“GG” Good Gaming Wordlists

Insight into single-player commercial off-the-shelf games and action-roleplay game vocabulary

By

Julian Sterling Heidt

A Thesis submitted to the faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master

In

Applied Linguistics and Discourse Studies

Carleton University

Ottawa, Ontario

© 2020

Julian Heidt

Abstract

The purpose of this study is to create two pedagogical wordlists for videogames to be used by ESL/EFL teachers and learners. Using a multimethod design, the present research utilizes both quantitative and qualitative measures. A keyness analysis using the Minimal Ratio index provided by the Keyworder software was conducted to investigate a game corpus' distinctive words. The corpus of ten videogame scripts from Rodgers and Heidt (in press) (5.7 million tokens) was compared to the SUBTLEXus TV/movie corpus (50.5 million tokens) to identify flemmas that are significantly more frequent in videogames. With Schmitt's (2019) call for a better understanding of the vocabulary found in games and game-genres in mind, two wordlists were created: a common game wordlist and an action-roleplay game (ARPG) wordlist to investigate variation in game vocabulary. To facilitate vocabulary acquisition, the listed words were coded for ludic/diegetic properties and patterns. Results regarding learning are discussed.

Acknowledgments

To my family and friends who have always been so supportive in my academic pursuits and have always been there willing to lend an ear when needed.

I wanted to sincerely thank both of my supervisors, Dr. Michael Rodgers, and Dr. Geoff Pinchbeck for all the support they have provided me throughout writing this thesis. Their insight and wisdom have been invaluable. I have learned so much by working with them, and I am in debt to them for their kindness, patience, and understanding.

Dr. Rodgers gave great insight into the inner workings of lexical frequency profiling. On top of his advice throughout my thesis, I was able to learn so much more through his special topic course on the subject of vocabulary. The content of his course helped with my understanding of the field which ultimately aided with writing this paper. Dr. Pinchbeck had also aided me greatly on how to conduct a keyness analysis, and his love for research on vocabulary had kept me motivated while conducting my research.

I also wanted to thank Jeremiah Bell, a Ph.D. student at Carleton University. We had lengthy discussions about the flow of the paper and his advice in general has been incredibly helpful.

And finally, I am very grateful to the department of the School of Linguistics and Language Studies, from my fellow peers, to staff, who have always been unconditional in their support.

Table of Contents

List of Tables

List of Appendices

Glossary

Commercial off-the-shelf (COTS) Game: Commercial games available to consumers from the point of sale whether that be in store or online.

Extramural Learning: Language learning that takes place on behalf of the learner outside of the classroom without any involvement from the learner's curriculum or teacher. It is an extramural activity that was taken up which has input in the L2.

Flemma: A base word counting measure. The Flemma is a hybrid between the *lemma* and the *word-family*. Flemmas are untagged parts of speech (i.e. not explicitly stated as noun, verb, or ajective) that share lemma inflections; refer to lemma, and word-family.

Game-Enhanced Learning: The use of what Sykes and Reinhardt (2013) refer to as the use of vernacular games (common games played) in the classroom.

High-Frequency Vocabulary: Vocabulary that occurs very frequently, common words.

Incidental Learning: Language learning that takes place without deliberate effort.

Intentional Learning: Language learning that is intentional is where there is a deliberate effort.

Involvement Load Hypothesis: Laufer and Hulstijn's hypothesis that states the more cognitively difficult an input is, the likelier acquisition will occur.

Keyness: An analysis conducted to identify a corpus' keyness: the distinctive words that are within the study corpus when being compared to the reference corpus.

Lemma: A base word counting measure that is organized by inflectional suffixes (example given on p.40 in detail).

Lexical Coverage: The amount of words necessary to understand a text.

Lexical Frequency Profile (LFP): A profile that demonstrates lexical coverage of a text. Used to report for adequate (95%) and ideal (98%) comprehension of a text to demonstrate the lexical difficulty and the level of student to be targeted by a text by using wordlists, such as the BNC-COCA 20. Example of software that can do this: AntWordProfiler and LexTutor.

Low-Frequency Vocabulary: More difficult and less used words.

Marginal Words and Proper Nouns (MWPN): A metric in LFP studies to refer to words that have meaning that is easily inferable from context, such as swears, and names of people, places, and things.

Massively Multiplayer Online Game (MMOG): An online video game that can be played often worldwide and simultaneously with many other players.

Massively Multiplayer Roleplaying Game (MMORPG). A roleplaying game that can be played often worldwide and simultaneously with many other players.

Multiplayer Game: A game that can be played together simultaneously. Can be cooperative (2-4) players or have more online.

Negotiation of meaning: Part of Michael Long's interactional hypothesis. Language learning occurs due to a linguistic gap that must be detoured through this construct known as negotiation of meaning. When a linguistic gap is present, the way that the interlocutors negotiate meaning is when learning of the target language occurs.

Off-list: A metric from LFP studies that is used to count for words that are not caught by wordlists and cannot be marked as a noun, marginal word, compound noun, or abbreviation.

Playthrough: A playthrough is the recording of a player play the game's story from start to finish that is usually uploaded on a video streaming platform such as YouTube or Twitch either synchronously or asynchronously**.**

Reference Corpus: The corpus that the study corpus is being compared to which determines the study corpus' "key" words from a keyness analysis.

Rheme: What is being said about a topic. He (theme) is a good student (rheme).

Roleplaying Game: Players play characters in a fictional setting and are put into narrative where they must make decisions.

Single-Player Game: Games meant for just one player.

Statistic (Stats): Often in roleplaying games, it is data that reflects an attribute of the character. For example, STR is strength. Depending on how high the STR stat is, could determine how much a character may be able to hold in their inventory or how much damage they do to enemies.

Study Corpus: The corpus being studied for keyness; the target corpus

Theme: The subject of a sentence; refer to Rheme.

Willingness to Communicate: The desire to communicate unprovoked.

Word Family: A base word counting measure that includes both inflectional and derivational forms, as per level 6 of Bauer and Nation (1993)

Chapter 1: Introduction

1.1 Current Landscape

Playing videogames has become a common form of recreation in Canada. The average Canadian spends 800 hours a year playing videogames in their free time (Statistics Canada, 2010) while spending only 300 hours on reading ("*NOP World Culture*", 2005). Playing videogames is not just popular among teens, but rather people of all ages play them (*"Essential Facts"*, 2016; Gee, 2007). According to the Entertainment Software Association of Canada (ESAC) (2018), 61% of Canadians define themselves as "gamers" with Canada having 23 million gamers overall. This number may in fact be larger than reported since not every game-enthusiast subscribes to the label of "gamer" due to the pejorative stereotypes that follow it, such as gamers are "socially awkward", "spend too much time playing games", "overweight", and have "negative life consequences" due to gaming (Bergstrom et al., 2014).

There seems to be this notion set out by those who game in their free-time, that if you do not sleep, breathe, and play games 24/7, that you are not a gamer (Juul, 2010). Rather, it is the enjoyment of games and seeing it as a medium for entertainment which is arguably more important. 80% of Canadians view videogames as mainstream entertainment similar to movies, and the average age of the typical Canadian gamer is 39 years. Keeping in mind how much time Canadians spend gaming, and how 64% of Canadians own a gaming console, and 89% own computers ("*The Canadian video game industry"*, 2018), it is apparent that researching gaming as an extramural activity to acquire vocabulary from is warranted due to the massive time commitment that Canadians undertake to play these games. As well, due to the recent *COVID19* pandemic, extramural learning and gaming is likely more frequent, and it is important to discover

how Canadians can harness gaming as an extramural tool to supplement classroom input as a result.

In 2016, the ESAC compiled their essential facts report that highlighted key gaming demographics based on game-genre. According to survey data on 270 teens aged 13-17, male gamers preferred shooting games (48%), action-adventure (44%), and fighting games (30%), whereas female gamers preferred puzzle games (42%), action-adventure games (29%), and social games (28%). Game-genre preference differed upon age and gender; however, it is important to consider that genre is difficult to attribute to games, as they can have multiple. For example, the game *Minecraft*, has three genres: *Sandbox*, *Action-Adventure*, and *Survival*. This is important to consider when discussing game demographics, as it may not always be clear where a game falls regarding the categorization of genre.

Considering how prevalent games are in Canadian society as a recreational activity, a review of the implications that videogames may have on vocabulary gain as a supplementary language input is therefore warranted, as playing commercial off-the-shelf (COTS) videogames, outside of the classroom, has become mainstream and applies to many, regardless of age. Since games are so widely played and are a source of motivating input for learners, it would be interesting to see their use both inside the classroom and as an extramural activity. Creating a bridge between learners' passion for games in their first language (L1) and the classroom could lead to exposure to meaningful language that may otherwise be a barrier to entry in the second language (L2).

To this day, there has not been a clear-cut answer to whether COTS videogames are beneficial in acquiring vocabulary. Videogames have been linked to proficient language usage in additional language learning (De Wilde et al., 2019; Sundqvist and Wikström, 2015; Sylvén &

Sunqvist, 2012) as well as being a highly motivating medium for language learners to acquire language from due to its perceived authenticity as a source of language; however, the vocabulary inherent in games is still unknown to many researchers. Recently, Schmitt (2019) in his vocabulary research agenda, created a call for investigations into the vocabulary that games contain. There has yet to be a comprehensive scientific inquiry into the vocabulary from games, and whether different game-genres affect the variety and type of words found in games. Recently, Rodgers and Heidt (in press) reported the lexical coverage thresholds for games, and the number of words required to reach adequate comprehension was found to be higher than other media, such as books and movies. As a result, creating gaming wordlists would then be ideal to create to provide educators insight into the most "key" characteristic vocabulary from these games. These words can then be used in the classroom by teachers and/or used to prime vocabulary for learners to use on a supplementary basis outside the classroom.

1.2 Gaming Vocabulary Affordances

When using the term 'affordance', I refer to the construct defined by Norman (1988): "the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used*"* (p. 9). The example he gives is that a chair affords support, and therefore affords sitting; in other words, affordances are what allow certain actions to occur, and that it is through something's specific properties that allows you to accomplish something. Regarding gaming, vocabulary from games may share unique affordances from game-design (Reinhardt, 2018), such as the use of *ludic* and *diegetic* words. Diegetic words are described as words used for dialogue and narration; they are part of the narrative experience (Castelvecchi, 2020), whereas ludic words are described as words that serve

an overall function where language is "manipulated" for play (Crystal, 2001). As an example, in gaming, if the word *melee* was used in this context: 'this item increases melee damage', melee would be then serving a ludic function; it is identifying that melee is a type of damage that the player of the game can perform. If it were part of dialogue or narration, such as 'there is a melee going on' that would be an example of diegesis; it does not refer to melee damage or a game element, but rather is a part of the dialogue which is attempting to create a setting. There are barriers to entry concerning acquiring language from video games which are important to keep in mind. If the learner is cognitively overloaded by the task they are performing, they are unlikely to have acquired new words (cf. Sundqvist & Wikstrom, 2015; cf. Hitosugi et al., 2014; cf. Chen & Yang, 2013; cf. deHaan, Reed, & Kuwada, 2010). Another barrier, is the player's familiarity with different game-genres based on the player's gender (cf. Sundqvist & Sylvén, 2014; cf. Hitosugi et al., 2014), or even playing games at all based on the player's gender (cf. Sundqvist, 2011). While videogames encompass a huge audience, this does not mean that one specific game is applicable to all learners; learners will have different preferences for games which may be grounded on their gender, age, and idiosyncrasies. Finally, perhaps the most obvious barrier, is simply not having what is known as "game literacy", the ability to play and learn from games (Gee, 2007). Based on prior research conducted by Rodgers and Heidt (in press) and these potential linguistic difficulties from the literature, I thought that identifying ludic words to target for teaching would be ideal to alleviate these issues. Depending on whether the educator is teaching ludic or diegetic words may shift the way the word should be approached in a teaching context.

1.3 Organization of Thesis

This thesis is organized into six chapters. Chapter 2 provides an in-depth literature review of how games have been used in language and vocabulary studies and what trends could be going forward through different video game affordances. Chapter 3 informs of the study's multimethod methodology, how it addresses the different research questions posed to different methods and lists the three procedures (phases) done to answer each research question. Chapter 4 lists the results of the present research. Finally, chapter 5 is on the research implications for extramural learning and teaching, the future research possibilities informed by gaps in the literature, as well as informs the reader of the limitations of the present study. Chapter 6 is the conclusion to the thesis, tying everything together.

Chapter 2: Literature Review

This paper is addressing the call put out by Schmitt (2019): There is a need to understand the "specialist" vocabulary in games [words indicative of games] and "Do different games [genre of games] promote different types of vocabulary?" (p. 268). In order to understand how vocabulary is realized through language functions, identification of words used in a ludic or diegetic fashion and providing them in context is necessary to provide language teachers understanding of how game-design comes into play with language-use in games. Through the conducted research, two comprehensive pedagogical gaming wordlists were compiled: 1) a common gaming wordlist for all ten games included in the study corpus, and 2) an action-roleplay game wordlist using four of the 10 games. Both wordlists can be used inside the classroom and outside the classroom to aid with the incidental or extramural use of gaming.

2.1 Gaming, Language Learning, and Motivation

Motivation is an essential component for successful language acquisition (Williams & Burden, 1997), and is the most important factor concerning the acquisition of the target language according to Krashen's affective-filter hypothesis (Krashen, 1981). If a student's affective filter is high, learners are not receptive to learning. If the filter is low, students are motivated and encouraged to learn because they feel comfortable. Reinders and Wattana (2015) reaffirm Krashen's theory in gaming based on qualitative research data; they concluded that videogames may be beneficial in additional language learning by increasing students' *willingness to communicate* (WTC). The study took place with 30 Thai students enrolled in a 15-week

university language course. After playing Ragnarok Online, a Massively Multiplayer Online (MMO) Role Playing Game (RPG) (see Appendix A for types of multiplayer games) regularly in the classroom, five students were randomly selected and asked questions in semi-structured interviews focusing on their WTC in the game. WTC, as operationalized by Reinders and Wattana, is "an individual's intention to initiate or participate in communication in English at a particular moment in a particular situation" (2015, p. 42). Four of the five students felt more engaged and demonstrated a higher WTC in the game and claimed that they had used English more because the setting had felt relaxed (Reinders & Wattana, 2015).

Furthermore, there is a link between motivated learners and gaming. In Sweden, Sundqvist and Sylvén (2014) asked students to agree or disagree with the following statement "English is interesting". Among 86 students in grades 4 – 6 in Sweden, the learners in the frequent gamer group (average of 6.6 hours/week) either 'agreed' or 'agreed strongly', whereas the largest portion of disagreement (19%) originated in the non-gamers group. The link between gaming and motivation is important, as foreign-language learning has been well-documented as a context where students lack motivating input (Dörnyei, 1994). Videogames can potentially be an answer to that dilemma by creating access to more motivating language input for additional languages. If it is the case where videogames promote motivation, then making foreign language videogames more available for language-learning consumers could facilitate language acquisition with a less deliberate effort made by the learner. Ideally, videogames should promote motivation and have pedagogical validity to support its use, both, in and outside the classroom.

As mentioned earlier, motivation is often cited to promote language acquisition; however, the efficacy of motivation is not unique to games, but other media as well. It may be hard to ascertain the benefit of studying any input's merit if it is only functioning well because of

motivation; one could wrongly assume that motivation is enough credence to use any type of input for language learning. There must be a certain threshold of comprehension, enough text to make inferences from, and ideally, good pedagogy surrounding the input. Playing videogames has been correlated with increased English proficiency scores (Sundqvist & Wikström, 2015; Sylvén & Sunqvist, 2012), but the present study is concerned with the direct relationship between the language used in games and the potential for comprehension through the developed wordlists. Gaming can be used to teach vocabulary through "spatial and temporal elements that can be adapted for use in foreign language learning and teaching" (Sykes & Reinhardt, 2013, p. 25) for intentional learning, as well as extramural learning where learners can "invest time in gaming for incidental vocabulary learning to happen" (Sundqvist, 2019, p. 104).

The cognitive complexity of games

It has been theorized by Gee (2005) that videogames are cognitively complex. Some instances where games are complex are when a) there is a requirement to communicate to succeed such as the game-genre of MMO role-playing games (Gee, 2007), b) when there is a puzzle that is blocking the player from advancing, and c) *fail states*—instances wherein the videogame penalizes the player and as a result, they learn to not perform the same erroneous action again (Sykes & Reinhardt, 2013). For example, if two players are playing a tactical game together and their initial strategy does not work due to unexpected enemy behaviour, they may learn from losing the battle, and attempt to counteract that enemy by playing differently the second time around. Different types of games have different elements that may aid in the acquisition of certain parts of the language. Mentioned earlier was the MMO game-genre which is highly focused on communicative competency and may be more related to *negotiation of meaning* or *pragmatic competence*. In interactive-movie based games, such as the ones

developed by *Telltale*, a player can communicate with an object more than once which may facilitate vocabulary acquisition through repeated exposure and imagery. Imagery has been theorized to play a role in how language is used in movies (Rodgers, 2018) and has been documented to be an effective way for learners to acquire target vocabulary (Al-Seghayer, 2001). In other words, the affordances that different videogames provide through play and game-design can be more conducive to acquiring vocabulary than traditional sources of foreign language input which makes it an interesting medium to investigate.

If games are cognitively complex, then it follows that they have the potential to be beneficial in acquiring vocabulary according to Laufer and Hulstijn's (2001) *Involvement Load Hypothesis,* which states that the more cognitively demanding a language input is (within reason), the more likely learners will acquire vocabulary from the input. Due to their cognitive complexity, videogames could offer an ideal scenario to learn vocabulary from. Games add a dimension that requires a certain amount of thinking. If a game is too easy, likely players would grow tired of it relatively fast. The same idea applies if the difficulty is too high. This is important to consider, especially when relating game-playing to *Flow* theory. Flow is an optimal state of being that refers to a person who goes through a "flow experience" which results in feeling more "capable and skilled" (Csikszentmihalyi, 1990). It is a positive state that is conducive to learning something new: "People who achieve flow more regularly pay close attention to the minute details of their environment, discover hidden opportunities for action, set goals, monitor progress using feedback and keep setting bigger challenges for themselves" (Csikszentmihalyi, 1990, p.3). Game players are likely reaching an optimal experience (i.e. flow) while playing, as they are taking everything into consideration to progress to the next stage of the

game. This is a state where focusing and learning occurs rapidly and is why players can finish games so quickly. This is one of the reasons why fail states are so important in understanding game-design and flow regarding gaming, as these states, which can be seen as blockades, are in fact motivators that are encouraging players to continue to play. For this reason, due to the cognitive complexity which affords flow, games can be used as a great tool for learning and acquisition to take place.

Lack of extended playing in experimental designs leads to uncertain research results

Theoretically, games provide rich input; however, there may be a certain threshold of time required to be spent before any meaningful acquisition can take place. Cobb and Horst's (2011) study is a cautionary tale that vocabulary acquisition may only be achieved in videogames by extended playing. Playing a game for only an hour or two in a sitting for vocabulary gain may be insufficient. While *Word Coach* is a pedagogical COTS game and can be likened to comparing apples to oranges at least in all intensive purposes for comparing to the games studied in this thesis, but this study was chosen as an example that any kind of input requires a significant amount of exposure to acquire vocabulary from regardless of game format. In Cobb and Horst's study, players had been exposed to the studied game for two hours per day which totalled 120 hours over 2 months. The game used in the study, Word Coach, published by *Nintendo*, was a word-based puzzle game. Similar to the idea of *extended reading* (Mason & Krashen, 1997), *extended playing* is the idea that there is a minimum amount of time required before there is any gain in the target language. Extended exposure is also supported in the literature to encourage vocabulary gains while playing videogames (Sundqivst & Wikstrom, 2015; Sylvén & Sundqvist, 2012). An issue with the game in this study is that it was a puzzle

game where naturalistic language was not present. Non-puzzle-based games may have a task dimension: to understand unknown words, inferences must be made. To make inferences a certain amount of language is then necessary, which the game in this study did not have. The Word Coach game is not a suitable game to promote naturalistic language acquisition due to lacking the element of "language-in-use".

Regarding extended exposure, a similar case occurred when Chen and Yang (2013) conducted two studies on students playing an interactive-movie videogame called "*BONE*" which contributed to small vocabulary gains in the L2. Their first study involved a two-hour intervention where 22 college freshmen learners in Taiwan had a vocabulary pre-test and post-test before and after playing the game. In study one, there were two groups of students: players and notetakers. Notetakers simply watched the players and took notes but received the same level of vocabulary gain as the players of the game according to the study. While there were vocabulary gains for both groups, there was no significant difference between them ($t(20) = -.99$, $p = .746$, 95% [-.134, .97]). Chen and Yang (2013) speculated that the experiment only being two hours was insufficient for there to be significant vocabulary gain. They argued that more exposure to the input would then be required, therefore adding weight to Cobb and Horst's (2011) caution, as well as Sylvén's, Sundqvist's, and Wikstrom's reaffirmation for extended exposure to input: "vocabulary learning is a by-product of another learning activity", "such as extensive reading" (Sylvén & Sundqvist, 2012, p. 306) and "a great deal of time probably needs to be investigated in gameplay if L2 words are to be stored in long term memory" (Sundqivst & Wikstrom, 2015, p. 67). The second part of Chen and Yang's study had students play the game in their free time. However, due to the students playing the game and having taken an English course concurrently where multiple language inputs were available along with other complex

social interactions found in the classroom, they did not investigate vocabulary gains for the second study.

Chen and Yang (2013) reasoned that it would be impossible to ensure that the gains were from the videogame, as opposed to other input available in their everyday school lives. How much learning that could take place in such a small amount of time is not just a concern for gaming, but for the classroom as well. Gaming and learning together must follow set criteria for them to be advantageous, and that is what is inferable from these studies; more time needs to be given to promote the acquisition of language. Simply using games through new technology does not increase pedagogical value; rather, it is their implementation, in other words, how they are used to promote learning that is important (Godwin-Jones, 2014). Good pedagogy will always take precedence and be at the foreground of any vocabulary acquisition that takes place.

Proposed research designs based on the literature on vocabulary gain and gaming

Overall, there appears to be a positive relationship with videogame playing and language proficiency, as noted by Sundqvist and Wikström (2015) who investigated the videogame habits of 80 teenagers in Sweden outside the classroom. The study revealed that learners who played games over five hours a week scored higher in their English class with a medium effect size. The main issue with Sundqvist and Wikström's study is that it is correlative and faces the chicken and the egg dilemma: do students who play videogames have a better vocabulary because they use videogames as an input, or is it because their language proficiency is already high that they play videogames? Perhaps a better way to approach the research issue of gaming and pedagogy would be to conduct more longitudinal studies to investigate the threshold for these vocabulary gains, as is mentioned by Sundqvist, who called for more research for what learners "do outside of school and achieve in school", and informed that "it would be interesting to map out

extramural English for young learners and study how it develops over time for them" (Sundqvist, 2011, p. 117).

