The Prevalence of Core Vocabulary in World of Warcraft's Written In-Game Quest Instruction

Miso Finne
Master's Thesis
English
The Department of Languages
University of Jyväskylä
Spring 2021

UNIVERSITY OF JYVÄSKYLÄ

Faculty Humanities and Social Sciences	Department Department of Language and Communication Studies
Author Miso Finne	
Title The Prevalence of Core Vocabulary in World of Warcraft's Written In-Game Quest Instruction	
Subject English	Level Master's Thesis
Month and year June 2021	Number of pages 56

Abstract

Tutkielmassa tarkastellaan massiviisten nettiroolipelien pelaamisen yhteyttä vieraan kielen sanaston oppimiseen. Verkkopelaamisen yleistymisen myötä myös siihen liittyvä tutkimus on lisääntynyt. Yleisimpinä tutkimuskohteina ovat olleet pelaajat ja heidän välinen kommunikaationsa. Tästä johtuen pelien muut aspektit, kuten ennalta tuotettu, pelimaailmaan sidonnainen staattinen kielisisältö, ovat jääneet vähemmälle huomiolle. Tämä korpustutkimus pyrkii hahmottamaan, millaisesta sanastosta *World of Warcraft* -nimisen massiivisen nettiroolipelin sisäiset tehtäväohjeistukset koostuvat. Tutkimuksen tärkein teoriallinen työkalu on Ronald Carterin keskeisen sanaston käsite, jonka pyrkimyksenä on löytää kielenkäytön kannalta olennaisimpia sanoja – sanoja joita käytetään muiden, harvinaisempien sanojen määrittämiseen.

Tutkimuksen omana aineistona toimii 282 958 sanan suuruinen korpus, eli lingvististä tutkimusta varten koottu kieliaineistokokoelma. Se koostuu 2011 tekstitiedostosta. Tätä internetistä kerättyä aineistoa tarkastellaan kaksivaiheisesti. Ensin verrataan sen sanaston esiintymisjakaumaa *AntConc* -nimisen korpusanalyysiohjelman avulla maailman suurimpaan, yli miljardin sanan englannin kielen korpukseen, *Corpus of Contemporary American English (COCA)*. Analyysin toisessa vaiheessa Carterin käsitteen avulla arvioidaan tutkimusta varten kerätyn korpuksen kymmenen yleisimmän verbin keskeisyyttä englannin kielessä.

Tulokset osoittavat, että *World of Warcraft* -nettiroolipelin sisäiset tehtäväohjeistukset koostuvat fantasiasanaston sijaan universaalisti, kaikissa eri diskurssimuodoissa käytetyistä englannin kielen sanoista. Analyysin toisessa osassa tutkitut, yleisimmät verbit ovat monikäyttöisiä, sillä niitä käytetään kielessä laajalti erilaisten fraasiverbien ja idiomien komponentteina. Tästä syystä ne voidaan todeta keskeiseksi sanastoksi myös pelin ulkopuolisissa konteksteissa.

Keywords – lexicology, corpus research, vocabulary, english language, video games, core vocabulary

Depository University of Jyväskylä

Additional information

TABLE OF CONTENTS

1 INTRODUCTION

The enormity of video games as an industry cannot be overstated. In 2015, there were an estimated 2 billion gamers around the globe, and by 2020 that number rose to 2.7 billion (Statista 2015; Statista 2020). With an ever-larger amount of consumers playing them, video games have consequently begun to receive more and more attention by academia and different fields of research. This has resulted in a wealth of studies that point to video games, particularly ones that are played online with other people, as potent and unique facilitators of language learning (Chotipaktanasook & Reinders 2018; De Aguilera & Méndiz 2003; Newgarden, Young & Zheng 2012; Piirainen-Marsh & Tainio 2009). These studies concluded that the spaced repetition of activities requiring recurrent language actions and the player-to-player interaction which results from co-operative gameplay are some of the aspects of online games that are beneficial to foreign language learners. (Chotipaktanasook & Reinders 2018, 379; Newgarden & Zheng 2016, 283; Piirainen-Marsh & Tainio 2009, 165). In other words, online games generate opportunities and scenarios for language learners to practice using a foreign language while working together with other players. The language play that results from such interaction is one of the driving factors for language learning (Piirainen-Marsh & Tainio (2009, 155). These findings are significant because they serve to highlight the unique strenghts of video games as conduits for learning.

