v-trel: Vocabulary Trainer for Tracing Word Relations - An Implicit Crowdsourcing Approach

Verena Lyding
Institute for Applied Linguistics,
Eurac Research Bolzano/Bozen
verena.lyding@eurac.edu

Christos T. Rodosthenous
Open University of Cyprus
christos.rodosthenous@ouc.ac.cy

Federico Sangati
Orientale University of Naples
federico.sangati@gmail.com

Umair ul Hassan
Insight Centre for Data Analytics,
National University of Ireland Galway
umair.ulhassan@insight-centre.org

Lionel Nicolas
Institute for Applied Linguistics,
Eurac Research Bolzano/Bozen
lionel.nicolas@eurac.edu

Alexander König
Institute for Applied Linguistics,
Eurac Research Bolzano/Bozen
alexander.koenig@eurac.edu

Jolita Horbacauskiene
Kaunas University of Technology
jolita.horbacauskiene@ktu.lt

Anisia Katinskaia
Department of Computer Science,
University of Helsinki
anisia.katinskaia@helsinki.fi

Abstract

In this paper, we present our work on developing a vocabulary trainer that uses exercises generated from language resources such as ConceptNet and crowdsources the responses of the learners to enrich the language resource. We performed an empirical evaluation of our approach with 60 non-native speakers over two days, which shows that new entries to expand Concept-Net can efficiently be gathered through vocabulary exercises on word relations.

We also report on the feedback gathered from the users and an expert from language teaching, and discuss the potential of the vocabulary trainer application from the user and language learner perspective. The feedback suggests that v-trel has educational potential, while in its current state some shortcomings could be identified.

1 Introduction

One of the major challenges for the NLP community is the continuing lack of comprehensive and high-quality language resources (LRs) for most languages. While LR creation can partly be approached in an automatic fashion (e.g., by harvesting a vocabulary from existing corpora), it often requires human intervention to reach a high level of quality and coverage. Crowdsourcing (Howe, 2006) is one promising approach that can be leveraged for the task of LR creation. However, any crowdsourcing ambition is faced with the challenge of attracting and retaining crowd-workers and with safeguarding the quality of results produced by the crowd (Daniel et al., 2018). EnetCollect[1] (Lyding et al., 2018; Nicolas et al., 2018) aims at exploring a solution to this challenge by combining the activities performed in language learning with approaches for crowdsourcing language-related datasets. Thus exploring a new path to address the NLP bottleneck of high-quality language resource creation.

In this paper, we present an application for vocabulary training that is specifically designed to crowdsource learners' answers to improve LRs like concept networks. The application builds on top of an architecture for crowdsourcing of language resources (Rodosthenous et al., 2019), which instantiates one of the core ideas of enet-Collect: the *implicit crowdsourcing paradigm* (Section 2). Accordingly, the vocabulary trainer aims at a two-fold purpose: serving automatically

---

[1]COST Action enetCollect: European Network for Combining Language Learning and Crowdsourcing Techniques, http://enetcollect.eurac.edu/

generated vocabulary exercises and gaining continuous input from the learners to improve LRs.

In the remainder of the paper, we first introduce the vocabulary trainer *v-trel* and describe its different modules and their technical implementation (Section 2). We describe how v-trel instantiates a generic prototype architecture for crowdsourcing language resources (Rodosthenous et al., 2019) and discuss technical decisions taken during the implementation process. Second, we describe the first experiment that has been carried out with a small crowd of advanced language learners[2] (Section 3) and discuss the results and their implications on the potential of the proposed approach. Third, we point out and discuss related work relevant to the presented approach (Section 4). Finally, we sum up preliminary conclusions and indicate directions for future work (Section 5).

2   Vocabulary Trainer

The vocabulary trainer builds on top of a prototype architecture for crowdsourcing language resources (Rodosthenous et al., 2019). It implements an *implicit crowdsourcing paradigm* which follows the idea that if a language resource (e.g., a lexicon) can be used to generate language learning exercises, then the answers of learners to these exercises can be used to improve the resource, which in turn will improve the quality and versatility of the exercises generated (Rodosthenous et al., 2019).[3]

2.1   Motivation and Design

The vocabulary trainer delivers interactive vocabulary exercises for learning word senses. The learner is asked to input words which are related to a given word by the semantic relation `RelatedTo`, and will in the future be extended to other relations such as `PartOf`, `AtLocation` etc. The learner input is collected and evaluated to enhance the LRs that it is generated from.

From the language learning perspective, vocabulary exercises play an important role in language learning (Nation and Hunston, 2013). Hulstijn (2013) notes that every word in a mental lexicon has formal as well as semantic associative features. Depending on the learner's level of language, vocabulary building may encompass single

lexical words with a specific meaning or formulaic sequences / lexical chunks where pedagogical relations are structured by a particular object representation or a part of a particular object (Aldabe et al., 2015). As noted by Schmitt (2013), vocabulary learning is a complex phenomenon that may be explored not only from the aspects of form, meaning and usage but also from a representation of different meanings in different contexts.