The recurring theme of this literature review is that it is difficult to find the source that videogame *play* can have on learners with a purely experimental design. This is due to the many social conditions that may contribute to vocabulary gain outside of playing the game and not in the input inherently itself. Indeed, it may only be possible to infer how much potential impact videogames may have on language learners by three types of studies:

1) Performing a form of lexical analysis, such as a Lexical Frequency Profile (Rodgers & Heidt, in press).
2) Investigate the use of gaming affordances through *play* in acquiring the target language (e.g. communication/speech acts in MMO games, exposure to recurring vocabulary in visual novel games, game language relating to problem-solving, etc.)
3) A corpus analysis to investigate the vocabulary that might be available to language-learning gamers. How does it compare to other types of language input texts (For example, fiction, textbooks, TV, and/or movies)

As # 1 has already been conducted, and # 2 is beyond the scope of this study, I will be addressing solely the third area of study to investigate the potential language input that is available from games to answer in which contexts games should be used in to promote vocabulary acquisition on a purely lexical basis.

2.2 Lexical Analyses

Vocabulary acquisition is arguably one of the most important aspects of language learning, as it is both the foundation of and a key piece of evidence for proficient and advanced language usage (O'Keeffe, 2012). Many studies have investigated various kinds of input regarding vocabulary acquisition and its authenticity through conducting *Lexical Frequency Profiles* (LFP) of texts such as books (Nation, 2006), songs (Tegge, 2017), television (Rodgers & Webb, 2011), and movies (Webb & Rodgers, 2009b). Conducting an LFP indicates how much a learner might understand with and without help in a given language input (Hu & Nation, 2000). Only very recently has there been an inquiry into the lexical profile of videogames (Rodgers & Heidt, in press). The intention of conducting an LFP for a language input is to separate possible confounding variables that may not play a causative role in vocabulary acquisition such as motivation.

Lexical Frequency Profiles

LFPs are needed to determine if an input is worth playing for learners by reporting how many words a student would need to know to reach adequate and ideal comprehension. In other words, LFPs can be used to infer whether an input text would be too easy or difficult based on a learner's vocabulary knowledge through their proficiency level. For example, an intermediate learner would likely know approximately 3000-word families, as inferred from Webb & Rodgers' (2009a) results for learners reaching 95% coverage at 3,000 word families in television and the authors stating that "[increased learning] is more likely to occur in an ESL setting with intermediate and advanced learners, who have a large enough vocabulary to reach 95%

[coverage]" (Webb & Rodgers, 2009a, p. 358). In essence, LFPs calculate the percentage of words that are likely to be understood by a learner in a text, which is known as *lexical coverage* (Nation, 2006). Nation (2006) identified that for a learner to be able to fully understand a text without guidance, the learner would need around 98% coverage of an entire text (i.e., 1 in every 50 words unknown), but with guidance, they might only need to know 95% of an entire text (i.e., 1 in every 20 words unknown).

LFP studies report how many word families are required to reach 95 and 98% comprehension from different types of language input. Studies often report both 95% and 98% while including both proper nouns (PN) and marginal words (MW) in their total word counts. Marginal words include interjections, such as words that resemble sounds such as "ahh", "grr", "yargh", and other words that do not contribute substantial meaning. The rationale for calculating PN/MW as known-words is that their meanings are easily inferable by context. Studies often show lexical coverage to both include and exclude PN/MW. Furthermore, LFPs show a text's words broken into different frequency bands known as wordlists. Wordlists are the groupings of words that list high frequency (used often) medium-frequency, and low-frequency vocabulary (words which are rarer and not used as often). Band 1000 is the most frequent 1,000 words in English, 2,000 is the following 1,000 most frequent, etcetera.

Rodgers and Heidt (in press) created an LFP for games which revealed the number of words necessary to be known for adequate and ideal comprehension (see Table 1 below); however, while this research is informative, it has not answered the question: what kind of vocabulary is used in videogames? Schmitt (2019) commented that little is currently known about the vocabulary that occurs in videogames, including the types of vocabulary that occur in

games and the frequency of occurrence of potentially learnable vocabulary. Ultimately, addressing this call is one of the goals listed in the present study.

According to Rodgers and Heidt (in press), videogames were reported to reach 95% coverage for games at the 5K band, and 98% at the 10K band. In Nation's study, by calculating the average for reading, 95% was reached at the 4K band and 98% at the 9K band. Books are the second most difficult lexically according to the literature, before games. Tegge (2017) investigated a corpus of songs that shows reaching 95 % coverage at the 3K band and 98% at the 8K band. Similarly, 95% was reached at 3K in both movies and television (Webb & Rodgers, 2013; Rodgers & Web, 2011), but movies and television differed on 98% coverage: 98% at the 8K band for TV, and 7K band for movies. See Table 1 for a summary of LFP studies conducted, and Table 2 for the coverage results of these studies.

Table 1: LFPs investigated

LFP: Medium Investigated	Study
Reading	Nation (2006)
Songs	Tegge (2017)
Movies	Webb & Rodgers (2013)
Television	Rodgers & Webb (2011)
Games	Rodgers & Heidt (in press)

Table 2: Coverage for different media reported by LFP studies

Frequency Bands	Reading 2006	Songs 2017	Movies 2013	TV 2011	Games In press
Tokens	121,099	180,892	2,841,887	1,330,268	5,744,388
1,000	82.93	88.96	89.76	89.16	83.64
2,000	90.14	93.18	94.02	93.56	90.19
3,000	93.28	**95.13[a]**	**95.83[a]**	**95.45[a]**	93.01
4,000	**95.06[a]**	96.32	96.89	96.51	94.79
5,000	96.13	97.15	97.52	97.21	**96.01[a]**
6,000	96.88	97.42	97.96	97.64	96.75
7,000	97.43	97.79	**98.22[b]**	97.91	97.23
8,000	97.90	**97.99[b]**	98.45	**98.17[b]**	97.69
9,000	**98.22[b]**	98.09	98.62	98.35	97.96
10,000	98.46	98.17	98.78	98.51	**98.16[b]**

**Coverage figures include PNs and MWs
** [a] reaching 95% coverage; [b] reaching 98% coverage

Artifact of methodology

These numbers inform that videogames contain more low-frequency vocabulary than that in other types of media, suggesting that videogames are more linguistically difficult, although this result might be an artifact of the study's methodology. Rodgers and Heidt (in press) included all language found in videogames in their LFP study. "All language" encompasses dialogue and narration as well as any language the player could experience in-game, such as the names of game items, the text of books available to read in-game, different types of vehicle modifications, and more. Any possible exposure to any form of language, including what Rodgers and Heidt referred to as *metalanguage* was present in their data. Metalanguage is the term used in the study to refer to instructional videogame language; it would include controller mappings (buttons),

tutorials, and prominent words and phrases found in games used for game functions. Upon realizing that metalanguage could be contributing to more words required for a learner to reach Nation's definition of adequate comprehension, an analysis of a *playthrough* of the game *GTA V* was warranted. A *playthrough* is the recording of a gamer who plays the game's story from start to finish, and it is usually uploaded on a video streaming platform such as *YouTube* or *Twitch* either synchronously or asynchronously.

By using a gamer's playthrough from a video platform for analysis, most of what remains are words from narration and story dialogue, which is a small amount of actual language from the game based on the sheer volume of tokens that were from extracted from these various games for this corpus. Diegesis from side-content in the games, such as books or extra readable materials (such as an in-game phone) that can be found and read in-game, likely contributed to more words that a player may have not otherwise experienced, which is an artifact of the methodology. To understand whether if it was the study's methodology of captivating language that contributed to such high lexical thresholds, an analysis of a playthrough of GTA V was performed. The GTA V playthrough showed that knowledge of the first 3,000 word families of English provided 95.74% coverage and knowledge of the first 8,000 word families provided 98.25% coverage which is in line with the previous research on television (Webb & Rodgers, 2009a) and movies (Webb & Rodgers, 2009b) that found 95% lexical coverage with knowledge of 3,000 word families.

This could be due to the way the LFP was constructed and categorized. For example, all the words from conlangs in the game *Skyrim* were labelled as "off-list" words, which are sorted as "unknown" in the software. Unknown words contribute to lower coverage because they are not captured by the wordlists used. In *Grand Theft Auto V*, there were low-frequency words

regarding vehicle modification in certain segments of the game that are behind menus that otherwise would not be seen or heard during the story mode of the game, but rather as optional-language that could be viewed when making adjustments to a vehicle (For example, the "off-word" *Splitter,* which is a bumper modification for cars). As mentioned earlier, some games had books or files that could be read, such as the games *Skyrim* and *Resident Evil 6*. How many players would read these texts is unknown. These texts are examples of optional language input from side-content, diegesis, that the player may not necessarily be exposed to and often included lower-frequency words than the actual narration found in the game. The methodology behind constructing the corpus is likely what created this difficult lexical threshold; however, the perspective of a lexical frequency profile for all language found from games is still an interesting point for analysis that may give a broader picture than just dialogue. This study aims to investigate both ludic and diegetic words. By using this corpus, which includes all language from games, the idea is to provide a better possibility for analyzing game words, as all language from the game is encoded and not just narration. Through analyzing the videogame corpus, the extent to which ludic and diegetic words are prevalent in these 10 games and whether they are learnable, and how they might contribute to overall game comprehension can be explored.

2.3 Corpus Research

From the literature reviewed, there does not seem to be a comprehensive corpus that has captured video game language. There has been a call for making a corpus of gamers' language while playing online games together to develop pedagogical strategies (Bawa, 2018), but not so far, the videogames themselves. As a result, there have not been any studies to draw a direct comparison to the present research. There has been research suggesting how to develop a

methodology for corpus-based game studies (Masso, 2009), and other studies that have used keyness for the analysis of specific vocabulary, such as key words from movie reviews and how those are distributed by genre (Castagno et al, 2018).

How would Massively Multiplayer Online Games (MMOGs) fit in researching game vocabulary through corpus methods?

Bawa (2018) encouraged the development of an MMO corpus for its educational value for a cross-discipline of fields, one of them being linguistics. Ryu (2013) also argued that there is a need to understand game culture and that this could be fulfilled by attending game sessions and identifying how learners engage in play through their community. Creating an MMO game corpus would provide insight into game-culture and also shed light on the second research study posed as outside the scope of this paper: "*Investigate the use of gaming affordances through play in acquiring the target language".* Thorne (2008) used the MMOG game *Word of Warcraft* to capture language learning exchanges between American and Ukrainian university students to show the social interactions that are conducive to learning through task-based reasoning. MMOGs are diverse and encompass mixed language groups, as not everyone playing together have English as their first language. Through investigating and tagging speech acts and fail states through a corpus, learner methods for acquiring target vocabulary can be idealized through negotiation of meaning. Developing an MMOG game corpus could also be fruitful for analysis to determine the vocabulary that is frequent in player communications (Bawa, 2018) which could be idealized through keyness and frequency approaches. It is especially worth-while to consider since MMO roleplaying games have an "end-game" where players focus on instances known as "raids" that are well documented to be difficult and require a lot of coordination and

communication from players (Ducheneaut & Moore, 2005; Ducheneaut, Yee, Nickell, & Moore, 2007).

It is important to not only investigate the language found in games, but the language that the players themselves use together while playing these games, which is an affordance that MMOs have over single-player games and normal multiplayer games. This would determine how sophisticated a vocabulary would be required to play these highly communicative games, as well as provide valuable insight into vocabulary acquisition taking place mid-game through gaming affordances and mechanics unique to the game-genre of MMO games.

The proposed methodology to build a game corpus

Masso (2009) suggested that to create any kind of videogame corpus, the researcher should play various games all the while capturing recordings of their sessions to later transcribe; however, this becomes difficult depending on the type of game and the researcher's ability to follow through such a massive time commitment. The main issue with such an approach is that there is a certain assumption that the average player will be playing in a manner similar to the researcher. If there are many diverging paths in the story, or if there are certain events that occur outside of the main game's story, these events may not be captured and are not providing a full picture of the target language. For example, a game such as *Skyrim* contains many, many, quests that are not required to finish the game, so how would this missing language be retrieved? There are strengths and drawbacks to developing a methodology based on playthroughs. To create wordlists indicative of gaming language, the justification was made to view all language present in a game to understand a word's use across the entire game, as opposed to a portion of it. If the goal was to understand a game's narration, then using a playthrough would be ideal, as Masso had originally recommended.

Reporting words' keyness results based on corpora

Chung and Nation (2004) conducted a study to demonstrate four different methods to create a technical wordlist using anatomy as the subject. They had four methods to creating technical words tried to determine their effectiveness: 1) a rating scale, 2) clues (diagrams of words), 3) a dictionary, and 4) using computer analysis (comparative corpus). For their computer analysis approach, the study text was taken from *Clinically Oriented Anatomy*, and their reference was the LOB corpus. They recommend that researchers use a similar cut-off of their proposed 1:50 ratio; in other words, the word must occur 50 times more in the studied text than the reference text for a distinctive technical word. For a technical wordlist, perhaps such a high cut-off is sensible; however, when comparing games to movies to identify game-specific vocabulary, to expect such a contrastive difference may not be as important. There are problems with considering such a high cut-off.

First, 1:50 would assume that the words are not being compared to something similar in nature (e.g. movies and games, language used with imagery in mind), and second, there would be an underlying assumption that each game in the total game corpus contains a very similar diction, as opposed to varying vocabulary due to game-genre, or words unique to that specific game (i.e. a skew in the data). Chung and Nation also mention that the reference corpus should not contain any text from "the specialised area being investigated" (p. 259) which is unclear. There is likely a lot of overlap between movies and games which is the reason why the SUBTLEXus movie corpus is used as a reference corpus in this study in order to separate that boundary and investigate specifically game words, and not words attributed to media with imagery in mind when comparing books to games, for example. Chung and Nation also caution that: "the most frequent words from a specialised [studied] corpus are not all true terms but

include many general words used across a wide range of subjects*"* (Nation, 2001, as cited in Chung & Nation, p. 259). Inclusion of many general words could be due to a reference corpus being too far isolated from what is being attempted to be investigated. For example, comparing a different corpus on medical information to the anatomy corpus may have shown more language indicative of physical properties and characteristics when considering for variation. Chung and Nation (2004) mention that once a corpus reaches the 2-million mark, fear of it being specialized is no longer a concern since it is more reflective of general language at such a volume of words. Both the game corpus in Rodgers and Heidt (in press) and SUBTLEX from Brysbaert and New (2009) are large corpora (5.7 million tokens, and 50.5 million tokens, respectively), while these may not reflect "technical words", the wordlists being made in this paper are not argued to be "technical" words, but rather "generalized" game words. As opposed to measuring game words' worth by their comparative frequency to the movie corpus, *minimal ratio* in tandem with *coverage* will determine how effective these words are conducive to reaching obtainable milestones for comprehension. If the wordlists can reach coverage levels similar to Coxhead's Academic Wordlist, then, game vocabulary should be considered a valid construct.

Keyness and "Genre"

Castagno et al (2018) investigated the key words derived from movie reviews of action, fantasy, and romance genres to determine their variety of vocabulary for designing classrooms lessons. Their second research question is of pertinent interest regarding the design of this paper, they asked: what are the keywords that distinguish each genre? For this study, they collected 79 real-life movie reviews from the website *Rotten Tomatoes.* In their analysis, they had set up reviews of each movie as a sub-corpus; for each sub-corpus, they used a combination of the other two sub corpora as a reference corpus to identify the words which make that genre unique. They

coded for connotation; red signified “positive meanings” and blue for “negative”, while black was connotatively neutral. Then through the results, they designed pedagogical activities for learners using movie reviews. I accomplished a near-identical analysis to view how words were used in context. The difference being, the coding done in this thesis was not for connotation, but for ludic and diegetic properties.

2.4 Extramural and Classroom Learning

Extramural Learning

It is likely impossible to acquire a new language by solely relying on exposure from the classroom. Research suggests that the average first-language (L1) speaker learns seven new words (lemmas) a day throughout their life to reach native-mastery of their L1 by the time they are 18 (Brysbaert, Stevens, Mandera, & Keuleers, 2016). It is therefore ideal that learners try outside of the classroom to improve their second language (L2) vocabulary so that they can attain higher proficiency and reach realistic linguistic milestones faster. Since gaming is considered a motivating and possibly an ideal source of language to learn English from, with research revealing that there is a relationship between gaming outside of the classroom and English language proficiency scores, the investigation of using games extramurally to promote English is warranted. Many studies that have investigated extramural learning concerning gaming have homed in on mostly younger learners.

Sundqvist (2019) recently studied the link between teenagers' time spent on extramural COTS gaming and vocabulary knowledge in Sweden. The study investigated 1,324 students in 61 classes. The Productive Levels Test (PLT) and Vocabulary Levels Test (VLT) were used to measure three variables concerning English proficiency and gaming outside the classroom: gender, type of game played, and hours spent gaming. Students who spent more time gaming scored significantly higher in both productive and receptive vocabulary and excelled at more infrequent words (such as *oath*, *vault*, *ledge*, *cavalry*, and *mature* - 5K band). Through a multiple regression analysis, it was revealed that *time spent gaming* was the most considerable variable for vocabulary learning. The study had four groups: 1) non-gamers, 2) single-player gamers, 3) multiplayer gamers, and 4) MMO players.

An interesting finding reported by the study is a medium-effect size for game type: players who had a preference for playing multiplayer or MMO games had reported spending more time on gaming compared to players who enjoyed single-player games ($\chi 2 = 1078.742$, $df = 9$, $p < .001$, $\varphi c = .626$). While the study itself did not make any strong assumptions as to why this might be, it could be due to the communicative orientation of multiplayer and MMO games. Gamers may enjoy playing with their friends, and these games provide them that affordance which may be more motivating than otherwise playing alone. It is also not gaming specifically that may have contributed to these scores, but rather, simply exposure to an extramural input, as exposure to varying extramural inputs aid with English (Sundqvist & Sylvén, 2016). Since some learners in the non-gamer group were reported to having similar proficiency to the advanced gamer group, it can be reasoned that it was not games specifically, but rather extramural English through games that aided with these learners' scores; however, if games are easier to promote

motivation and aid with linguistic growth, as seen in this study, then the distinction of extramural gaming is still warranted.

Before investigating how teenagers interacted with games, Sylvén and Sundqvist (2012) had observed the extramural habits of Grade 4-6 students (n = 86) in Sweden. The students were put into three groups: non-gamers (n = 9, 0 hours), moderate gamers (n = 17, > 0 to < 5 hours, mean 1.5 hours, SD 1.4), and frequent gamers (n = 13, > 5 to <= 23 hours, mean 9.7 hours, SD 5.1). According to questionnaire data, the frequent gamers group had a larger ratio of students claiming to have learned English mainly outside of school compared to the other two groups (Chi-2 testing, $p < .01$). More than half of the frequent gamers responded that they spoke English in their spare time. A national English test is administered in Sweden in March, and the scores improved with each digital game group with the frequent gamers having the highest scores for both reading and listening comprehension.

It is important to remember that gaming is not specifically an extramural activity that should be adopted solely for younger language learners, as video games provide skills for language that are positive to all language learners, regardless of age. Gee (2006) posited that knowledge of *situated meaning* can be improved through gaming, which is the idea that words have transformative properties in different contexts, as Gee gives the example of the word *coffee* which can be used to refer to the liquid, the grain, or coffee in cans depending on the context (Gee, 2006). Games provide a variety of contexts in their narrative where words are used differently depending upon the situation. Situated meanings may be more prevalent in games than books or movies since these contexts are not just construed by surrounding words and imagery, but rather the experience of being embodied into the game (Gee, 2006) along with the experience of dialogue, narration, images, and actions that go along with the word in use (Gee,

2004). As of my knowledge, concerning gaming, no studies have investigating incidental learning in relation to adult language learners. It would be interesting to see how adult learners spend time gaming and how it could help them acquire new words and their situated meanings. There is a demand for an understanding of how games could be helpful for language learning outside of school, as is demonstrated by the results from Reinhardt's recent small scale study.

Reinhardt (2019) conducted an interesting study that investigated Reddit and Quora to discover how games are being used to promote extramural English. Reinhardt developed a corpus from the two websites after scoping the net with the question "how can I learn a foreign language from videogames?" (p. 39). With a corpus of questions and game suggestions from two Quora posts with 8 comments, and two Reddit posts with 137 comments, a grounded analysis was conducted to reveal general advice of extramural gaming to improve the interested parties' foreign language competency which led to the compiled list below (see Table 3).

Table 3: Reinhardt's results (2019)

(1) users advocate a wide variety of vernacular game titles and genres for L2 learning,
(2) as long they make the learner-player use language,
(3) in a casual and enjoyable way.
(4) Users advise, playing a game at the right proficiency level,
(5) whose rules are not too unfamiliar,
(6) and that include a lot of language use
(7) and features that provide time to read, re-read, listen, and re-listen.
(8) They suggest, using subtitles, mimicking voices, using dictionaries, making vocabulary lists, using related media, and interacting with other players, and
(9) being sure to have the right console, region code, and server connection.

43 different titles and games were mentioned throughout, and the languages were diverse. With many learners subscribing to experiencing games and operationalizing their learning through these beliefs, it could be an interesting starting point to observe how the average language learner may engage in a foreign language input through extramural learning. While a lot of research has been done regarding youths' extramural use for gaming in the L2 (Sundqvist & Wikstrom, 2015; Sundqvist & Sylvén, 2014; Sylvén & Sundqvist, 2012), it is unknown what adults and older learners do and think while gaming. This compiled list from the internet is anonymous, so there is no clear indicator of how old the posters from the comments were, but there may have been older learners as well in these comments. Conducting the present research, then provides insight into the words required for a portion of these games' comprehension and may provide these learners insight into the vocabulary they may potentially be able to learn from or acquire while gaming.

Classroom Learning

Using games for L2 acquisition in the classroom has remained scarce (Blume, 2019), and investigations regarding their efficacy have been scant (Hitosugi et al., 2014) despite there being documentation and suggestions on how to promote language acquisition using games in the classroom (Sykes & Reinhardt, 2013). This problem is exacerbated when considering COTS games, as it may be hard for teachers to draw the link between pedagogy and playing games for enjoyment which has led to this underdeveloped area of research. With many stakeholders present and parameters that are set onto various teaching contexts in mind, it may be difficult to incorporate games and justifying their usage in the classroom. Despite this, some studies have investigated game-enhanced learning.

Coleman (2002) used a flight simulator video game that is based on SIM CITY – the game was repurposed to help learners engage in writing composition and rhetoric in an ESL pre-freshman composition course, and the targeted features for improvement were grammatical and mechanical accuracy concerning written discourse. The course met 3 hours a week and the development of the students' rhetoric was a precursor for entering the freshman composition course. Learners were given two scenarios: Scenario 1: they had to play the role of a pilot where they had to fly to a destination first and then later give instructions to a VIP visitor on how to get to that location. Upon completing Scenario 1, they were given Scenario 2: The students had to then play the role of the VIP visitor and must use the prior-made instructions to go to the desired destination. This contrast was done to inform learners of the difference between the constructs of *writer* and *audience*, especially when considering the directions for the VIP's target location are being made from above in a helicopter; they had to consider how the VIP would perceive the instructions by walking.