Much of the research that has revolved around online games and language learning, including the studies cited above, has centered around the player-driven, human-to-human social interaction that the game genre is known for. By looking at both text- and voice-based player communication, it has sought to identify how the language output produced by non-native speakers manifests itself in cyberinteraction, as well as the unique opportunities for social and linguistic interaction that online game worlds provide. While these aspects of online games should and successfully have been studied, doing so has meant that other facets of these games have not recieved nearly as much attention. One such facet is the lexical examination of the language content and in-game instructions provided to the players by the creators of video games. In the massively multiplayer online role-playing game *World of Warcraft*, much of that language content comes in the form of

in-game tasks, known as quests, that players must successfully complete in order to further strenghten their in-game avatars and earn various rewards (Peterson 2013, 30). As these quests constitute a core component of players' engagement with the game and are often encountered during regular play, it makes sense to examine the type of vocabulary that their textual elements consist of. Uncovering, examining, and highlighting these vocabular components is designed to serve as research which is complementary to the pre-existing findings which show that playing video games has a significant, positive impact on language learners' vocabulary knowledge in a foreign language (Behbahani & Vahdat 2013; Chen & Yang 2012; Ovalle & Vásquez 2019). At the heart of this corpus-based investigation lies Ronald Carter's (1987) conceptualization of *core vocabulary*, which provides an analytical template with which to measure the relative prominence of a word in a language. Using it, the present study seeks to establish a link between the most prominent vocabulary found within the in-game quest texts of *World of Warcraft* and their utility in the English language.

This inquiry is divided into the following sections. The first of these details the theoretical constructs that serve as tools to both frame this research and later analyze and discuss its findings. The second section of the study presents the research question, as well as describes its methodology, including the gathering and the processing of the data which was used to construct the study's corpus. This is followed by a lengthy, two-phased analysis, which first examines the structural assembly of the corpus by analyzing the numerical and percentual distribution of its most-often occurring vocabulary. The last phase of the analysis consists of a detailed examination of ten of the most prevalent, i.e., frequently occurring, verbs in that corpus which intends to demonstrate that they also see extensive use in other contexts that extend beyond the confines of virtual game worlds. The second-to-last section of the study is dedicated to the summary and presentation of its findings. The present study concludes with a section that aims to succinctly answer its two main research questions, discuss its limitations, as well as consider potential research objectives for future studies.

2 THEORETICAL BACKGROUND

2.1 The nature of high-frequency words and core vocabulary

It is difficult to overstate the importance of vocabulary in language learning. Not only do children with larger vocabularies acquire and gain mastery of new words faster, but also it has been proven that the size of their lexical repertoire is inextricably linked to academic success (Marzano & Simms 2013, 7). Not all words are created equal, however, as according to Milton (2009, 22) the frequency of occurrence is one of the central ways in which words differ from one another. In other words, some words are far more present in a language and more integral than others. He further extrapolates on the idea by going on to state how the large amount of text coverage provided by a small amount of very highly-frequent words appears to be a language agnostic phenomenon (Milton 2009, 67). Nation (2001, 9) echoes this line of thought by making mention of language learners' potential to achieve much with a comparatively modest amount of key vocabulary.

Word frequency is important because it allows language instructors to make more well-informed choices about which words to dedicate their classroom time to. Thus allowing them to spend what precious little time they have on teaching words that matter most.

Knowing how often a single lexical unit occurs as part of a language also turns out to be of use when trying to gauge how "core" that item is. This is because, according to Carter (1987, 187), the frequency of occurrence is one of the key factors that govern a word's coreness, as core vocabulary sees much more regular use than words which cannot be deemed as such. He is the leading voice when it comes to defining the identity of core vocabulary. Therefore, what follows can only be attributed to him. There are ten other factors that should be considered when attempting to identify core vocabulary: *syntactic substitution*, *antonymy*, *collocability*, *extension*, *superordinateness*, *freedom from culture-specific uses*, *summary*, *associationism*, *neutral field of discourse* and *neutral tenor of discourse* (Carter 1987, 180-185).

Firstly, core words can be used to describe other, non-core words, but the same does not work in reverse. Hence, words such as *gobble*, *dine*, *devour*, *stuff* and *gormandize* can all be described as manners of eating, but to describe the verb *to eat* as devouring something solely accounts for a very specific method or style of ingestion. In other words, words that are core possess a greater degree of *syntactic substitutability* than ones that are not (Carter 1987, 180).

Second, it is much harder to find antonyms for non-core words. Hence, words like *big* and *small*, which are core words, are clear antonyms of one another, but trying to find a similar pairing for words such as *sob, emaciated, corpulent* and *guffaw* is a far more challenging, assuming that it is even possible (Carter 1987, 181).

A distinct characteristic of core vocabulary appears to be that they share a greater number of partnerships with other vocabulary than words that cannot be classified as such. Carter (1987, 181) illustrates this by showing how the word *bright* in the English language has a greater number of partnerships with other words than the semantically similar word *radiant*. The *sun, light, sky, idea, colors, future, prospects* and a *child* can all be described as bright but calling an *idea* or a *child radiant* would sound more than a little odd. Therefore, the word *bright* could be classified as possessing a greater magnitude of *collocability* than words like *radiant*.

Similarly to *collocability*, core words also find themselves as part of a far greater amount of idioms, phrasal verbs, multi-word verbs and other similar constructs than other words. Core verbs like *run* are commonly 'extended' to expressions such as *run-of-the-mill, run about* and to be *in the running* (Carter 1987, 182). Because of this, they can be argued to be far more integrated into the language and that their distinct capability of *extension* is what separates them for other far-less used vocabulary.

