Wormingo: A 'true gamification' approach to anaphoric annotation

Doruk Kicikoglu
Queen Mary Univ. Of London
United Kingdom
o.d.kicikoglu@qmul.ac.uk

Richard Bartle
University Of Essex
United Kingdom
rabartle@essex.ac.uk

Jon Chamberlain
University Of Essex
United Kingdom
jchamb@essex.ac.uk

Massimo Poesio
Queen Mary Univ. Of London
United Kingdom
m.poesio@qmul.ac.uk

ABSTRACT

In this paper we present *Wormingo*, [1] a new Game-with-a-Purpose for anaphoric annotation. It introduces the motivation-annotation paradigm which uses linguistic puzzles and other widely known gamification techniques and word game mechanics to motivate players to carry out anaphoric annotation tasks. In a preliminary experiment, the game was tested on 270 players recruited through the Reddit platform, achieving promising results.

CCS CONCEPTS

• **Applied computing** → **Computer games**; • **Human-centered computing** → *Web-based interaction.*

KEYWORDS

GWAPs; anaphora resolution (coreference); gamification; gaming motivation; word games

ACM Reference Format:
Doruk Kicikoglu, Richard Bartle, Jon Chamberlain, and Massimo Poesio. 2019. Wormingo: A 'true gamification' approach to anaphoric annotation . In *The Fourteenth International Conference on the Foundations of Digital Games (FDG '19), August 26–30, 2019, San Luis Obispo, CA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3337722.3341868

1 INTRODUCTION

Games with a purpose (GWAPs) [18] are a sub-genre of serious games aimed to produce the data required by Artificial Intelligence as the games' by-product [27]. Well-known examples include *ESP Game* [28] and *FoldIt* [8]. Popular NLP GWAPs include *Phrase Detectives*[23], *Jeux De Mots* [17], and *Zombilingo* [3].

[1] https://wormingo.com/

FDG '19, August 26–30, 2019, San Luis Obispo, CA, USA

ACM ISBN 978-1-4503-7217-6/19/08…$15.00
https://doi.org/10.1145/3337722.3341868

The first GWAPs achieved very promising results. Both the *ESP Game*, that came out in 2004 [28], and *Peekaboom* in 2006 [27] collected more than 1,200,000 annotations from approximately 14,000 players within one month only. Both GWAPs also had more than 80% of their players coming back for another session, which today would be remarkable for even a non-serious casual game to achieve, as even 30% rate is considered a big success [12]. Other successful examples include *Galaxy Zoo* and *Foldit*. *Galaxy Zoo* has attracted 100,000 players on its first 9 months and collected a fascinating number of 40,000,000 annotations [19]. Similarly, *Foldit* on its first 3,5 months has attracted 721 players to produce 158,682 annotations, or "recipes" per their terminology [16].

Despite these initial successes, few GWAPs were as successful in recent years [18]. One reason for this may be that the offer of online games has soared tremendously along with the competition. But more significantly, many GWAPs developed afterwards lacked the element of fun which is essential for a GWAP to engage the player into the task[14, 18, 27]. Many of these GWAPs are gamified versions of annotation tools [12]. This especially applies to NLP GWAPs, as their interfaces are largely text-based and can hardly avoid resemblance to an annotation tool [14]. NLP GWAPs are further disadvantaged by the fact that participants interact with images better than with text [13, 20, 25], meaning that such GWAPs are also considered less attractive than platforms such as *Galaxy Zoo* or *Foldit* which are also gamified interfaces but labelling 2D or 3D images [16, 19]. So text-based games remain as a more niche taste compared to the taste of the average crowd [1]. An example of data collection platform for NLP straddling the boundary between true GWAP and simple gamification is *Phrase Detectives* [23], launched in 2009. *Phrase Detectives* is an anaphoric annotation tool deploying gamification techniques such as scoreboards, level-up mechanisms, experience points, etc. Although *Phrase Detectives* has been very successful, collecting over 4 million judgments, it is arguable whether employing these techniques is sufficient to call a platform a game [30]; what GWAPs seek out is amplifying engagement and thus increase quantity and quality of data produced [18].

In this paper we present *Wormingo*, an anaphoric annotation platform aiming to amplify engagement via the "fun" element found in proper games [15]. In *Wormingo* we have experimented with a new technique that we call the **motivation-annotation** paradigm. In *Wormingo* players are first involved with a purely game-like phase, before being asked to carry out a short annotation task that

allows them to earn points to be used in the game phase. We carried out a first test of this technique with players recruited mainly from Reddit, obtaining promising results. In this paper we present the **motivation-annotation** paradigm, the architecture of the game, and the outcomes of this first test.