In Scenario 2, when the students are the VIP, they must adjust the directions should there be any errors or any steps that require additional clarification. Such activities can be done by simply using a map and paper in real life; however, the affordance of a virtual 3D space allows for more precise language concerning giving and receiving directions without leaving the classroom. This is a stellar example of taking a game and repurposing it for an educational context due to its affordance of being a 3D space to allow for precise writing and following specific instructions. Such an activity could be useful for learning new vocabulary as well for such chunks as "that building (X) is adjacent to (Y), perpendicular to" etc. Using gaming affordances and remodelling them for the educational context is a possibility that exists with

COTS games, but it will depend on what the target feature of language being observed is, and then finding the most appropriate game to fit that target.

Due to the nature of task-based language teaching and content-based language instruction, it seems that single-player games are seldom used in the classroom. Studies that have used single-player games have used different ways to keep the non-playing students busy. Some of the ways that single-games have been repurposed to keep students involved has been by putting them into groups and have them do various activities, such as take notes or watch (cf. Chen & Yang, 2013; deHaan, Reed, & Kuwada, 2010), conduct research (cf. Ranalli, 2008), journal or write about their game experiences (cf. Sykes & Reinhardt, 2013), encourage or support the player in control (cf. Hitosugi et al., 2014), or after a certain allotted time, the player in control can swap roles with the backseat gamer (cf. Ranalli, 2008). There also seems to be this dichotomous categorization for learning according to the literature, that either learning takes place in the classroom, or as part of an extramural activity. Games could be used outside of the classroom by learners and perhaps some important details regarding how these games can be played effectively with learning strategies can be taken up in the classroom by the teacher. Similar to how a teacher may give special attention to a particular book's themes or vocabulary before assigning chapters for homework, games can be primed as well.

Another example of a game being adopted to the classroom context is Hitosugi et al's (2014) study, which had used the game *Food Force* (FF) published by the *United Nations* to investigate vocabulary acquisition regarding the theme of "world hunger" in two intermediate Japanese classrooms (study 1: n = 9; study 2, n = 11). While the game was created to serve educational purposes, the authors had argued that the design associated with this game is akin to a COTS game, and therefore the pedagogy surrounding the design of this study was game-

enhanced. Five 50-minute class sessions were assigned for the FF activities for both classrooms and they did not play the game until the fifth day of the experiment. Prior to the study, students were given a pre-test, and then they would later have a post-test, and finally, they would then later be administered a delayed post-test (5 weeks) to observe for retention in vocabulary. Students were seated in groups of two to three that were assigned by the teacher, and for each group, one student was at a computer at a time. For the first study, no vocabulary list was given to students, but they had regular tasks that embedded the key vocabulary. There was no end of term test on the vocabulary found in the game for study 1; however, for study 2, the target vocabulary was focused, and students were given a list of 32 new words. Students in study 2 took a vocabulary pop quiz as well, had an end-of-unit test on the vocabulary found in the game, and had a textbook unit test on a similar theme that the students were made apparent to.

On the fifth day, both groups were told to complete the first two missions of the game and attempt to achieve the highest score possible within the allotted time. For both groups, there was a significant vocabulary gain on mean scores on the post-tests. On average, students were able to correctly write the meanings of eight more Japanese words in a 70-item vocabulary test upon completion of the FF unit. Students in study 1, where vocabulary teaching was NOT the focus, had acquired eight more Japanese words in the 70-item vocabulary test (immediate post-test) and that number diminished to 7.89 on the delayed post-test. Students in study 2, where vocabulary teaching and acquisition was highlighted, acquired an additional 11.36 words on average at the end of the unit and the number increased to 12.18 on the post-test. Students were receptive to the game and were more engaged as a result.

It may be important to also consider not only gaming's use in the foreign language context but the actual language being taught as well. Japanese, for example, has three distinctive writing systems: two syllabaries (Hiragana and Katakana), and a logographic writing system (Kanji). It may have been harder for language learners in study 1 for a more incidental approach in acquisition because the words themselves are written in logographic characters; it may be more time-consuming for an English student learning Japanese to acquire a Japanese character-word incidentally. It is still indicative that regardless of the approach that there was a meaningful gain in vocabulary. It is then warranted that when games are used meaningfully with a pedagogical purpose that they can provide a substantial gain in the classroom. It is important to not only consider gaming affordances, but language affordances as well. For example, McArthur (2020) gave insight in how the affordance of Japanese characters can be used in an augmented reality application through digital story telling and historical Kanji pictographs to help English learners of Japanese with Kanji recall. While it was not a game itself, it gives insight on how there are more factors than just the medium itself when considering development for pedagogical applications.

When it comes to games and the development of additional learning materials, it cannot be expected for the game to take the reins of pedagogy. For example, the game *Sims* was used in a similar game-enhanced manner in Ranalli's (2008) study. There were nine participants in the study, all intermediate-level ESL learners enrolled at a major research university in the United States. The nine students were divided into dyads ("manager" and "controller" roles) in which both members had a similar proficiency level but had different L1s to encourage them to speak in English. For the purpose of this study, a companion website that contained supplementary materials was created to serve as a compendium to the game and also a way to administer the

study's experimental conditions. The website was divided into different "stations". "Managers" used the stations while "Controllers" played the game, and then the roles would then swap. Station 1 included vocabulary information and quizzes, culture notes, and instructions for each game session of Sims. Station 2 included a link to an online dictionary and the same culture notes and instructions as found in station 1, but it contained no vocabulary-related materials. Station 3 contained only the gaming instructions. This compendium was accessed from a resource computer located next to each gaming computer used by the participants. Each dyad had experienced all three experimental conditions. Participants in stations 2 and 3 had access to an online dictionary. The culture notes found in stations 1 and 2 were compiled explanations to inform of some of the cultural references that the participants would encounter in the game, which was situated in a suburban American neighbourhood.

A pre-test was conducted to evaluate prior knowledge of target words to help group students into similar levels of proficiency, and to create a point of comparison for the post-test (k = 30). Using a one-way Anova analysis, the results suggest a link between the combined effects of the materials made for the class being used together with the game. There was a statistically significant difference between stations 1 and 2 (MD = 2.78, p = 0.01*, Station 1: M = 8.56, SD = 1.33, Station 2: M = 5.78, SD = 2.22), however, no such difference existed when being compared to other stations. While learners had access to these additional materials, it may have been insufficient to just provide access to them, as they likely require additional training. It would have been interesting to note how much time is spent by the "Manager" and if they stayed on-task or if they were engaged with the game just as much as the "Controller". Training and/or a more engaging learner's compendium may be required when developing additional materials to promote game-enhanced learning.

2.5 Research Purpose and Questions

The intended purpose of this thesis is to create two specialized vocabulary wordlists that can help educators and learners acquire key single-player game vocabulary found in games both in and outside of their classroom depending on the way the teacher approaches gaming and pedagogy (intentional learning), or even for learners that game as an extramural activity (incidental learning). The present research is as well answering the research call laid out by Schmitt (2019) contributing to the research area of vocabulary within the field of Applied Linguistics. With all this in mind, the research questions are the following:

1. What are the words with the highest keyness in commercial off-the-shelf video games that could be used by teachers in English teaching classrooms or for learners engaging in extramural learning?
    a. How many of these words are ludic and/or diegetic? To what extent could these high keyness words from the total game corpus be taught in the L2 classroom?
2. How does keyness change in the game-genre of action-roleplay games compared to the total game corpus?
    a. How many of these words are ludic and/or diegetic? To what extent could these high keyness words from the action-roleplay game-genre be taught in the L2 classroom?

This study follows a multimethod design where both quantitative and qualitative measures are used to answer the multifaceted questions posed.

2.6 Chapter Summary

This chapter has provided an extensive view on gaming and its relation to vocabulary and language acquisition in general through various game-designs and media. The literature presented was to showcase potential future research as well as focusing on the research most pertinent to this study's use case. By having supplementary wordlists when using a COTS game for either extramural or in-classroom usage, the hope is to provide learners and educators with a comprehensive insight into the variety of vocabulary these games contain and whether they vary based on game-genre, and finally, whether these words serve a ludic or diegetic function and how they reflect those properties.

Chapter 3: Methodology

This chapter presents the research design of this mixed-methods study. In this chapter, the corpora used will be justified along with providing insight into the research tools operated and methods undertaken. Steps taken in choosing which aspect of the data to analyze will be explained, and the approach to coding adopted in this study will be elaborated on while mentioning measures taken to ensure inter-rater reliability for coding.

3.1 Study Design

The goals of this thesis are the following: (1) to answer Schmitt's call for research of variation in game vocabulary, (2) provide teachers with two tangible wordlists to use in the teaching of English in commercial off-the-shelf games as well as learners having access to these lists for extramural learning, and (3) further disseminate these words into ludic and diegetic properties to observe their differences in teaching and acquiring them. In this case, as this study is addressing three different concerns, its design is varied, and as such, a multimethod design was adopted. As Brewer and Hunter (2006) define multimethod research, it is "any research that contributes in any way to gaining a multimethod view of social phenomenon" and that it is "an attempt to apply our individual work lessons learned at the level of the discipline as a whole and thereby to enrich the collective effort to which we each contribute" (p. 11). Using wordlists to teach is a practice in the field of language teaching and is learned in the discipline of Applied Linguistics, and by investigating these lists and how they affect teaching brings the application of these words and games into the classroom; as well, understanding extramural vocabulary

learning and acquisition, these words could help learners. In the table below, the methods are mapped to the study's research questions (see Table 4).

Table 4: Methods Mapped to Research Questions (Multimethod Design)

Research Questions	Methods
1. What are words with the highest keyness in commercial off-the-shelf video games that could be used by teachers in English teaching classrooms?	• Keyness Analysis (Quantitative) • Lexical Frequency Profile of Summative Keyness Wordlist (Quantitative)
1a. To what extent could these high keyness words from the total game corpus be taught in the L2 classroom?	• Evaluation Coding (Qualitative) • Pattern Coding (Qualitative)
2. How does keyness change in the game-genre of action-roleplay games compared to the total game corpus?	• Keyness Analysis (Quantitative) • Lexical Frequency Profile of Summative Keyness Wordlist (Quantitative)
2a. To what extent could these high keyness words from the action-roleplay game-genre be taught in the L2 classroom?	• Evaluation Coding (Qualitative) • Pattern Coding (Qualitative)

To investigate what words are more attributable to games, the decision was made to conduct a keyness analysis, so that words that appear more frequently in games when compared to movies could be identified. To do so, a game corpus was constructed. This was the same video game corpus used in Rodgers and Heidt (in press); a 5.7 million token corpus that encompasses all language from 10 games: examples of "all-language" include language from the user interface, side quests, dialogue, additional language from books in the games, and any language that the player could possibly experience. Since the lexical thresholds required for comprehension were high in the previous study done by Rodgers and Heidt (in press), the initial assumption was made that there could be game-specific words that were contributing to more infrequent words which warranted further investigation and shifted the research paradigm from a

lexical frequency perspective to a multi-method corpus study; this thesis seeks to discover what words are prevalent by using *Minimal Ratio* (MR) to identify keyness (quantitative) and then discover how these words are used in games through evaluation and pattern coding (qualitative). The keyness analysis was conducted using flemmas as base words. The game corpus from Rodgers & Heidt (in press) was used as *the study corpus* and the SUBTLEXus Movie Corpus (Brysbaert & New, 2009) as the *reference corpus*. These terms are used to refer to the nature of how keyness is conducted.

The reference corpus is what the study corpus is being compared to by using MR. MR functions not by identifying frequency, but rather the disparity or similarity in frequency between two corpora, and then computes the result giving an "MR value". MR is the value that attributes how prevalent the word is in the study corpus compared to the reference corpus. The higher the MR value, the larger the difference between the corpora's use of the word, which is indicative of a "key" word in the study corpus. The smaller the MR value, the closer the words are in frequency ranking, meaning there is no perceivable difference of the word or very little. This study is concerned with positive key words, in other words, the distinctive key words of the study corpus. The intended purpose of the study is to give teachers and learners insight into the variety of vocabulary used specifically in commercial off-the-shelf (COTS) games and to what extent the teaching of vocabulary is required for learners of English as a Second/Foreign Language in this context.

3.2 Corpus Approach: Why Keyness?

The end goal of the study is to produce two wordlists for teachers and learners to aid with vocabulary learning and acquisition in COTS games. Similar to how students are tasked in classrooms to write journals and report events that transpire within their daily lives, Sykes and Reinhardt (2013) mention that video games can be used similarly by journaling what occurs in the game as a way to improve students' English. There are a wide variety of activities that teachers can employ using COTS games, so it is therefore important to consider the contextual vocabulary learners may need to know before or while playing COTS games. Evidence of a good wordlist is that it serves the context it purports to serve (Brezina & Gablasova, 2015). Simply reporting for frequency in this context would not be ideal, as it is not indicative of game-specific words, but rather overall words used in games. The keyness method has been used in various other fields to create pedagogical wordlists, one of them being business English words (Nelson, 2000). The struggle with a pure frequency and range-based perspective is that it fails to consider sub-topics in its data. For example, the Academic Wordlist is noted for not giving enough salience to different academic disciplines within its wordlist (Paquot, 2007). By using keyness, the game-genre of games can be investigated, and their words can also be identified when comparing to the total corpus.

Overall, corpus research can be divided into two main camps of theory: corpus-based and corpus-driven (Biber, 2019). The former involves the researcher preselecting a target language feature and analyzing it in a corpus to discover how that feature is used; whereas the latter is the feature of language becomes apparent from analyzing the corpus inherently in itself. Keyness analysis is more corpus-based than it is driven, as keyness aims to identify a corpus' "aboutness"

– the key words derived from a keyness analysis illustrate a corpus' linguistic profile by comparing it to what is known as a reference corpus; in other words, a keyness analysis is used to identify similarities between the frequency of word-forms in two corpora (Gabrielatos, 2018). Doing so allows researchers to uncover the descriptive nature of the words in the corpus. The corpus being researched on is referred to as the study corpus, and the corpus being compared to/against is known as the reference corpus. By using a reference corpus, programs that analyze for keyness have a sample of language to base its differences and similarities.

Gabrielatos (2018) recommends effect-size metrics to approach keyness analysis. This does not refer to effect-size as a way to illustrate the strength of a cause and effect relationship, but rather, normalizing the data and then calculating for differences between that normalization thus accounting for differences in corpora size. It is not the frequency, rather it is word ranking according to the frequency in the corpora (Gabrielatos, 2018), and this is used to find if words are different from the reference corpus.

An emerging measure in the field of corpus linguistics is *minimal ratio*: the minimal possible ratio between the relative frequencies in the corpora (Milička, 2012). It can be utilized for keyness analysis by using confidence intervals to calculate for similarities/differences determined by the user's given threshold. For example, if a researcher were interested in the difference of a word between two texts with a threshold of 95%, it would calculate for 95 % of all samples by removing high and low-frequency vocabulary in the corpora and provide an upper and lower confidence interval. If the value of words falls between the upper and lower limits of the confidence interval, there is no difference in relative frequency between the two corpora (Milička, 2012). By using confidence intervals, minimal ratio is an ideal measure, as it is an effect-size measure. It has three main advantages over other effect-size measures:

1) The results are very easy to interpret
2) It can be used for more than just word types (such as collocations)
3) It can select different 'partitions' of the data and determine what the correct threshold of significance is.

3.3 The Data

Study corpus

The study corpus is the corpus used in Rodgers & Heidt (in press). The corpus includes 10 games across four different game-genres and totals 5.7 million tokens (see Table 5). This corpus was selected to expand on the previous study's findings and to create two pedagogical wordlists. Furthermore, to date, there are no other comprehensive corpora that contain COTS games. This thesis focuses on how well videogames can be used as an accompaniment to learn English both in and out of the classroom. Due to these reasons, this corpus was selected. The language in this corpus was extrapolated using various software decompilers to gain both overall language used and subtitle data. This data was then modified using a string-searching algorithm known as *Regular Expression* (RegEx) to attempt to remove most programming codes, field notes, and language not visible to the players of these games. These ten games were chosen due to two main reasons: 1) the convenience of obtaining their string data due to existing decompilers made for these specific games, and 2) these ten games were selected as well due to their variety; a deliberate effort was made to diversify the game genres to attain a general-game perspective vocabulary-wise.

Table 5: Videogame corpus from Rodgers and Heidt (in press)

Videogame	Genre	Release Year	Tokens
Sam & Max	Graphic Adventure	2007	139,719
Walking Dead	Graphic Adventure	2012	221,526
Phoenix Wright Trilogy	Life Simulation	2019	751,347
Sims 4	Life Simulation	2014	323,555
Divinity: Original Sin 2	Action role-playing	2017	1,095,755
Monster Hunter: World	Action role-playing	2018	205,219
Fallout 4	Action role-playing	2015	1,029,969
Skyrim	Action role-playing	2017	960,322
Grand Theft Auto V	Action-adventure	2013	949,165
Resident Evil 6	Action-adventure	2019	67,811
		Total Tokens	5,744,388

Reference Corpus

The SUBTLEXus corpus was compiled by Brysbaert and New (2009). It is a 51 million word corpus of American television shows and films. Similar to the extrapolation of subtitle data from the studied game corpus, the data gathered in SUBTLEXus is the subtitle data from these various television shows and movies. The justification to use SUBTLEXus as the reference corpus was due to the perception that television, film, and games may have a similar vocabulary due to their affordances as a visual aural input. It was thought that television may serve as a better reference than books or other media, as the results of the analysis would return vocabulary more representative of a game, as opposed to a result which demonstrates a difference in media choice: “There are a number of different factors involved in understanding movies such as the overlap between imagery and dialog” (Webb & Rodgers, 2009b, p. 425), and naturally imagery applies to games as well. Since both the game corpus and the television and film corpus are also designed with imagery in mind, it is possible that dialogue constructed in this media could be less detailed than what would normally be required in books. Using the SUBTLEXus corpus as the reference corpus is then a better way of measuring these perceived differences. In other words, because there are images on the screen when a person is watching a show or playing a

game, the dialogue may be shifted to accompany what is occurring on screen, and therefore may have different word choices than a book that may rely on more descriptive language.

3.4 Procedure

Phase 1: Modifying the pre-existing corpora for Keyness

To identify videogame key words, it was first deemed necessary to remove language that is key to each individual game. This was done as an effort to remove language that is thematic and specific. First, a flemma wordlist was developed for each game, then the entire game corpus and the SUBTLEXus corpus. The flemma is a hybrid between the *lemma* and the *word-family* (Pinchbeck, 2017); they are untagged parts of speech that share lemma inflections. Word-Families contain both derivational and inflectional affixes, whereas lemmas and flemmas only contain inflectional suffixes. The distinguishing characteristic of a lemma versus a flemma is that a lemma is tagged for its part of speech (PoS), meaning that the words are labelled as noun, verb, adjective, or adverb, whereas flemmas are not tagged. To illustrate the difference between a word-family and a lemma/flemma, consider the word “go”. Under word-families, “go” would contain “go, goer, goes, goin, going, goings, gone, goner, goners, gonna, gunna, went”, whereas the flemma/lemma entry of “go” would contain “go, going, goes, went, gone”. flemmas were selected for this segment of the procedure due to three main reasons:

1) Word families have been criticized to encompass too many derivations that a language learner may not have the morphological skill to decode (Gardner & Davies, 2014).

2) Lemmas have been shown to have more promising results when conducting a keyness analysis (Paquot, 2007); however,
3) Lemmas consider words' parts of speech which may overestimate the learning burden for words that are closely related in meaning (Nation, as cited in Dang, Coxhead, & Webb, 2017). Words close to each other, such as run, as in a run, noun, and to run (verb), may not confuse a learner's understanding of the word "run".

That is why Flemmas were decided to be used out of all the base-word counting metrics. The flemma wordlists for each game were then used in Milička's (2012) *Keyworder* software using Minimal Ratio to calculate a 95% confidence level to determine what language is key to each game when comparing to the entire game corpus. Words key to each game, in other words, a MR value of 2 or higher, would then be excluded from the keyness calculation when comparing the entire game corpus to the SUBTLEXus corpus. For example, in the game, *the Walking Dead*, the word "walker" has an MR value of 13.06 (see Appendix B for analysis results of walking dead when compared to total game corpus). Walker is the term the characters use in the game to refer to enemies: zombies. It is a word that is very specific to this game according to the MR value, and not all games. Language specific to each individual game does not answer this thesis's research questions and intended purpose, and therefore words with an MR value of 2 or higher were removed for individual games to consider general game language (see Appendix C for a visual representation).

When creating the total game wordlist, it was decided that the words should have a minimum MR value of 2, and that the words should appear in a minimum of across 6 games to be considered a part of the total wordlist, and all games (n = 4) for the ARPG wordlist. The confidence level was set to 95% in Keyworder for both the total game and ARPG corpora. 100

was the number arbitrarily decided to be the cut-off for both wordlists (total game wordlist, last word: ‘capacity’, MR value of 2.89, ARPG wordlist, last word: ‘sadly’, MR value of 3.33). Laurence Anthony’s (2019) AntConc was then used to go into the corpus and search by flemma inflections to retrieve concordances of the words in use. Upon completing the wordlists, the words were re-entered as word families into AntWordProfiler (2014) to ascertain its validity as pedagogical wordlists by identifying how much coverage they provide to their respective corpora (total game corpus and ARPG corpus)

Phase 2: Game-Genre words: ARPGs

This is a replication of Phase 1. Schmitt (2019) mentions in his vocabulary research agenda that the next segment of vocabulary research in games should research whether game-genre affects the distribution of vocabulary used. For example, would the word recoil be more apparent in games that involve action, or shooting games? To be able to answer this question, a similar step to removing thematic language must be employed. To receive a tangible result of the variety of vocabulary used for a specific genre of games, it was decided to investigate the action-roleplay game-genre from the total game corpus (4 of 10 games) by creating an ARPG corpus (i.e. a separate corpus of just the 4 games). The action-roleplay game-genre games are Divinity 2, Fallout 4, Monster Hunter World, and Skyrim. These games are being compared against SUBLTEXus. Most of the corpus consists of language from Divinity 2, Fallout 4, and Skyrim. All four games contain 3,291,265 tokens of the 5,744,388 tokens, in other words, 57% of the entire corpus. The corpus favours this game-genre over the others and provides a lot more data to compare against the movie corpus which was the reason why this game-genre was selected over the others. The reported words for this segment must have had a MR value of 2 or higher AND be present in all 4 games. Game-specific words were removed just as in procedure 1.