Another factor that commonly demarcates a word as being core is *superordinateness*, which highlights how many core words serve as categorical superordinates for other vocabulary, commonly referred to as hyponyms. As an example of this, Carter (1987, 182) notes how less common and more specific words like *tulip* and *rose* are both part of the generic superordinate grouping *flower*, which itself is a more common word in the English language. Evaluating words

in this manner does come with some drawbacks however, as superordinates like *vehicle* and *furniture* appear to be far less core than their respective hyponyms like *car* and *chair* (Carter 1987, 183). Despite this, there remains a clear connection between *superordinateness* and core vocabulary.

The next aspect shared by core vocabulary is that they are culturally neutral. This *freedom from culture-specific uses* means that unlike words such as *pouffe* and *divan*, which have been borrowed from other languages, and by extension other cultures, core vocabulary, used as more generic descriptors, is markedly culture-free (Carter 1987, 184.)

Along with serving as generic descriptors, core vocabulary is also preferred by speakers when recounting or summarizing a sequence of events. What this finding suggests is that within the *summary* genre information should preferably be relayed in a manner which omits any stylistic, rhetoric and evaluatory flair (Carter 1987, 184). Hence, when talking about a dog, speakers are far more likely to simply call it a *dog* without mentioning its breed, unless such information is crucial in order to accurately describe the canine or what happened to it.

The final three factors, *associationism, neutral field of discourse* and *neutral tenor of discourse*, all deal with core vocabulary's tonal neutrality in language. The first of these, *associationism,* relates to how the more 'core' a vocabulary item is, the easier it is to insert it into various, even tonally conflicting, contexts. Brown & Jule's (1983, 125) discussion on the interlinked nature of discourse utterances demonstrates this characteristic in the adjectives *tall* and *thin*. They manage to show that these attributive words will be perceived as either desirable or undesirable depending on the context they are in. Thus, the sentence "*She's tall and thin and walks like a crane."* will be thought of as either a positive or a negative remark, depending on if it is preceded by either the sentence "*I like Sally Binns.*" or "*I hate Sally Binns.*" This kind of tonal amorphism is also present in how core words, according to Carter (1987, 185), do not, on their own, point to any particular field of discourse. He mentions that while words such as *starboard* and *galley* can be identified as belonging to different nautical and seafaring lexicons, more generic vocabulary like *left* and *right* cannot. They are, by themselves, athematic. Finally, core vocabulary is also neutral in terms of formality. This is why the majority of informants rate words like *fat* as being the most neutral

in tests when paired with other similarly characterizing words like *podgy* or *corpulent* (Carter 1987, 185).

The ten factors that have been detailed above are important in the context of the present study, because they facilitate a kind of evaluation of the vocabulary in its corpus which goes beyond the notion that words' frequency of occurrence is the only factor which determines their prevalence in a language. They jointly form an additional metric with which to evaluate the pertinence of words in a language, which may consequently be used to aid in the selection of the most relevant vocabulary for language instruction. Additionally, establishing a link between the lexical content of the study's corpus and *core vocabulary* will serve to further validate the efficacy of online games as tools for learning a language.

2.2 The characteristics of play

A close examination of the lexical composition of in-game quests in an MMORPG (Massively Multiplayer Online Roleplaying Game) would certainly be of little to no use without a proper delineation of play, its character, and the reasons that drive humans to engage in it.

According to Caillois (2001, 9 - 10), there are 6 distinct characteristics that set it apart from any other kind of human activity. The first of these is that it is non-obligatory, meaning that those who participate in it are free to continue or forfeit doing so at any time and for any reason without being penalized for it. Were this not the case, it would lose its recreational allure. The second characteristic of play is that its duration is fixed, existing outside the realm of mundane, everyday obligations. It is therefore separate from other aspects of human life. The third underlines its aleatoric nature, that is, there is an inherent quality of uncertainty to it. If its outcome could be forecast and determined in advance, it would just be a completely deterministic affair, incapable of surprise. The fourth aspect of play is that it is unproductive. No goods, wealth, or new elements are born of it. It also does not end in a state that is different from its moment of inception.

Unlike labor, which has a predefined goal and produces a measurable, observable outcome, play exists for its own sake, unburdened by a rigid objective. Its penultimate feature is that there are rules to it, which are subconsciously or openly agreed upon by any number of participants. These rules can be made up from elements that carry no meaning as part of ordinary life. For example, the boundaries that delimit a tennis court are determined to be of certain dimensions only by a shared agreement between the players and the umpire overseeing the match. No part of the natural world impedes a tennis ball from moving and existing outside such constraints. The game exists only for as long as the rules are being adhered to.

The final characteristic of play defined by Caillois is the element of make-believe. Anyone taking part in play is aware of the second reality separate from real life that play happens in. This grants a high degree of freedom to whoever is playing because they can make choices and act in ways which their real life roles do not allow them to. This naturally engenders experimentation, for it

becomes possible to play out certain behaviours and choices to see the simulated outcomes that they most likely would lead to. These experiences, while not entirely representative of their real-world counterparts, do offer opportunities for learning.

