Making Text Annotation Fun with a Clicker Game

Chris Madge
Queen Mary Univ. Of London
United Kingdom
c.j.madge@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

Udo Kruschwitz
University Of Essex
United Kingdom
udo@essex.ac.uk

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

ABSTRACT

In this paper we present *WordClicker*, a clicker game for text annotation. We believe the mechanics of 'Ville type Free-To-Play (F2P) games in general, and clicker games in particular, is particularly suited for GWAPs (Games-With-A-Purpose). *WordClicker* was developed as one component of a suite of GWAPs meant to cover all aspects of language interpretation, from tokenization to anaphoric interpretation. As such, *WordClicker* is intended to have a dual function as part of this suite of GWAPs: both for parts-of-speech annotation and for teaching players about parts of speech so that they can go on and play GWAPs for more complex syntactic annotation. Therefore, game-based language learning platforms also had a strong influence on its design.

CCS CONCEPTS

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

KEYWORDS

Game Design; Games With A Purpose; Incremental Games; User Training; Language Resourcing

ACM Reference Format:
Chris Madge, Richard Bartle, Jon Chamberlain, Udo Kruschwitz, and Massimo Poesio. 2019. Making Text Annotation Fun with a Clicker Game. 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, 6 pages. https://doi.org/10.1145/3337722.3341869

1 INTRODUCTION

Games-With-A-Purpose (GWAPs) for creating language resources [11, 25, 36, 44] have shown promise in terms of their ability to gather high quality annotations and in terms of scalability. However, player recruitment and retention remains a challenge with such games, that have yet to acquire or retain players at a scale comparable to the most successful GWAPs [9, 45]. The original GWAPs for AI by von Ahn, such as The ESP Game, were effective in presenting their tasks, as per the original definition, in such a way that the labels gathered were a byproduct of play [45]. In contrast, it has been said that language resourcing games such as PhraseDetectives [36], are not entirely GWAPs as annotations are not a byproduct, but rather it is evident that the player is annotating text [26]. This can be said of the majority, if not all language resourcing GWAPs. Wordrobe for example, unlike PhraseDetectives, is a game which deliberately aims to hide the true nature and linguistic complexity of the tasks by presenting them as multiple choice questions and removing linguistic terminology [44]. However, it remains evident the player is annotating text. Similarly for other well-known game-like approaches to NLP resource creation such as Jeux-de-Mots and Zombilingo [11, 25]. Proper GWAPs have been proposed, but never really used for resource creation or reported high levels of player acquisition [19].

The approach to making text annotation GWAPs more game-like followed in this work is based on the general principle of starting from a pre-existing and engaging game mechanics, just as done in some of the most interesting GWAPs for AI [19, 47]. The question we addressed was: what type of existing game 'hides' the mechanics of text annotation more easily?

Our contribution *WordClicker*, the adaptation of so-called 'Ville game [21] mechanics for text labelling. We believe this type of game design addresses a lot of the challenges to be addressed by text annotation GWAPs (or indeed, GWAPs for any type of annotation). We believe this type of design, in which entertaining games are created out of intrinsically repetitive activities, is uniquely suited for annotation games, in which the objective is to keep players performing unentertaining activities for a long period.

WordClicker was developed as one component of a suite of GWAP meant to cover all aspects of language interpretation, from tokenization to anaphoric interpretation, not unlike the *WordRobe* suite of games [3]. *WordClicker* is intended to have a dual function as part of this suite of GWAPs: to be usable to label parts of speech, but also, and equally importantly, to train players in preparation for other games for syntactic annotation which presuppose the ability to recognize nouns or verbs. As such, its design drew inspiration not only from clicker games, but also from work on language learning platforms.

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/10.1145/3337722.3341869

2 RELATED WORK

2.1 Text Labelling In Games With A Purpose

Games With A Purpose (GWAPs) were originally proposed by Von Ahn [45] as a method of soliciting human computation as a byproduct of users playing a game. The original GWAPs targeted image labelling tasks [46, 48, 49], but the concept was later deployed tackle far more ambitious tasks [9], including language resourcing (LR) [11, 25, 36, 44].