Phase 3: Coding for Total Game Corpus and ARPG Wordlists

First Cycle of Coding

Both wordlists, the Total Game wordlist, and the ARPG wordlist, are coded for ludic and diegetic functions in this phase. The words were searched in AntConc with their given flemma inflections and the retrieved information was stored into an excel database for coding. It is very noticeable if a word is used in dialogue, as it is often a longer utterance, and there are character interactions involved (i.e. someone speaking and being spoken to). Ludic words are therefore determined by their abruptness, surrounding information in the data (inferring, for example, that they are from a settings menu), and short imperatives. Some ludic words, as I have observed while conducting this research, lack a *theme* and are simply *rhemes*, as in the example "greatly reduces recoil after shooting" [Resident Evil 6]. Recoil is used in the ludic perspective of a gun's recoil while aiming in a game. There is no theme attached to the utterance. The word's function can also be distinguished by close-collocates, in this instance, recoil and shooting together illustrate the use of a gun or a bow. Evaluation coding was used as the coding approach to categorize dichotomously if a word functioned as LUDIC or DIEGETIC. Five was the determined number to be the threshold for whether a word was ludic. In other words, if the words were used in a ludic context five times, it would be considered LUDIC. If there are five instances of the word being used in a ludic context across the majority of the games, it would mean that it is prevalent enough to be apparent in other games outside of the game corpus as well.

In qualitative research, evaluation coding is an assessment that stems from the researcher's evaluative perspective (Saldaña, 2015). It is a coding system that reflects initial questions that structured the evaluation. This method of coding was selected to assign *ludic* or

diegetic properties to a word and use the researcher's evaluation as well as peers' evaluation for re-evaluation to ensure interrater reliability. 10 concordances for each word on the list would be used for the analysis of whether a word was primarily diegetic or ludic in those contexts.

A systematic sampling method was employed to obtain the data to use for coding. The initial step was to "thin" the data by a value of three. In other words, every third concordance would be collected and used for the analysis. This was done initially to ensure a certain extent of random sampling; however, it was later decided that a minimum of 50 concordances to choose from would be a requirement. The rule decided was if there were under 50 concordances given from a word being thinned three times, the thinning would then be reduced to 2 or remain at 1 to ensure that a preferable minimum of 50 concordances was generated from the corpora. This was due to two main reasons: 1) to include a variety of games from the coded data to ensure a variety of vocabulary usage, and 2) to have more concordances to narrow down if the word's category (ludic or diegetic) was hard to infer, or was not chosen due to the language not being present to the player.

The final number of concordances decided to be used for analysis was 10 concordances for each word on the game wordlist. The literature in qualitative research indicates that there is no clear-cut answer to how much data is required to be coded for comprehensive results. Elliott (2018) stated that to obtain tangible results, the researcher must consider the study's research questions, design, and intended purpose when determining how much data is required to code. 10 concordances were chosen as the saturation point for this analysis. This wordlist is intended to inform teachers and learners of how a word can be ludic or diegetic in games and provide an insight into an easily inferable scale. When the concordances retrieved were thinned to three, two, or remained the same, a logic chart was devised to choose the 10 concordances for each

word to be investigated (refer to Appendix D for TOTAL corpus logic chart, Appendix G for ARPG corpus). For both corpora, the parameters set were the following: Before selecting 10 concordances, 50 concordances must be present from the data-thinning process. If there were under 50 concordances, then the data thinning would be reduced or excluded, as necessary to retain 50 or more concordances. The language must be visible to the player. If I, as the researcher was unsure, the item was thrown out from the analysis. If the item still did not include 50 concordances after thinning, it remained as it were and was still included for analysis.

Total Game Corpus

For the total game corpus, concordances were selected in a top to down order. The first six concordances chosen are all from different games. This is to ensure a variety in the samples collected. If all 10 unique games are in the 50 or so concordances, then the 10 concordances would automatically be decided (i.e. the 10 unique instances of concordances were chosen), 1 from each unique game. If not, continue past the last unique game (marked in yellow) and begin a new round of selecting concordances. When I use "round" for coding concordances, I am not referring to the coding cycles in qualitative research, but rather beginning a second round of coding games internally within the excel database. Mentions of rounds regarding coding concordances is referring to the steps I took to ensure a variety of games were included to be investigated. After completing the first round of concordance coding, the same games from the first round can now be selected again in order descending from the last unique game, but no duplicates in the second round can be used (see Appendix E). If there were insufficient concordances from the last unique game descending, then the order for selection was reversed, and I went upward from the last unique game, attempting to have no duplicates. If there were not enough unique games for the last round in ascending order, then the next game ascending is

chosen, regardless if it is a duplicate game or not. Since not all the data is thinned, this procedure was adopted to ensure that not all data fell within the same partition to provide some random sampling while in a controlled manner to ensure a variety of games present, as opposed to pure random sampling that could possibly provide all concordances from the same game.

In Appendix E, the yellow highlights are the unique games in descending order. There were six unique games. Since there were not 10 unique games in the retrieved concordances, the order continues to descend. The second round is illustrated in cyan. Duplicate games between round one and two are accepted (yellow – cyan), but no duplicate games in the same colour. When 10 concordances are selected, the coding begins. The concordances are fully expanded by retrieving the key word from the original text file. It is then evaluated by not only a single concordance line present in Appendix E, but the text that comes before and after it in the original text file as well (see Appendix F as an example of querying the corpus for context).

ARPG-specific Words

Words unique to the ARPG genre wordlist (n = 37) were coded, five of them being ludic and the remaining being diegetic. The reason why not all words in the list were coded (n = 100) was due to them being repeated from the total game corpus (n = 63). The purpose behind this paper is to provide insight in how these words can be used, and since that objective has already been fulfilled by looking at the repeated words in the total game wordlist, only words distinctive to ARPGs within the 100 wordlist were coded. ARPG-specific concordances were chosen in a very similar manner to the general corpus. The data-thinning rule and systematic sampling remained the same, however, since there would always only be 4 unique instances, the logic (see Appendix G) devised for choosing concordances was as follows: The first selection of concordances was 4 unique games, and then a repetition of those games for a second round of

unique games (n = 8) and the remaining 2 concordances taken from the two upmost games in alphabetical order to total 10 concordances chosen. This was done again to ensure that there was a systematic sampling method and an example of how this was done can be seen in Appendix H.

Second Cycle of Coding

Upon completion of the coding, I went through all the data again and recoded for patterns: "A pattern is repetitive, regular, or consistent occurrences of action/data that appear more than twice" (Saldaña, 2015, p.5). A second cycle of coding is done in qualitative research to ensure reliability in the data and to mark for the construct in a revised way or in a fashion that addresses gaps in the initial coding (Saldaña, 2015). The initial gap being resolved through pattern coding is how the words are being used ludically. Pattern coding was used to see how the usage of these words led to ludic functions. Pattern coding "organizes the corpus into sets, themes, or constructs and attributes meaning to that organization" and that it is "appropriate for second cycle coding" (Saldaña, 2015, p. 6). Words with a ludic score of 5 or higher were categorized and observed for patterns using a key word in context (KWIC) Excel Database and the Total Game Corpus text file. The patterns were then investigated out of the total occurrences of ludic function within that word. For example, the word "melee" had nine of the ten occurrences as ludic, so the total number of concordances queried to view for patterns is then nine for the word melee. The coding for patterns was done to mark any overt features that reflect the words' usage with the sentence in consideration. At the end of the coding for individual words, a culminating list of features is compiled to define what a ludic word is based on the data obtained from the corpus for both the total game and ARPG wordlists.

Interrater Procedure for Coding

Three inter-raters were tasked with evaluating 20% of the coded data (28 words) which contained words from both the total game wordlist and the ARPG wordlist (10 concordances for each word, totaling 280 concordances). The retrieved concordances contained information from the Excel database PLUS the surrounding context from the corpus if that data was available and requested on behalf of the rater. The words were selected at random (see Table 6) and were pulled from both the total game corpus and the ARPG corpus. The researcher and inter-raters had played some of these games within the corpus; their playing profiles can be viewed in Appendix I. The games were not required to be completed but enjoyed casually. Game names were removed from the extracted samples for rating.

The inter-raters were trained with 5 words, 4 concordances each (20 concordances) to make them think on the construct of ludic and diegetic concepts. The inter-raters were also provided with the researcher's rationale for categorizing those five words as ludic or diegetic. They were permitted to ask questions or disagree with any of the evaluations performed in the training session. These words were also marked for context where they would appear if the surrounding words may have been insufficient on their own to evaluate, for example, the concordance "Turn water and blood **SURFACES** into oil" had an additional comment included "(skill description)" at the end to signify that the phrase was retrieved from the description of a usable skill the player can use in the game. Key words were highlighted and capitalized. Inter-raters were told to evaluate the word in context, and not the word isolated by itself for ludic and diegetic functions. There were three evaluation responses available for each concordance in the interrater excel sheet: "A: LUDIC, B: DIEGETIC, or C: REQUEST CLARITY". If "REQUEST CLARITY" was selected, and if more information could be retrieved from the corpus, it would

be delivered to the inter-rater. The inter-rater would then have to decide from both the original concordance and the additional information if the word was used in a ludic or diegetic context. Completion of the inter-rater (IR) reliability check (n = 4, Researcher, IR 1, IR 2, and IR 3) resulted in 87.86% agreement with a free-marginal kappa value of 0.76 [0.70, 0.81] and a fixed-marginal value of 0.73 [0.67, 0.79]. This figure is in line with the literature's recommendations for reporting a high rate of reliability (Mackey & Gass, 2015; Saldaña, 2015). Since the rate of agreement was high, a discussion session was not pursued, but rather a revisiting of the disagreed upon items and recoded the same or differently based upon reviewing the various words in context for a second time.

Table 6: Randomly Selected Words for Inter-rater Coding

Randomly Selected Words for Interrater Coding
Ensure, Glare, Poison, Sate, Autosave, Otherworldly, Vase, Rend, Loot, Surface, Weaponry, Silver, Encounter, Armoury, Teleportation, Composure, Mist, Path, Current, Gameplay, Tutorial, Artifact, Beneficial, Melee, Bastion, Highlight, Resistant, Spear

3.5 Chapter Summary

This chapter reported the mixed methods designed used in this study. Minimal Ratio is used to create a pedagogical wordlist by calculating the difference in frequency between the game *study corpus*, and the SUBTLEXus *reference corpus*. The methodology combines quantitative measures (MR) with qualitative coding (evaluation and pattern coding) and those processes are described in the examples given.

Chapter 4: Results

This chapter presents the results of Phases 1 to 3 of the study: A wordlist generated from the total game corpus using a keyness analysis (Phase 1) a wordlist generated from the ARPG corpus using keyness (phase 2), and finally, evaluation and pattern coding of both wordlists (Phase 3). The results of phase 1 will indicate the variety of key words present in the total game corpus (n = 10 games) which will generate a common game wordlist; it will also identify the coverage that these key words provide for the entire corpus. Phase 2 will reveal an action-roleplay game-genre wordlist from the 4 ARPG games to illustrate game-genre specific vocabulary as well as show coverage of the corpus. Phase 3 will reveal the strength of ludic and diegetic words of both wordlists through evaluation coding to give insight into whether they serve a narrative or gameplay function, and ludic words will be expanded upon using by pattern coding for the second cycle of coding within the phase.

4.1 Phase 1 Total Game Corpus

The 100 game wordlist was created using Milička's (2012) keyworder software. The two corpora described in the methodology chapter (the studied and reference corpus) was used to create a common game wordlist as described in the previous section using Minimal Ratio (MR). This study investigates strictly positive key words (words that occur more frequently in the study corpus when compared with the reference corpus), more specifically an MR score of 2 or higher. Before applying filters to the data, using the Keyworder software, a keyness list of 151,936 words (types) was generated. Entering the filter parameters generated 484 types. To generate a

comprehensive wordlist, the cut-off rate for the list was 100 words containing the highest MR scores and the key words themselves must have appeared a minimum in 6 of the 10 games. The final wordlist can be seen in Table 6. Marginal words and common acronyms were removed from the original 100 words (this list of removed words can be seen in Table 7).

The KWIC database was then compiled in Excel and included all the words used in the 100-wordlist; it listed the key words within a word boundary of 20 words (i.e. there are 20 words before the key word and 20 after as well). The database was used as a reference tool and each key word in context was further queried into the total game corpus in the software *Notepad++* for further context. This analysis was used to target teachable words for games, thus the justification to have the words come across 6 games was made to remove any words that were tailored to a specific set of games. While the language key to each unique game was removed, there was still a concern for having a group of games contain specific words that fall within that game category; for example, graphic novels may have a specific vocabulary over roleplaying games, therefore, the justification for 6 games was valid and enforced. The purpose of a wordlist is to generate a teachable list that is representative for that category (e.g. common games).

Table 7: Total Game Corpus Key Words

Rank	Word	MR Score	Range												
1	Melee	6.54	6	**31**	Lemme	3.65	8	**61**	Décor	3.21	9	**91**	Madness	2.95	9
2	Recoil	5.22	7	**32**	Composure	3.63	6	**62**	Weaponry	3.21	6	**92**	Poison	2.94	9
3	Tutorial	5.20	6	**33**	Courtyard	3.60	7	**63**	Glow	3.18	8	**93**	Reveal	2.94	10
4	Teleport	5.09	7	**34**	Flee	3.59	8	**64**	Power	3.17	10	**94**	Beneficial	2.93	6
5	NPC	5.00	6	**35**	Glare	3.58	9	**65**	Formidable	3.15	6	**95**	Remain	2.92	10
6	Trader	4.96	6	**36**	Grumble	3.53	6	**66**	Quarry	3.15	6	**96**	Resurrect	2.92	7
7	Artifact	4.91	7	**37**	Thankfully	3.50	9	**67**	Craftsmanship	3.14	6	**97**	Loot	2.91	7
8	Nearby	4.83	9	**38**	Return	3.47	10	**68**	Feat	3.13	7	**98**	Skilled	2.90	7
9	Gameplay	4.60	7	**39**	Shard	3.45	8	**69**	Regeneration	3.13	6	**99**	Grime	2.90	7
10	Retrieve	4.45	9	**40**	Brewing	3.45	8	**70**	Sigh	3.12	9	**100**	Capacity	2.89	10
11	Optional	4.44	6	**41**	Culprit	3.44	7	**71**	Cleanse	3.12	8				
12	Foe	4.26	7	**42**	Dialog	3.41	8	**72**	Creation	3.12	9				
13	Teleportation	4.22	6	**43**	Highlight	3.41	6	**73**	Upwards	3.11	6				
14	Autosave	4.14	6	**44**	Cobble	3.40	6	**74**	Amethyst	3.11	6				
15	Mastery	4.11	6	**45**	Jagged	3.40	7	**75**	Otherworldly	3.11	7				
16	Infuse	4.11	6	**46**	Cloak	3.38	8	**76**	Fury	3.10	7				
17	Successfully	4.03	9	**47**	Attacker	3.37	9	**77**	Delve	3.10	7				
18	Stealth	4.02	7	**48**	Masterwork	3.37	6	**78**	Path	3.09	10				
19	Crimson	3.92	6	**49**	Skeletal	3.35	6	**79**	Volume	3.09	10				
20	Wary	3.89	8	**50**	Deity	3.35	6	**80**	Hunger	3.08	9				
21	Pause	3.88	9	**51**	Armoury	3.34	6	**81**	Tier	3.07	6				
22	Current	3.87	10	**52**	Unleash	3.32	9	**82**	Rogue	3.06	8				
23	Useful	3.86	10	**53**	Potent	3.32	7	**83**	Leather	3.05	8				
24	Overwrite	3.77	7	**54**	Assassin	3.31	8	**84**	Skull	3.04	10				
25	Stagger	3.76	8	**55**	Flame	3.30	9	**85**	Pillar	3.02	7				
26	Disable	3.75	6	**56**	Scrawl	3.30	6	**86**	Ensure	3.01	8				
27	Relic	3.72	8	**57**	Hideout	3.29	7	**87**	Fiery	3.00	8				
28	Fearsome	3.70	7	**58**	Vigilant	3.29	7	**88**	Navigate	3.00	7				
29	Scythe	3.69	6	**59**	Mist	3.25	7	**89**	Projectile	3.00	8				
30	Agility	3.66	6	**60**	Rename	3.23	7	**90**	Vase	2.98	6				

Table 8: Removed words from Total Corpus List

Word	MR Score	Range	Reason for removal
Haha	4.65	9	Marginal Word
Hehe	3.75	6	Marginal Word
Iv	3.416666667	9	Acronym + abbreviation for the number “four”
Bah	3.363636364	7	Marginal Word
Gah	3.142857143	8	Marginal Word
Pfft	3.117647059	9	Marginal Word
Urgh	2.928571429	6	Marginal Word

The pedagogical validity for the summative key wordlist (484 words, see Appendix J) was confirmed by re-entering the key words into AntWordProfiler as *word families* to view how much lexical coverage they provided (see Table 8). Word families were used instead of lemmas for this portion of the analysis so that comparisons can be made to research in lexical frequency studies. The 484 words provided 6.06% coverage for the 5.7 million token corpus. This number was lower than originally expected. This could be due to the language that I was attempting to captivate being too general.

Table 9: Lexical Frequency Profile of Total Corpus

Word Band	Tokens	Coverage	Cumulative Coverage
1,000	4364211	75.97	75.97
2,000	320723	5.58	81.55
3,000	141253	2.46	84.01
Total GG Wordlist (484 words)	310719	5.41	89.42
4,000	102254	1.78	91.2
5,000	70008	1.22	92.42
6,000	42624	0.74	93.16
7,000	27671	0.48	93.64
8,000	26275	0.46	94.1
9,000	15674	0.27	94.37
10,000	11691	0.2	94.57
11,000	10286	0.18	94.75
12,000	8209	0.14	94.89
13,000	6182	0.11	95
14,000	4335	0.08	95.08
Proper Nouns	65798	1.15	96.56
Marginal Words	52873	0.92	97.48
Compounds	23103	0.4	97.88
Abbreviations	7872	0.14	98.02
Off List	113984	1.98	100
Total Tokens: 5744388			

The more interesting figure, perhaps, is how much the coverage is after the 3K band, as this is where intermediate learners' vocabulary knowledge begins to decline. The 6.06 % falls to 5.41% when put after the 3K band list. This illustrates that these 484 words hold pedagogical validity for learners wanting to use vocabulary in games, as that one band of 484 words, which is half of a normal band, holds 5.41% coverage whereas each full 1K band over 3K covers less. 0.65% of the game wordlist covers the 1-3K bands, meaning that 0.65% of the total game words fall within the 3,000 most common word families in English (see Table 8). Most words are covering the 4K band and higher, which is indicative of a good list to acquire vocabulary from. To create a comprehensive common game pedagogical wordlist, 100 words was decided as the cut-off point.

EFL learners know words from the 4K band considerably less than the 1-3K word bands but have similar levels of knowledge of 5K+ bands as the 4K band (Brown, 2012). This suggests that there is not a difference in difficulty or rate of exposure for acquisition of vocabulary past the 3K band; in other words, a word from the 6K band (a word from the most common 6,000 word families) has a similar chance of being known by a learner as a word from the 4K band. By having a list that is more informative and captures these higher levels into one band for vocabulary acquisition, a quicker way to acquire these words becomes available for use by teachers and informed learners. As an example, words from the total game wordlist can range from the 1K band all the way to the 20K band. There is no need to scavenge through all 20 bands to find the appropriate words. By having all the proper words grouped into one band provides a better idea of the coverage that the vocabulary these games contain, as opposed to being dispersed among the word bands.

4.2 Phase 2 ARPG Corpus

The Action-Roleplay game-genre makes up 3.2 million of the 5.7 million tokens. 38 of the 100 listed words are distinctive of the ARPG genre (see Table 9 for ARPG wordlist, bolded and underlined are distinctive ARPG words from the 100 words). Removed words are listed in Table 11. The notable difference between the ARPG wordlist and the total game wordlist is that with 543 words (see Appendix K) (510 families) provided 9.63 % coverage of the four games. 0.47% of the 9.63% falls within the first 1-3K bands (see Table 11). 9.63 percent of coverage through a wordlist is in line with Coxhead's academic wordlist coverage which has 570 word families that covers 10% of her developed academic corpus of 3.5 million tokens (Coxhead, 2000). Considering that the academic corpus is nearly identical in size to the four games (3.2 million) and contains a near identical amount of coverage, this is a strong find for indication that action-roleplay games have a specific vocabulary. This could also mean that other game-genres have their own specific vocabulary as well, and that bundling them all together when conducting an analysis may not show specific vocabulary because the game language is far too generalized, as it is coming from a variety of different game-genres. This finding supports Schmitt's (2019) hypothesis that vocabulary is dependent upon the game-genre.

Table 10: Action-Roleplay Game Wordlist

Rank	Word	MR									
1	Melee	8.68	**31**	**Golem***	4.29	**61**	Volume	3.80	**91**	**Encounter***	3.44
2	Recoil	7.64	**32**	Unleash	4.25	**62**	**Grip***	3.80	**92**	Grime	3.43
3	Optional	6.50	**33**	Feat	4.24	**63**	Current	3.79	**93**	**Captive***	3.42
4	Foe	6.25	**34**	Thankfully	4.22	**64**	Regeneration	3.78	**94**	**Presence***	3.40
5	Nearby	5.91	**35**	Gameplay	4.20	**65**	Fiery	3.77	**95**	**Damned***	3.39
6	Retrieve	5.69	**36**	Deity	4.20	**66**	**Surface***	3.72	**96**	**Unknown***	3.38
7	Wary	5.41	**37**	Successfully	4.18	**67**	**Flourish***	3.71	**97**	**Resistant***	3.36
8	Infuse	5.36	**38**	Poison	4.17	**68**	**Task***	3.71	**98**	Skull	3.35
9	Relic	5.30	**39**	Sigh	4.17	**69**	Skilled	3.68	**99**	**Mettle***	3.35
10	Stagger	5.23	**40**	Rogue	4.15	**70**	**Grasp***	3.68	**100**	**Sadly***	3.33
11	Npc	5.20	**41**	Stealth	4.13	**71**	**Remnant***	3.67			
12	Useful	5.06	**42**	Glare	4.08	**72**	Leather	3.63			
13	Flee	5.05	**43**	Pillar	4.05	**73**	**Bastion***	3.62			
14	Flame	4.95	**44**	Path	4.04	**74**	**Ale***	3.62			
15	Agility	4.88	**45**	Autosave	4.02	**75**	Return	3.62			
16	Vigilant	4.81	**46**	Composure	4.00	**76**	**Within***	3.61			
17	Masterwork	4.80	**47**	Hunger	4.00	**77**	**Brow***	3.56			
18	Tutorial	4.77	**48**	Beneficial	4.00	**78**	**Pale***	3.55			
19	**Scholarly***	4.77	**49**	Quarry	4.00	**79**	**Makeshift***	3.53			
20	Power	4.62	**50**	Skeletal	4.00	**80**	Otherworldly	3.50			
21	Jagged	4.61	**51**	**Consuming***	4.00	**81**	**Watcher***	3.50			
22	Mist	4.60	**52**	Pause	4.00	**82**	**Rend***	3.50			
23	Fury	4.53	**53**	Madness	3.97	**83**	Rename	3.50			
24	Cloak	4.53	**54**	Remain	3.92	**84**	**Welp***	3.50			
25	Potent	4.51	**55**	**Shimmer***	3.89	**85**	**Silver***	3.50			
26	Brewing	4.50	**56**	**Brute***	3.88	**86**	Ensure	3.47			
27	**Cowl***	4.46	**57**	**Roar***	3.88	**87**	Vanish	3.45			
28	Assassin	4.40	**58**	**Sate***	3.86	**88**	**Monstrosity***	3.45			
29	Cleanse	4.37	**59**	**Attribute***	3.83	**89**	**Spear***	3.45			
30	Glow	4.34	**60**	**Anew***	3.80	**90**	**Truly***	3.45			

**Bolded and underlined signify distinctive ARPG words. Differences when compared to the Total Corpus Wordlist.*

Table 11: Removed Words from ARPG List

Word	MR Score	Reason for removal
Haha	6.18	Marginal Word
Iv	4.53	Acronym + Abbreviation of the word "four".
Pfft	3.45	Marginal Word

Table 12 Coverage of Action-roleplay game words

Word Band	TOKEN	TOKEN%	Cumulative Coverage
1,000	2401135	72.95	72.95
2,000	164537	5	77.95
3,000	80375	2.44	80.39
ARPG Wordlist (543 words)	301626	9.16	89.55
4,000	63031	1.92	91.47
5,000	42721	1.3	92.77
6,000	26780	0.81	93.58
7,000	16978	0.52	94.1
8,000	13597	0.41	94.51
9,000	9993	0.3	94.81
10,000	7327	0.22	95.03
11,000	6202	0.19	95.22
12,000	5080	0.15	95.37
13,000	3055	0.09	95.46
14,000	1781	0.05	95.51
TOTAL Tokens 3291265			

4.3 Phase 3 Coding

Evaluation coding is customized for specific studies (Pitman & Maxwell, 1992). The evaluation coding undertaken in this study is to look at the function of a word: whether it serves a game-purpose or narrative function. To do this, keywords were observed in context to determine their purpose. This was done by identifying the theme, the rheme, close-collocates,

and searches in the games themselves for what the purpose of the word used is by inferring from the surrounding text.