Vocabulary exercises based on words' semantic relations are considered to be effective activities. Rosenbaum (2001) shows that background knowledge, context and morphology are essential in vocabulary instruction to enable the learner to understand and disambiguate word meanings effectively. The richness of acquired vocabulary depends not only on the number of learned lexical items but also on the ability to connect and share semantic networks of similar concepts. Hadley et al. (2018) argue that *"word learning is not simply the process by which isolated object-label associations are added to the mental lexicon one by one but also involves the learning of interrelated clusters of concepts, in which the knowledge of one concept supports the learning of another"* (p. 42).

From the crowdsourcing perspective, learners are used as crowdworkers to enhance the LR underlying the vocabulary trainer, namely the common sense ontology *ConceptNet*[4] (see Section 2.2). While using the vocabulary trainer for learning word senses the learners are providing their knowledge of related words which is collected and evaluated in order to validate and enhance the LR.

The vocabulary trainer is composed of four modules which are presented in the following subsections: 1) The *exercise generation module* that retrieves words from ConceptNet and generates exercise content (Section 2.2), 2) the *exercise and result storage dispatcher* that ingests the previously created exercise content (Section 2.3), dispatches it to the various learner interfaces and handles the responses from the learners, 3) the *evaluation module* that is responsible for evaluating if learners' contributions are fit for expanding the language resource and assign points to each learner according to the response given (Section 2.4), and 4) the *user interaction module*, where users are presented with the exercises and submit their responses (Section 2.5). In Figure 1, a high-

---

[2]For the initial experiment we involved proficient non-native speakers of English (see Section 3.2 for details).

[3]The related article discusses the paradigm in more detail and points out strategies to counter the risk of collecting wrong or low-quality data from non-proficient learners.

[4]http://conceptnet.io/

level diagram of the vocabulary trainer's architecture is depicted along with the exchange of data between the core modules of the system.

Interested readers are invited to also browse the project repository [5].

2.2 The Exercise Generation Module

The exercise generation module is responsible for content retrieval from language resources like ConceptNet (Speer et al., 2017) which is a large semantic network that describes general human knowledge and how it is expressed in natural language. ConceptNet provides a large set of background knowledge for different terms that not only describes them but also connects them with other terms using relations such as `RelatedTo`, `AtLocation`, `PartOf`, `IsA`, etc.

In its current version, the exercise generation module is able to search ConceptNet for terms connected with the relation `RelatedTo`, `AtLocation` and `PartOf` and generate exercises using a template such as "Name one thing that is <RELATION> <TERM>". For instance, if a search for knowledge that is `RelatedTo` the term "dog" is initiated, ConceptNet will yield results such as "bark", "pet", "animal", etc. The generation module processes these by removing stopwords, duplicates and terms that have a language other than English. The relevant information is stored in a database to be processed later by other modules along with the exercise data.

Searching ConceptNet is a straightforward process since the knowledge base offers a number of APIs to query it. In our case we use the typical search query[6] to get `relatedTo` terms. The search term is provided in canonical form in ConceptNet, e.g., `/c/en/cat` and offset is the number of records to skip and show the next, as ConceptNet API has a limit of 1000 results per call.

An example of a generated exercise is "Name one thing that is related to dog", where the learner is expected to enter a word that exists in the results retrieved from the knowledge base. In cases where new words are entered by the learner, the evaluation module handles whether they should be added to the knowledge base or not while a specific user feedback strategy is used to account for the unknown correctness of the new answers (see Section 2.4).

2.3 The Exercise and Results Storage Dispatcher Module

Transactions between modules are handled using web services through API calls. Data are presented in a JSON[7] format that can be consumed by any programmatically created interface. Specification of the API is available at the project repository using the Swagger[8] Opensource API management tool. This abstraction layer allows the exploitation of the system from various interfaces without developers having to know its underlying functionality. Currently, the system offers web services for registering users, retrieving exercises from the exercise generation module, checking learners' contributions, assigning points/awards, storing presented hints and showing a leaderboard.

For the latter part, the evaluation module is employed. The outcome of the evaluation module is used to update both the learner's points and awards and the knowledge base list of answers for that specific exercise.

The dispatcher is also directly connected with the database for storing/retrieving data in/from tables and provide another abstraction layer for information handling workflows.

2.4 The Evaluation Module

The evaluation module processes the learner's answers in order to both update the knowledge base with new words and to assign points and award badges to the learners, which are transformed into feedback messages and leaderboards in the user interface.