From their inception it was evident there were various challenges to GWAPs, for some of which, Von Ahn offered some design solutions [47]. Naturally, there were additional challenges encountered when attempting to apply the paradigm to different areas, particularly when the tasks were more challenging than the original image labelling tasks. However, as pointed out by Tuite, as games present a multitude of opportunities for teaching their users, in contrast with other crowdsourcing approaches, this challenge is also GWAPs opportunity to excel [42]. GWAPs for language resourcing (LR) is perhaps one of the most challenging yet opportunistic domains.

One of the problems with all GWAPs, from a design perspective, is the conflicting interest of tools and toys. Challenge in a game is artificially introduced in the form of internal goals, for the sake of the game. Tools however, are designed to reduce the challenge of achieving the external goal or purpose [32]. By improving the game one could negatively impact achieving the purpose ("orthogonal game mechanics" [42]), or by improving the users ability to achieve the purpose, make the game less entertaining.

It has been said that Phrase Detectives [36] (shown in Figure 1), was more of an application of gamification than a true GWAP [26], retaining its tool like focus. Another approach, at the other end of the spectrum, has been to make the GWAP as game-like as possible [19, 43]. A third approach, somewhere in between, is to try to and hide the linguistic nature of the task to give a game like appearance whilst retaining the tool like utility [44]. None of these GWAPs has succeeded yet in providing a truly game-like experience as originally envisaged by von Ahn [45], although Phrase Detectives at least has been able to collect very large amounts of annotations if not perhaps on the scale originally envisaged by GWAP proponents.

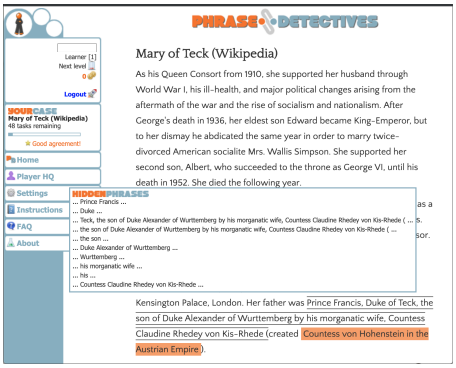

Figure 1: Text Labelling in PhraseDetectives

Cost is another constraint that features repeatedly in the discussion of GWAPs. Modern games have very large development [13, 31, 50] and marketing budgets [13, 16]. However, reduced cost over manual labelling was the main motivation given for the development of the original GWAPs [46, 49] and has been given as a motivation for LR GWAPs [44]. Multiple LR GWAPs use it as a metric by which they evaluate their success [19, 36] and have shown it is possible to use a GWAP to gather annotations more cost effectively than other methods [36]. However, for the majority of projects this constraint rules out development of traditional expensive modern games for GWAPs, as they would likely be more expensive than crowdsourcing [36, 39].

GWAPs require a game that supports infinite gameplay, this is part of the most popular games [19, 36, 46, 48, 49]. This is not a feature of many game genres, particularly those that have a limited story.

Another challenge is that the GWAP player bases contribute very unevenly. Phrase Detectives reports a Zipfian distribution, with 1.6% of its players made 89% of its annotations in the Facebook version of their game[6]. There's a very similar situation with ZombiLingo players [12]. This means catering to two very different types of player.

2.2 Text Labelling and Language Learning

As it happens, very popular Mobile Assisted Language Learning (MALL) apps such as Duolingo [1], have a strong labelling focus that is often token based, as shown in Figure 2.

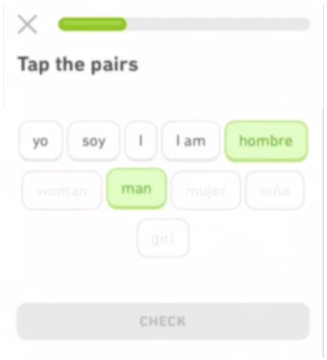

Figure 2: Text Labelling in Duolingo

Learning using drills, flashcards, or generally learning by rote [33], remains the preferred method in modern Mobile Assisted Language Learning (MALL) with the most popular apps [17] such as Duolingo [37] and Memrise [20] using spaced repetition algorithms that calculate the optimal interval with which to test a player to ensure long-term memory.