This sentiment is strongly echoed by Gee (2008), who in his article, *"Cats and Portals: Videogames, Learning and Play",* draws parallels between the production of knowledge in video games and how it is produced in the real world. He argues that the tools offered by video games and the tools that are used to make discoveries in real life both change the way that individuals approach and consider the possibilities in their environment (Gee 2008, 234). In other words, play and games can be seen as framing devices for problem-solving, distilling the complexities of reality into more condensed and manageable frameworks. This is especially advantageous for language learning, as according to Gee (2008, 243), words care very-little about the context that they find themselves attached to, as they still allow us to discover, experiment, and test hypotheses, which may lead to discoveries in our reality.

2.3 The characteristics of a game

Games exhibit a lot of the same qualities as play, which is not too surprising given that they can be thought of as frameworks for play to occur. Of these shared qualities, rules are the single-most quintessential pillar that all games are built upon. Adams & Dormans (2012, 1) define games as an activity, which takes place in pretended reality, during which those who take part try to achieve at least a single arbitrary goal, while adhering to a set of rules. Peterson (2013, 35) points out that rules are an intrinsic element of computer games and force their players to pursue certain paths. What is key is that rules give players challenges that require effort to overcome. (Juul 2005, 5).

The idea of a goal is also important, for it distinguishes objective-driven games, such as football and gambling from more freeform play characteristic of young children. Caillois (2001, 13) terms these two opposites as *ludus* and *paidia*, respectively. While most modern video games employ a variable mixture of the two, it is undeniable that in particular *ludus,* the goal-centric type of play is an integral part of most computer-based gaming experiences. In *World of Warcraft*, players are tasked with gaining levels by successfully completing tasks known as quests (Peterson 2013, 30). From a game design point of view, quests can be thought to fall under the umbrella of *mechanics of progression* instead of *mechanics of emergence* (Adams & Dormans 2012, 24). This is because their state does not allow for any sort of gradation. They are either in-progress or have been completed and act as a way to bar players from progressing further until they have been completed. Players also do not choose what objectives they must achieve in order to complete them. *Mechanics of progression* allow game designers to have complete control over the sequence of events that take place. (Adams & Dormans 2012, 24).

Just as there are 6 distinct characteristics of play outlined by Caillois (2001, 9 - 10), There are an equal number of different features of that, according to Juul (2005, 36) and his *classic game model*, are necessary in order for something to be classified as a game.

1. All games are rule-based.
2. They have mutable, measurable outcomes.
3. Each of the possible outcomes is assigned a value, either positive or negative.

4. Players must work in order to influence the outcome.
5. Players are emotionally attached to the outcome.
6. They can be played with or without real-life consequences.

There are a number of shared traits and similarities between the characteristics of play and a game. For one, they can both be practiced with or without any real-life consequences. This is because games, just like play, hold no innate sway over our physical world. Their rules possess no corporeal entity that could act on their behest. The act of playing card games is on its own meaningless; somewhat resembling child's play, rather than a game. Only a shared agreement on what happens when one player wins or loses assigns material or immaterial value to it. This touches on the second trait that playing and a game share: rules. The degree to which rules govern each activity varies, as pretending to be a pirate generally does not involve as many restrictions and regulations as engaging in a game of Backgammon. Nonetheless, they cannot be omitted, as even a child with the wildest of imaginations abides by some semblance of rules, or alters them in order to continue playing. The last commonality shared by both play and a game is the need for an element of unpredictability. While most games are played under far stricter time constraints and follow some sort of predefined format, one can never ascertain their outcome with unequivocal certainty. As Adams & Dormans (2012, 2) point out, games with easily expectable outcomes are not fun.

The single most notable dissimilarity between play and a game is the presence or, conversely, the absence of an outcome. The goal-centric nature of games is what makes them distinct from other forms of play. It is no wonder that four of the six features of the *classic game model* involve the concept of an outcome. Peterson (2013, 36) asserts that outcomes function as a measure of progress towards a goal, while simultaneously establishing whether or not a player has been victorious. The emotions that stem from encountering these different outcomes further enhance their appeal. Without this, a player would not care about the game or what happens as a result of it.

2.4 Computer games and second language acquisition

A study conducted in 2006 by Rankin, Gold & Gooch involving a game called *EverQuest 2*, a massively multiplayer online roleplaying game just like *World of Warcraft*, concluded that second language learners saw a 40% increase in their vocabulary as a result of playing the game (Rankin, Gold & Gooch, 2006). While it is almost certain that the social elements and the need to cooperate with other players partly explain such a result, there are also other important factors at play. The chief among these is situated meaning. Gee (2003, 15) describes how words possess more than just dictionary-like meanings. These vary from one scenario to another. According to him, video games allow language to be placed into context, through experience, images and actions. They also do not bombard players with lots of words prior to being needed or useful (Gee 2003, 17.) Players therefore do not need to exert as much mental energy trying to retain large amounts of data, that may or may not prove to be of use at a later time. Reinhardt (2018, 116) makes a not too dissimilar remark by mentioning how in games, vocabulary is contextualized within narratives and that it is this contextualization that allows players to make and learn to form links between them. In essence, computer games help to clarify the meaning of words and concepts by having them be directly tied to action (Peterson 2013, 41).