4.4 Total Corpus Coding

Evaluation Coding (1st Cycle of coding)

In the common game wordlist, there were far more diegetic words than there were ludic (see Appendix L for a list of ludic and diegetic words from the total corpus). This study labelled a word ludic if it had five or more occurrences as a ludic function. To give a visual representation of the words' properties, it was decided that this information would be made into a scale of ludic versus diegetic properties (see Appendix M for a visual representation and the number of concordances used in a ludic or diegetic function for the TOTAL corpus, and Appendix N for ARPGs). The red bar illustrates the ludic function, and the green, diegetic. The last column at the end informs if the word served primarily a ludic or diegetic function in the 10 samples, and the number beside it is the total tally of that word category. There were many less ludic words than originally anticipated (n = 20), whereas key words were used a lot more often for the purpose of diegesis (n = 80). For overtly-marked game words such as autosave, recoil, and gameplay, it is obvious that they will be recurring frequently in a ludic context as these are very prominent terms used in games to refer to constructs that are outside of the narrative.

The lack of ludic words likely suggests that gaming does not have a highly sophisticated technical wordlist that can be described, but more of general game language as mentioned earlier. It could also be that ludic functions are realized through other means than just the words

used in context. It could be possible that there is syntax, semantics, and unique game affordances that could dictate whether a word fits a ludic or diegetic use. Despite all this, only having 20 ludic words is good and simple for teaching and learning, especially for those students who are unfamiliar with gaming or are used to different game-genres. A main finding of this study is that most words can be used in a ludic context, but it is not the ability of them being used as a ludic word is what categorizes them, but rather, it is how frequent they appear in a ludic context that does. It is in fact arguable that these words are less warranted to be investigated, as frequency of exposure will allow learners to acquire these ludic words incidentally; rather, it is the words that are seldom used in a ludic context, but frequent enough to cause confusion that might warrant further investigation.

4.5 ARPG Coding

Evaluation Coding (1st Cycle of coding)

The only ARPG words that were coded were distinctive to the ARPG corpus. The ARPG-coded ludic words were very small in comparison (n = 5) (see Table 13 for ARPG ludic and diegetic words); however, this is because overlapping vocabulary from the total game wordlist was not investigated and because only the first 100 words were observed. If these words were further investigated, it would be possible to see if game-genre affects the variety of ludic words as well. For example, “attribute” is a very strong word associated to spendable skill points in the ARPG game-genre to invest and make characters stronger. How these words react and appear also differ. For example, different words, such as *consuming* and *resistant*, take different arguments: “consuming” is associated with items and their effects, whereas for the word

"resistant" could be used in relation to an item, a piece of equipment *equippable* by the player, or even a spell. These words have different arguments they can take within their ludic parameter.

Table 13: ARPG Ludic and Diegetic Words

Ludic	Diegetic
Consuming	Scholarly, Cowl, Golem, Shimmer, Brute, Roar,
Attribute	Sate, Anew, Grip, Flourish, Task, Grasp,
Surface	Remnant, Bastion, Ale, Within, Pale, Makeshift,
Brow	Watcher, Rend, Welp, Silver, Monstrosity, Spear,
Resistant	Truly, Encounter, Captive, Presence, Unknown, Damned, Mettle, Sadly

There were more diegetic words (n = 32) for the ARPG corpus as well which mirrors the result of the previous total game corpus coding. This makes sense, as regardless of the game-genre, games likely consist more of diegesis than ludology. A game would not be very fun for an audience if it were all parses of strings referring to game functions. As such, it is likely that the increased threshold of coverage needed for games reported by Rodgers and Heidt (in press) is likely due to the artifact in their methodology, and not due to the presence of game *metalanguage*. In other words, it is likely due to the holistic method of capturing all game text-data.

4.6 Pattern Coding (2nd Cycle) Total Corpus and ARPGs

For the pattern coding cycle, the words were re-evaluated, and their usage was coded for the ludic function it was serving and out of the total number of ludic concordances, how often it would be used in that sense (see Table 14 for these categorizations). When separating the words into functions, three general themes became apparent: 1) system-related vocabulary ('system'), 2) game-property related vocabulary ('game property'), and 3) game-information vocabulary

where the game is attempting to relay information to the player ('game information'). *System* vocabulary informs the player regarding the systems in place, in other words affordances from the medium that games provide, for example: saving the adventure in a save file, or disabling certain graphical settings, etc. *Game property* vocabulary are words that are universal game constructs that can be found in particular game-genres, for example, such as "melee damage" and the "recoil" of a bow or a gun. Finally, *game information* vocabulary are words that are being used to inform the player to do an action in relation to the game mechanics, such as informing the player that they can steal from an NPC engaged in dialogue. It is an indication made to the player through the game's voice informing of them of a *game mechanic* that can be performed. Game mechanics are features in games made by developers of the game to have players interact with, such as the above example of stealing an item that a non-playable character holds in their possession. By utilizing pattern coding, the usage of these words from the lengthy amount of concordances can become inferable and be used as a teaching material.

It is important to consider how the same word can be used differently, even when concerning two different ludic contexts. In the game *Monster Hunter World*, surfacing is a state that certain monsters can enact (to come out of the ground) and certain items known as *Screamer Pods* can force them to surface. However, in the game *Divinity 2*, *surface* is used to refer to the environment that the players are on which can be changed by skills and items. For example, a surface can be made icy and an enemy can slip, or a puddle of water can be turned into oil which can then be ignited. Based on the evaluations from the pattern coding, it became evident that it is not entirely the word that dictates a ludic or diegetic function but the context in which the word is used in the game as well. It may be harder for teachers to target ludic words for games they might have not played themselves.

While the distinction between ludology and diegesis serve an important understanding of game structure concerning vocabulary, how much it plays a role in comprehension is unknown. In the two instances above, simply knowing the diegetic context of the word *surface* is likely sufficient to understand their ludic functions for a monster surfacing, or a changeable environment. The concordances analyzed were but a small comparison to the entire corpus, but the purpose of this coding was to demonstrate how words can be come to be understood in these different contexts. If a player is unfamiliar with COTS games, they may not be able to comprehend adequately, how much this can be confirmed based on the existing literature is unknown.

Table 14: Pattern Coding of LUDIC game words

ORIGIN	Word type	Word	Additional Information	Occurrences
TOTAL GAME CORPUS	SYSTEM	Current	System informs the state of controls (buttons) or settings.	3/5
TOTAL GAME CORPUS	SYSTEM	Overwrite	System informs of a consequence for an intended action regarding saving or game settings/alterations.	10/10
TOTAL GAME CORPUS	SYSTEM	Disable	System informs that certain settings can be disabled.	7/8
TOTAL GAME CORPUS	SYSTEM	Rename	System informs items and objects can be renamed.	3/8
TOTAL GAME CORPUS	SYSTEM	Navigate	System informs on choosing options between controls and menus.	4/6
ARPG CORPUS	SYSTEM	Brow	System informs of parameters for a created character's brow.	5/5
TOTAL GAME CORPUS	SYSTEM	Autosave	System informs the player about the state of their progress.	10/10
TOTAL GAME CORPUS	GAME PROPERTY	Melee	A game property: melee damage. Collocates with "attacks" and "damage".	8/9
TOTAL GAME CORPUS	GAME PROPERTY	Recoil	A game property: gun/bow knockback. Associated to gun recoil. Collocates with close-gun related terms such as "scope, shooting, aim, suppressor, reload time".	7/7
TOTAL GAME CORPUS	GAME PROPERTY	Attacker	A game property: Refers to an in-game enemy.	6/6
TOTAL GAME CORPUS	GAME PROPERTY	Potent	A game property: Refers to strength of an effect or potion.	4/8
TOTAL GAME CORPUS	GAME PROPERTY	Regeneration	A game property: Has to do with the gradual recovery of a certain stat, such as "health" or "magic".	6/7
TOTAL GAME CORPUS	GAME PROPERTY	Resurrect	A game property: The ability to revive a player or NPC.	4/6
TOTAL GAME CORPUS	GAME PROPERTY	Projectile	A game property: A ranged weapon or ability – informs that an attack is aerial and distant.	5/5
TOTAL GAME CORPUS	GAME INFORMATION	Tutorial	Game informs player regarding tutorial settings and tooltips.	10/10
TOTAL GAME CORPUS	GAME INFORMATION	NPC	Game informs player on non-playable characters and their functions.	10/10
TOTAL GAME CORPUS	GAME INFORMATION	Gameplay	Game informs player about gameplay options and experiences.	9/9
TOTAL GAME CORPUS	GAME INFORMATION	Optional	Game informs player that the goal is optional and not required to progress the main campaign.	6/6
TOTAL GAME CORPUS	GAME INFORMATION	Dialog	Informs player about the state of "being in dialogue" and how this affects skills or inhibits actions.	4/7
TOTAL GAME CORPUS	GAME INFORMATION	Highlight	Game informs player about selectable objects.	5/8
TOTAL GAME CORPUS	GAME INFORMATION	Tier	Game informs player about beating a tier to unlock a reward.	3/6
ARPG CORPUS	GAME INFORMATION	Attribute	Game informs player on allocable skill points for character improvement or stats. Subject is the player.	6/6
ARPG CORPUS	GAME INFORMATION	Consuming	Game informs the player of an aftermath of item consumption	4/5
ARPG CORPUS	GAME INFORMATION	Surface	Game informs the state of a monster "surfacing" Game informs of the ground which is a changeable environment.	2/5 3/5
ARPG CORPUS	GAME INFORMATION	Resistant	Game informs to stats/status/equipment/item that has a resistant quality to physical damage, elements such as fire, and other ailments or statuses that are detrimental to the player.	7/7

Chapter 5: Discussion & Future Research

5.1 Implications for extramural activities

It has been reported that gamers will play games despite them not being available in their first language (L1) (Chik, 2011). There is a high intrinsic motivation to play games even if they are not within the player's L1 capabilities. This motivation stems from wanting to play the game and that could be harnessed as a motivating agent to involve more learners to engage in extramural English activities. Regardless if the learning is taking place because of a game, or because it is an extramural activity, the evidence suggests that it is still helpful and actively sought as a means to promote language acquisition. Considering that there has been prior research conducted showing how varied games are in their language (Thorne et al., 2012) and that the present research corroborates this claim, the argument can be made that games are more linguistically diverse than what is typically found in the classroom (Reinhardt, 2019).

With the research presented in this thesis suggesting game-genre does affect the variety of vocabulary found in games, it then suggests that Reinhardt's (2019) notion is correct and that gaming may include more diverse language that otherwise may be harder to encompass in the traditional classroom or even in other media such as books, songs, movies, television, etc. It may be more beneficial to use gaming as an extramural activity over other media for intermediate to advanced language learners to expand their vocabulary, as games have a wider range of lower frequency of vocabulary that may not be present in other media. Longitudinal studies should be conducted on adult language learners who are most likely to engage in this variety of extramural input and how it could be conducive to the acquisition of vocabulary. With the provided wordlists, learners can use them to target salient words. However, since there is evidence to

suggest that games are a good extramural activity by the literature and the research presented, the next logical question is: “What kind of games should learners play?”

The answer is: it depends. Games that do not have a lot of vocabulary, such as the popular Multiplayer Online Battle Arena (MOBA) game *League of Legends (LoL)* would likely not be ideal to play alone. The affordance that a game like LoL has is the communication that takes place during the game with the player’s teammates which could help develop language problem-solving and more fluent communication; as well, the threshold that is required for comprehension may be lower than single-player games that contain a lot of text, as spoken-conversation contains more high-frequency words as it requires knowledge of approximately 2,000 word families to reach 95% comprehension of spoken discourse (Adolphs & Schmitt, 2003). In other words, it is not the vocabulary itself found in the game, but rather, the act of communicating and negotiation of meaning through fail states that are helpful (Sykes & Reinhardt, 2013); therefore, when a learner picks up a COTS game, they must first consider whether they will be playing alone or with others for the intensive purpose of acquiring vocabulary. Other multiplayer games may be both rich lexically and communicatively, such as the MMOGs *World of Warcraft*, and *Final Fantasy 14*. However, the question of how players enjoy these games persist. It could very well be that dialogue and narration are skipped by players to simply enjoy the gameplay.

How learners approach unknown vocabulary is incredibly important. Do the learners have enough knowledge to infer from context the meaning of unknown words (i.e. 95 or 98 % comprehension of a text) (Hu & Nation, 2000), or would they use a dictionary to search for an unknown word? Would this disrupt their gaming experience? Without knowing how players would enjoy or play these games, especially those of whom are language learners, it becomes

very difficult to posit any hypotheses regarding of how beneficial they are in acquiring vocabulary outside of a theoretical potential. How players enjoy their games gives credence to Bawa's (2018) request for the development of gamers playing an MMOG corpus to study. More longitudinal studies regarding language learners and how they play games while taking into consideration their language learning beliefs and practices of any kind of game, whether it be single-player, multiplayer, or MMOG, is incredibly important and warranted to better understand the actions taken on behalf of learners for vocabulary acquisition and growth.

What is observable from this research is that learners who favour a particular game-genre, such as ARPGs, may benefit more over a longer period of time in terms of acquisition of words, since there is evidence to suggest by the present research that certain game-genres share a certain level and type of vocabulary. If a learner plays all sorts of different games, they may not receive the same benefit of vocabulary transfer from the previous game they played. For example, while ARPGs may share a certain amount of vocabulary that is similar, there is still a lot that are not shared in the game-genre between games. By playing the same game-genre, ideally, learners should become acclimatized to this vocabulary which can then help them reach 95% and 98% milestones faster, as opposed to shifting to a different game each time. The potential benefits of learning through related game play has been established in research into the language in related television programs (Rodgers & Webb, 2011). This is not to say that it would not be beneficial to still play different game-genres as an extramural activity, but the rate of progression for acquiring new vocabulary may be slower if a focus is not given to the same particular game-genre. This idea falls into the extended playing idea that stemmed from Krashen's extended reading hypothesis (Mason & Krashen, 1997). By being exposed to an input with considerable time involved, acquisition of vocabulary is likely to take place. However, if a

player enjoys a wide variety of games and does not stick to one game-genre consistently as a language learner, there may not be enough time spent to actively acquire specific game vocabulary, such as ARPGs.

How much time a player spends on a game is important to consider as well. If the learner is spending two hours a day on gaming, how much of an advantage they are getting over studying could be minimal. While it may seem unhealthy to encourage more extended game playing, that negative connotation may stem from false perceptions regarding games as time-wasting activities. Conversely, if it were reading, many teachers and proficient learners would likely actively encourage more reading, despite evidence suggesting that gaming and social media usage are stronger indicators for vocabulary proficiency (De Wilde et al., 2019). Interestingly enough, reading was once thought to be a time-wasting activity as can be seen from a published opinion piece in the *Telegraph* entitled "too many young people are reading novels" ("too many young people are reading novels", 1907).

As is with anything, how the medium is used is at the pinnacle of good vocabulary acquisition, and the affordances that the media possess can be molded for language learning and teaching. Although, some considerations are worth nothing: the extended exposure necessary to obtain significant vocabulary gains may be more difficult with gaming, since at least with reading, there is constant exposure to vocabulary, whereas not every scene contains words in games. This is not universal to all games, but rather depends on the specific game and the game-genre; however, like movies, games benefit from imagery which may aid in acquiring key vocabulary faster than other media, especially considering other game-affordances that may be available, such as inspecting an item for further description which may give collocates, synonyms, or other words or linguistic hints that may be conducive in acquiring vocabulary. It

would be beneficial in a User Experience (UX)-design perspective study to see how games can be utilized to help with the acquisition of vocabulary; that is to say, how the game is laid out through its menus and interfaces to potentially help with acquiring target vocabulary, as is indicated by trending game user and UX research where "novel game user research methods are promising for getting insights into learning and decision making in games" and "where the goal of the game is not to play, but to learn something or motivate a behaviour" (Drachen et al., 2018, p. 514).

5.2 Implications for teaching

While video games are an ideal source of input to learn language from according to the literature, its practice of being incorporated in the classroom remains difficult. Studies such as Coleman (2002), and researchers who have adapted video games from pedagogical principles, such as Reinhardt and Sykes (2013), have offered suggestions on how to use their tasks and activities. With various stakeholders, curricula, and other variables that are outside the teacher's control, it may be difficult to justify "playing games" in the classroom despite being an ideal source of language input to acquire vocabulary from. More pedagogical materials may need to be designed while considering curriculum goals. In certain English classes, learners may better benefit from the nuance of some games' vocabulary. For example, learners in the field of Information Technology who are learning ESL/EFL may benefit from the game *PC Building Simulator* on how to create technical reports and manuals, or Business students from *OpenTTD,* which is a business simulation game. With more research on the horizon of gaming and its efficacy in L2 acquisition, more niche uses for games should come to mind when designing activities to take advantage of gaming affordances while still considering lexical coverage

needed for 95% and 98%. The current research adopted a corpus approach and revealed unique vocabulary that can be attributed to game-genre. Now that there is knowledge of potential lexical merit associated to gaming, video gaming affordances in classroom teaching should be further investigated to give practitioners a better idea of how to incorporate gaming in not only a pedagogically-grounded method, but in an efficient and easy way to encourage their use in the classroom.

One important aspect that seems to have been very apparent in the literature is to not cognitively overload learners who are playing games in the classroom. Some studies that have involved single-player games have either had backseat players benefit more from the game than the actual players playing the game, or had not received results of the game being extremely conducive to learning over other forms of media (Hitosugi et al., 2014; Chen & Yang, 2013; Ranalli, 2008). When picking a game there are many important elements to consider: 1) What is the language being taught? Is there any associated difficulty to incidental or intentional acquisition from the marked features of the language being taught? And what is the goal of introducing the game through the language, is there a specific set of vocabulary that is ought to be acquired? 2) Is the game within a reasonable lexical difficulty or simplicity for learners? In other words, is it too easy or difficult? 3) Are learners overburdened with tasks or learner compendiums? Are they conducive to aiding learners with language, or are they a distraction (i.e. are they being helpful)? 4) Are backseat gamers staying on task? Is the player staying on task? When in single-player environments, is the single-player the only one playing, or are players who are in the backseat playing as well? 5) is the game thematically related to what is being taught, or can it be molded to be used for what needs to be taught? (cf. Coleman, 2004).

In foreign-language learning environments, learners may benefit more from multiplayer games or MMOGs that can work on communicative competency and negotiation of meaning through tasks. In an L2 learning environment where these occur naturally outside the classroom, and learners may be more proficient, perhaps encouraging students in the classroom to journal about their gaming experiences might get them more involved with a second language input to more depth which may be more conducive to their journey as second language learners. With extended playing being the idea behind good vocabulary acquisition for reading, watching, and playing, games may benefit more as an extramural input if there are some serious considerations given to their adaptation in the classroom. Just as reading can be encouraged in the classroom from altering a passage in a book into a task, or watching a movie into facilitating a discussion, games can be used similarly. For example, the affordance of a 3D space can be used to incorporate a stronger pedagogical component to using a game.

5.3 Potential Research Directions

How much language do players engage with in games? It is difficult to ascertain vocabulary growth if there is this shared perspective that learners all play games the same way. Contemporary game-design allows for open ended branching of paths in the game, and this as a result, contributes to players interacting in *play* through different ways. This means that how one player acquires target vocabulary and which variety of vocabulary is dependent upon not only the game but how they engage with the game. If one player decides to read the books present in a game versus skips over them, the player will be exposed to different amount of words. Understanding how players interact with games is paramount and should be at the foreground of

this kind of research. This can be done by using qualitative reports, eye-tracking, and quantifying time spent in different segments or areas of the game. Understanding how game time is spent is key to understanding how gaming affordances can be used for learning.

5.4 Ludic and Diegetic words in the CLASSROOM and EXTRAMURAL activities

Some of the limitations regarding how ludology and diegesis are ascertained in this study are mentioned in the limitations segment of this chapter, and more specifically how the distinction may not be as important as originally conceived. Arguably, exemplar game-design disguises ludic functions as a part of diegesis. I refer to this as the *ludic narrative*, but it is more commonly referred to as the *ludonarrative* of a game (Aarseth, 2012). The ludic narrative is the idea that game-design and storytelling are two sides of the same coin; this construct can range from being how space is organized in a game, for example whether characters follow a linear path (as an example), to the way that language is used, such as tutorial segments of a game. An example of which being that the tutorial segment is segued as part of the game's narrative for seamless immersion of teaching the player while still being part of the narrative. There has been contention whether games fall into camps of ludic gameplay or if they can even be referred to containing a narrative according to narrative theory (Aarseth, 2012); however the research conducted suggests that it is an amalgam of the two, as words can be found from off-lists such as vehicle modifications from GTA V that incorporate game-features, and words on-list that highlight diegesis such as lore-building from books in Skyrim and dialogue from Divinity 2. The high-rate of reliability for coding among the IRs demonstrated a high understanding for ludology among gamers, however, the data did not show a high amount of ludic words, rather suggestive that games in the corpora are more narrative-based than ludic-based. The listed ludic words in

the present research could be used by learners who are unfamiliar to extramural gaming and since there are only 20 key ludic words, they can be acquired relatively fast.