2 BACKGROUND

2.1 Anaphora

Wormingo gathers data about **anaphora**. One important aspect of language interpretation is building a so-called **discourse model**: recording the entities that have been mentioned, and recognizing subsequent references to these entities [22]. For instance, in:

Sherlink Holmes went to the shop. He got some tobacco for his pipe. He liked it.,

the pronouns *he*, *his*, and *he* are all mentions of the entity first introduced in the discourse via proper name *Sherlink Holmes*. Anaphora is nominal reference to entities that have already been mentioned in a discourse [22]. The interpretation of anaphoric expressions is not always obvious: e.g., in the previous example pronoun *it* could refer to either the tobacco or the pipe.

Anaphora resolution involves first of all deciding whether a nominal phrase refers to a discourse entity (new or old) or whether it is **non-referring**, like pro-form *it* in *it's five o'clock*, which is semantically vacuous. In case a noun phrase refers, anaphora resolution requires specifying which discourse entity it refers to.

2.2 Gamified tools for anaphora annotation

The best-known anaphoric annotation GWAP, *Phrase Detectives*, was developed to ask the interpretation of anaphoric expressions to the crowd [7]. *Phrase Detectives* requires players to make the judgments about nominal phrases discussed in the previous Section (deciding whether a nominal phrase is non-referring or referring, marking its antecedent if required). Players can also carry out advanced annotation tasks such as marking plurals [7].

Figure 1: PhraseDetectives Anaphoric Annotation Interface

Phrase Detectives employs gamifications mechanics such as leaderboards and level-ups, and offers prizes of several kinds [23]. However, no further ludic mechanics are employed–hence, *Phrase Detectives* also falls somewhat between "gamified tools" and proper games.

2.3 Other GWAP designs for NLP

2.3.1 WordRobe. *WordRobe*[26] is a platform designed to collect annotations about different aspects of language processing. Along

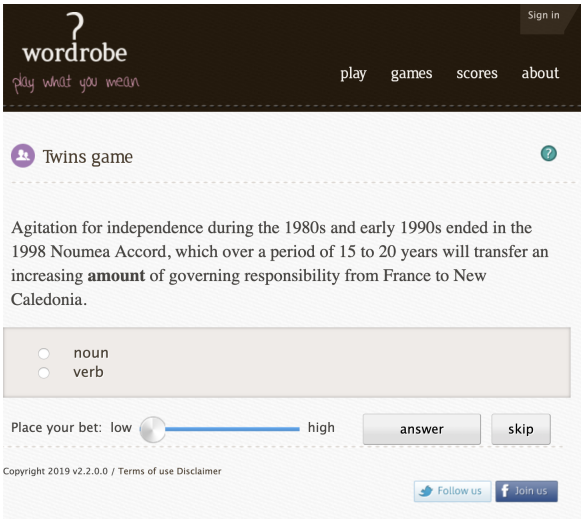

Figure 2: *Wormingo* WordRobe annotation interface

with other widely used gamification mechanics such as earning points, achievements etc. *WordRobe* employs an interesting device, "bets". If a player is very confident about their answer to a question, they can increase a bet on the question to earn more points. The "bet" data is then cleverly used to extract stronger judgements, assessed along with player's reliability as an annotator. But despite the integration of gamification mechanics, the games in WordRobe are still mainly gamified tools, as the gameplay consists of the repetition of labelling tasks.

2.3.2 PlayCoref. *PlayCoref* [13] is the only design idea we are aware of for a proper GWAP for anaphoric annotation. *PlayCoref* adds 2 different mechanics from *Phrase Detectives*. First, is a 2-player game. *Phrase Detectives* and *Wormingo* are both single-player. Second, in *PlayCoref* the text is presented to the players bit by bit, to alleviate frustration and boost reading comprehension. This technique is employed in *Wormingo* in the form of "chunks". As far as we are aware, *PlayCoref* was never published so we do not know how successful its design would have been.

2.3.3 Puzzle Racer and Ka Boom! *Puzzle Racer* and *Ka-boom!* are two other recent GWAPs that also attempted to return to von Ahn's original idea to produce data as a by-product of ludic activities, rather than gamifying an annotation tool [14].

Puzzle Racer is a real-time race game similar to *Mario Kart*. Since this game is in real-time, players need to take their actions/decisions quickly. Otherwise they might run out of time and lose the level. The game collects annotations in its "golden gates" phase, during which the player, who is initially given a theme about the level, sees 3 gates that they can choose from. Each gate displays an image and only one of them is related to the level's given theme, hence is the correct gate. If the player chooses to go through the correct gate, they earn points (time, in this case).

