

# WORDGUESS

## Using Associations for Guessing, Learning and Exploring Related Words

Cennet Oğuz<sup>1</sup>, André Blessing<sup>2</sup>, Jonas Kuhn<sup>2</sup>, Sabine Schulte im Walde<sup>2</sup>

<sup>1</sup>German Research Center for Artificial Intelligence (DFKI), Saarland Informatics Campus

<sup>2</sup>Institute for Natural Language Processing, University of Stuttgart

ceog01@dfki.de,

{blessiae, jonas, schulte}@ims.uni-stuttgart.de

### Abstract

This paper presents WORDGUESS, a game-with-a-purpose vocabulary training where –in order to guess a target word (such as *snow*)– the player is offered associations of that target word (such as *winter*, *white*, *cold*). The game relies on existing association norms and co-occurrence information to establish an entertaining way of deepening the player’s learning and understanding of vocabulary and of associative relatedness between words in the vocabulary. WORDGUESS comes with data in English and German and can be extended with data from further languages. From an application-oriented point of view the players’ data enables us to induce conditions and weights for word association and to quantify contextual relationships, which is useful for many NLP purposes such as ontology induction and anaphora resolution.

### 1 Introduction

Games-with-a-purpose (GWAP) offer enjoyable entertainment to players and at the same time allow researchers in Natural Language Processing (NLP) to collect and explore cognitive and (computational) linguistic facets of human-generated data. While the term GWAP has been coined by Von Ahn and Dabbish (2008), the underlying idea has been pursued across linguistic levels and across NLP purposes for much longer. To provide a few examples across research fields, the adventure and interactive fiction games by Gabsdil et al. (2002) rely on natural-language question-answering dialogues to explore inference systems with reference resolution, syntactic ambiguities, and scripted dialogues; Chamberlain et al. (2008) exploit collaborative work to identify relationships between words and phrases in web data; *OntoGame* (Siorpaes and Hepp, 2008) matches classes in an ontology with Wikipedia articles; Hladká et al. (2009) propose a

gamified annotation approach for coreference resolution; Guillaume et al. (2016) design *ZombiLingo*, a game for syntactic dependency annotation.

We present a GWAP-style game implementation called WORDGUESS<sup>1</sup> where associations of a target word are offered to players in order to guess the target word. For example, associations such as *winter*, *white*, *cold* provide hints to players when guessing the target word *snow*. Our game is a web-based and mobile-based application whose aim is to learn and understand word-association and word-context relationships: previous research has shown that associations and corpus co-occurrence are related (Church and Hanks, 1990; de Deyne and Storms, 2008a; Schulte im Walde et al., 2008, i.a.); we plan to explore their connections and differences in more depth. In this vein, (i) we vary associations obtained from humans, and context-based words induced from corpus co-occurrence; (ii) we provide a multilingual gaming environment in order to understand the conditions across languages and relational patterns between native and second languages; and (iii) we offer the players to choose between levels of difficulty (i.e., providing more or less cues). The obtained data enables us to induce conditions and weights for word association and to quantify contextual relationships, taking native language, age and gender into account.

Regarding the technical setup, we use *Angular*, a TypeScript-based open-source web application framework, for the implementation of the user interface (UI), while *MongoDB*, a cross-platform document-oriented database program, is utilized for storing and organizing the game constituents (e.g., defined games, users, and played games). We also provide a responsive UI design in order to make the game usable on different devices such as phones, tablets and computers.

---

<https://wordguess.ims.uni-stuttgart.de>

## 2 Related Work

Gamification is a common way to make a wide variety of tasks entertaining in NLP. In the following, we provide an overview of NLP-oriented games for data collection, language learning and linguistic analysis, which are the purposes closest to ours.

**Data Collection** Lafourcade (2007) proposes a gamification approach by making people play with associative words in order to memorize associations with *JeuxDeMots*, a two-player game based on agreement. Kuo et al. (2009) present an interactive community-based game for collecting question-answering data in order to provide a report about data quality, collection efficiency, player retention, concept diversity, and game stability for future community-based games. Lafourcade et al. (2017) use games to create and enrich weighted lexical resources from crowdsourced data by investigating existing rich lexical networks that can be used to infer linguistic coercion.

