. Moreover, in our case, the game rooms for the collocation module are created randomly, and since the dataset is very large, this means that the likelihood of the same headword and collocates being offered more than once are rather low. In fact, we realized that it was the Thematic mode that showed the most potential for crowdsourcing since it was the easiest way to control collocational data, and the best way to motivate large groups of players. The synonym data, on the other hand, is much better controlled, with the only problem being that the order of levels is fixed; considering that the number of players completing a level decreases with each level, lower levels will get annotated more often than higher ones.

As the example of the gaming vs. lexicographic perspective above shows, crowdsourcing purposes (and the data they are related to) and optimal playability often contradict each other. An example of this was our experience when designing different games, as many features that would make the game more attractive, e.g. the first and last letter of the collocate shown, could not be used as the data had to be cleaned first. The entire process of game development became one great balancing act between the lexicographic intentions, numerous possibilities of using language information, and the enjoyment and motivation of playing the game.

6 Conclusion and Future Work

The gamification of lexicographic data in Slovenia is still in its infancy; however, the experience gained during the development of the Game of Words is invaluable for our community. The initial evaluations and analyses have shown promising results, but have at the same time pointed out several mistakes in our approach, which we aim to rectify in future versions of the game.

For those interested in developing games of this type, it should be stressed that the development needs to involve a very interdisciplinary team, i.e. not only linguists/lexicographers and computational linguists (for data preparation), but also mathematicians (for scoring etc.), graphic designers (for app design) and developers (for app programming). Crucially, a great deal of continuous proactiveness (i.e. dissemination) after the launch of such an app is required due to a plethora of different types of apps, not only linguistic ones, available on the market.

There is undoubtedly a lot of room for improvement of the game, both on the side of playability and crowdsourcing procedure. For instance, in addition to the already mentioned problem of low control over data annotation, we have found it difficult to get easy access to user logs - as the developers finished their work and moved to other projects, it has been often hard to get a person to export the data in the desired format. This of course means that the 'responsiveness' of our analyses, and relatedly dictionaries, has not been as quick as we would have liked.

The future of the Game of Words has become much brighter recently, as the game has attracted the attention of the European Lexicographic Infrastructure (ELEXIS), a Horizon 2020 project, which has one of the activities focused on the development of techniques and tools for crowdsourcing lexicographic data. This resulted in the development of the next version of the game, which will bring the game to other languages (beginning with English, Estonian, and Dutch) and address many crowdsourcing-related shortcomings mentioned in this paper. It will introduce a solo format for collocations, admin tools for easier uploading/downloading of the data and games, and dynamic data selection (e.g. non-annotated collocates will be given priority over already annotated ones). At the same time, the game attracted interest from the researchers involved in the EnetCollect COST Action aimed at connecting crowdsourcing and language learning (Lyding et al. 2018), and there are already plans to develop a module for marking corpus examples that could be potentially problematic for didactic purposes, e.g. due to the presence of sensitive issues or vulgar/derogatory vocabulary (Dekker et al. 2019) - a module that could serve the purposes of both language teachers/learners and lexicographers. Therefore, by widening the community working on the development and dissemination of the game, we can hope that the potential of crowdsourcing in lexicography and language learning can finally be fully exploited.

7 References

Abel, A., Meyer, C. (2013). The dynamics outside the paper: user contributions to online dictionaries. In *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference*. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut, 179-194.

Arhar Holdt, Š., Čibej, J., Dobrovoljc, K., Gantar, A., Gorjanc, V., Klemenc, B., Kosem, I., Krek, S., Laskowski, C., Robnik Šikonja, M. (2018). Thesaurus of Modern Slovene: By the Community for the Community. Čibej, Jaka et al. (eds.) *Proceedings of the 18th EURALEX International Congress: lexicography in global contexts.* Ljubljana: Ljubljana University Press, Faculty of Arts. 401-410.

Arhar Holdt, Š. (2020). How Users Responded to a Responsive Dictionary: The Case of the Thesaurus of Modern Slovene. In *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje*, 46(2): in print.

Benjamin, M. (2015). *Crowdsourcing microdata for cost-effective and reliable lexicography*. No. CONF, pp. 213-221. Accessed at: https://infoscience.epfl.ch/record/215062 [07/05/2020].

Čibej, J., Arhar Holdt, Š. (2019). Repel the syntruders! A crowdsourcing cleanup of the thesaurus of modern Slovene. In I. Kosem, S. Krek (eds.) *eLexicography in the 21st century: proceedings of eLex 2019 Conference, 1-3 October 2019, Sintra, Portugal.* Brno: Lexical Computing, 338-356. Accessed at: https://elex.link/elex2019/wp-content/uploads/2019/10/eLex-2019_Proceedings.pdf.

Dekker, P., Zingano Kuhn, T., Šandrih, B., Zviel-Girshin, R., Arhar Holdt, Š., Schoonheim, T. (2019). Corpus filtering via crowdsourcing for developing a learner's dictionary. In I. Kosem, T. Zingano Kuhn (eds.) *eLexicography in the 21st century (eLex 2019): smart lexicography: book of abstracts.* Brno: Lexical Computing, 84-85. https://elex.link/elex2019/wp-content/uploads/2019/10/eLex_2019-Book_of_abstracts.pdf.