The evaluation module operates on pairs of exercise and result (see Figure 1) produced each time a user completes one exercise. It checks whether the user's answers are known or new to the knowledge base and evaluates if new answers are valid candidates to enhance it. According to the evaluation points received, badges are assigned to the learners.

More specifically, the evaluation module checks for each user's answer, whether it is part of the results set for that exercise or whether the answer is new. If the answer is part of the knowledge base, the user receives one point. If the answer is new, the user receives *potential* points, and the answer is stored as a candidate answer for that exercise together with the user id. The feedback message
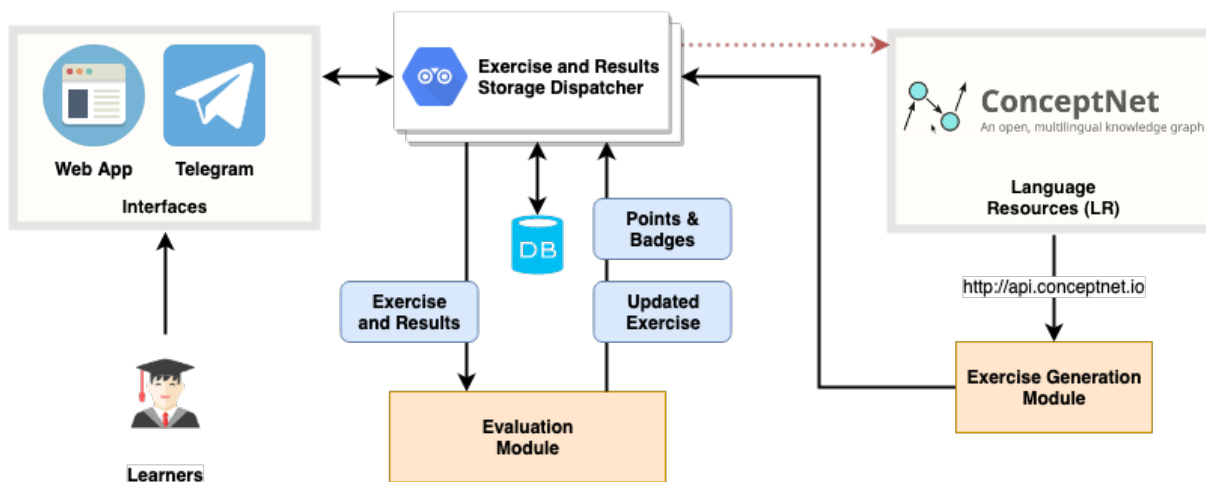

Figure 1: An overview of the vocabulary trainer architecture, depicting the main modules of the architecture, the data interchange between them and user interaction at the interface level.

to the learner informs him that the answer is new and that he is gaining *potential* points, which may transform into actual points if the answer is validated by several users over time.

Every time a predetermined number *K* of candidate answers has been accumulated, the evaluation process is triggered:

- All new answers are ranked by their occurrence frequency (i.e., how many mentions).

- The top-ranked answer is selected (given that it was mentioned at least *N* times).

- All learners that gave the selected answer get informed that it was a correct answer and receive two points (transforming *potential* points into actual points).

- In addition, the learner who was the first to give the selected answer receives a badge.

- The knowledge base is updated with the new word for that exercise.

- All occurrences of the selected word are removed from the candidate list.

2.5 User Interaction Module and Prototypical User Interfaces (UI)

In the current version of the vocabulary trainer, two interfaces are implemented: a chatbot on Telegram and a web application. Both interfaces allow the learner to receive and complete exercises on the `RelatedTo` relation. Learners get an immediate response back on the correctness of their answer and when this is not possible (i.e., in cases

of unknown answers), they first receive potential points and get notified asynchronously, once the unknown answer was confirmed by other learners.

Currently, the vocabulary trainer forces the user to input a word in order to move to the next exercise. In order to support the learner in case he runs out of ideas for related words on a given exercise, the "I Don't Know" feature provides functionality for requesting a "hint". The hint provides the user with a correct related word, which is taken from the knowledge base. After reading the hint, the learner is free to input the hint word or any other word that he/she deems fitting. Whenever a learner uses that feature, there is an underlying mechanism that stores the presented hint in the database.

Within both interfaces, learners can see their points and badges gained and for the Telegram chatbot, they can also access a leaderboard.

Chatbot interface The chatbot interface was implemented on Telegram, a very versatile messaging system. Telegram is available both as a native application for mobile phones and desktop computers for all operating systems and as a web application. It enables the implementation of chatbots via the Telegram Bot APIs[9]. In the implemented chatbot[10], users interact with the system via a standard dialogue chat interface (see Figure 2). Apart from textual input, the interface provides a button area that changes during the dialogue flow to simplify the interaction.