Ka-boom! is based on a very similar design. It is a GWAP adaptation of the well known mobile game *Fruit Ninja*. However in this game images are thrown on the screen instead of fruits and

the player, who is again initially given a theme, must only cut the images that are not related to the theme.

Both GWAPs seem to be fun, as they are adaptations of fun-proven games. However, they carry a flaw impacting the quality of the annotations gathered. Real-time games may be more engaging for the player depending on the type of activity [24, 29], however the time pressure on the player may lead them to do false annotations [14]. Similarly, players in the usability tests conducted for *Phrase Detectives* had reported feeling frustrated if they were constrained by time limitations, leading to poor judgements when annotating [6, 23]. The player should have time to think about the annotations they make, hence GWAPs should either be turn-based games or pause for a little respite during the moments of annotation.

2.3.4 RoboCorp. *RoboCorp* [11] starts as an ordinary arcade game (a platform game, specifically) and employs F2P (Free to Play) mechanics in order to motivate players to annotate more. These mechanics are usually used by the games belonging to the casual market [11, 30], and are tuned to maximise players' retention and therefore overall revenue; e.g., players can only play for a limited period of time and if they want to play more, they need to buy more energy. *RoboCorp* uses the same mechanics to gather more annotations, asking for players' work instead of their money.

3 THE MOTIVATION-ANNOTATION PARADIGM

GWAP were originally meant by von Ahn [27] to be entertaining games designed in such a way that the required labels could be collected as a by product of the game mechanics. And indeed, the design of (some of) von Ahn's original games satisfies this requirement. But the experience with GWAPs since, particularly with GWAPs for NLP, has been that designing games whose mechanics naturally lends itself to collecting the required labels seems very hard. So as discussed in the previous Section, most NLP GWAPs are little more than annotation tools with added gamification mechanics (leaderboards, level-up, badges etc.) to provide some fun [18].

In this paper we propose a game, *Wormingo*, based on a different approach to combining playing with annotating, that we call the **Motivation-Annotation Paradigm**. Instead of implementing a different game for each annotation task whose mechanics is designed to collect labels for a particular task, we propose to give players a fun moment between annotations (the **motivation phase**) by presenting them with intervening games -puzzles, in fact. These puzzles are, while being language-related puzzles, are not closely tied to the particular annotation intended. In fact, no annotations are collected in this phase. After players solve the puzzle in the motivation phase, they are represented with the **annotation phase** of the chunk. Once they have done the annotation for the chunk, next chunk will follow from its motivation phase, then that chunk's annotation phase and so on. Players are allowed to skip at any phase (motivation or annotation) in this version. When a player skips a phase, they do not lose or earn points; they are presented with the next phase and the game continues.

We argue that separating the game into the motivation-annotation phases solves a number of problems encountered by NLP GWAPs:

- Compared to other types of GWAPs, game design in NLP GWAPs generally face *comprehension* as an extra challenge. To make a successful annotation, a player has to first comprehend the task and its related content in question [13]. NLP GWAPs usually investigate textual data, whereas the discussed examples such as *ESP Game*, *FoldIt* point their questions towards images or imagery content [16, 28]. Image data being easier to absorb than textual data [25], leaves more room for focusing their ludic elements towards annotation tasks. For example, in ESP Game, it takes an insignificant amount of time for a player to absorb an image; hence it makes sense for the game to focus on gamifying the annotation task. In NLP labelling however, text comprehension resides as an initial challenge for the player and for the game design. By turning the reading part into a game, *Wormingo* aims to make the reading part easier for the players.
- The players might become distracted when they are reading the text and have attention slips, which in the end leads them to do wrong or low-quality annotations. The puzzles in the motivation phase help ensuring that the text has been comprehended, as players cannot possibly solve them without knowing what is going on in the text.
- The cold start problem, meaning the system not knowing the answer to an asked question yet, renders the scoring mechanism vague and frustrating to the player [2, 21]. By removing parts of the text and creating puzzles for the players, *Wormingo* generates questions that it knows the answer to. This allows immediate scoring, alleviating the negative effects of the cold start problem.
- In general, the puzzles add more ludic elements and uncertainty [9], hopefully making the game more fun for the players, thus more engaging [15].

The Motivation-Annotation Paradigm was inspired from the *annotation moment* mechanic introduced in *RoboCorp* [11], as they both feature interference of annotation tasks into an ongoing motivational game. Additionally, *Wormingo* utilizes its motivational phase to supplement the annotation phase by boosting text comprehension. Word games were chosen for the motivational phase, as they are more integrable into the text-based interface of annotation phase, keeping the player on similar interfaces compared to the switch of tasks in *RoboCorp*. We also assumed that players who would feel attracted to text-based annotation tasks might be more inclined to favor word games due to their text-based nature -an assumption that could be subject for a future study.