**Language Learning** Advances in NLP techniques have been used to investigate students' learning situations and behaviour patterns in a wide range of learning practices and studies. Mart (2012) claims that guessing new words presented in isolation is hard but words in context help learners to deduce meaning from the context, whereas Crow and Quigley (1985) demonstrate that an approach to vocabulary studies based on semantic organization is productive. These theories indicate that computer games are powerful tools for educational aims (Malone, 1980). Therefore, we suggest our game for vocabulary learning in both the native and a second language to attract user motivation.

Many games aim to improve teaching methods for language learning and other educational environments. Jung and Graf (2008) build a word-association game and show that word association supports more effective and attractive vocabulary learning. Madge et al. (2019) offer a text-tagging and language-learning game for enhanced syntactic annotation and language resources.

**Linguistic Analysis** NLP techniques have also helped to identify students' behaviour and learning models by explaining complex linguistic patterns that occur in the games' language data in order to provide enhanced education methodologies. Goodman (2014) uses a guessing game in order to understand whether reading is a series of guesses informed by graphic, semantic and syntactic cues

while substituting the words. Picca et al. (2015) show NLP utilization for understanding children's language development by gathering data from a pedagogical *Serious Game* which is designed for a primary purpose other than pure entertainment.

We offer a novel gamified approach that is inspired by (Goodman, 2014; Jung and Graf, 2008) for word guessing by using word-association and word-context pairs. Our game aims to create opportunities for both players and researchers: players go for it for learning and entertaining purposes, while researchers may analyze the cognitive and linguistic inferences.

## 3 WORDGUESS: Motivation, Design, Architecture

**Motivation** Associations, i.e., words spontaneously called to mind by a stimulus word, have served as a tool in cognitive science research for decades to investigate the mechanisms underlying semantic memory, making use of the implicit notion that associates reflect meaning components of words. Accordingly, a large number of data collections of associations is available, such as the *Edinburgh Association Thesaurus* (Kiss et al., 1973), the University of South Florida norms (Nelson et al., 2004), the *Database of Noun Associations for German* (Melinger and Weber, 2006), norms for German nouns and verbs (Schulte im Walde et al., 2008) and for Dutch words (de Deyne and Storms, 2008b), and the *Small World of Words* norms (de Deyne et al., 2019), among others.

For many NLP purposes such as ontology induction and anaphora resolution, it is crucial to define and induce semantic relations between words or contexts, and according to the *co-occurrence hypothesis* (Miller, 1969; Spence and Owens, 1990) semantic association is related to the textual co-occurrence of stimulus-associate pairs. Therefore, a number of studies have exploited the connection between co-occurrence distributions and semantic relatedness, and used association norms as a test-bed for distributional models of semantic relatedness (Church and Hanks, 1990; Rapp, 2002; de Deyne and Storms, 2008a; Schulte im Walde et al., 2008, i.a.).

**Game Idea** The aim of WORDGUESS is to exploit a gamification environment in order to deepen the understanding of associative relatedness. Differently to previous approaches, we do not directly