3 'VILLE GAMES

The hypothesis behind the development of *WordClicker* is that so-called **'Ville Games** are a particularly apt model for GWAP design. In this Section we first introduce F2P games before going on to discuss 'Ville Games and the particular form of 'Ville games we used, **clicker games**.

[1] https://www.duolingo.com

3.1 Free-to-Play Games

The F2P (Free-to-Play) revenue model has become a popular method of reaching casual gamers on web and mobile platforms [1, 29]. These audiences would not necessarily consider committing to an initial purchase, but may consider small purchases to enhance experience as they progress in the game [29]. For games that employ this revenue model, F2P motivates a specific set of design objectives [28], For example, these small purchases take place over a long period of time, so F2P games, like GWAPs, are often designed to be infinite or long lasting. The lack of initial financial investment means the game needs to appeal to the player right from the start, as there is nothing to discourage the player putting it down if it is too difficult to master for a casual gamer, or is not immediately entertaining. Consequently F2P games commonly feature a shallow learning curve. One of the most important parts, at the heart of the design of Free-to-Play games, is their core game loop [29], $\hookrightarrow Action \rightarrow Wait \rightarrow Reward \rightarrow Upgrade \rightarrow (loop)$.

Commonalities between the interests of F2P and GWAPs have been pointed out before, which has led to the adaptation of F2P metrics for GWAPs [7, 30], and the translation of their microtransaction system into a "microwork" system in a GWAP [10].

3.2 'Ville Games

So-called "'Ville Games" are, we would argue, a particularly relevant category of F2P games for GWAP design. Following the advent of social networks [5], various organisations, particularly Facebook [15], opened their platforms to the embedding and distribution of third party applications. Of these, was a set of multiplayer games that allowed friends to play together known as social network games (SNGs). Often using the Free-to-Play revenue model [34], costing comparatively less to develop than their conventional counterparts [38], offering inclusive play to casual gamers (being easy to pick up and put down in short sessions) [34], and being web-based (accessible on a variety of devices [18]), they quickly shot to success [8, 23, 24]. One particularly successful group of such games is the 'Ville group of games created by Zynga that share the "Ville" suffix to their names (e.g. *FarmVille*, *FishVille*, *YoVille*). Over time, led by the successful 'Ville titles, the design of popular SNGs began to homogenize into a common set of studied design patterns [23, 27].

There are reoccurring design patterns that appear in these games [27]. For example, in the gameplay, a player action typically results in gathering an in-game resource which develops over time. A further player action realises a reward from **harvesting** the resource as in-game currency. Resources that the player has failed to convert to in-game currency after some given timer period **wither**. They also have a variety of resources, aside from the previously mentioned **game-specific resource** and **in-game currency**.

The core game loop of action (purchase resource), waiting (resource appreciates in value), reward (resource converted to currency), upgrade from F2P, is very evident in the design patterns of 'Ville games. [29]

3.3 Clicker Games

The aforementioned **'Ville Games** and their wider SNG genre, have been the subject of satire with critics creating games with deliberately bland core game mechanics, such as "Cow Clicker", that involves simply clicking once every six hours [4]. This widely mocked, but undeniably successful [40] game design pattern spawned a subgenre of games that distilled the 'Ville paradigm known as "clicker games" in which the player repeatedly clicks to earn points which they can use to purchase items that enable them to earn more points [21].

There are now many variations of the clicker game design, collectively known as "Incremental games", with some research proposing a taxonomy [2]. The key defining factor that separates Incremental games is the spectrum of interactivity with the player [21]. The previously mentioned "clicker game" variety is among the highest level of interactivity and the lowest "zero player games" in which the player's role is reduced to that of a spectator for the majority of the game [2].

Many games have continued these ideas more seriously and there have already been successful, entirely text based "clicker games", such as "A Dark Room" [14]. Exploiting behavioural psychology and decision making, these games appear to have, in part, changed our definition of what we believe a "good" game is [27].