It was originally argued that ludic words were important for gaming, or that was the initial assumption at least – however, words are still words. Learning any new word, regardless if it is used in a ludic context, as part of diegesis, or even part of the ludic narrative, it is still a word that is potentially unknown to a learner. The only contention that can be made regarding this point is if ludology can interfere with learners' understanding of the game's objective or goals. If ludology impeding comprehension can be proven, then more attention may need to be given to the teaching of ludic words for language learners who wish to engage in extramural games for language learning or for teachers who wish to incorporate COTS games into the classroom. The perceived cognitive overload phenomenon mentioned in previous studies could be evidence that there is a dissonance between ludology and the narrative, where learners are confused by the goals of the game, as opposed to being overburdened by the demands of the game and task (cf. Sundqvist & Wikstrom, 2015; cf. Hitosugi et al., 2014; cf. Chen & Yang, 2013; cf. deHaan, Reed, & Kuwada, 2010). More research should be conducted in how ludology and diegesis affect vocabulary retention for language learners and whether it affects comprehension and if this warrants further investigation.

5.5 Limitations

There are many limitations present within the current study. However, being an innovative study in vocabulary research and gaming, I believe that my thesis has provided valuable insight into how these wordlists can be further refined and included into gaming

research. Since the original corpus was encoded in word families, to make an accurate comparison of how much coverage these words provide for the entire corpus, word families must be used even if it assumes an overage in the amount of coverage. One of the issues working with script data from a game is that not all of it derives from the game itself. There are rare occurrences while working throughout the data where programming notes from the game developers had remained in the text files despite parameters set through RegEx. Due to the sheer volume of the corpus, it was impossible to go through all of these concordances, identify, and remove all these notes given the time-frame for this thesis. Another limitation is that some common nouns were also used as proper nouns, such as "Brute" for a name, or "Shimmer" as a name of a blade. This, however, is a common issue with corpus and lexical research.

The coding was done with the assumption that a word served a ludic or diegetic function, however, it is more likely that a formulaic utterance serves such a function. For example: "Your Argonian blood is 50% resistant to disease" [Skyrim] (see Table 15). These words are working together to identify that there is a reduction of a status ailment that occurs within the game, known as *disease* (i.e. reducing a negative effect, known as disease, on your character). Another example from the same game: "Nords are 50% resistant to cold damage" and follows the exact same pattern.

Table 15: Example of [resistant] used in a syntactic context [Skyrim]

1. Your argonian blood {parameter} is {copula – to be, linking} 50% {quantifying number} resistant to {reduces affliction} disease {type of affliction}
2. Nords {parameter} are {copula – to be, linking} 50% {quantifying number} resistant to {reduces affliction} cold damage {type of affliction}

It can be inferred when using words from surrounding context that there is a syntactic trend (see Table 16). There is likely more to ludology and diegesis than the word level. In the example given in the introduction, ‘recoil’ does not infer to the recoil of a gun automatically. It has close-collocates such as ‘aim, shooting, scope, suppressor’, etc. Keyness can calculate for collocates, but it is unable to calculate for longer formulaic chunks and syntactic breakdowns. Investigating these concepts are warranted to better understand how language serves a game and/or narrative function, as opposed to observing it on the word-boundary level. Keyness is helpful for true collocates, but not close-collocates. Close-collocates in the sense that they are words that occur within proximity to one another and not directly next to each other. This is beyond the scope of this paper, and as well my area of expertise.

Table 16: Example of [infuse] used in a syntactic context [Monster Hunter World]

used to give this iron sword & shield used to imbue this iron great sword used to infuse this iron bow with {verb} {verb infinitive} {determiner} {item} ({property})

Words from game language scripts were at times hard to discern, since they represent all language, present at any place in the game. For example, a Heads-Up Display (HUD) is a visual indicator showing the player’s status, health, and perhaps other variables pertinent to the game at hand. In the files, there may just be a word isolated by itself listed as “health”. It is not possible to determine where it is in the game, as there are no surrounding words or syntactic structures that are related to that word. Naturally, if the word’s location was not inferable, it was cut from the analysis. This is a limitation based on the data; it was sometimes impossible to ascertain how the word was used.

There is also an issue with how I had demonstrated coverage in this paper. Originally, the Lexical Frequency Profile conducted by Rodgers & Heidt (in press) used word families to compare to other LFP studies. When conducting keyness, this study had switched to flemmas with the intended principle behind it being to give a more reasonable idea of coverage; however, when conducting coverage of these key words, they were used in their word family forms again. There is a big disparity between the consistency of how this study was conducted, but there were two main reasons in why this was done.

1) The literature stipulates that keyness is better conducted through Lemma form than word form (Paquot, 2007). Lemma and Flemma suffixes are the same, while the only difference being PoS tagging.
2) The original study Rodgers & Heidt (in press) and other LFP studies report in word families. This research cannot be adequately compared if it remained in flemmas.

There has been recent debate regarding the discrepancy in coverage between word families and lemmas. Laufer & Cobb (2019) state that the discrepancy between the two is "exaggerated and further that there is little reason to reconsider the large amount of useful and influential research that is based on the word family as the unit of counting" (p.26) when considering for common affixes; however, a recent article has emerged from Stoeckel et al (2020) that informs the state of assessing for receptive vocabulary knowledge and provides evidence that affixes beyond the 2K band are not easily inferable by learners. The point of contention regarding word counting metrics is beyond the scope of this thesis, as the conversion from word family to flemma was done to have a more accurate keyness analysis, and the converting back to word families was done was to create a point for comparison with other research.

Another limitation was the lack of tags in the corpus. Had the corpus been tagged for its parts of speech (i.e. lemmas), it would have been clearer on how these words were used in different contexts across different games and their genres. Consider the word ‘surface’. In Divinity 2, ‘surface’ is used as a noun, but in Monster Hunter World, it is used as a ‘verb’. By having these words tagged for their PoS, it would give insight into their different meanings. It would be able to list surface as a ludic noun to refer to the ground in Divinity 2, and list surface as a ludic verb to refer to the state of rising out of the ground in Monster Hunter World. By doing so, words could be more accurately pinpointed to their usage across the games.

The final limitation of this study is that it considered the distinction between ludic and diegetic words as an absolute certainty across different game-genres. For example, the word “flee”: “A monster will flee when hit” [Monster Hunter World]. This utterance illustrates that the monster will run away if it is attacked. However, while “flee” may serve a ludic function in Monster Hunter World, the use of the word was part of the narrative for others. It is important to consider the game-genre, and perhaps even the specific game, when analyzing for ludic or diegetic properties. Another example, in the game Divinity 2, there are weapons and equipment that have a skill name added to their equipment name. For example, “spear of bloodletting” [Divinity 2]. In this context, the word *spear* does not only indicate to the player that the spear is a weapon that can be used, but bloodletting is a usable skill that is given from using the spear. Normally, a player would have to spend attribute points to obtain the skill; however, with the spear equipped, regardless of the attributes and stats, the player can use the skill. Bloodletting is a common skill in the game through the name alone it is easily inferable to an experienced player that this weapon has a ludic function that allows them to use the skill if the weapon is equipped. To a novice, it is unlikely that they would make the connection.

Chapter 6: Conclusion

By conducting the present research, I was able to find evidence that suggests there is a distinctive vocabulary associated to ARPG games and have presented how various game words fall into ludic and diegetic contexts to aid with teaching and learners' incidental acquisition of the target vocabulary. This research has answered Schmitt's (2019) call and has opened a lot of doors regarding suggestions on how to conduct future research concerning gaming. Through the multimethod used in this study, it was also possible to see how the distinction between these words can be used in games and that this distinction warrants further investigation in whether it affects learner comprehension when playing games. When comparing all the games together, there was no indication to a specific "game metalanguage" vocabulary or language indicative of games; however, when researching the ARPG genre – it became inferable that there may be a distinctive vocabulary as hinted by the lexical coverage numbers. By giving some context to the ludic and diegetic words, practitioners can use these lists to teach ludic words to learners who may not have had frequent exposure to games and can encourage their use, as games have been shown to be a good extramural source of language to acquire vocabulary from.

Further research should be conducted on how learners engage in games, how communication is conducted in multiplayer games between players, and whether ludology or diegesis affect learners' comprehension of a given game or game-genre. This research can be done by performing longitudinal studies both in the classroom and outside through extramural activities. With short experimental-design game studies being too contrived without understanding the broader processes of game-design and its affordances, we will never be able to fully understand how beneficial games are for learners as an extramural activity or to be used as

a teaching material in the classroom. We first must better understand the games themselves and how they can aid in vocabulary acquisition and retention. While research suggests a strong link between gaming and language proficiency, we do not have a good enough understanding yet to ascertain gaming's true value. What has been inferable from this study is that 1) games provide a rich range of vocabulary as is seen by their frequency profiles, 2) they are likely more advantageous for intermediate to advanced learners of the language, 3) that if they are adapted into the classroom they must follow careful planning on how they can be utilized to promote vocabulary acquisition, 4) that they are not simply "ludic games", and that 5) certain varieties of games may have their own kind of vocabulary that can be acclimatized if players play games of a similar game-genre over time.

References

Aarseth, E. (2012). A narrative theory of games. Proceedings from*: The International Conference on the Foundations of Digital Games*. 129-133. https://doi.org/10.1145/2282338.2282365

Adolphs, S., Schmitt N., Lexical Coverage of spoken discourse. *Applied Linguistics*, *24*(4), 425–438. https://doi.org/10.1093/applin/24.4.425

Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology*, *5*(1), 202-232.

Anthony, L. (2019). AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/softwareKeyworder person, 2012.

Anthony, L. (2014). AntWordProfiler (Version 1.4.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, *6*(4), 253–279. https://doi.org/10.1093/ijl/6.4.253

Bawa, P. (2018). Massively multiplayer online gamers' language: Argument for an m-gamer corpus. *The Qualitative Report*, *23*(11), 2714-2753.

Bergstrom, K., Fisher, S., & Jenson, J. (2016). Disavowing 'that guy' identity construction and massively multiplayer online game players. *Convergence*, *22*(3), 233-249. https://doi.org/10.1177/1354856514560314

Blume, C. L. (2019). Investigating digital game-based language learning: applications, actors and issues of access (Unpublished doctoral dissertation). Leuphana University of Lüneburg, Germany.

Brezina, V., & Gablasova, D. (2017). How to produce vocabulary Lists? Issues of definition, selection and pedagogical aims. A response to Gabriele Stein. *Applied Linguistics*, *38*(5), 764–767. https://doi.org/10.1093/applin/amx022

Brown, D. (2012). The frequency model of vocabulary learning and Japanese learners. *Vocabulary Learning and Instruction*, *1*(1), 20-28. http://dx.doi.org/10.7820/vli.v01.1.brown

Brysbaert M., Stevens M., Mandera P., & Keuleers E. (2016) How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology 7*, 1116. https://doi.org/10.3389/fpsyg.2016.01116

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990. https://doi.org/10.3758/brm.41.4.977

Castagno, R., Shin, M., & Tanaka, Y. (2018). Teaching English for real-life communication with movie reviews: A corpus approach. *TESOL Working Paper Series*, *16*, 19-39

Castelvecchi, S. (2020). On "diegesis" and "diegetic": Words and concepts. *Journal of the American Musicological Society*, *73*(1). https://doi.org/10.1525/jams.2020.73.1.149

Chen, H., & Yang, T. (2013). The impact of adventure video games on foreign language learning and the perceptions of learners. *Interactive Learning Environments*, *21*(2), 129–141. https://doi.org/10.1080/10494820.2012.705851

Chik, A. (2011). Learner autonomy development through digital gameplay. *Digital culture & education*, *3*(1), 30-44.

Chung, T. M., & Nation, P. (2004). Identifying technical vocabulary. *System*, *32*(2), 251-263. https://doi.org/10.1016/j.system.2003.11.008

Csikszentmihalyi, M. (1990). *Flow: the psychology of optimal experience* (1st ed.). Harper & Row.

Drachen, A., Mirza-Babaei, P., & Nacke, L. (2018). Frontlines in Games User Research. In Drachen A., Mirza-Babaei, P., & Nackle, L. (Eds.), *Games User Research*. Oxford University Press (pp 510 – 520). https://doi.org/10.1093/oso/9780198794844.001.0001

Cobb, T., & Horst, M. (2011). Does Word Coach coach words? *CALICO Journal*, *28*(3), 639–661. https://doi.org/10.11139/cj.28.3.639-661

Coleman, D. W. (2002). On foot in Sim City: Using Sim Copter as the basis for an ESL writing assignment. *Simulation & Gaming*, *33*(2), 217-230. https://doi.org/10.1177/1046878102332010

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–238. https://doi.org/10.2307/3587951

Creswell, J. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches.* SAGE Publications.

Crystal, D. (2001). *Language Play.* University of Chicago Press.

Dang, T., Coxhead, A., & Webb, S. (2017). The academic spoken wordlist. *Language Learning*, *67*(4), 959–997. https://doi.org/10.1111/lang.12253

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2019). Learning English through out-of-school exposure. Which levels of language proficiency are attained and which types of input are important? *Bilingualism, 23*(1), 171–185. https://doi.org/10.1017/s1366728918001062

deHaan, J., Reed, W., & Kuwada, K. (2010). The effect of interactivity with a music video game on second language vocabulary recall. *Language Learning & Technology*, *14*(2), 74–94. http://dx.doi.org/10125/44215

Dongwan., R. (2013). Play to learn, learn to play: Language learning through gaming culture. *ReCALL, 25*, 286-301. https://doi.org/10.1017/S0958344013000050

Dörnyei, Z. (1994). Motivation and motivating in the foreign language classroom. *The Modern Language Journal (Boulder, Colo.)*, *78*(3), 273–284. https://doi.org/10.1111/j.1540-4781.1994.tb02042.x

Ducheneaut, Nicolas & Moore, Robert. (2005). More than just 'XP': Learning social skills in massively multiplayer online games. *Interactive Technology and Smart Education, 2*, 89-https://doi.org/10.1108/17415650580000035

Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. J. (2007). The life and death of online gaming communities: a look at guilds in world of warcraft. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 839-848. https://doi.org/ 10.1145/1240624.1240750

Elliott, V. (2018). Thinking about the coding process in qualitative data analysis. *The Qualitative Report*, *23*(11), 2850–2861.

Entertainment Software Association of Canada (2018). The Canadian video game industry. http://theesa.ca/wp-content/uploads/2018/10/ESAC18_BookletEN.pdf

Entertainment Software Association of Canada (2016). Essential Facts. http://theesa.ca/wp-content/uploads/2016/11/2016_booklet_Web.compressed2.pdf

Gabrielatos, C. (2018). Keyness analysis: Nature, metrics and techniques in book: Taylor, C. & Marchi, A. (eds.) Corpus Approaches to Discourse: A critical review. Oxford: Routledge., Chapter: 11, Routledge, pp. 225-258

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, *35*(3), 305–327. https://doi.org/10.1093/applin/amt015

Gee, J. (2007). *What video games have to teach us about learning and literacy.* Palgrave Macmillan.

Gee, J. P. (2004). Situated language and learning: A critique of traditional schooling. London: Routledge.

Gee, J. P. (2006). Are video games good for learning? *Nordic Journal of Digital Literacy*, *1*(03), 172-183.

Godwin-Jones, R. (2014). Games in language learning: Opportunities and challenges. *Language Learning & Technology*, *18*(2), 9-19. http://dx.doi.org/10125/44363

Hitosugi C. I., Schmidt., M., & Hayashi K. (2014). Digital game-based learning (DGBL) in the L2 classroom: The impact of the UN's off-the-shelf videogame, Food Force, on learner affect and vocabulary retention. *CALICO Journal*, *31*(1), 19–39. https://doi.org/10.11139/cj.31.1.19-39

Hu, M., & Nation, I.S.P. (2000). Vocabulary density and reading comprehension. Reading in a Foreign Language, *13*(1), 403–430

John Brewer, & Albert Hunter. (2006). The Multimethod Approach and Its Promise. In *Foundations of Multimethod Research*. SAGE Publications, Inc. https://doi.org/10.4135/9781412984294.n1

Juul, J. (2010). *A casual revolution: Reinventing video games and their players*. MIT press.

Krashen, S. (1981). *Second language acquisition and second language learning* (1st ed.). Pergamon Press.

Laufer, B. & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: the construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1–26. https://doi.org/10.1093/applin/22.1.1

Laufer, B., & Cobb, T. (2019). How Much Knowledge of Derived Words Is Needed for Reading? *Applied Linguistics, 0(*0*),* 1–29. https://doi.org/10.1093/applin/amz051

Mackey, A., & Gass, S. (2005). Second language research: Methodology and design. In Second Language Research. Erlbaum. https://doi.org/10.4324/9781410612564

Mason, B. & Krashen, S. (1997). Extensive Reading in English as a Foreign Language. System *25*(1), 91-102. https://doi.org/10.1016/s0346-251x(96)00063-2

Masso, C. I. (2009). Developing a methodology for corpus-based computer game studies. Journal of Gaming & Virtual Worlds. https://doi.org/10.1386/jgvw.1.2.143/7

McArthur, V. (2020). Kanji Monogatari: Design considerations for a visual storytelling approach to Japanese language learning in augmented reality. *REPLAYING JAPAN*, (2), 95-104. https://doi.org/10.34382/00013366

Milička, J. (2012). Minimal Ratio: An exact metric for keywords, collocations etc. *Czech and Slovak Linguistic Review 1*.

Nation, I. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*(1), 59-82. https://doi:10.3138/cmlr.63.1.59

Nelson, M. (2000). *A Corpus-based study of business English and business English teaching materials*. Unpublished PhD Thesis. Manchester: University of Manchester.

NOP World Culture. NOP World Culture Score index examines global media habits: Uncovers who's tuning in, logging on and hitting the books. (2005). *PR Newswire*.

Norman, D. (1988). *The design of everyday things.* Doubleday.

O'Keeffe, A. (2012). Vocabulary instruction. In Anne Burns & Jack C. Richards (Eds.), *Pedagogy and Practice in Second Language Teaching*. Cambridge Press.

Pinchbeck (2017). Vocabulary use in academic-track high-school English literature diploma exam essay writing and its relationship to academic achievement (10630422) [Doctoral Dissertation, University of Calgary] ProQuest.

Pitman, M. A., & Maxwell, J. A. (1992) Qualitative approaches to evaluation: Models and methods. In M. D. LeCompte, W. L. Millroy, & J. Preissle (Eds.), The handbook of qualitative research in education (pp. 729–70). San Diego: Academic Press.

Ranalli, J. (2008). Learning English with The Sims: Exploiting authentic computer simulation games for L2 learning. *Computer Assisted Language Learning*, *21*(5), 441-455. https://doi.org/10.1080/09588220802447859

Reinders, H., & Wattana, S. (2015). Affect and willingness to communicate in digital game-based learning. *ReCALL*, *27*(1), 38–57. https://doi.org/10.1017/S0958344014000226

Reinhardt, J. (2019). Gameful L2 Learning. In *Gameful Second and Foreign Language Teaching and Learning* (pp. 101-139). Palgrave Macmillan, Cham.

Rodgers, M. & Heidt, J. (in press). Levelling up comprehensible input and vocabulary learning: The lexical profile of video games. In Werner, V. & Tegge, F. (Eds.), *Pop culture in language education*. Routledge.

Rodgers, M., & Webb., S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly*, *45*(4), 689-717. https://doi:10.5054/tq.2011.268062

Rodgers, M.P.H, (2018). The images in television programs and the potential for learning unknown words. ITL-International Journal of Applied Linguistics*, 169*(1), 191-211. https://doi.org/10.1075/itl.00012.rod

Saldaña, J. (2015). *The coding manual for qualitative researchers* (3rd edition). SAGE.

Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching*, *52*(2), 261–274. https://doi.org/10.1017/S0261444819000053

Statistics Canada (2010). General Social Survey: Time use. https://www150.statcan.gc.ca/n1/daily-quotidien/110712/dq110712b-eng.htm

Sundqvist, P. (2019). Commercial-off-the-shelf games in the digital wild and L2 learner vocabulary. Language Learning & Technology, 23(1), 87–113. https://doi.org/10125/44674

Sundqvist, P., & Sylvén, L. (2014). Language-related computer use: Focus on young L2 English learners in Sweden. *ReCALL*, *26*(1), 3–20. https://doi.org/10.1017/S0958344013000232

Sundqvist, P., & Sylvén, L. (2014). Language-related computer use: Focus on young L2 English learners in Sweden. *ReCALL (Cambridge, England)*, *26*(1), 3–20. https://doi.org/10.1017/S0958344013000232

Sundqvist, P., & Sylvén, L. K. (2016). *Extramural English in teaching and learning: From theory and research to practice*. Springer.

Sundqvist, P., & Wikstrom, P. (2015). Out-of-school digital gameplay and in-school L2 English vocabulary outcomes. *System*, *51, 65-76*. https://doi.org/10.1016/j.system.2015.04.001

Sundqvist, P. (2011). A possible path to progress: Out-of-school English language learners in Sweden. In *Beyond the language classroom* (pp. 106-118). Palgrave Macmillan, London.Sykes, J, & Reinhardt, J. (2013). *Language at play.* Pearson.

Sykes, M.J., Reinhardt J. (2013). *Language at Play*: *Digital games in second and foreign language teaching and learning*. Pearson Education.

Sylvén, L., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL*, *24*(3), 302–321. https://doi.org/10.1017/S095834401200016X

Stoeckel, T., McLean, S., & Nation, P. (2020). Limitations of size and levels test of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, 1–23. https://doi.org/10.1017/S027226312000025X

Tegge, F. (2017). The lexical coverage of popular songs in English language teaching. *System*, *67*, 87-98. https://doi.org/10.1016/j.system.2017.04.016

Thorne, S., Fischer, I., & Lu, X. (2012). The semiotic ecology and linguistic complexity of an online game world. *ReCALL (Cambridge, England)*, *24*(3), 279–301. https://doi.org/10.1017/S0958344012000158

Thorne, S. (2008). Transcultural Communication in Open Internet Environments and Massively Multiplayer Online Games. In S. Magnan (ed.), Mediating Discourse Online (pp. 305-327). John Benjamins.