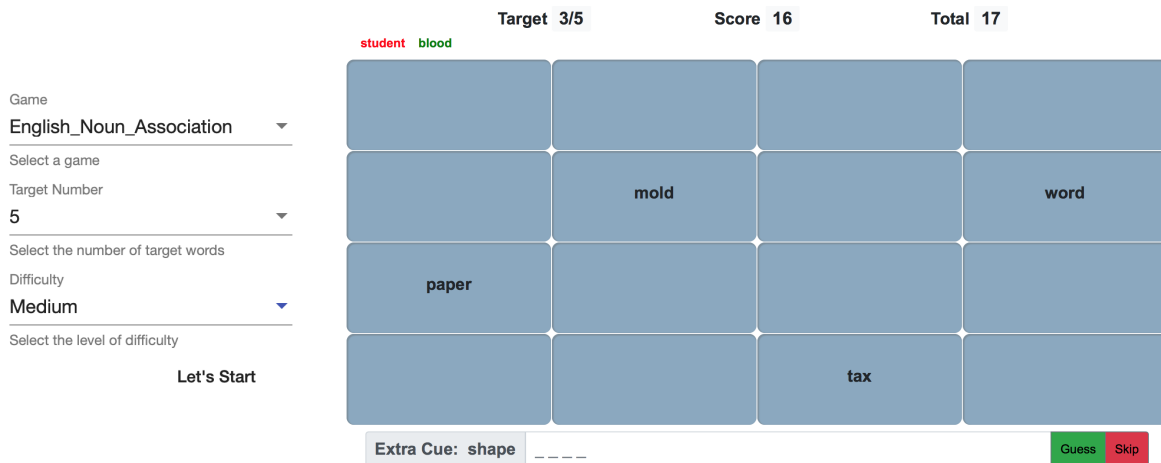


Figure 1: Sample page views of the game: (a) On the left you can see the decision page where the player can choose a predefined game, the number of targets and the difficulty level. (b) On the right you can see a game page at some point during the game where (top row: left to right) the player is currently working on the third out of five targets (3/5), currently scores 16 for this target (out of a maximum of 20) because four cue words have been selected already, and currently holds a total score of 17 from the past two target word guesses. On the bottom left we see an extra cue word (*shape*) and five underscores \_\_\_\_\_ to indicate that the target word is five characters long; this information corresponds to the medium difficulty level. The player can either write a guess and press "Guess" or give up on this target at any point and press "Skip" (bottom right). The red+green words above the grid refer to previously guessed (green) or not guessed (red) target words.

analyse and quantify existing norms of human association in comparison to corpus co-occurrence. Rather, we investigate their relatedness by *utilising* them alternately in the same gaming environment and under the same conditions, so that comparing the ease or difficulty of players can inform us indirectly about their similarities and differences.

In this vein, the game relies on existing association norms and co-occurrence data across languages to establish an entertaining way of collecting human associations. The players see a grid with empty cells which they can click in any order (see Figure 1). Each click reveals a cue. The less cues the player needs to guess the correct target word, the higher the score; a wrong guess decreases the score by three.

For each player who is registered<sup>2</sup> we keep track of the order of the chosen cues, the correct and wrong guesses, and the required time for a correct guess. In addition, we can relate those parameters to the players' profile including age, gender and native language. The underlying cues are either based on existing association norms or on corpus-induced co-occurrences, so that we can use the data to obtain a clearer picture for association-target relations, co-occurrences and the interplay of both.

<sup>2</sup>Players can choose between playing with or without an account. We only keep track of players who create an account.

**Game Implementation** WORDGUESS works in two different modes: (1) the *project mode* for the researcher to set up a new game, and (2) the *player mode* for the player to play a game from the available set of games in the project mode.

In the project mode, a researcher defines a game and specifies the game properties. Our system accepts data collections in JSON file format. Each file corresponds to a game setup, i.e., the data is read from a file and establishes a new game. Each target-association pair is presented as a JSON object with *target*, *context*, and *score* (key,value) pairs. After uploading the game file, the researcher defines the game settings, such as the name of the game, target order, context order, context number, cue selection and game definition (see examples in Appendix). The name of the game is the attribute seen by the player for the game selection. Target-context pairs may be selected according to their order, or randomly. In addition, the context number defines how many contexts are provided to the players; the cue selection determines the choice of extra cues (random, highest, lowest, none).

The player mode presents the game to the players. A player may register for playing a game, or alternatively skip the registration and play the game anonymously for entertainment and learning. The registration is performed as explicit agree-

ment between researcher and player to provide data. Without that agreement, the player is able to play the game without sharing any information. Game-related functions are not affected by this decision. Before starting the game the player has to determine the game’s attributes. As can be seen from Figure 1 (left), the decision page represents the game-related options such as game, target number, and difficulty level. Afterwards, the player can start playing the game according to the selected options. WORDGUESS currently offers two game languages: German and English. At the same time, target words distinguish between two word classes, nouns and verbs. The player can select the number of target words to guess as either five or ten. We present two different types of cues, i.e., target cues and context cues for each target word to guess. *Target cues* are derived from the target word. In the easy mode, we provide the number of characters of the target word together with one of the characters (e.g., `__r_` for *farm*), whereas in the medium mode we only show the target’s character count (e.g., `----`). We do not show any target cue in the hard mode. *Context cues* for target words are the main focus of our game and research, either human associations or corpus co-occurrence words. They are shown to the player in the grid, as illustrated in Figure 1 (right), plus one bonus cue (bottom left corner). The extra context cue which is the most associated context word of the target is available right from the beginning when the context cues in the grid are still hidden. The player clicks the boxes in the grid one by one to find the context cues, and tries to guess the target. Previous correct and wrong guesses for the current target word are displayed on the same page. At the same time, the player is able to skip a target word and to move on to guessing the next target word. The game continues for the chosen number of target words, and at the end of the game the player sees a summary of wrong and correct answers, and their scores. Registered players can track their scores and vocabulary development across games on their profile page.