---

[9]https://core.telegram.org/bots/api
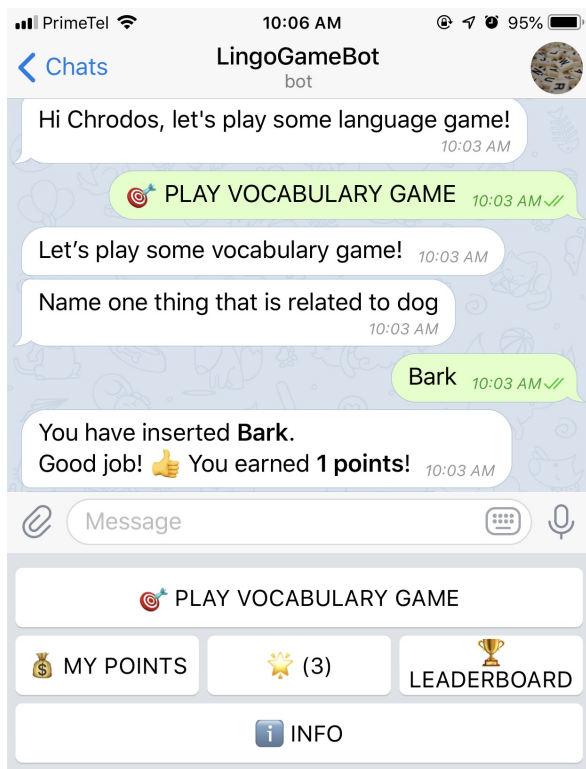[10]https://t.me/LingoGame_bot

677

Figure 2: A screenshot from the chatbot interface, where a user is interacting with the system.

Web interface The web interface[11] (see Figure 3) currently offers three exercise types for learners to practice with. It was implemented using a free bootstrap template[12] to ensure a stable interface that works well both on computers and mobile devices.

3 Empirical Evaluation

To evaluate the potential of the v-trel architecture for the purpose of (1) gathering commonsense knowledge and populating the language resource used to generate the exercises, and (2) delivering a meaningful application for vocabulary learning, we designed an experiment as follows.

For the initial experiment, the focus was put on objective (1), evaluating the quality of collected data. However, a feedback questionnaire and the manual analysis of the gathered data by an expert in language teaching served to gain first insights also related to (2), the educational value of such a vocabulary trainer, which need to be expanded on in a follow-up study (see Section 5).
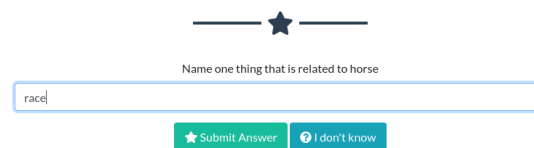
---

[11] http://cognition-srv1.ouc.ac.cy/vtrel/

[12] https://startbootstrap.com/themes/freelancer/

Figure 3: A screenshot from the web interface.

Term	RelatedTo term	#terms
cat	animal, dog, house...	94
dog	friend, puppy, bone...	135
bird	chicken, chick, canary...	219
cow	sacred, animal, steak...	96
fish	water, creature, lure...	221

Table 1: An example of the terms retrieved to generate the experiment exercises. The subject is RelatedTo the object. In the last column the number of filtered terms is depicted.

3.1 Setup of the Experiment

First we generated 26 exercises using the RelatedTo relation from ConceptNet and 26 terms (see Table 1) that fall under the A1 and A2 language learning level. For each of these terms, we acquired more than 20 terms that are related to them. We used the conceptnet.io API to retrieve all terms connected with the RelatedTo relation where the language is set to English. We filtered out terms that had less than 20 RelatedTo terms in order for the learners to have a plethora of possible answers for the "I Don't Know" feature.

We also implemented a recording mechanism that captures the answers presented to the learner when clicking the "I Don't Know" button for a specific exercise. This served the purpose for us to analyze how the learner used this feature and reacted to the words presented to them while contributing and completing exercises.

For the experiment the evaluation parameters were set to a threshold of $K=5$ candidate answers to trigger the evaluation and a minimum limit of $N=2$ words for including the candidate into the knowledge base was applied.

3.2 Implementation of the Experiment

The initial experiment was conducted with people from our peer group, about 60 non-native speakers
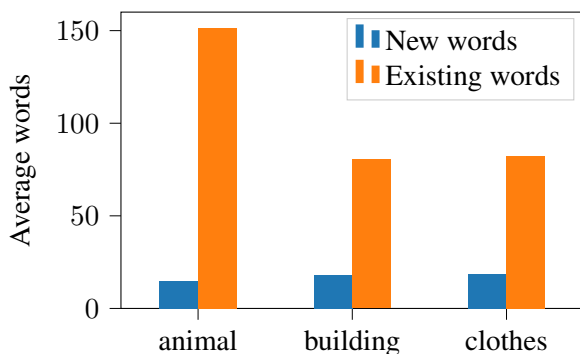
Figure 4: The average number of new and existing words per exercise for each category.

of English, with a high proficiency level.[13] Each of these users received an email with a link to both the Web interface and the Telegram chatbot and was asked to try any of the two for more than ten minutes within a period of two days. At the end of this period, each user received a link to an on-line questionnaire[14] to provide feedback on both the interfaces and the presented questions.