Figure 3: Clicker Game - "A Dark Room" [14]

Of all incremental games, "Clicker games" are of particular interest as they have both a high level of interaction and an almost tool-like "high ludic efficiency" that matches the interests of GWAPs and serious games. The player doesn't have to have fast reactions, good coordination or be practised, they simply click. [2, 21] Additionally, whilst SNGs have been shown to offer an inexpensive method of gathering a large casual player base, "Clicker games" are cheaper still, being created by very small teams or individuals in very little time [2, 22, 40].

Additionally, Clicker games provide an entertaining way of presenting a highly repetitive task. Aside from being suitable for labelling, this may also be of interest for learning by rote (as commonly used in MALL apps).

4 *WORDCLICKER*

WordClicker is a web-based [2], desktop and mobile friendly, one-player game in which a player learns the classes of words by playing a baker that gets her/his ingredients by clicking on words associated with those ingredients. The core game mechanics is simply classifying individual words into classes (associated with ingredient jars) by clicking on them, a mechanic that should be transferable to the majority of word-labelling tasks. If the player is correct, after clicking they get ingredients, that are used to make the cakes. The game is very simple, taking approximately two weeks for one person to develop.

4.1 Gameplay

To begin with, the player is shown details of the task they will be performing with a short explanation.

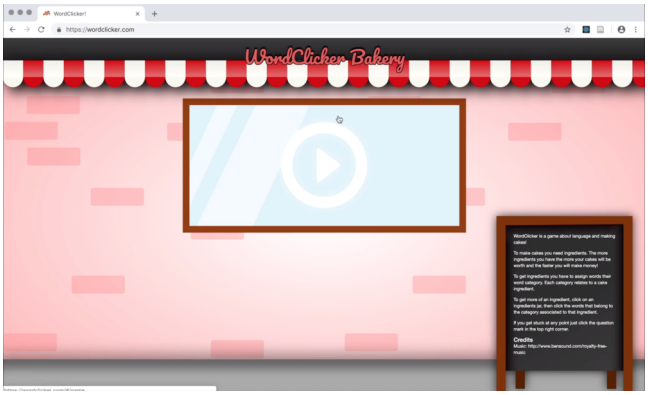

Figure 4: *WordClicker*- Introduction

When they press play they are presented with an interactive tutorial that takes them through basics of the game (shown in Figure 5).

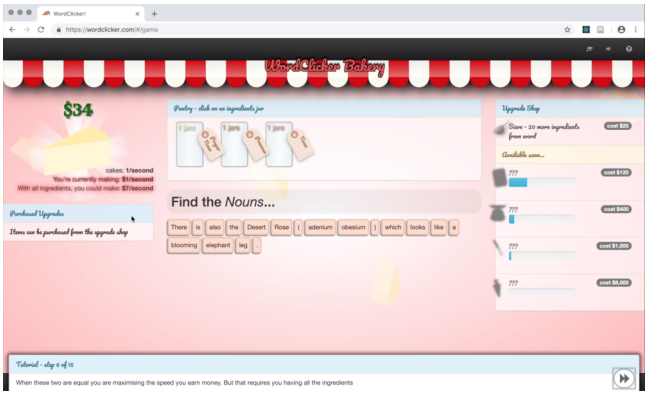

Figure 5: *WordClicker*- Tutorial

They can repeat this tutorial and view additional instructions regarding the classes at any point (shown in Figure 6).

[2]https://wordclicker.com/

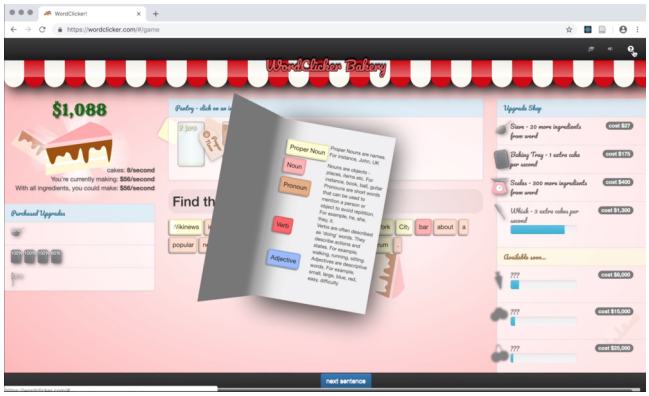

Figure 6: *WordClicker*- Instructions

During gameplay, the player is shown a single sentence at a time (see Figure 8). They can advance to the next sentence by using the "Next sentence" button. Once players have earned a sufficient amount of in-game credits they unlock and are offered the opportunity to progress onto a language resourcing game (see Figure 7).