4 GAME DESIGN

4.1 Chunks

Wormingo divides documents into sequential **chunks**: segments of a text at most 50 words long. (Chunks never divide a sentence in two; a chunk stops at the ends of a sentence, if adding the next sentence would mean exceeding the 50-word limit.) The purpose of chunks is to avoid overwhelming the player by presenting lengthy texts at one time [13, 25], while overcoming one of the problems with *Phrase Detectives*, which is that players only see parts of a text so may not be able to correctly classify some markables as

discourse new. In *Wormingo*, a whole text is presented to a player, but only one small portion at a time.

4.2 Motivation puzzles

Wormingo currently has 3 different puzzles in its motivation phase:

(1) Fill the blanks,
(2) Hangman, and
(3) Crosswords.

All puzzles follow the "blanks" pattern: some parts of the text are removed from the user's display and the players are asked to find the words that should go into the blanks. Players can also set the frequency of the puzzles, allowing them to play a puzzle more if they like it.

The blanks are chosen pseudo-randomly by the algorithm, but all players play the same set of blanks on a given chunk of a text. We chose this method simply to prevent cheating. Players could open another account or view the same question on a friend's computer to see the answers, as a blank on one chunk could be on display on another player's view.

Puzzles get more difficult if the players increase the difficulty setting of the game. For higher difficulty, longer and rarer words are chosen from the chunks. Rarer words are currently determined in comparison with the other documents in the game, an external corpus has not been scanned for this purpose. Words are assigned a frequency value based on how many times they appear in the corpus and gain a higher difficulty score if they appear more seldom.

The choice of whether a word is a blank is maintained in each difficulty level; if a word is in a blank in the easiest level, it is still in a blank in the most difficulty one. This strategy has again been chosen to prevent cheating. Otherwise a player could try different difficulties of a chunk to view the blanks' answer.

We discuss each puzzle in turn.

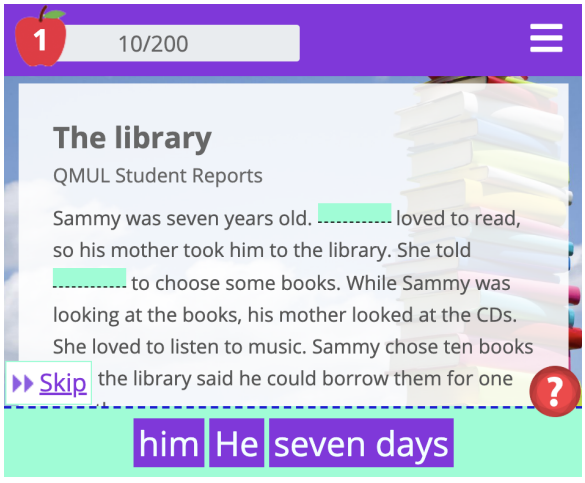

Figure 3: Fill the blanks

4.2.1 Fill the blanks. The first and the most basic puzzle of *Wormingo* is the "fill the blanks" puzzle. Players see at least one blank and try to fill the blanks with a word chosen among the options contained in the menu at the bottom. To increase the challenge, this puzzle also has an extra choice that is picked randomly from a nearby paragraph in the text.

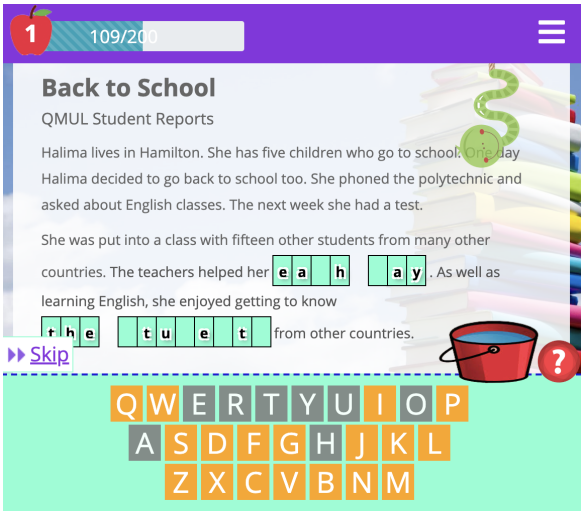

Figure 4: Hangman

4.2.2 Hangman. Hangman, referred to in the game as "Hang-Worm", is our implementation of the classical hangman puzzle. Again, the text displayed contains blanks, but this time the players choose a letter from the keyboard at the bottom. If the letter is contained in the blanks, those letters are revealed. If not, the player loses a life -presented as a worm character tied to a string getting closer to a bucket of water. After 6 failed tries, the worm reaches this bucket meaning that the player has failed the puzzle. The game moves on to the next task without awarding points.