To summarize, during the 2 days period we managed to gather 4533 contributions to 26 exercises presented to the user in random order. The contributions were collected in 44 distinct user sessions, of which 17 were on the Telegram chatbot and 27 on the Web interface. Presumably, the sessions mostly relate to unique users, although we know of at least one user, who accessed both interfaces (see Section 3.4).[15] We also captured 683 possible answers presented to learners through the "I Don't Know" feature. Out of the 4533 contributions, 449 new words were crowdsourced from the users based on the evaluation mechanism with the parameter settings described above (see Section 2.4 and 3.1). In terms of questionnaire feedback, we gathered 17 fully-completed and 17 partially-completed responses.

3.3 Results Analysis

Characteristics of new words. Overall, the experiment allowed to gather 449 new words, di-

---

[13]We are aware that involving speakers with a high English proficiency implicates that the crowd does not represent genuine language learners. By being composed of non-native speakers, we however assume them to resemble advanced learners to a degree that allows to draw meaningful conclusions for the scope of this first evaluation.

[14]LimeSurvey, a Free Opensource online tool was locally deployed: https://www.limesurvey.org

[15]For simplicity, we refer to user sessions as *users*.

New word	Frequency	Level
grass	15	basic
calf	6	advanced
meat	5	basic
cowboy	4	basic
farmer	4	basic
herd	4	advanced
horn	3	moderate
pasture	3	advanced

Table 2: Top new words for the term "cow", their frequency of mentions and proficiency levels.
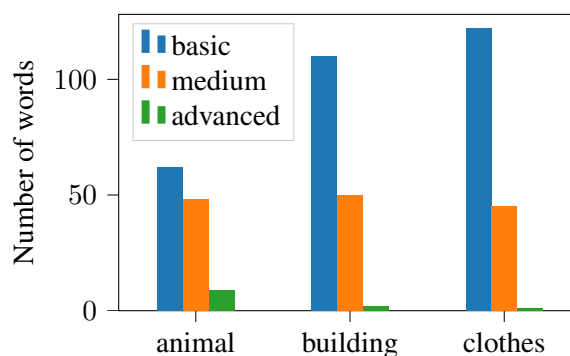

Figure 5: Level of proficiency of new words by exercise category.

vided by exercise category as follows: 119 words on animals, 168 words on clothing, and 162 words on buildings. The lower number of new words for the category animal relates to the higher number of existing words in the knowledge base for that category (see Figure 4). For example for the term "cow", 15 new words were gathered. Table 2 shows the 8 new words that were named three or more times, while seven words[16] met the minimum threshold of two mentions. This shows that ConceptNet has empty spots even for basic vocabulary like "grass" or "farmer", which could be gathered through the learners. Also, it shows that learners were able to propose advanced vocabulary such as "ruminate" or "pasture".

A manual analysis of the proficiency level of all new words was carried out by an expert from language teaching. It showed that the vast majority of new words is part of basic vocabulary (65% of all added words), while 32% are of moderate level and only 3% belong to advanced vocabulary.

---

[16]New words for term "cow" with two mentions only: *bell, burger, methane, ox, ruminate, sheep* and *veal*
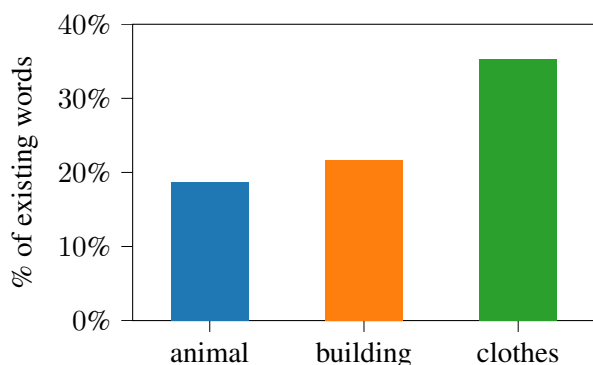
Figure 6: Average percentage of existing words, provided by users, per exercise for each category.

Interestingly, for the animal category the ratio between basic and moderate vocabulary is far more balanced than for the other two categories, also more advanced vocabulary is found (see Figure 5). This could be explained by the higher number of existing words in the knowledge base for the category animal. Most basic words are probably already part of ConceptNet, which implies that new words necessarily need to be more advanced.