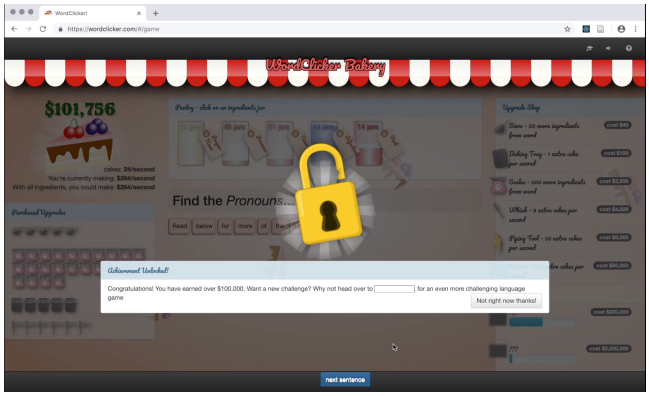

Figure 7: *WordClicker*- Progression

4.2 Mechanics

In this section we zoom in at a more granular level on the gameplay focussing on the core game mechanics, organised by their place in the incremental game design loop.

Action step. In the action step of the game loop, the player collects the resources (ingredients). With the goal of **high ludic efficiency**, the user interaction involves simply selecting the appropriate ingredients jar (category), then selecting one or more words in the sentence that are of that ingredient (category). The incremental game design choices negates the need for adding orthogonal mechanics. Accumulating ingredients are shown in their respective jars. The correctly marked token is then shown with a shimmering effect.

Wait Step. In the wait step of the loop, cakes are automatically produced and sold in the quantity specified by the current multiplier (in the generator) giving the player a reward. Resources that the player has gathered (cake ingredients), are consumed synchronously and added to the cakes when available. This relationship is illustrated to the user through an animation that shows ingredients leaving the jars and moving to the cake and the ingredients being shown on the cake itself. The more ingredients a cake has, the more it is worth. The player is shown the cakes potential worth and their current worth in the game. This is designed to encourage the player to explore all of the labelling categories currently available to them, to maximise their potential gain by leveraging the notion that players do not want to waste their purchases, known as the *sunk-cost fallacy* [27, 41]. Here we are using an *avoidance fixed interval schedule* with *fixed avoidance schedule* (known to be suitable for a slow but steady response) [27] underneath to soften it. That is, the players receive a reward based on their investments regardless, but they receive far less reward unless they manage to continue to steadily find ingredients. Here we are directing players towards marking labels.

Reward. We require no action (e.g. harvesting) on the part of the player to receive their reward. However, there is a deliberate disconnect (cakes) between the resources that are gathered (ingredients) and the virtual currency (dollars). The purpose of this is to add an additional opportunity for control that is utilised, as described in the wait step, to motivate the player to label all the categories.

Upgrade. Upgrades are purchased from the shop by the player investing their primary reward and affect the game in two ways. They can either, increase the generator multiplier or increase the quantity of resources produced by a correct. These purchases effectively either increase game speed/reward/difficulty, or slow down the game whilst preserving the speed/reward. The cost of each upgrade increases infinitely, providing potentially **infinite gameplay**, and exponentially, with each purchase (in line with typical idle game formulas [35]).

Leveraging the *goal-gradient hypothesis* that players exert more effort when approaching a reward [27], upgrades are obscured in the store until the player has almost sufficient funds, and a progress bar shows how close the player is to being able to purchase that reward (Figure 8). Here again, we direct players towards marking labels.

As the game progresses the player also has the opportunity to purchase additional labelling categories. This allows for a configurable, self-paced player progression.