2.5 The anatomy of a quest in World of Warcraft

As was already previously mentioned in the section *The characteristics of a game*, in *World of Warcraft*, like in other MMORPGs, players must complete tasks known as quests in order to advance in the game and receive rewards. These quests come in a wide variety of forms, ranging from solving puzzles to trading in-game items and more (Peterson 2013, 29; Reinhardt 2018, 92). Despite the fact that they come in all shapes and sizes and a multitude of different scenarios, they all share common structural elements that unify them. They can all be viewed as self-contained narratives, and are often framed as such (Reinhardt 2018, 96). This means that they all have a starting point, a midpoint and an end point. In-game, these different states are indicated to players by three different instructional text segments. The first of these, titled *Description*, acts as a way to set the stage, explaining to the player what kind of a situation they are getting into and what they must do in order to successfully resolve it. As an example, the quest "*Goldtooth",* found inside an in-game area called *Elwynn Forest*, tasks the player with recovering a necklace owned by a woman called Bernice Stonefield that a boy called Billy Maclure lost while playing near *Fargodeep Mine*. As part of the quest *Description*, he says he saw a big, gold-toothed kobold (a fictional creature) pick it up and run inside a nearby mine. If the player chooses to help Billy, the quest will move onto the next phase, *Progress*. Until the player has found the necklace, talking to Billy will offer them a new text segment, Greeting the player by name and asking them if they have found the necklace. This exists to reinforce the notion that the quest will not advance any further until the player has done what the task demands of them. Once the player has found and retrieved the item, in this case confronted the creature known as *Goldtooth*, he can return it to Bernice Stonefield which brings the quest to its conclusion and supplies the player with one final text segment titled *Completion*. In the case of the example quest, Bernice is happy that the player has managed to find the necklace and says she wants to give them her husband's lucky gloves that he apparently forgot to take with him on his last military campaign. In addition to receiving an item, players are also awarded in-game currency and experience points. Once a certain game-defined threshold of experience has been reached the player character's level increases and they must reach a new threshold to increase it once more, leading to further play.

This cycle of *Description*, *Progress* and *Completion* forms the backbone of every quest in the game and their textual elements are what form the data used in the present study.

2.6 Corpus linguistics

The present study's corpus centricity necessitates an overview of what corpus linguistics is and what kind of terminology is related that field. Firstly, a corpus, according to Weisser (2016, 23), is a set of written or spoken data that has been gathered and converted into text based on certain predefined criteria, which can be used as part of linguistic analysis. The primary motivator for their employment in language research is to search for patterns in language use (Scott & Tribble 2006, 6). Instead of relying on a researcher's preconceived notions related to language and grammatical structure to try and intuit such patterns, a large corpus enables the adoption of a statistical approach, one backed by mathematical evidence (Scott & Tribble 2006, 3). Put another way, corpora may be seen as a finite set of language data taken from the infinite amount of possible utterances that speakers can generate (Lüdeling & Kytö 2006, 777). The adoption of such a view shifts the focus of research from studying how humans produce language to what those products are like and what conclusions may be drawn from what has already been said.

Corpora, based on the nature of their data, can be classified as either *general* or *specific*. When labeled as *general*, a corpus is viewed as being representative of a given language in its entirety (Weisser 2016, 32). Such data can be used in all kinds of language research because of its non-specificity. Their counterparts, *specific* corpora, which are also called *domain-specific*, see much more marginal use, as they are constructed to fit highly specific needs, such as studying registers (Weisser 2016, 32). The corpus constructed for the present study can be labeled as being *domain-specific* because, while it exhibits many of the traits of natural language use, the texts have been constructed with a specific use case in mind, both in terms of how they are to be interacted with and what kind of a fantastical setting they are trying to portray.

Another key classification of corpora that should be covered here is the distinction between *synchronic* and *diachronic*. Basically, if the data is contemporary, the corpus is *synchronic*. Alternatively, if it the data is historic and reflects the language or style of a specific era, the corpus is considered *diachronic* (Weisser 2016, 24.) The present study's corpus may be classified as *synchronic*, for even though the game in question is almost 17 years old, the data is not historic in the same sense that, for example, letters from World War 2 reflect the historical period they were

written in. Lastly, the *World of Warcraft* quest corpus is *static* rather than *dynamic*. This is because the data that it contains has been created once and has not or will not be added onto. According to Weisser (2016, 37), corpora, described as being either *static* or *dynamic*, are sometimes alternatively called *snapshot corpora* or *monitor corpora*. This denotes both a difference of usage and research goals, where the former is concerned with examining linguistic data from a specific moment in time whereas the latter seeks to monitor and record change over a lengthier time frame.

Another factor that needs to be considered as part of corpus-based research is the domain of a corpus. It may seem self-evident that if one seeks to study the grammatical and lexical fabric of Shakespearean plays, those plays should be what their data consists of. In other words, all corpora should reflect the target domains of use so that, when analyzed, the findings can be generalized to the domain in question (Lüdeling & Kytö 2006, 1287). In the context of this study it means that, in order to provide relevant insight into the how prevalent core vocabulary is as part of World of Warcraft's written in-game quest instruction, the study's domain must be the written text instructions that accompany the quests so that whatever findings are made from that data are in service of what is being inquired.