Figure 6 shows the average percentage of existing words (per exercise) that were entered by the learners, divided by category. It shows that learners named less than 40% of the words present in ConceptNet[17], while they can still gain knowledge on more than 60% of the words, e.g. by requesting hints through the "I don't know" feature.

Responses to "I Don't Know" hints. Overall, 683 hints were provided to the learners by means of the "I Don't Know" button. In response, 365 words were entered by the learners of which 331 were a direct repetition of the hint word. The lower number of entered words in relation to hints is due to multiple "I Don't Know" clicks before entering a word (up to 17 clicks in a row).

In 34 cases, the word entered in response to the hint(s) was different from the hint(s), which indicates that the hint activated the learner's knowledge on related words. A look at the words shows that:

- 2 times a variation of the lemma was entered

- 11 times an analogy of the hint was entered

- 16 times a different word class was entered

---

[17]The lowest coverage were found for category animal, the category with the biggest set of available words.

Existing relations in ConceptNet. To get a better idea of the type and quality of words learners contributed regarding our initial LR, i.e., ConceptNet, we queried what other relations might exists in ConceptNet between the search word used for generating the exercise and the contributed words other than the `RelatedTo` relation. From the 26 exercises learners contributed, on average 14.15 words have no other direct relation (not bidirectional) with the original word from ConceptNet, i.e., <SUBJECT> <RELATION> <CONTRIBUTED_WORD>. When searching also for bidirectional relations, 100% of contributed words have such a relation between them in ConceptNet, including `RelatedTo`, which can be taken as indication for the appropriateness of the words added by learners. On average 4.54 new relations (other than direct `RelatedTo`) were identified between the contributed terms and the subject used to generate the exercise.

3.4 Feedback Questionnaire

After the experiment, a feedback questionnaire with six items was sent to all participants:

- Level of English: [A1/A2; B1/B2; C1/C2]

- Interface used: [Chatbot; Web; both]

- What did you notice regarding the UI? [open]

- What did you think of the questions? [open]

- What did you like about this approach to a vocabulary exercise? [open]

- Any other comments? [open]

Out of 34 users, 22 completed the closed questions and 9 to 14 also responded to the open questions; 18 of 22 respondents indicated an advanced level of English; 9 users used the Chatbot, 12 used the Web interface, and one user used both interfaces.

User interface. 9 respondents perceived the interface as clear, easy and pleasant to use, while 9 users criticized unclear navigation and pop-ups interrupting the workflow.

Questions. 9 users remarked that words repeated too often and that the phrasing of the question "name one thing" can be misleading in terms of which word class is requested (3 respondents).

Approach to vocabulary exercise. 12 users evaluated positively the interactivity and simplicity of the exercise, the opportunity to learn new

words by means of the "I don't know" function, and its effect to reactivate words and to incentivize brainstorming. Still, five users suggested that it is no real vocabulary exercise, that it would be difficult for beginners and that the processing of the answers and assignment of points/feedback was unclear, too open-ended, and lacking negative feedback.

Other comments. The criticism about the way to award points and the overall educative value were stressed further.

3.5 Discussion of the Experiment Results

The analysis of new words shows that our approach is promising for extending ConceptNet with meaningful words, in particular the more advanced level of new words added for the category animal indicates that relevant new terms can be gathered in a progressive fashion (e.g., basic vocabulary is added first). Given that this first experiment was carried out with advanced learners it needs to be evaluated to which extent similar positive results can be achieved with beginner and intermediate learners. Also, the analysis indicates that ConceptNet is ample enough to propose a wide set of new vocabulary to the user (in average more than 60% of the words per exercise). Furthermore, results suggest that this approach can also identify new relations between terms in ConceptNet. Learners managed to reproduce what was already coded in a different manner in the resource and thus improved its completeness.

User feedback verifies that the vocabulary trainer as a User Interface can be improved but the idea behind it is interesting and can be applied in different language learning scenarios.

4 Related Work

Approaches for crowdsourcing language resources can be divided into two broad categories: 1) implicit crowdsourcing, i.e., users carry out any activity of their interest while their data is crowdsourced as a byproduct, and 2) explicit crowdsourcing, i.e., the crowd is actively engaging.