Too many novels are read by young people. (1907, January 7th). The Telegraph. Retrieved from: https://news.google.com/newspapers?id=DhRAAAAAIBAJ&sjid=LqQMAAAAIBAJ&pg=1195%2C5718774

Webb, S., & Rodgers, M. (2009b). The lexical coverage of movies. *Applied Linguistics*, *30*(3), 407-427. https://doi.org/10.1093/applin/amp010

Webb, S. & Rodgers, M.P.H. (2009a). Vocabulary demands of television programs. *Language Learning*, *59*(2), 335-366. https://doi.org/10.1111/j.1467-9922.2009.00509.x

Williams, M., & Burden, R. (1997). Motivation in language learning: a social constructivist approach. *Cahiers de l'APLIUT*, *16*(3), 19–27. https://doi.org/10.3406/apliu.1997.1201

Appendices

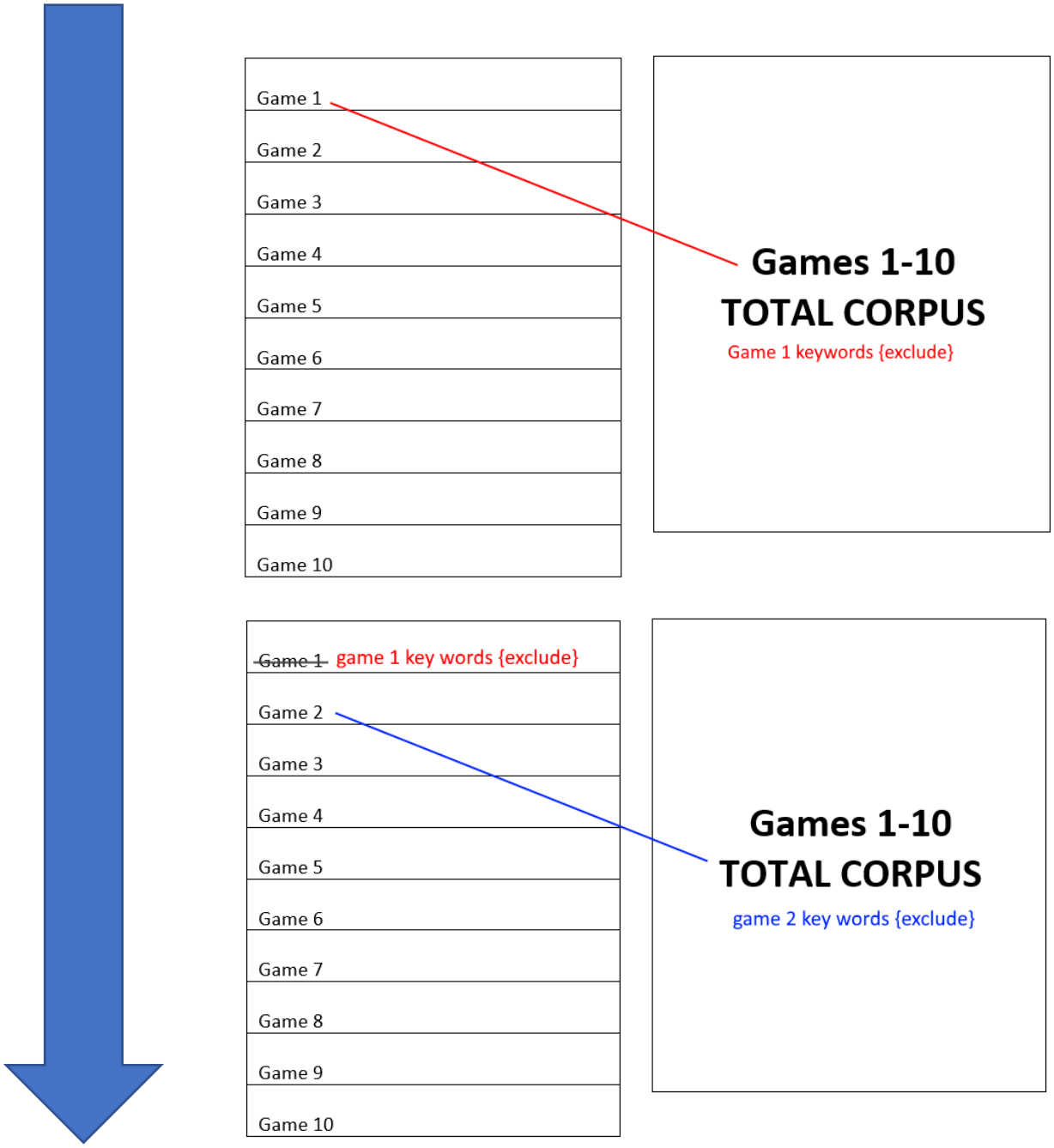
Appendix A: Types of Multiplayer Games

Label	*Physical Space*	*Electronics*	*Player Limit*	*Affordance*
Cooperative Play	Same immediate physical space (i.e. same room)	Same electronic device	2 – 4 players	Personal Computer or Console (Hardware)
Local Cooperative Play	Same physical space (e.g. same room, same house)	Different electronic devices	4 or so players	Local Area Network (LAN) – shared connection through the same Internet modem/router
Online Cooperative Play	Same or different physical space	Different electronic devices	4 or so players	The Internet
	Different physical space (e.g. different houses)	Different electronic devices	Practically limitless	The Internet

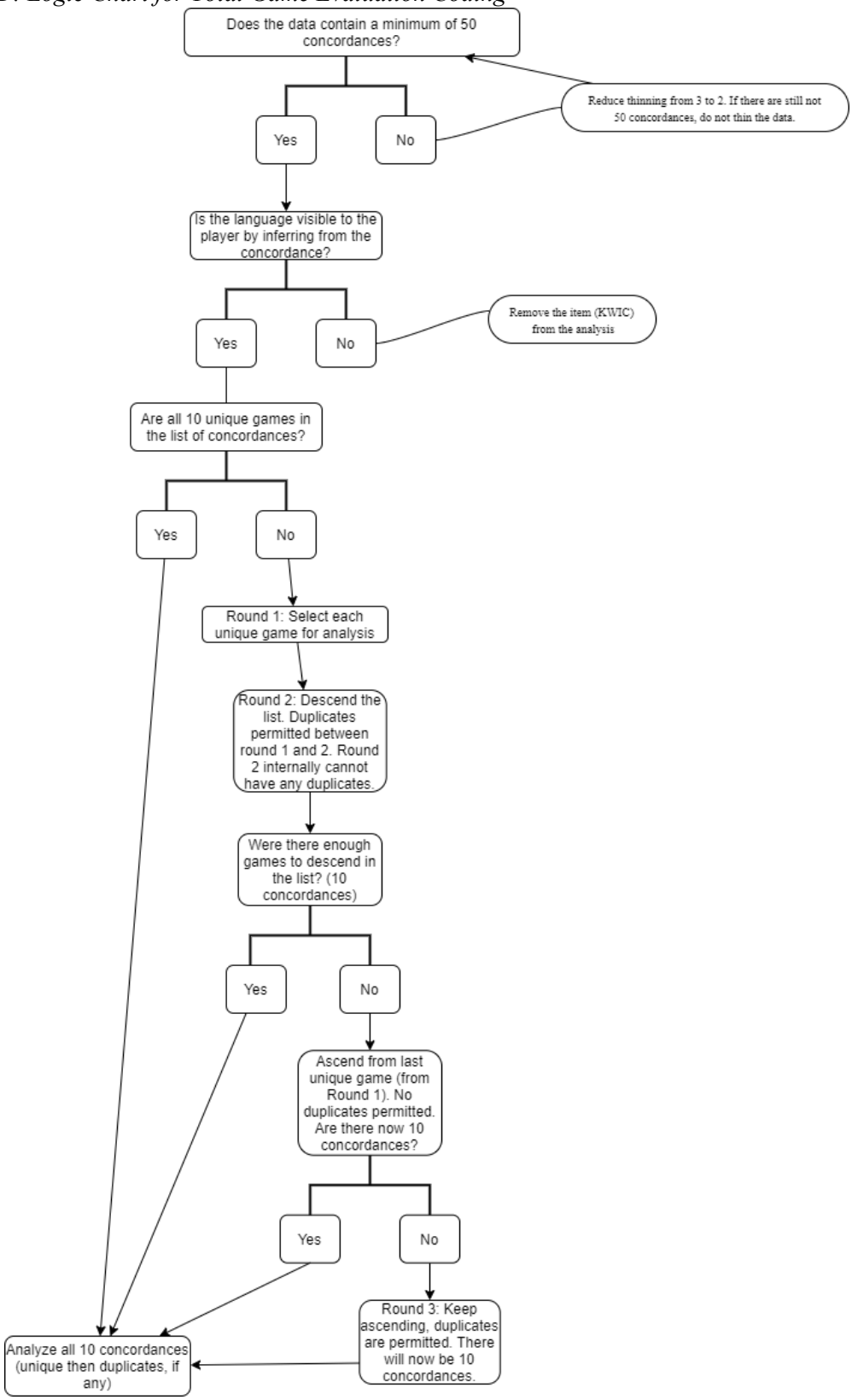
Appendix B: Walking Dead (study corpus) compared to Total Corpus

Type	Minimal Ratio	Total Game Corpus Freq 5.7 million tokens	Walking Dead Freq 221,531 tokens	Lower confidence Limit	Upper confidence Limit
Lee	19.03	20	1085	30	57
Kenny	17.85	2	607	14	34
Clementine	17.66	0	583	13	33
Lilly	15.00	2	255	3	17
Clem	14.53	0	218	2	15
Omid	13.33	0	160	1	12
Walker	13.06	29	209	3	16
Crawford	13.00	1	169	1	13
Christa	12.33	0	148	1	12
Katjaa	11.63	0	128	0	11

Appendix C: Visual representation of procedure to exclude game-specific key words



Appendix D: Logic Chart for Total Game Evaluation Coding



Appendix E: Example of excel sheet for coding (not thinned): infuse (Flemma)

Concordance	Left side (before key word)	Right side (key word + after key word)	File (game)
1	ricles O sacred isle Artaeum, where rosy light	infuses air, O'er towers and through flowers, gentle b	Skyrim
2	wer washed over the army, crackling, hissing, and	infusing all with their ghostly force. When they arrived a	Skyrim
3	Library.* He was also quite the Sourcerer. Magic-	infused blood can be used to produce the most vivid	Divinity 2
4	the mind and could only have come from an	infused brew.r Brainy Brew Only a truly talented entrepre	Sims 4
5	treason. It is also said that her madness so	infused Castle Solitude that it infected the next king to	Skyrim
6	.* Ornate Book The loss is all of ours. Source-	Infused Cellar Key And every lizard could be a threat.	Divinity 2
7	pillar. Braided through thick iron rings, Source-	infused chains bind the dragon tightly.* You pay the pri	Divinity 2
8	stability, be it in their relationship or career.	Infuse change often! Noncommittal Lesson is Creative. I	Sims 4
9	BLIND? Poor dear, not fit to shake my Source-	infused claw. BOW to me! BOW to the most powerful	Divinity 2
10	agisters searching the rotten place! This Source-	infused copper ring makes the bearer shine with confidenc	Divinity 2
11	sailed the seas with his trusty crew. Vitamin C	infused cotton candy kept the crew free of scurvy and	Sims 4
12	enters the crown of your head and melts downward,	infusing every inch of you with a feeling of great	Divinity 2
13	02 11/09/2076 Set OFF time Botany Terminal Plasma	Infused Hey Mom, they're letting me send out a	Fallout 4
14	sses; count my blessings instead. Blessed Source	Infused I couldn't agree more. I saw what you	Divinity 2
15	grind it down so fine that it can be	infused into the lattice of packed snow. They're almost..	Skyrim
16	used with falcons' tears, this large-sized elixir	infuses its drinker with superior speed. *Wonder how Ifa	Divinity 2
17	sed with falcons' tears, this medium-sized elixir	infuses its drinker with superior speed. *She disappeare	Divinity 2
18	Y SHOE! Infused with falcons' tears, this elixir	infuses its drinker with superior speed. Come quietly, S	Divinity 2
19	nce, they think of the process of invention. The	infusing of charms and spells into objects. The creation	Skyrim
20	ersized Stuffed Animal) (From Finding Muse) (From	Infused Outfit)4 (From Gaining a Skill) (From Trying to L	Sims 4
21	fit Customize Board Everyday Outfit Formal Outfit	Infuse Outfit With Mood Party Outfit* Sleep Outfit@ Swim	Sims 4
22	luck in the next life. We found a Source-	infused ring. I know who you are. I know what	Divinity 2
23	weakly and smiles.* Perhaps drawn by the Source-	infused ring, Voidwoken attacked. We're not inmates! We'	Divinity 2
24	* *Let him off the hook. Take your leave.* Soul-	Infused Skull when a resentful courtier stabbed him to d	Divinity 2
25	not me. A small container, made of the magically	infused snow that first fell on the Throat of the	Skyrim
26	look of her, too much ferocity was had.* Void	Infused That not even I can do. Though I do	Divinity 2
27	in the eyes of the granite statue seems to	infuse the air around this shrine.* Farewell Letter Pr	Divinity 2

Concordance	Left side (before keyword)	Right side (keyword + after keyword)	File (game)
28	as truly something. Meet the twins! The desire to	infuse the blood stream with ethanol is truly perplexing	Fallout 4
29	time to get rich flavor like this! We pre	infuse the leaves with steam before brewing. I knew	Phoenix Wright
30	time to get rich flavor like this! We pre	infuse the leaves with steam before brewing. I k	Phoenix Wright
31	ng. Not until the sun burned through her window,	infusing the light wood and flesh colors of her chamber	Skyrim
32	Better off alone than with- oh very well. Source-	Infused The Magisters are all I know, all I want!	Divinity 2
33	the gold he gave to the Guild, Gorgos could	infuse the quill with the highest-price soul available,	Skyrim
34	rful magics, and the transformations they undergo	infuse their entire beings with some element of that pow	Skyrim
35	. Romantic Sims tend to be Flirty and need to	infuse their lives with Romance often.i*J Romantic Lesso	Sims 4
36	Source of spirits. I devoured my god's Source,	infusing them within my soul. My Source powers were now	Divinity 2
37	Source of spirits. I devoured my god's Source,	infusing them within my soul. My Source powers remain unde	Divinity 2
38	sword used to imbue this iron hammer used to	infuse this iron bow with used to infuse this iron	Monster Hunter W
39	used to infuse this iron bow with used to	infuse this iron heavy used to infuse this iron light	Monster Hunter W
40	with used to infuse this iron heavy used to	infuse this iron light used to set a trap. used	monsterhunter7.txt
41	We gave the tablet to Almira. Hamfists Source	Infused Vampiric Physical I hunger for you, Sebille. Th	Divinity 2
42	come? Here, let me see... This cute little arrow	infused with a charming potion is quite something. You re	Divinity 2
43	badges must be worn at all times. They are	infused with a low-yield radioactive isotope that will pr	Fallout 4
44	on A mushroom from the ancient forest. A mushroom	infused with A mushroom that glows faintly like the moon.	Monster Hunter W
45	strange aura surrounding it. It's like it's	infused with a spirit of both anger and tragedy. Well,	Phoenix Wright
46	eye lands on a section about Idols of Rebirth.]	Infused with ancient magic, these small statues allow the	Divinity 2
47	uses steel instead of iron, but the steel is	infused with Corundum to make the metal inserts stronger.	Skyrim
48	d from the great and ancient Queen Lysandra, then	infused with deadly poison. Connection failed: Server ha	Divinity 2
49	lood lie across one cheek.* Flowing Light Pants	Infused with falcons' tears, this large-sized elixir infu	Divinity 2
50	quiet for my liking. Then again, it usually is.	Infused with falcons' tears, this medium-sized elixir inf	Divinity 2
51	! YOU ARE NOTHING! YOU ARE SHITE UNDER MY SHOE!	Infused with falcons' tears, this elixir infuses its drin	Divinity 2
52	of drops fall on her hand, but instead of	infusing with her body, they quickly evaporate on her scal	Divinity 2
53	get caught in the act. Your skeleton has been	infused with indestructible metal, reducing limb damage b	Fallout 4
54	freely to create the first Divine, a mortal we	infused with our united strength. To bathe in the lake,	Divinity 2
55	huge... How do they form? Some parts have been	infused with Some scholars say it's the most typical	Monster Hunter W
56	ient amulet that contains five jewels ready to be	infused with Source; each blessed with a kiss of Astarte.	Divinity 2

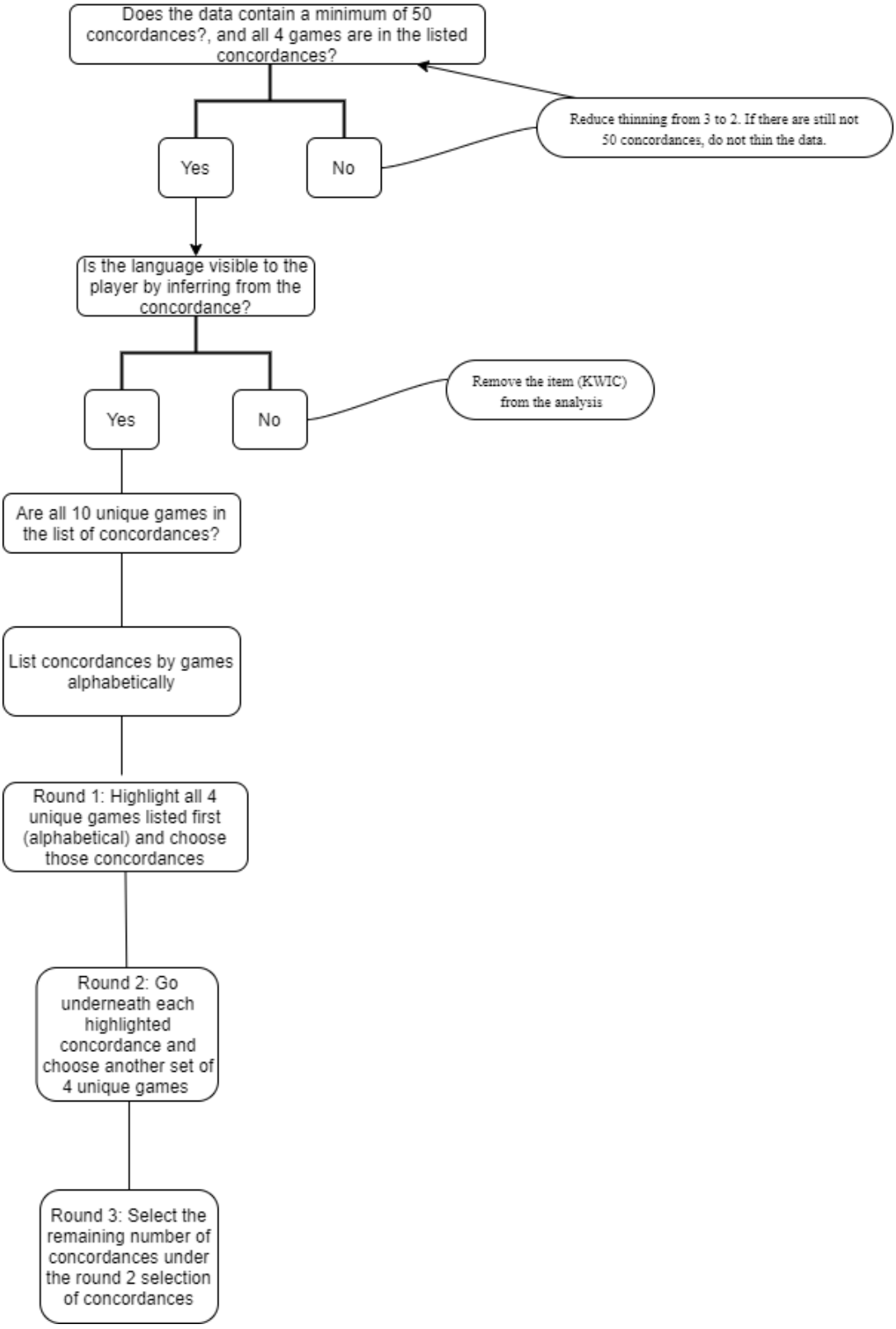
Concordance	Left side (before keyword)	Right side (keyword + after keyword)	File (game)
57	ient amulet that contains five jewels ready to be	infused with Source; each blessed with a kiss of Astarte.	Divinity 2
58	s over each of them.* Character's is temporarily	infused with Source. Klaud *Sounds like a challenge. Yo	Divinity 2
59	! The ritual was a failure as we were not	infused with Source. We should make use of the Source	Divinity 2
60	of Homare Strength of Sabi Strengthened by being	infused with Strengthens explosive attacks like Strider S	Monster Hunter W
61	bends to the King's every whim. An arrow	infused with the fury of the air. If you listen	Divinity 2
62	from around her neck. "It's a soul gem,	infused with the spirit of a great werewolf your father	Skyrim
63	beast, pacing. "It's a soul gem," she said. "	Infused with the spirit of a great demon werewolf. It	Skyrim
64	Reset Progress Set Complete The Outfit is already	infused with this Mood. Confident Moods Energized Moods F	Sims 4
65	With the Adamantium Skeleton perk, your bones are	infused with unbreakable metal, which reduces - and may e	Fallout 4
66	legs! If that wasn't enough, each pair is	infused with 5000 watts of energy before leaving our fact	Sims 4
67	knees and bows before you in reverence.* Source-	Infused *You're sorry to hear that. Say that you	Divinity 2
68	coming. Dead Silent Monk *A warmth cocoons you;	infusing your every muscle with the will to relax. A	Divinity 2
69	hieves.* Civil Abilities *A warmth cocoons you;	infusing your every muscle with the will to relax. A	Divinity 2
70	must see that you are fit. First, we must	infuse your poor bones with power. *Drily say you guess	Divinity 2

Appendix F: Example of in-text query of concordance 38 for "infuse" in total corpus (Notepad++)

used in select gear.
used in this armor was treated
used to fortify gear. Said to be
used to fortify gear. Said to be quite
used to give these iron dual blades
used to give this bone great sword
used to give this bone insect glaive
used to give this bone lance
used to give this bone sword & shield
used to give this iron charge blade
used to give this iron great sword
used to give this iron hammer
used to give this iron hunting horn
used to give this iron insect glaive
used to give this iron lance
used to give this iron switch axe
used to give this iron sword & shield
used to imbue this iron great sword
used to imbue this iron hammer
used to infuse this iron bow with
used to infuse this iron heavy
used to infuse this iron light

**One big text file. Highlighted is the result of the search. Other text is surrounding data.*

Appendix G: Logic Chart for APRG Evaluation Coding



Appendix H: Example of Excel sheet for coding (not thinned): heft (Flemma)

Concordance Left	Key Word + Concordance Right	Game
hear an original - something with a little more	heft than the same old tavern jingles.* Oh hi.	divinity2.txt
're under arrest! Off to prison you go. *	Heft your weapon. You're ready to kill this	divinity2.txt
soul breaks bone as well as a well-	hefted axe.* The Black Ring appreciates yore service, L	divinity2.txt
s. *Without looking once in your direction, Kemm	hefts his sword.* *He glances up, and you see	divinity2.txt
? Don't answer that. I do. I! [IFAN] *	Heft your weapon and tell her bad news is	divinity2.txt
one. *He grabs hold of the plank and	hefts his great bulk backward. The rusted nail pops	divinity2.txt
ommand. Place ping beacon Severed Head *Conway	hefts a heavy sack toward you, head slightly tilted	divinity2.txt
you've had a good conversation.* *The skeleton	hefts Huwbert's dog-eared encyclopedia, its weight heav	divinity2.txt
, stands Rhalic: a twisted old man struggling to	heft his own sword. With difficulty he lifts it	divinity2.txt
Armor, and then you'll be able to	heft that minigun. Prepare to be shocked: not every	fallout(clean) 4.txt
I did? Baseball history right here. Feel that	heft... I have... I just wanted to say that	fallout(clean) 4.txt
r of... Damn. Great swords, mostly. Watching them	heft Great work lately, Hunter! While you're here,	monsterhunter7.txt
absolute power of the tyrant's The added	heft of this Uragaan The admiral was here!? The	monsterhunter7.txt
you on your artistry, and the balance and	heft of your daggers. The knife blade is whisper	og_skyrim_clean.txt
de with ornately decorated handle. The weapon had	heft, and Emmeg realized on brandishing it that the	og_skyrim_clean.txt
ed steel, Sheogorath strode into the clearing. He	hefted Emmeg groKayra's disembodied head and bundled it	og_skyrim_clean.txt
lift these days, but I still enjoy the	heft of a good hammer. Vampires are utter abominations	og_skyrim_clean.txt
to speak my mind. It was the blade	hefted by Ysgramor when he returned to drive the	og_skyrim_clean.txt