**Technical Setup** We utilize Angular (v9.2.3) as the application framework as well as TypeScript (v3.2.4) as the programming language. MongoDB (v4.2.3), a document-oriented database program, is used for managing the stored data. Additionally, Express (v4.16.1) is exploited as the server framework for Node (v14.2.0) which is our runtime application environment.

**Motivating Users** Malone (1980) indicates that players are willing to master long-term activities (challenge), pursue informative games (curiosity), and let games invoke their imagination (fantasy). Challenges require a maximum level of difficulty whereas curiosity needs an optimum level of complexity in the game. WORDGUESS enables vocabulary improvement abilities as an informative reason to activate curiosity during the game, as well as different difficulty levels to enable a challenge.

We provide a very simple user interface to the players to keep their attention to only the game. Additionally, the necessary actions to play the game are not complicated or tiring such as long reading, learning additional techniques, or checking the accuracy of the current knowledge. We use harmonious colors to create serious perception whereas simple actions (to click the boxes) make the application game-like.

**Data Privacy and Ethics** As mentioned before, the player registration is performed as explicit agreement between researcher and player to provide data. Without that agreement, the player is able to play the game without sharing any information. Game-related functions are not affected by this decision. Information about age range, gender, and native language are kept if the player registers to the system. Furthermore, we encode the usernames by applying hashing algorithms. Players are able to delete their accounts whenever they want.

## 4 Conclusion and Future Works

This paper presented WORDGUESS, a game-with-a-purpose vocabulary training where associations of a target word are offered to players in order to guess the target word. From an application-oriented point of view, the gamification provides data that enables us to induce conditions and weights for word association and to quantify contextual relationships, which is useful for many NLP purposes such as ontology induction and anaphora resolution.

As regards future work, we plan to implement age-based user interfaces like colorful pages for children. The multiplayer, score-based competitions with enriched context based on descriptive information are on the agenda as well. Finally, data from further languages will be added to enable cross-lingual studies.

## Acknowledgments

We thank the three anonymous reviewers for feedback on the game and on the paper. Furthermore, Cennet Oğuz was supported by the CRETA center funded by the German Ministry for Education and Research (BMBF), FKZ 01UG1601 and FKZ 01UG1901; André Blessing was supported by funding from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) through MARDY (Modeling Argumentation Dynamics) within SPP RATIO.