The Duolingo platform (von Ahn, 2013) presents a most similar approach to our work, based on language learning, as it generates language exercises, allowing the crowdsourcing of language-related data (i.e., translations). Other research work based on implicit crowdsourcing has utilized the Games With A Purpose (GWAP) approach (von Ahn and Dabbish, 2008). Among GWAP, in particular the JeuxDeMots game by Lafourcade (2007) shows several similarities to our approach. The game is designed for constructing lexical data in French in a playful way. In order to gain points two players have to agree on words related to given terms. Also, the more recent TileAttack game by Madge et al. (2017) builds on player agreements to gather annotations for text segmentation tasks. In addition, Rodosthenous and Michael (2016) developed a platform that combined automated reasoning with games for acquisition of knowledge rules. Moreover, in work of Guillaume et al. (2016) a game titled ZombiLingo was developed, for annotating dependencies in French data. In work of Chamberlain et al. (2008), the Phrase Detectives game is presented, where players contribute relationships between words and phrases, aiming to create a resource that is rich in linguistic information. Yet another indirect form of crowdsourcing has used large collaborative knowledge bases like Wikipedia to create multilingual resources such as YAGO3 (Mahdisoltani et al., 2015) and DBpedia (Lehmann et al., 2015). Recently, Meurers et al. (2019) have proposed a web-based workbook, which can be integrated into classroom curricula and offers instructive feedback to students while also gathering data on learning processes for Second Language Acquistion (SLA) research.

Research works with explicit crowdsourcing often employ Amazon Mechanical Turk to collect data. For instance, Biemann (2013) created the Turk Bootstrap Word Sense Inventory of frequently used nouns in English and Ganbold et al. (2018) localized the WordNet in the Mongolian language. Related to SLA, MacWhinney (2017) proposes a collaborative platform for collecting and sharing learner data from corpora, online tutors, and Web-based experimentation.

Our research presents an implicit crowdsourcing approach implemented as a vocabulary learning application with open-ended questions for language resource augmentation using multiple user interaction methods (i.e., chatbots and web apps).

5 Conclusion and Future Work

In this paper, we presented the v-trel vocabulary trainer for crowdsourcing language resources and delivering exercises to language learners. V-trel can be accessed through two interfaces, a Tele-

gram chatbot and a Web application.

Moreover, we presented the results of an empirical evaluation and a user satisfaction survey for the vocabulary trainer and provide a relevant discussion of these results. Feedback from users and the analysis of the contributed words are taken into account for updating v-trel with new features such as a more attractive interface, and new exercise types[18]. In addition, we aim at including links to pictures and definitions or examples of use of the "hint" words, to support the learner not only in refreshing their existing vocabulary, but also to acquire new words. These are first steps to strengthen the learning effect of the tool, while in the midterm we foresee to intensify further the collaboration with language teaching experts in order to tailor the offered exercises more closely to specific learning goals.

Also, in future work we aim at implementing more strategies for safeguarding the quality of the acquired data, e.g., control mechanisms for active misuse, further evaluation strategies for new words, which could also involve dynamic retrieval of knowledge from ConceptNet or evaluation cycles re-proposing new words in new exercises, and strengthened gamification elements. In addition, we intend to evaluate the crowdsourced words and their difficulty levels in relation to established reference data such as the *English Vocabulary Profile*[19] or similar resources.

As proposed in the feedback questionnaire, the educational value needs to be validated and improved further. Accordingly, a follow-up user study with focus on the educational aspect is foreseen and will be designed and carried out with authentic learners of different proficiency levels.

Last but not least, the vocabulary trainer will be integrated into the setup of the Revita online system for language learning (Katinskaia et al., 2018) to reach a larger audience. In particular, it can be implemented as a part of a testing mode where crowdsourced questions do not influence the learner's final score. We will also work towards integrating v-trel into Games With A Purpose. Previous work of Rodosthenous and Michael (2016, 2014) suggests that crowdsourcing and GWAPs, in particular, can be used to gather background knowledge

---

[18]E.g., various new types of vocabulary exercises such as "fill the gap", or "select all verbs among the given words"

[19]https://www.englishprofile.org/wordlists

References

Itziar Aldabe, Mikel Larrañaga, Montse Maritxalar, Ana Arruarte, and Jon A. Elorriaga. 2015. Domain module building from textbooks: Integrating automatic exercise generation. In Cristina Conati, Neil Heffernan, Antonija Mitrovic, and M. Felisa Verdejo, editors, *Artificial Intelligence in Education*. Springer International Publishing, Cham, pages 521–524.

Chris Biemann. 2013. Creating a system for lexical substitutions from scratch using crowdsourcing. *Language Resources and Evaluation* 47(1):97–122. https://doi.org/10.1007/s10579-012-9180-5.

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

Florian Daniel, Pavel Kucherbaev, Cinzia Cappiello, Boualem Benatallah, and Mohammad Allahbakhsh. 2018. Quality control in crowdsourcing: A survey of quality attributes, assessment techniques, and assurance actions. *ACM Computing Surveys (CSUR)* 51(1):7.

Amarsanaa Ganbold, Altangerel Chagnaa, and Gábor Bella. 2018. Using crowd agreement for wordnet localization. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC-2018)*.

Bruno Guillaume, Karën Fort, and Nicolas Lefebvre. 2016. Crowdsourcing complex language resources: Playing to annotate dependency syntax. In *International Conference on Computational Linguistics (COLING)*.