In corpus linguistics research, a distinction is made between what are called *types* and *tokens*. *Types* are a label for distinct words within a corpus. *Tokens*, for their part, refer to the amount of instances such words appear as part of the data (Lüdeling & Kytö 2006, 805.) The terms are important because by using them, it becomes possible to make statements about the general structure of all corpora. For instance, it has been observed that in a natural corpus, regardless of its size or scope, most *types* occur very infrequently, perhaps only once. There always exists a small subset of *types*, which have an extremely large amount of *tokens,* that make up a significant portion of all language use in a corpus (Lüdeling & Kytö 2006, 818; Milton 2009, 46.) Such an observation is noteworthy, because it coincides with statements that were posited in an earlier section of the present study when advocating for the use and significance of core vocabulary. In both cases, a small number of very words appear very frequently and thus form most of the language used on a regular basis.

Concordancing is the name given to an analysis technique afforded to corpus-based research. Similarly to how corpus data bestows statistical backing to any findings from its research, *concordancing* lets researchers examine words and their multiple forms in real-life contexts and recorded use, rather than solely relying on native speakers' intuition about the innerworkings of a language (Weisser 2016, 80). This means observing and researching actual manifestations of language, a process during which any hypotheses can be tested against recorded and quantifiable language output instead of individuals' interpretations and preconceptions.

3 THE RESEARCH QUESTION

The purpose of the present study is to examine the lexical makeup of game-developer-written quest texts in the massively multiplayer online roleplaying game *World of Warcraft*. It focuses on trying to identify what is the kind of vocabulary players learn while playing the game, and thus further evidence the language learning potential inherent in video games, which was covered in the *Computer Games and Second Language Acquisition* -section of the study. The study tries to provide research into the exact vocabulary that gamers engage with as part of in-game quests when they step into various virtual realms, in this case *Azeroth* where the events of *World of Warcraft* take place. At its core, it seeks to provide an answer to the following questions:

What is the kind of vocabulary that players most often encounter as part of in-game quest texts in World of Warcraft?

Does the vocabulary that is encountered possess characteristics of core vocabulary?

As these are two separate but closely interlinked questions, it makes sense to seek answers to them in two distinct phases of analysis. The first of these involves identifying the most often occurring vocabulary in the study's domain-specific quest text corpus and seeing how well those findings mirror the structural composition of a much larger general corpus that better reflects the lexical makeup of the English language in its entirety. In order to answer the second research question, the second phase of the analysis sees a small selection of the corpus vocabulary examined with Carter's ten factors, which can be used to identify the how core each word within that small selection is. These ten factors: *syntactic substitution*, *antonymy*, *collocability*, *extension*, *superordinateness*, *freedom from culture-specific uses*, *summary*, *associationism*, *neutral field of discourse* and *neutral tenor of discourse* have been discussed at length in *The nature of high-frequency words and core vocabulary* -section of the present study. The reason why the two phases are interlinked is because frequency of occurrence of words in the study's corpus, which the initial phase of the analysis uncovers in conjunction to its effort to compare and contrast two corpora, is one of the key factors that is used to identify core vocabulary (Carter 1987, 187). By combining the findings of each of the two phases of analysis, it becomes possible to determine the kind of

vocabulary that players of *World of Warcraft* encounter while completing quests and ascertain the prevalence of core vocabulary in *World of Warcraft*'s written in-game quest instruction.

4 METHODS AND DATA COLLECTION

This study's data consists of a corpus of 282,958 words within 2011 individual text documents. Each of these documents houses the dialogue of a singular quest that has been gathered from a website called *Wowhead*. The site hosts individual links to all of the quests featured in the standard edition of the game. While nothing theoretically impedes a researcher from using the site to access and gather all of the quest texts that have ever been written for the 16 year-old game, the sheer amount of manual labor required to do so makes it an ill-fitting endeavor for the scope of the present study. As such, quests from the game's expansions featuring new areas and additional content such as *The Burning Crusade* and *Battle for Azeroth* have been excluded from its data. The data that has been gathered is already abundant enough to be of use as is. A complete, all-encompassing quest text corpus featuring the game in its entirety is a task best left to a team of multiple researchers.

Besides accessing each individual quest link on the *Wowhead* -website, the data gathering process also involved copying and pasting the text segments from the three *Description, Progress* and *Completion* sections of each quest on the website into separate quest text documents stored on a local computer. The documents were then organized into folders named after the game's geographical zones where the tasks are located. Despite the fact that the three descriptors *Description, Progress* and *Completion* are always visible subheadings to players in-game and on the website, they have each been replaced by a single line of two dashes (/) within the text documents. This has been done in order to not pollute the data by artificially inflating the amount of times the three words seem to occur within the actual quest texts. This kind of an approach is not uncommon in corpus linguistics, as according to Weisser (2016, 47), this kind of metadata rarely contributes to the meaning potential of the text that it accompanies. In fact, their omission bestows a greater inspectional acuity for when they do appear naturally as part of the quest narratives, as such occurrences will better reflect their actual prevalence within the text data.