Dobrovoljc, K., Krek, S., Erjavec, T. (2018). The Sloleks Morphological Lexicon and its Future Development. In V. Gorjanc, P. Gantar, I. Kosem, S. Krek (eds.) *Dictionary of Modern Slovene: problems and solutions.*Ljubljana: Ljubljana University Press, Faculty of Arts, 42-63.

Fišer, D. and Čibej, J. The potential of crowdsourcing in modern lexicography + Crowdsourcing workflows in lexicography. (2017). In V. Gorjanc et al. (eds.). *Dictionary of Modern Slovene: problems and solutions*, (Book series Prevodoslovje in uporabno jezikoslovje). 1st ed., e-ed. Ljubljana: Ljubljana University Press, Faculty of Arts. Accessed at: http://www.ff.uni-lj.si/sites/default/files/Dokumenti/Knjige/e-books/dictionary_of_modern_slo.pdf

Gorjanc, V., Gantar, P., Kosem, I., Krek, S. (eds.) (2015). *Dictionary of Modern Slovene: problems and solutions*. Ljubljana: Ljubljana University Press, Faculty of Arts, 2017.

Guillaume, B., Fort, K., Lefebvre, N. (2016). Crowdsourcing Complex Language Resources: Playing to Annotate Dependency Syntax. *Proceedings of the International Conference on Computational Linguistics (COLING).* Accessed at: https://hal.inria.fr/hal-01378980/ [07/05/2020].

Kosem, I., Krek, S., Gantar, P., Arhar Holdt, Š.; Čibej, J., Laskowski, C. (2018a). Collocations dictionary of modern Slovene. In J. Čibej et al. (eds.) *Proceedings of the 18th EURALEX International Congress: lexicography in global contexts.* Ljubljana: Ljubljana University Press, Faculty of Arts, 989-997.

Kosem, I., Gantar, P., Krek, S., Čibej, J.. Arhar Holdt, Š. (2018b). The Good, the Bad and the Noisy? An Analysis of Inter-Annotator Agreement on Collocation Candidates in Different Grammatical Relations. In J. Čibej et al. (eds.) *The XVIII EURALEX International Congress: Lexicography in Global Contexts Book of Abstracts*. Ljubljana: Ljubljana University Press, Faculty of Arts, 71-72.

Krek, S., Laskowski, C., Robnik Šikonja, M. (2017). From translation equivalents to synonyms: creation of a Slovene thesaurus using word co-occurrence network analysis. In I. Kosem et al. (eds.), *Proceedings of eLex 2017: Lexicography from Scratch, 19-21 September 2017*, Leiden, Netherlands.

Lafourcade, M. (2007). Making people play for Lexical Acquisition with the JeuxDeMots prototype. SNLP'07: 7th International Symposium on Natural Language Processing, Dec 2007. Pattaya, Chonburi, Thailand, 7.

Lew, R. (2013). User-generated content (UGC) in online English dictionaries. In A. Abel and A. Klosa (eds.) *Ihr Beitrag bitte! – Der Nutzerbeitrag im Wörterbuchprozess (OPAL – Online publizierte Arbeiten zur Linguistik)*. Mannheim: Institut für Deutsche Sprache, 9-30.

Logar, N., Grčar,M., Brakus, M., Erjavec, T., Arhar Holdt, Š. and Krek, S. (2012). *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES: gradnja, vsebina, uporaba*. Ljubljana: Trojina, zavod za uporabno slovenistiko; Fakulteta za družbene vede.

Lyding, V., Nicolas, L., Bédi, B., Fort, K. (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.

Poesio, M., Chamberlain, J., Kruschwitz, U., Robaldo, L. and Ducceschi, L. (2013). Phrase Detectives: Utilizing Collective Intelligence for Internet-scale Language Resource Creation. ACM Transactions on Interactive Intelligent Systems 3, 3:1–3:44.

Pori, E., Kosem, I., Čibej, J., Arhar Holdt, Š. (2020). User study: The attitude of dictionary users towards automatically extracted collocation data. In I. Kosem and P. Gantar (eds.) *Slovenščina 2.0*, in print.

Šorli, M., Grabnar, K., Krek, S., Košir, T. (2006). Oxford-DZS comprehensive English-Slovenian dictionary. *In Proceedings of the XII EURALEX International Congress.* Edizioni dell`Orso: Universita di Torino: Academia della Crusca, 631–637.

Acknowledgements

The research presented in this paper was conducted within two projects titled "The promotion of a language mobile app" and "The Thesaurus of Modern Slovene: By the Community for the Community", which were financially supported by the Ministry of Culture of the Republic of Slovenia (2018–2019). The authors acknowledge the financial support from the Slovenian Research Agency (research core funding No. P6-0411, Language Resources and Technologies for Slovene). The authors also acknowledge the project Collocations as a Basis for Language Description: Semantic and Temporal Perspectives (J6-8255) was financially supported by the Slovenian Research Agency. The ELEXIS part of the research received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. Many ideas result from the framework of the CA160105 eNetCollect COST Action. We thank all that made our work possible.