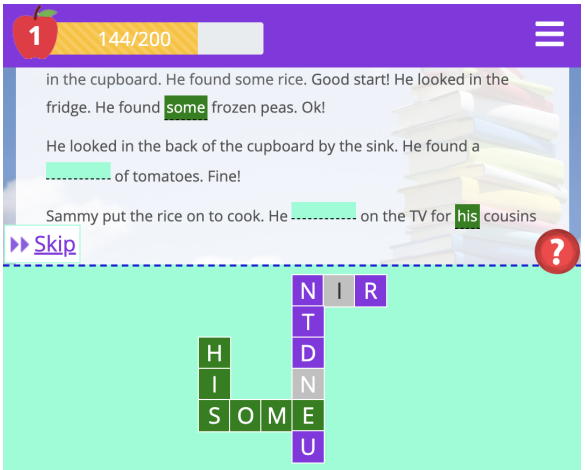

Figure 5: Crossword

4.2.3 Crossword. The third puzzle is crossword. The player sees each blank word in a crossword puzzle at the bottom of the screen,

however the letters are shuffled. The player has to swap the letters to bring each letter to their correct position. Some letters are displayed in grey and cannot be swapped; these are the clue letters that were added to make the puzzle somewhat easier. (In its raw version the puzzle was tested to be too hard and therefore too frustrating for the players.)

Players can also hover on a letter to see which blank that this letter's word is linked to. The linked blank is highlighted in colour red to point the relation out.

The blanks in Crossword puzzle are always composed of one word whereas "fill the blanks" and Hangman puzzles can have blanks of multiple words. This is because it would introduce the space character into the puzzle which would complicate the game for the player.

4.3 Annotation Phase

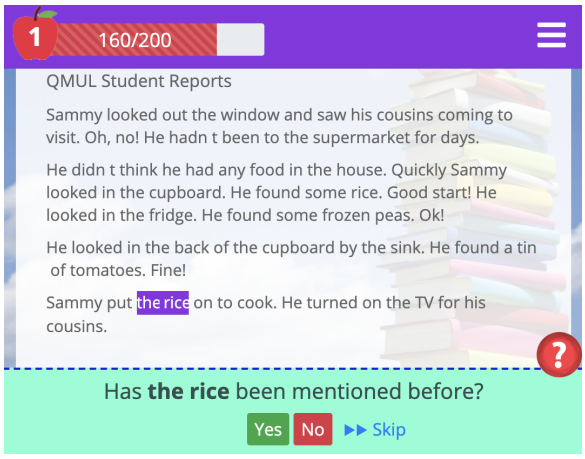

Figure 6: *Wormingo*'s anaphoric annotation interface

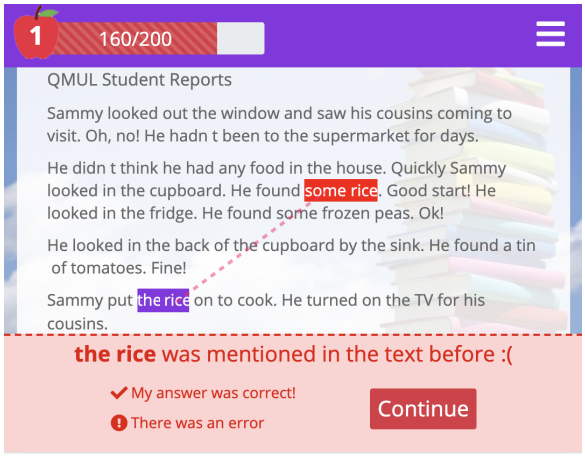

Figure 7: *Wormingo*'s anaphoric annotation interface displaying a wrong answer

After playing a puzzle, the players have to do one annotation in order to move to the next puzzle. *Wormingo* uses a similar interface to *Phrase Detectives* in annotation mode (Figure 1, 6, 7). One markable is highlighted, and the players have to choose "yes" if they believe the entity in question has already been mentioned, and then tag the coreferring mention. They should choose "No" if they think the entity has not been mentioned. Finally, if for some reason they cant decide they can choose "Skip", in which case the system displays a popup interface that allows the players input what kind of problem they have encountered.

4.4 Scoring

Players earn points in the game for each puzzle they solve and each correct annotation they make. Puzzles earn as many points as the number of letters in the blanks, plus a slight boost if player plays in increased difficulty.