The freeware corpus analysis kit *AntConc* was used to study the corpus. It allows its users to scan and combine large amounts of text files of various sizes and generate data based on different search queries inputted by the user. By specifying a file directory, the program can rank words based on

the number of instances, also known as *tokens*, exist of those words in the files of that directory. Users can also choose to view those instances in their original context and draw conclusions based on where in the text they are and how they are used. For example, the noun *rider*, which is ranked as the 896th most frequent word occurs in the corpus 36 times and never refers to a person riding a horse, but rather to characters that ride winged creatures, such a gryphon. Save for three instances, it always co-occurs with the noun wind that precedes it, creating the compound noun *wind rider*.

The present study uses *AntConc* to generate all of the base numerical data related to the contents of its corpus, token count being the chief metric to be accessed in order to measure the word frequency of each type in relation to others within the corpus. Such data is impossible to generate manually, which explains the necessity of the software. Once gathered, the data was fed into *Microsoft Excel*, which facilitated the creation of graphical representations of the data and where it could be further processed into percentages to be used as part of those graphical representations.

The corpus of the present study was analyzed in two distinct phases. First of these was to see how well the study's domain-specific corpus mirrors the structural composition of large general corpora that are thought to reflect the English language as a whole. As was already stated in the section covering corpus linguistics, in such corpora most types are very infrequent, while a small subset of them have a very large amount of tokens and thus account for the majority of all language use in those corpora (Lüdeling & Kytö 2006, 818; Milton 2009, 46.) Carrying out such an analysis involves comparing the type to token distribution and ratios within the quest text corpus to the observation cited above and other similar ones related general English corpora. Finding any parallels will not only give credence to core vocabulary as a veritable, conceptual tool in vocabulary prioritization, but also testify the efficacy of MMORPGs, and in this case *World of Warcraft*, as vocabulary learning tools which feature language that is authentic.

The second phase of the analysis focuses on the coreness of the vocabulary within the corpus. There are numerous different ways to test that attribute, and one might find themselves overwhelmed by the sheer number of different word types and grammatical elements that are housed within a regular corpus. A study could look at hundreds of different words, and separately

evaluate its adjectives, nouns, verbs, prepositions, and articles. However, doing so requires an amount of time and resources that far exceeds the scope of the present study. In order for the analysis to be feasible, the ten most token rich verb types will be evaluated based on the ten different factors used to measure a words coreness that have been explained in detail in *The nature of high-frequency words and core vocabulary* -section of this study. *Syntactic substitution*, *antonymy*, *collocability*, *extension*, *superordinateness*, *freedom from culture-specific uses*, *summary*, *associationism*, *neutral field of discourse* and *neutral tenor of discourse* will serve as the factors used to determine how core the verbs are. The following factors: *superordinateness*, *freedom from culture-specific uses*, *summary*, *associationism*, *neutral field of discourse* and *neutral tenor of discourse* are covered jointly and summarily since they are metrics which are better suited for the analysis of other word types, but none-the-less deserve to be discussed briefly.

Such an analysis will by no means offer an exhaustive answer to the exact degree of coreness that all of data within the corpus holds, but it does give a strong indication as to what that degree may be.

5 ANALYSIS

5.1 Phase one

There is a broad selection of corpora available online, each of which has been constructed to fill a different need. For the purposes of the present study and the first focal point of analysis, its best to choose a general, non-domain specific corpus that is as large as possible. For that reason, and the fact that the MMORPG being analyzed in the present study was made by a North American company, the Corpus of Contemporary American English, or COCA for short, seems like the best option. It seems like a suitable point of reference because it is a genre-balanced corpus, consisting of more than one billion words in 485,202 texts from the last 30 years (Word frequency data, 2021).

Taking the token distribution data of the top ten most frequent words in both the *World of Warcraft* -quest text corpus and COPA and calculating their percentual weight in their respective corpora, using the formula below, produces the figures, FIGURE 1.1 and FIGURE 1.2, featured on the following page.

$$\frac{Part}{Whole} \times 100 = percentage$$

(Formula used to calculate the percentage, where *Part* represents the amount of tokens of a word and *Whole* represents the total number of tokens in the corpus)