Appendix I: Game Profiles of Coders

Games	**Researcher**	**Interrater 1**	**Interrater 2**	**Interrater 3**
Divinity 2	X	X	X	X
Fallout 4	X			X
GTA V	X			X
Monster Hunter World	X	X	X	
Phoenix Wright	X			X
Resident Evil 6	X	X		
Sam and Max	X			
Sims 4			X	X
Skyrim	X	X	X	X
Walking Dead	X			X

Appendix J: All Common Game Words (484) – In Order

melee recoil tutorial teleport npc trader artifact nearby haha gameplay retrieve optional foe teleportation autosave mastery infuse successfully stealth crimson wary pause current useful overwrite stagger disable hehe relic fearsome scythe agility lemme composure courtyard flee glare grumble thankfully return shard brewing culprit dialog iv highlight cobble jagged cloak attacker masterwork bah skeletal deity armoury unleash potent assassin flame scrawl hideout vigilant mist rename decor weaponry glow power formidable quarry craftsmanship gah feat regeneration sigh cleanse creation upwards pfft amethyst otherworldly fury delve path volume hunger tier rogue leather skull pillar ensure fiery navigate projectile vase madness poison reveal beneficial urgh remain resurrect loot skilled grime capacity task recharge brute accurse contain fend pale consuming bastion advanced grip famed siphon ability width recover dunno within surface hmmmm harden resistant grasp erupt remnant ai walkway enrage interior ornamental eliminate attribute silver dispel quell sniff heirloom lest brow flaming roar mundane hide construct treasure metallic security rage fulfil previous victorious staff allow encounter captive truly eager strengthen damnit electrify controller unknown commonly seed worthy appraise decline scent mettle overrun provide boast instantly imprison various etc continue intimidate indeed request engulf intimidation glimmer nemesis revere selection entrance reminder presence gleam sadly array amiss surroundings infiltrate despite valuable contents fatigue vanish drain pristine ol sever carnage bounding incoming fail thug illuminate chamber freely trading potency elite untouched wander begin expansion appear descend recent ominous corrupt emerge prevention generic widely uncover sight grate incite unusually muster vendor map longer note countless replenish greater revert fist annihilate patron discount crop reinforcement unwavering resilience dank transpire incandescent assign perhaps watcher plank entity manage confront shadowy axis nefarious brightly damned mysterious feast black attempt dire budge ahem intruder traitor extinguish vermin blood scribble nears brim venture im brawn ravenous tamer detail leader contraband puppet unwanted keen stumble glory stunning assist elsewhere enlightenment offer knowledge roam menace visitor pile cunning claim unsettling bestow withstand meagre stash toll dangerously recruit sharpen wrath marker angrily sa weave shady stat fewer trade effort husk facade travel tapestry perch align contraption voice mixture existence surge slink experimentation invaluable discover value oddly cough prone treacherous fluster site scour symbol crumble restrict query gawk rubble silence exceedingly unsure rust key lightning inaccessible rethink despot undetected others ponder sneak chat scurry mechanism order mere primary protect underling yield able search likely latter expand payment scourge mourn resist safely meditation spotlight impressive though survive ignore limitless farm deadly shortly sinister peril proclaim slightly small arm supply cloud quite reflect secure quietly ceremonial idle revel unclear surround pointless receive trap manipulation rebuild boon finally jest rightly greet journey sharp persist castle lack kindness graveyard multiple console surprisingly adventure become land action massive non greeting regain lightly defend empty larger point bolt seem radiance bloodthirsty patrolling aquatic misdeed exceptional namesake trove icy renown wondrous alongside weaker pinnacle tedious stride chit absorption squarely restock assortment havoc mercilessly perilous topple handiwork newcomer nary squalor terrifying

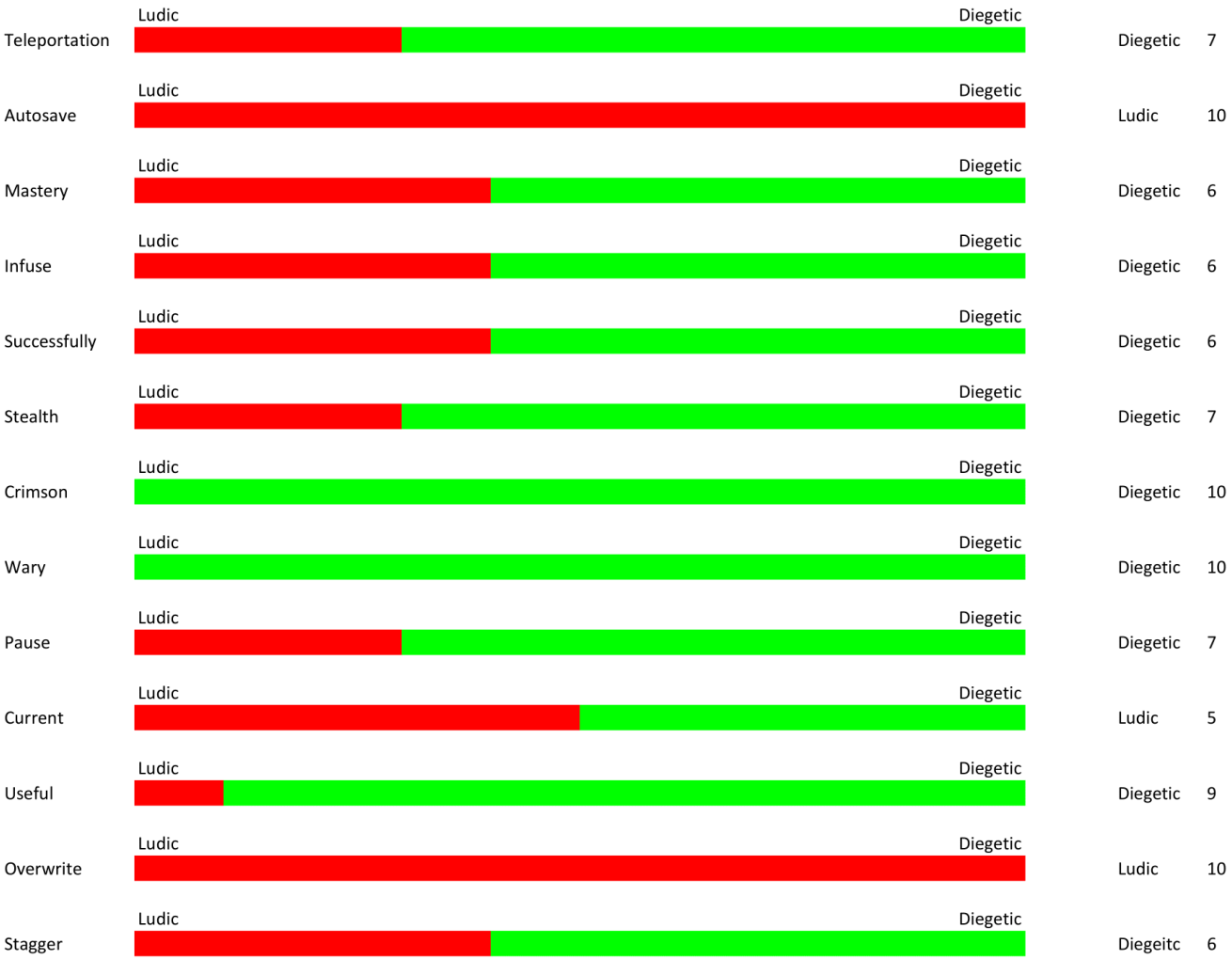
Appendix K: All ARPG Words (543) – In Order

melee recoil optional foe haha nearby retrieve wary infuse relic stagger npc useful flee flame agility vigilant masterwork tutorial scholarly power jagged iv mist fury cloak potent brewing bah cowl assassin cleanse glow golem unleash feat thankfully gameplay deity successfully poison sigh rogue stealth glare pillar path autosave composure hunger beneficial quarry skeletal consuming pause madness remain shimmer brute roar sate attribute anew volume grip current regeneration fiery surface flourish task skilled grasp remnant leather bastion ale return within brow pale makeshift otherworldly watcher rend rename welp silver ensure vanish pfft monstrosity spear truly encounter grime captive presence damned unknown resistant skull mettle sadly hideout reveal creation greater eager worthy glory drain scent blood chamber widely emerge shadowy indeed feast atop treacherous staff carrion knowledge potency contain muster sight sniff amiss boast recover wander famed resist gloom seed commonly rage instantly despite advanced bestow ability dire mixture fist keen uncover recharge unequip respite heft hrm carnage venture restorative invaluable durability freely valuable yield peril contraption lemme wrath capacity confront provide untouched sever leader countless appear attempt note travel kingdom existence recruit stumble graveyard assist sharpen constitution vermin cunning steadfast rebuild though wisdom mechanism mere bind castle weaker effort eliminate freedom surroundings overwrite ominous metallic venomous journey menace cure greeting quite flaming lightly nobility boon fear treasure etc toll significantly mistress scour among icy surely continue sneak quell deadly mutton stride adventure ahem cough intact perhaps regain discretion gah silence crop survive inflict various bolt acidic tier despot howl devotee itself entrance strength deem likely mysterious hall wondrous persevere nimble ravenous absorption serve wretched weak decipher manage trap trading brighter supply discover free lightning chit akin aloft devourer sinister exceptional patrolling filth impressive decade beyond loyal vision cruelty learn seem approach lack controller guidance perish order topple trade cloud gladly fierce primary unsettling ol tread withstand detection gratitude search downward root greet map berserker fewer effortlessly savagery memory claim stronger hide illuminate plunder clutch secure themselves selection death brawn thick resolve stout faintest muzzle safely persist ignore contract drinker frenzy firmly wager capable stalwart trove slightly dunno lash recent sea suffice utmost vessel become warning infested dangerously previous swift incinerate dwell further budge eternity sharp tangle thirst subject chat errand simply stunning quietly glorious offer hound merely contents observe disposal admiration fully visitor endless preserve forth perilous resilience stranger craven harm fight navigate once against unfamiliar strife loyalty puppet marvel stir attacker narrow allow usher beset bloodthirsty grit immeasurable radiance fearful oblivious procure hasten execute seal hatred horizon spare potential voice begin harsh unclear unwanted brethren moment incoming manner site refuse value surround request pinnacle surge latter symbol fail finest ground weakness unusually demise undertake restock amok murky black disappointing key ghastly taunt ongoing effective interior torch savage limitless flood retreat pile unaware assure succeed devastation entirely quickly defend closely expand weakling avail shelter surprisingly assessment speak thrive thereafter proper alongside manipulation unfortunate rejoin yearn charmer above bare apparent wide churn hero dangerous hail lad curl tunnel command action movement intriguing finer barely inform winter cautious per larger deepest below trust detail arm end far burst curious doom cruel squarely nefarious crippling heavily hoard susceptible docile whence flawless futile regroup plentiful debt mutt vanguard glimpse vi tenacious radiant accumulate ghoulish freshest reflect unthinkable scattering

Appendix L: Ludic and Diegetic Words of Total Corpus

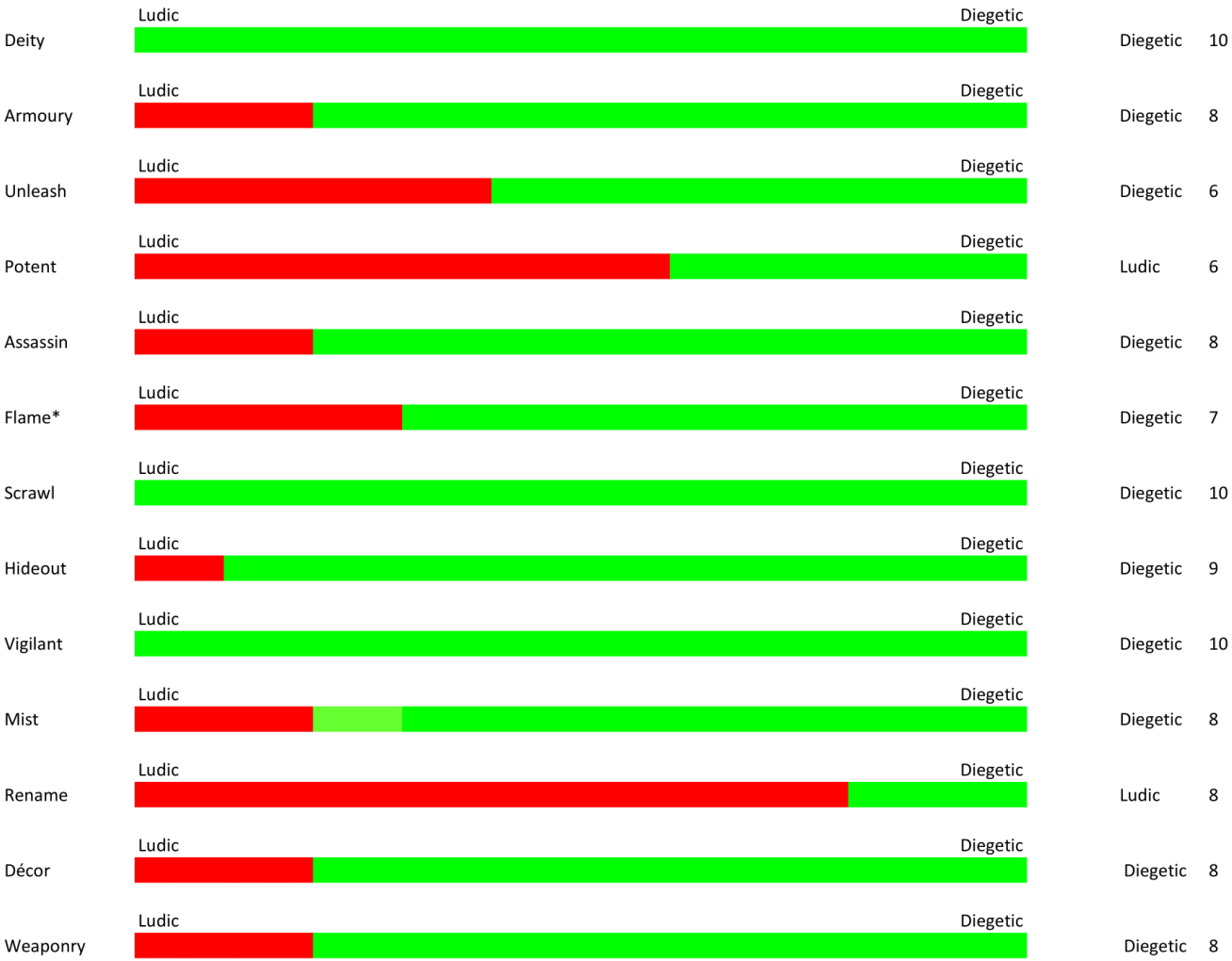
Ludic Words N = 20	Diegetic Words N = 80			
1. melee	1. Teleport	22. Composure	43. Hideout	64. Hunger
2. recoil	2. Trader	23. Courtyard	44. Vigilant	65. Rogue
3. tutorial	3. Artifact	24. Flee	45. Mist	66. Leather
4. NPC	4. Nearby	25. Glare	46. Décor	67. Skull
5. Gameplay	5. Retrieve	26. Grumble	47. Weaponry	68. Pillar
6. Optional	6. Foe	27. Thankfully	48. Glow	69. Ensure
7. Autosave	7. Teleportation	28. Return	49. Power	70. Fiery
8. Current	8. Mastery	29. Shard	50. Formidable	71. Vase
9. Overwrite	9. Infuse	30. Brewing	51. Quarry	72. Madness
10. Disable	10. Successfully	31. Culprit	52. Craftsmanship	73. Poison
11. Dialog	11. Stealth	32. Cobble	53. Feat	74. Reveal
12. Highlight	12. Crimson	33. Jagged	54. Sigh	75. Beneficial
13. Attacker	13. Wary	34. Cloak	55. Cleanse	76. Remain
14. Potent	14. Pause	35. Masterwork	56. Creation	77. Loot
15. Rename	15. Useful	36. Skeletal	57. Upwards	78. Skilled
16. Regeneration	16. Stagger	37. Deity	58. Amethyst	79. Grime
17. Tier	17. Relic	38. Armoury	59. Otherworldly	80. Capacity
18. Navigate	18. Fearsome	39. Unleash	60. Fury	
19. Resurrect	19. Scythe	40. Assassin	61. Delve	
20. Projectile	20. Agility	41. Flame	62. Path	
	21. Lemme	*42.* Scrawl	*63.* Volume	

Appendix M Ludic / Diegetic Visual Chart for Total Corpus


Ludic
Diegetic
Teleportation
Diegetic 7
Ludic
Diegetic
Autosave
Ludic 10
Ludic
Diegetic
Mastery
Diegetic 6
Ludic
Diegetic
Infuse
Diegetic 6
Ludic
Diegetic
Successfully
Diegetic 6
Ludic
Diegetic
Stealth
Diegetic 7
Ludic
Diegetic
Crimson
Diegetic 10
Ludic
Diegetic
Wary
Diegetic 10
Ludic
Diegetic
Pause
Diegetic 7
Ludic
Diegetic
Current
Ludic 5
Ludic
Diegetic
Useful
Diegetic 9
Ludic
Diegetic
Overwrite
Ludic 10
Ludic
Diegetic
Stagger
Diegeitc 6


Ludic
Diegetic
Disable
Ludic 8
Relic
Diegetic 9
fearsome
Diegetic 10
Scythe
Diegetic 10
Agility
Diegetic 6
Lemme
Diegetic 10
Composure
Diegetic 9
Courtyard
Diegetic 10
Flee
Diegetic 9
Glare
Diegetic 10
Grumble
Diegetic 10
Thankfully
Diegetic 10

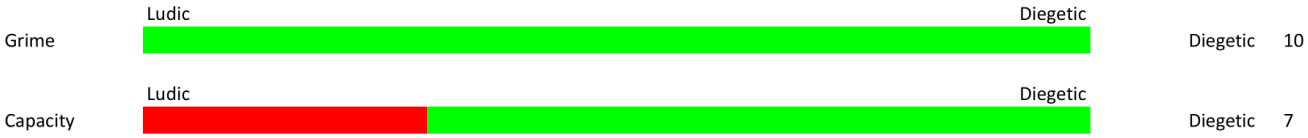

Ludic
Diegetic
Return
Diegetic 10
Ludic
Diegetic
Shard
Diegetic 10
Ludic
Diegetic
Brewing
Diegetic 10
Ludic
Diegetic
Culprit
Diegetic 10
Ludic
Diegetic
Dialog
Ludic 7
Ludic
Diegetic
Highlight
Ludic 8
Ludic
Diegetic
Cobble
Diegetic 10
Ludic
Diegetic
Jagged
Diegetic 10
Ludic
Diegetic
Cloak
Diegetic 9
Ludic
Diegetic
Attacker
Ludic 7
Ludic
Diegetic
Masterwork
Diegetic 10
Ludic
Diegetic
Skeletal
Diegetic 10


Ludic
Diegetic
Deity
Diegetic 10
Armoury
Diegetic 8
Unleash
Diegetic 6
Potent
Ludic 6
Assassin
Diegetic 8
Flame*
Diegetic 7
Scrawl
Diegetic 10
Hideout
Diegetic 9
Vigilant
Diegetic 10
Mist
Diegetic 8
Rename
Ludic 8
Décor
Diegetic 8
Weaponry
Diegetic 8


Ludic
Diegetic
Glow
Diegetic 9
Power
Diegetic 8
Formidable
Diegetic 10
Quarry
Diegetic 10
Craftsmanship
Diegetic 10
Feat
Diegetic 10
Regeneration
Ludic 7
Sigh
Diegetic 10
Cleanse
Diegetic 9
Creation
Diegetic 8
Upwards
Diegetic 9
Amethyst
Diegetic 10


Ludic
Diegetic
Otherworldly
Diegetic 10
Ludic
Diegetic
Fury
Diegetic 10
Ludic
Diegetic
Delve
Diegetic 10
Ludic
Diegetic
Path
Diegetic 8
Ludic
Diegetic
Volume
Diegetic 6
Ludic
Diegetic
Hunger
Diegetic 8
Ludic
Diegetic
Tier
Ludic 6
Ludic
Diegetic
Rogue
Diegetic 9
Ludic
Diegetic
Leather
Ludic 4
Ludic
Diegetic
skull
Diegetic 7
Ludic
Diegetic
Pillar
Diegetic 6
Ludic
Diegetic
Ensure
Diegetic 10


Ludic
Diegetic
Fiery
Diegetic 7
Navigate
Ludic 6
Projectile
Ludic 5
Vase
Diegetic 10
Madness
Diegetic 9
Poison
Diegetic 6
Reveal
Diegetic 10
Beneficial
Diegetic 7
Remain
Remain 9
Resurrect
Ludic 7
Loot
Diegetic 9
Skilled
Diegetic 7


Ludic
Diegetic
Grime
Diegetic 10
Ludic
Diegetic
Capacity
Diegetic 7

Appendix N: Ludic / Diegetic Visual Chart for ARPG Corpus


Ludic
Diegetic
Flourish
Diegetic 9
Ludic
Diegetic
Task
Diegetic 10
Ludic
Diegetic
Grasp
Diegetic 10
Ludic
Diegetic
Remnant
Diegetic 10
Ludic
Diegetic
Bastion
Diegetic 10
Ludic
Diegetic
Ale
Diegetic 10
Ludic
Diegetic
Within
Diegetic 8
Ludic
Diegetic
Brow
Ludic 5
Ludic
Diegetic
Pale
Diegetic 10
Ludic
Diegetic
Makeshift
Diegetic 6
Ludic
Diegetic
Watcher
Diegetic 9
Ludic
Diegetic
Rend
Diegetic 10


Ludic
Diegetic
Welp
Diegetic 10
Silver
Diegetic 10
Monstrosity
Diegetic 10
Spear
Diegetic 9
Truly
Diegetic 10
Encounter
Diegetic 7
Captive
Diegetic 10
Presence
Diegetic 10
Damned
Diegetic 10
Unknown
Diegetic 9
Resistant
Ludic 7
Mettle
Diegetic 10


Ludic
Diegetic
Sadly
Diegetic 10