An annotation can have 3 different results: (1) a correct answer, (2) a wrong answer, (3) the system may not know the answer for items which have not been completely annotated yet. Wrong answers do not earn any points, but correct answers earn 25 points. We have found 25 to be an optimal number as it is usually more than what motivations score but rarely more than its double. We intended the annotations to score higher (but not too much more) than motivation puzzles, to motivate the player into marking the annotations as well. In the third case, players earn 2 points plus an "egg". This is an analogy as once the system learns the correct answer after gathering enough annotations, the egg will hatch and the player will earn their 25 points if it was actually correct. So there is a minor boost in score to contemplate the frustration the player can experience when they do not receive immediate feedback to their answer [10].

5 RESULTS

Wormingo was released on Reddit on Saturday 30 Mar 2019, and we observed the results over the course of a week, evaluating it using the adaptation of F2P metrics for GWAPs proposed in [5]. For simplicity, the post on reddit featured only a URL to *Wormingo* and asked for the players to play the game for 10 minutes. Players were not asked to give any personally identifiable info (unless they chose to subscribe to our emailing list), and were provided with a link that explained about the NLP purposes of the project. During this week we did not repost the game or did not do any marketing ads or campaigns whatsoever. Also at the time of the experiment, *Wormingo* was not available to mobile devices so the visitors were coming only from desktop computers.

270 visitors started playing, producing 2416 annotations, so on average *Wormingo* has 8.94 LTJ (Lifetime judgements) [5]. Retention-wise, 5 players arrived after their first visit -however this low number is due to the limitation that for the purposes of piloting, *Wormingo* currently contains only 18 documents, and a player can finish these documents in 30 minutes. So players' retention rate cannot be correctly measured within the current limitations as players do not have documents to return to. Also, another set of 10 players have run through all the available documents on their first (and only) session, hence these players are also out of the retention equation. However, if we were to assume these players would be

retained, 15 players out of 270, a 5.5% could be called the loyal ones. Considering the 5% threshold in casual games for the paying customers [5, 11], and its interpretation into GWAPs [5] this could be considered as satisfactory.

In GWAPs usually a small portion of contributors produce most of the annotations [4]. In other words, players who play extensively produce much more than the total of the other players. We did not observe this effect in our experiment, again due to the lack of available documents. Since the players with the tendency to play more did not have many documents to run through, they could not go too much in front of the other players to garner this effect.

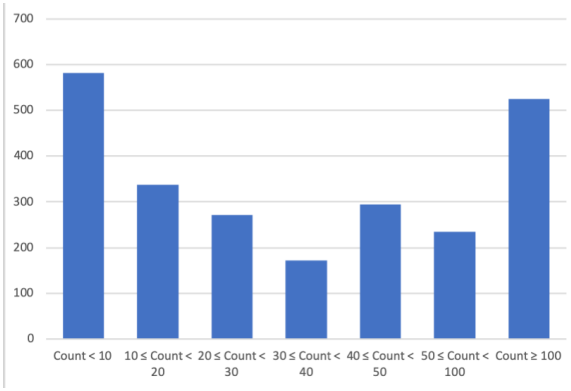

Figure 8: Number of annotations by band of annotation count

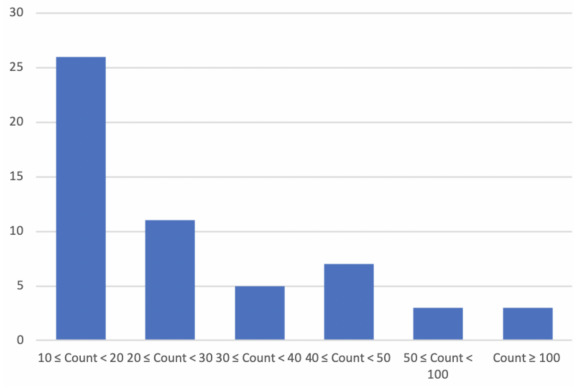

Figure 9: Number of players by band of annotation count

215 of the players produced less than 10 annotations per player, and 582 annotations in total. This portion of players will be ignored from the analysis as the first few annotations are pointed by the tutorials, and 10 annotations is not enough to judge the players' competence. Players who made between [10-20) annotations per player produced 338 annotations; [20-30) band produced 271, [30-40) band produced 172, [40-50) band produced 294, [50-100) band produced 234 and finally the players producing more than or equal to 100 annotations produced 525 annotations (Figure 8). Each band has 26, 11, 5, 7, 3 and 3 players respectively (Figure 9). Their accuracy