Penalizing incorrect responses. When the player labels incorrectly one of their purchases, if available, becomes damaged. This negative reinforcement leverages the players loss aversion to encourage considered annotation. Feedback is given in the form of a text notification message that appears in the bottom left hand corner and a flashing red outline on the token (shown in Figure 8).

4.3 Corpora

WordClicker requires a corpora, or large body of annotated data. There needs to be a sufficiently permissive copyright licence that we could present the texts in the context of a game; the part-of-speech tags needed to be of particularly high quality and whilst not a requirement, ideally the texts would be interesting to read. The GUM corpora [51] was selected as a corpora matching these criteria.

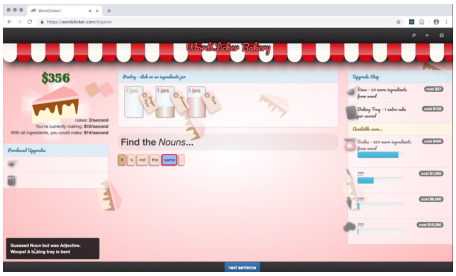

Figure 8: *WordClicker*- Gameplay with errors and feedback

5 CONCLUSION

To be considered a truly effective approach and useful for integration into a GWAP, a text labelling game and its constituent mechanics should be capable of acquiring players alongside traditional games, but face considerable design constraints including cost, the requirement of high ludic efficiency and tasks that are not a natural fit for many existing game genres. In this work we present an adaptation and application of clicker game mechanics to address the challenging design space of text labelling in games.

Future research will carry out a complete experiment exploring *WordClicker* ability to recruit and retain players. discussed, to test *WordClicker* directly as a GWAP.

ACKNOWLEDGEMENTS

This research was supported in part by the EPSRC CDT in Intelligent Games and Game Intelligence (IGGI), EP/L015846/1; in part by the DALI project, ERC Grant 695662.

REFERENCES

[1] Kati Alha, Elina Koskinen, Janne Paavilainen, and Juho Hamari. 2014. Free-to-Play Games: Professionals' Perspectives. (2014), 14.
[2] Sultan A. Alharthi, Olaa Alsaedi, Zachary O. Toups, Joshua Tanenbaum, and Jessica Hammer. 2018. Playing to Wait: A Taxonomy of Idle Games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–15. https://doi.org/10.1145/3173574.3174195
[3] Valerio Basile, Johan Bos, Kilian Evang, and Noortje Venhuizen. 2012. A platform for collaborative semantic annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 92–96.
[4] Ian Bogost. 2010. Cow Clicker: The Making of Obsession (2010). http://bogost.com/writing/blog/cow$_{c}$licker$_{1}$/.
[5] Danah M Boyd and Nicole B. Ellison. 2007. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication* 13, 1 (2007), 210–230. https://doi.org/10.1111/j.1083-6101.2007.00393.x
[6] Jon Chamberlain. 2014. Harnessing the intelligence of the crowd for problem solving and knowledge discovery. In *AAAI conference on Human Computation and Crowdsourcing*.
[7] Jon Chamberlain, Richard Bartle, Udo Kruschwitz, Chris Madge, and Massimo Poesio. 2017. Metrics of games-with-a-purpose for NLP applications. (2017), 2.
[8] Sande Chen. 2009. The Social Network Game Boom. http://www.gamasutra.com/view/feature/132400/the$_{s}$ocial$_{n}$etwork$_{g}$ame$_{b}$oom.php
[9] 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.

[10] Dagmara Dziedzic and Wojciech Włodarczyk. 2016. Making NLP games with a purpose fun to play using Free to Play mechanics: RoboCorp case study. In *Proceedings of EACL-Games4NLP Workshop*. 187–197.

[11] Karen Fort, Bruno Guillaume, and Hadrien Chastant. 2014. Creating *Zombilingo* , a game with a purpose for dependency syntax annotation. In *Proceedings of the First International Workshop on Gamification for Information Retrieval - GamifIR '14*. ACM Press, Amsterdam, The Netherlands, 2–6. https://doi.org/10.1145/2594776.2594777

[12] Karën Fort, Bruno Guillaume, and Nicolas Lefebvre. 2017. Who wants to play Zombie? A survey of the players on ZOMBILINGO. In *Games4NLP 2017-Using Games and Gamification for Natural Language Processing*. 2.