Elizabeth Hadley, David Dickinson, Kathy Hirsh-Pasek, and Roberta Golinkoff. 2018. Building semantic networks: The impact of a vocabulary intervention on preschoolers' depth of word knowledge. *Reading Research Quarterly* 54:41–61. https://doi.org/10.1002/rrq.225.

Jeff Howe. 2006. Crowdsourcing: A Definition. https://www.wired.com/2006/06/crowds/.

Jan H. Hulstijn. 2013. *Incidental Learning in Second Language Acquisition*, Wiley-Blackwell, volume 5, pages 2632–2640). https://doi.org/10.1002/9781405198431.wbeal0530.

Anisia Katinskaia, Javad Nouri, and Roman Yangarber. 2018. Revita: a language-learning platform at the intersection of its and call. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC-2018)*.

Mathieu Lafourcade. 2007. Making people play for Lexical Acquisition with the JeuxDeMots prototype. In *SNLP'07: 7th International Symposium on Natural Language Processing*. Pattaya, Chonburi, Thailand, page 7. https://hal-lirmm.ccsd.cnrs.fr/lirmm-00200883.

Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, Sören Auer, et al. 2015. Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web* 6(2):167–195.

Verena Lyding, Lionel Nicolas, Branislav Bédi, and Karën Fort. 2018. Introducing the european network for combining language learning and crowdsourcing techniques (enetcollect). In Peppi Taalas, Juha Jalkanen, Linda Bradley, and Sylvie Thouësny, editors, *Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018*, Research-publishing.net, pages 176–181.

Brian MacWhinney. 2017. A shared platform for studying second language acquisition. *Language Learning* 67(S1):254–275. https://doi.org/10.1111/lang.12220.

Chris Madge, Jon Chamberlain, Udo Kruschwitz, and Massimo Poesio. 2017. Experiment-driven development of a gwap for marking segments in text. In *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play*. ACM, pages 397–404.

Farzaneh Mahdisoltani, Joanna Asia Biega, and Fabian M. Suchanek. 2015. Yago3: A knowledge base from multilingual wikipedias. In *Proceedings of the 7th Conference on Innovative Data Systems Research (CIDR)*.

Detmar Meurers, Kordula De Kuthy, Florian Nuxoll, Björn Rudzewitz, and Ramon Ziai. 2019. Scaling up intervention studies to investigate real-life foreign language learning in school. *Annual Review of Applied Linguistics* 39.

I. S. P. Nation and Susan Hunston. 2013. *Learning Vocabulary in Another Language*. Cambridge Applied Linguistics. Cambridge University Press, 2 edition. https://doi.org/10.1017/CBO9781139858656.

Lionel Nicolas, Verena Lyding, Luisa Bentivogli, Federico Sangati, Johanna Monti, Irene Russo, Roberto Gretter, and Daniele Falavigna. 2018. Enetcollect in italy. In *Proceedings of the 5th Italian Conference on Computational Linguistics (CLiC-it 2018)*.

Christos Rodosthenous, Verena Lyding, Alexander König, Jolita Horbacauskiene, Anisia Katinskaia, Umair ul Hassan, Nicos Isaak, Federico Sangati, and Lionel Nicolas. 2019. Designing a prototype architecture for crowdsourcing language resources. In Thierry Declerck and John P. McCrae, editors, *Proceedings of the Poster Session of the 2nd Conference on Language, Data and Knowledge (LDK 2019)*. CEUR, pages 17–23.

Christos Rodosthenous and Loizos Michael. 2014. Gathering background knowledge for story understanding through crowdsourcing. In *Proceedings of the 5th Workshop on Computational Models of Narrative (CMN 2014)*. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Quebec, Canada, volume 41, pages 154–163. https://doi.org/10.4230/OASIcs.CMN.2014.154.

Christos Rodosthenous and Loizos Michael. 2016. A hybrid approach to commonsense knowledge acquisition. In *Proceedings of the 8th European Starting AI Researcher Symposium*. pages 111–122. https://doi.org/10.3233/978-1-61499-682-8-111.

Catherine Rosenbaum. 2001. A word map for middle school: A tool for effective vocabulary instruction. *Journal of Adolescent & Adult Literacy* 45(1):44–49.

Norbert Schmitt. 2013. *An Introduction to Applied Linguistics*. Hodder Arnold Publication. Taylor & Francis. https://books.google.com.cy/books?id=5gIvAgAAQBAJ.

Robert Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Proceeding of the 31st AAAI Conference on Artificial Intelligence*.

Luis von Ahn. 2013. Duolingo: learn a language for free while helping to translate the web. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. ACM, pages 1–2.

Luis von Ahn and Laura Dabbish. 2008. Designing Games With a Purpose. *Communications of the ACM* 51(8):57. https://doi.org/10.1145/1378704.1378719.