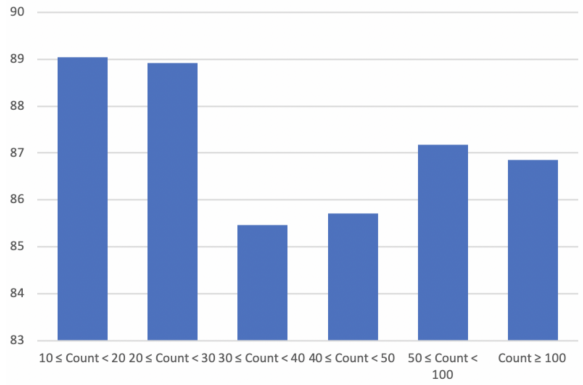

Figure 10: Accuracy by band of annotation count

looks to be higher on the lower bands (Figure 10), but this is due to the bias created by the tutorials. As the bands get bigger in annotation count, their accuracy goes higher which is expected - however their accuracy drops in the "Count ≥ 100" band. This is unexpected but might have occurred due to the small size of the sample.

6 CONCLUSIONS

Wormingo is the first GWAP for anaphoric annotation based on the motivation-annotation paradigm, and only the second GWAP of this type overall. Its results both in terms of player satisfaction and player accuracy look promising.

It might be argued that the ideal GWAPs should resemble the original *ESP Game* and *Peekaboom*, where the annotations are produced naturally by the ludic activity, so gameplay is not interrupted at all, be it with annotation moments. However, for at least the type of NLP annotation considered here, interrupting the gameplay fairly inobtrusive as the actual annotation task only takes a very short time. and the task of comprehending the text becomes part of the motivational game. Gamifying the reading process in this manner should make the annotation less boring. Our usability tests suggest that users do not think of the annotation moments following the puzzles as overly intrusive.

We are also convinced that our implementation of the motivation-annotation paradigm in terms of turns and "word games" is on the right track. The experience with *Phrase Detectives* suggests that GWAPs for this type of annotation should not rush the player -at least during the annotation tasks. Word games also, in our experience, align well with the taste of players who are attracted by NLP GWAPs.

ACKNOWLEDGEMENTS

This research was supported by the DALI project, ERC Grant 695662.

REFERENCES

[1] Tyler Baron and Ashish Amresh. 2015. Word towers: Assessing domain knowledge with non-traditional genres. In *Proceedings of the European Conference on Games-based Learning*, Vol. 2015-January. Dechema e.V., 638–645.

[2] Luke Barrington, Damien O'Malley, Douglas Turnbull, and Gert Lanckriet. 2009. User-centered Design of a Social Game to Tag Music. In *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP '09)*. ACM, New York, NY, USA, 7–10. https://doi.org/10.1145/1600150.1600152

[3] Karën Fort Bruno Guillaume and Nicolas Lefebvre. 2016. Crowdsourcing complex language resources: Playing to annotate dependency syntax. In *Proc. of COLING*.

[4] J. Chamberlain. 2016. *Harnessing Collective Intelligence on Social Networks*. Ph.D. Dissertation. University of Essex, School of Computer Science and Electronic Engineering.

[5] Jon Chamberlain, Richard Bartle, Udo Kruschwitz, Chris Madge, and Massimo Poesio. 2017. Metrics of games-with-a-purpose for NLP applications. (2017), 2.

[6] Jon Chamberlain, Karën Fort, Udo Kruschwitz, Mathieu Lafourcade, and Massimo Poesio. 2013. *Using Games to Create Language Resources: Successes and Limitations of the Approach*. 3–44. https://doi.org/10.1007/978-3-642-35085-6$_1$

[7] Jon Chamberlain, Massimo Poesio, and Udo Kruschwitz. 2008. Phrase Detectives - A Web-based Collaborative Annotation Game. In *In Proceedings of I-Semantics*.

[8] Seth Cooper, Firsas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popovic, and the Foldit Players. 2010. Predicting protein structures with a multiplayer online game. *Nature* 466 (2010), 756–760.

[9] Greg Costikyan. 2013. *Uncertainty in Games*. The MIT Press.

[10] Anca Dumitrache, Lora Aroyo, Chris Welty, Robert-Jan Sips, and Anthony Levas. 2013. "Dr. Detective": combining gamification techniques and crowdsourcing to create a gold standard in medical text. In *Proceedings of the 1st International Workshop on Crowdsourcing the Semantic Web, Sydney, Australia, October 19, 2013*. 16–31. http://ceur-ws.org/Vol-1030/paper-02.pdf

[11] Dagmara Dziedzic. 2016. Use of the Free to Play model in games with a purpose: The RoboCorp game case study. *Bio-Algorithms and Med-Systems* 12 (11 2016), 187–197. https://doi.org/10.1515/bams-2016-0020

[12] L. Gu and A. L. Jia. 2018. Player Activity and Popularity in Online Social Games and their Implications for Player Retention. In *2018 16th Annual Workshop on Network and Systems Support for Games (NetGames)*. 1–6. https://doi.org/10.1109/NetGames.2018.8463415

[13] Barbora Hladka, Jirí Mírovský, and Pavel Schlesinger. 2009. Play the Language: Play Coreference. 209–212.