[13] John Funk. 2009. How Much Did Modern Warfare 2 Cost to Make? | The Escapist. http://www.escapistmagazine.com/news/view/96227-How-Much-Did-Modern-Warfare-2-Cost-to-Make

[14] Doublespeak Games. 2013. A Dark Room. http://adarkroom.doublespeakgames.com/.

[15] Doug Gross. 2010. The Facebook games that millions love (and hate) - CNN.com. http://www.cnn.com/2010/TECH/02/23/facebook.games/index.html

[16] Ronald Grover and Malathi Nayak. 2014. Activision plans $500 million date with 'Destiny'. *Reuters* (May 2014). https://www.reuters.com/article/us-activision-destiny-idUSBREA4501F20140506

[17] Catherine Regina Heil, Jason S. Wu, Joey J. Lee, and Torben Schmidt. 2016. A Review of Mobile Language Learning Applications: Trends, Challenges, and Opportunities. *The EuroCALL Review* 24, 2 (Sept. 2016), 32. https://doi.org/10.4995/eurocall.2016.6402

[18] David Helgason. 2008. Thoughts On Browser Plugin Penetration âĂŞ Unity Blog. https://blogs.unity3d.com/2008/03/31/thoughts-on-browser-plugin-penetration/

[19] David Jurgens and Roberto Navigli. 2014. It's all fun and games until someone annotates: Video games with a purpose for linguistic annotation. *TACL* 2 (2014), 449–464.

[20] Clara Herlina Karjo and Wiwik Andreani. 2018. Learning Foreign Languages With Duolingo and Memrise. In *Proceedings of the 2018 International Conference on Distance Education and Learning - ICDEL '18*. ACM Press, Beijing, China, 109–112. https://doi.org/10.1145/3231848.3231871

[21] Imran Khaliq and Blair Purkiss. 2015. A study of interaction in idle games & perceptions on the definition of a game. In *Games Entertainment Media Conference (GEM), 2015 IEEE*. IEEE, 1–6.

[22] Roisin Kiberd. 2016. Cookie Clicker, the Internet's most pointlessly addictive game, is also its most subversive. https://kernelmag.dailydot.com/issue-sections/staff-editorials/15694/cookie-clicker-capitalist-dystopia/

[23] Ben Kirman. 2010. Emergence and playfulness in social games. In *Proceedings of the 14th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '10*. ACM Press, Tampere, Finland, 71. https://doi.org/10.1145/1930488.1930504

[24] Zoe Kleinman. 2009. Free games attract high prices. (Nov. 2009). http://news.bbc.co.uk/1/hi/technology/8376392.stm

[25] Mathieu Lafourcade. 2007. Making people play for Lexical Acquisition with the JeuxDeMots prototype. (2007), 8.

[26] Mathieu Lafourcade, Alain Joubert, and Nathalie Le Brun. 2015. *Games with a Purpose (GWAPS)*. John Wiley & Sons.

[27] Chris Lewis, Noah Wardrip-Fruin, and Jim Whitehead. 2012. Motivational game design patterns of 'Ville games. In *Proceedings of the International Conference on the Foundations of Digital Games*. ACM, 172–179.

[28] Pascal Luban. 2011. The Design of Free-To-Play Games: Part 1. https://www.gamasutra.com/view/feature/134920/the$_{\text{d}}$esign$_{\text{o}}$f$_{\text{f}}$reetoplay$_{\text{g}}$ames$_{\cdot}$php

[29] Will Luton. 2013. *Free-to-play: Making money from games you give away*. New Riders.

[30] 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 - CHI PLAY '17 Extended Abstracts*. ACM Press, Amsterdam, The Netherlands, 397–404. https://doi.org/10.1145/3130859.3131332

[31] Eddie Makuch. 2017. Dead Space 2 Dev Says Game Cost $60M To Make And Sold Millions, But Failed Commercially. https://www.gamespot.com/articles/dead-space-2-dev-says-game-cost-60m-to-make-and-so/1100-6454150/

[32] Thomas W Malone. 1981. Toward a theory of intrinsically motivating instruction. *Cognitive science* 5, 4 (1981), 333–369.