## References

- 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*, pages 42–49.
- Kenneth W. Church and Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1):22–29.
- John T. Crow and June R. Quigley. 1985. A semantic field approach to passive vocabulary acquisition for reading comprehension. *Tesol Quarterly*, 19(3):497–513.
- Simon de Deyne, Danielle J. Navarro, Amy Perfors, Marc Brysbaert, and Gert Storms. 2019. The “Small World of Words” English word association norms for over 12,000 cue words. *Behavior Research Methods*, 51:987–1006.
- Simon de Deyne and Gert Storms. 2008a. Word associations: Network and semantic properties. *Behavior Research Methods*, 40(1):213–231.
- Simon de Deyne and Gert Storms. 2008b. Word associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods*, 40(1):198–205.
- Malte Gabsdil, Alexander Koller, and Kristina Striegnitz. 2002. Natural language and inference in a computer game. In *Proceedings of the 19th International Conference on Computational Linguistics*, pages 1–7.
- Kenneth S Goodman. 2014. Reading: A psycholinguistic guessing game. In *Making Sense of Learners Making Sense of Written Language*, pages 115–124.
- Bruno Guillaume, Karèn Fort, and Nicolas Lefebvre. 2016. Crowdsourcing complex language resources: Playing to annotate dependency syntax.
- Barbora Hladká, Jiří Mírovský, and Pavel Schlesinger. 2009. Play the language: Play coreference. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 209–212.
- Jaeyoung Jung and Sabine Graf. 2008. An approach for personalized web-based vocabulary learning through word association games. In *2008 International Symposium on Applications and the Internet*, pages 325–328.
- George R. Kiss, Christine Armstrong, Robert Milroy, and James Piper. 1973. An associative thesaurus of English and its computer analysis. In *The Computer and Literary Studies*. Edinburgh University Press.
- Yen-ling Kuo, Jong-Chuan Lee, Kai-yang Chiang, Rex Wang, Edward Shen, Cheng-wei Chan, and Jane Yung-jen Hsu. 2009. Community-based game design. In *Proceedings of Human Computation*, pages 15–22.
- Mathieu Lafourcade. 2007. Making people play for lexical acquisition with the jeuxdemots prototype. In *Proceedings of the 7th International Symposium on Natural Language Processing*.
- Mathieu Lafourcade, Bruno Mery, Mehdi Mirzapour, Richard Moot, and Christian Retoré. 2017. Collecting crowd-sourced lexical coercions for compositional semantic analysis. In *LENLS: Logic and Engineering of Natural Language Semantics*.
- Chris Madge, Richard Bartle, Jon Chamberlain, Udo Kruschwitz, and Massimo Poesio. 2019. Making text annotation fun with a clicker game. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*, pages 1–6.
- Thomas W Malone. 1980. *What makes things fun to learn? A study of intrinsically motivating computer games*. Ph.D. thesis, ProQuest Information & Learning.
- Çağrı Tuğrul Mart. 2012. Guessing the meanings of words from context: Why and how. *International Journal of Applied Linguistics and English Literature*, 1(6):177–181.
- Alissa Melinger and Andrea Weber. 2006. Database of Noun Associations for German. URL: <http://www.psycholing.es.uni-tuebingen.de/nag/>.
- George Miller. 1969. The organization of lexical memory: Are word associations sufficient? In George A. Talland and Nancy C. Waugh, editors, *The Pathology of Memory*, pages 223–237. Academic Press, New York.
- Douglas L. Nelson, Cathy L. McEvoy, and Thomas A. Schreiber. 2004. The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36(3).
- Davide Picca, Dominique Jaccard, and Gérald Eberlé. 2015. Natural language processing in serious games: A state of the art. *International Journal of Serious Games*, 2(3):77–97.

- Reinhard Rapp. 2002. The computation of word associations: Comparing syntagmatic and paradigmatic approaches. In *Proceedings of the 19th International Conference on Computational Linguistics*, Taipei, Taiwan.
- Sabine Schulte im Walde, Alissa Melinger, Michael Roth, and Andrea Weber. 2008. An empirical characterisation of response types in German association norms. *Research on Language and Computation*, 6(2):205–238.
- Katharina Siorpaes and Martin Hepp. 2008. Ontogame: Weaving the semantic web by online games. In *Proceedings of the European Semantic Web Conference*, pages 751–766.
- Donald P. Spence and Kimberly C. Owens. 1990. Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research*, 19:317–330.
- Luis Von Ahn and Laura Dabbish. 2008. Designing games with a purpose. *Communications of the ACM*, 51(8):58–67.

## A Appendix: Project Setup for English/German & Example Grid/End for German

**English\_Verb\_Context**

Choose File

random  The target word order

order  The context word order

16  The number of contexts of a target

highest  The score of extra cue of a target

The game description should be in the game langu...

**Guess an English verb based on typical context words!**

The game description

**Set Game**

**Deutsch\_Nomen\_Kontext**

Choose File

Zufällig  Die Auswahl von Zielwort

Zufällig  Die Auswahl von Kontextwörtern

16  Die Anzahl der Kontextwörter festlegen

Zufällig  Das extra Kontext-Wort festlegen

Die Spielbeschreibung sollte in der Spielsprache sein!

**Rate ein deutsches Nomen auf Basis typischer Kontextwörter!**

Die Spielbeschreibung

**Set Game**

Zielwort **5/5**
Punkte **13**
Gesamt **24**

Schmerz
Gesundheit
Kirche
Erkrankung
Erfahrung

<b>Gebiet</b>			
	langjährig		
		sagen	
		erfahren	

Tipp: international | 
Raten
Überspringen

**Das Spiel ist aus!**

Zielwort	Punkte
Schmerz	6
Gesundheit	0
Kirche	18
Erkrankung	0
Experte	12
<b>Gesamtpunktzahl</b>	<b>36</b>

**Neues Spiel!**