[14] David Jurgens and Roberto Navigli. 2014. It's All Fun and Games until Someone Annotates: Video Games with a Purpose for Linguistic Annotation. *Transactions of the Association for Computational Linguistics* 2 (2014), 449–464. https://www.aclweb.org/anthology/Q14-1035

[15] Jesper Juul. 2008. The Magic Circle and the Puzzle Piece. (01 2008), 56–67.

[16] Firas Khatib, Seth Cooper, Michael D. Tyka, Kefan Xu, Ilya Makedon, Zoran Popović, David Baker, and Foldit Players. 2011. Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences* 108, 47 (2011), 18949–18953. https://doi.org/10.1073/pnas.1115898108 arXiv:https://www.pnas.org/content/108/47/18949.full.pdf

[17] Mathieu Lafourcade. 2007. Making people play for Lexical Acquisition with the JeuxDeMots prototype. In *Proc. of SNLP*.

[18] Mathieu Lafourcade, Alain Joubert, and Nathalie Le Brun. 2015. *Games with a Purpose (GWAPS) (Focus Series in Cognitive Science and Knowledge Management)*. Wiley-ISTE.

[19] Chris J. Lintott, Kate Land, Kevin Schawinski, Anže Slosar, Daniel Thomas, Robert C. Nichol, Steven Bamford, Alex Szalay, Jan Vandenberg, M. Jordan Raddick, Dan Andreescu, and Phil Murray. 2008. Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey*. *Monthly Notices of the Royal Astronomical Society* 389, 3 (09 2008), 1179–1189. https://doi.org/10.1111/j.1365-2966.2008.13689.x arXiv:http://oup.prod.sis.lan/mnras/article-pdf/389/3/1179/3325962/mnras0389-1179.pdf

[20] Allan Paivio, T. B. Rogers, and Padric C. Smythe. 1968. Why are pictures easier to recall than words? *Psychonomic Science* 11, 4 (01 Apr 1968), 137–138. https://doi.org/10.3758/BF03331011

[21] José Pedro Pinto and Paula Viana. 2013. TAG4VD: A Game for Collaborative Video Annotation. In *Proceedings of the 2013 ACM International Workshop on Immersive Media Experiences (ImmersiveMe '13)*. ACM, New York, NY, USA, 25–28. https://doi.org/10.1145/2512142.2512154

[22] Massimo Poesio. 2016. Linguistic and Cognitive Evidence About Anaphora. In *Anaphora Resolution: Algorithms, Resources and Applications*, M. Poesio, R. Stuckardt, and Y. Versley (Eds.). Springer, Chapter 2.

[23] Massimo Poesio, Jon Chamberlain, Udo Kruschwitz, Livio Robaldo, and Luca Ducceschi. 2013. Phrase Detectives: Utilizing Collective Intelligence for Internet-scale Language Resource Creation. *ACM Trans. Interact. Intell. Syst.* 3, 1, Article 3 (April 2013), 44 pages. https://doi.org/10.1145/2448116.2448119

[24] Pramila Rani, Nilanjan Sarkar, and Changchun Liu. 2005. Maintaining optimal challenge in computer games through real-time physiological feedback. In *Proceedings of the 11th international conference on human computer interaction*, Vol. 58. 22–27.

[25] Maria Rasmusson and Monica Eklund. 2013. "It's easier to read on the Internet—you just click on what you want to read…". *Education and Information Technologies* 18, 3 (01 Sep 2013), 401–419. https://doi.org/10.1007/s10639-012-9190-3

[26] Noortje J. Venhuizen, Valerio Basile, Kilian Evang, and Johan Bos. 2013. Gamification for Word Sense Labeling. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Short Papers*. Association for Computational Linguistics, Potsdam, Germany, 397–403. https://www.aclweb.org/anthology/W13-0215

[27] Luis Von Ahn. 2006. Games with a purpose. *Computer* 39, 6 (2006), 92–94.

[28] Luis von Ahn and Laura Dabbish. 2004. Labeling Images with a Computer Game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. ACM, New York, NY, USA, 319–326. https://doi.org/10.1145/985692.985733

[29] Luis von Ahn and Laura Dabbish. 2008. Designing Games with a Purpose. *Commun. ACM* 51, 8 (Aug. 2008), 58–67. https://doi.org/10.1145/1378704.1378719

[30] Gabe Zichermann and Christopher Cunningham. 2011. *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps* (1st ed.). O'Reilly Media, Inc.