[33] Richard E. Mayer. 2002. Rote Versus Meaningful Learning. *Theory Into Practice* 41, 4 (Nov. 2002), 226–232. https://doi.org/10.1207/s15430421tip4104$_4$

[34] Janne Paavilainen, Juho Hamari, Jaakko Stenros, and Jani Kinnunen. 2013. Social Network Games: Players' Perspectives. *Simulation & Gaming* 44, 6 (Dec. 2013), 794–820. https://doi.org/10.1177/1046878113514808

[35] Anthony Pecorella. 2016. The Math of Idle Games, Part I. https://blog.kongregate.com/the-math-of-idle-games-part-i/

[36] Massimo Poesio, Jon Chamberlain, Udo Kruschwitz, Livio Robaldo, and Luca Ducceschi. 2013. Phrase detectives: Utilizing collective intelligence for internet-scale language resource creation. *ACM TiiS* 3, 1 (2013), 3.

[37] Burr Settles and Brendan Meeder. 2016. A Trainable Spaced Repetition Model for Language Learning. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Berlin, Germany, 1848–1858. https://doi.org/10.18653/v1/P16-1174

[38] Dong-Hee Shin and Youn-Joo Shin. 2011. Why do people play social network games? *Computers in Human Behavior* 27, 2 (March 2011), 852–861. https://doi.org/10.1016/j.chb.2010.11.010

[39] Rion Snow, Brendan O'Connor, Daniel Jurafsky, and Andrew Y Ng. 2008. Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In *Proceedings of the conference on empirical methods in natural language processing*. Association for Computational Linguistics, 254–263.

[40] Jason Tanz. 2011. The Curse of Cow Clicker: How a Cheeky Satire Became a Videogame Hit. *Wired* 20, 1 (Dec. 2011). https://www.wired.com/2011/12/ff-cowclicker/

[41] Richard Thaler. 1980. Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization* 1, 1 (March 1980), 39–60. https://doi.org/10.1016/0167-2681(80)90051-7

[42] Kathleen Tuite. 2014. GWAPs: Games with a Problem. (2014), 7.

[43] Daniele Vannella, David Jurgens, Daniele Scarfini, Domenico Toscani, and Roberto Navigli. 2014. Validating and Extending Semantic Knowledge Bases using Video Games with a Purpose.. In *ACL (1)*. 1294–1304.

[44] Noortje Venhuizen, Kilian Evang, Valerio Basile, and Johan Bos. 2013. Gamification for word sense labeling. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)*.

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

[46] Luis von Ahn and Laura Dabbish. 2004. Labeling images with a computer game. In *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*. ACM Press, Vienna, Austria, 319–326. https://doi.org/10.1145/985692.985733

[47] Luis Von Ahn and Laura Dabbish. 2008. Designing games with a purpose. *Commun. ACM* 51, 8 (2008), 58–67.

[48] Luis von Ahn, Shiry Ginosar, Mihir Kedia, and Manuel Blum. 2007. Improving Image Search with PHETCH. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*. IEEE, Honolulu, HI, IV–1209–IV–1212. https://doi.org/10.1109/ICASSP.2007.367293

[49] Luis von Ahn, Ruoran Liu, and Manuel Blum. 2006. Peekaboom: a game for locating objects in images. In *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06*. ACM Press, Montréal, Québec, Canada, 55. https://doi.org/10.1145/1124772.1124782

[50] Cevat Yerli. 2008. GC 2008: Crysis Cost 22 Million to Make. https://www.ign.com/articles/2008/08/19/gc-2008-crysis-cost-22-million-to-make

[51] Amir Zeldes. 2017. The GUM corpus: creating multilayer resources in the classroom. *Language Resources and Evaluation* 51, 3 (Sept. 2017), 581–612. https://doi.org/10.1007/s10579-016-9343-x