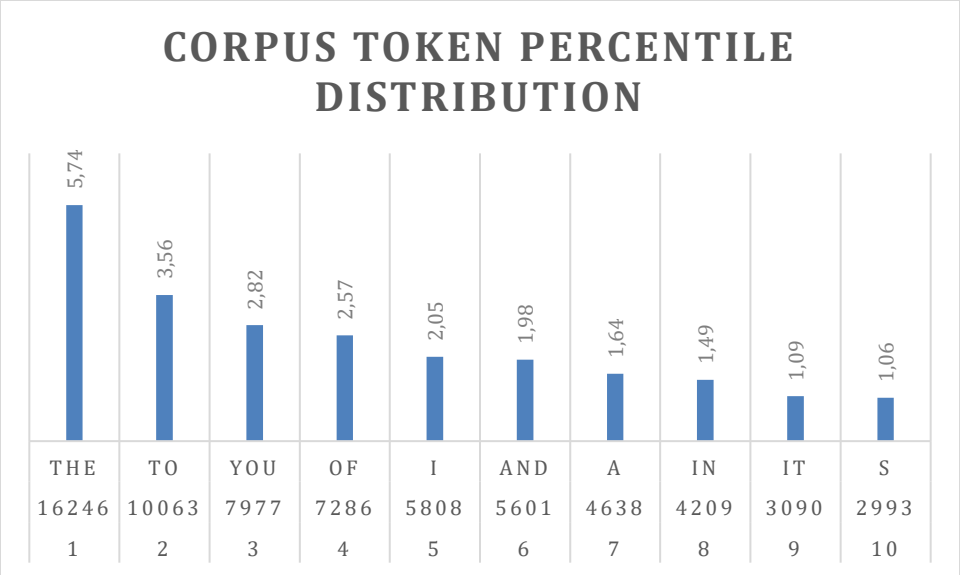

FIGURE 1.1 Corpus token percentile distribution of the present study

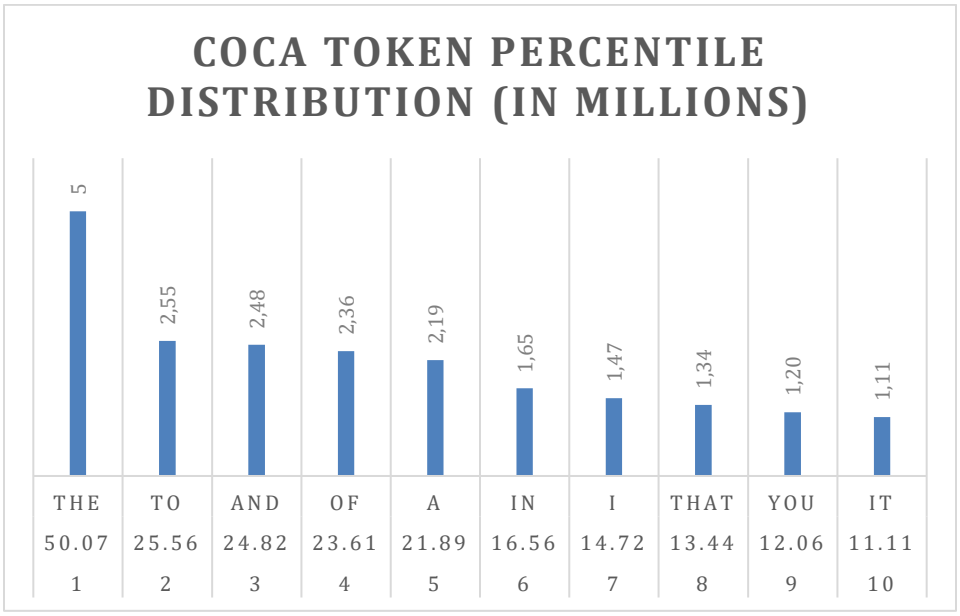

FIGURE 1.2 Corpus token percentile distribution of COCA

Before properly analyzing FIGURE 1.1 and FIGURE 1.2 and their similarities, a few things need to be discussed. First of these is how to read them. The ten blue vertical bars in each graph show the percentual weight of each word within their respective corpus. For example, in FIGURE 1.2, 2.356799742 percent of the billion-word corpus is made up of the word *of*. Below the blue vertical bars are, from top to bottom, the word, i.e., type, in question, its numerical frequency in the corpus, and finally its ranking based on the numerical frequency, with the most common word in the corpus occupying the number one rank. While the percentual weights in FIGURE 1.2 have been calculated using raw, unmodified data, the numerical frequency of each word has been rounded to two decimal places in order to render the graph more readable. Lastly, the letter '*s'* occupying the number ten spot in the graph of the present study's corpus includes all the contractions and possessives that are in the data. The next natural word, in the number 11 spot, is the word *is*. Trying to untangle the contractions presents several problems. Without an elaborate algorithm, it is impossible for the computer to know what word a contraction refers to. For sentences like," *He's made of money.*" and "*He's made a fortune.*", computers require very specific instructions in order to be able to decipher what words have been omitted, since as far as the computerized data string is concerned, "*He's*" are equivalents in both example sentences, human beings just interpret them differently based on the sentence construction.

The second issue that arises has to do with acknowledging that in some cases, a sequence of an apostrophe followed by the letter "*s*" denotes a possessive, in which case it does not mark any sort of omission, and should, perhaps, be categorized as a grammatical construct in the corpus data. The freeware corpus analysis software, AntConc, used in the present study may contain functionality to solve these issues, but no such features were uncovered during research. Finally, an argument could be made that, in the absence of a toolkit that could properly account for the issues that have been raised, the data would cease to be authentic if it were manually altered to display the contractions.

Moving on to the analysis and observations, it is perhaps not too surprising how similar the two figures are. Even though there is small variance in the distribution percentages, the graphs showcase a similar overall trend in vocabulary distribution. They both feature the all the same tokens, save for '*that*', which only appears in the top ten of the Corpus of Contemporary American

English, and the letter '*s*', which has been discussed in the two paragraphs above. While not in the top ten, the word '*that*' also features prominently in the present study's corpus, occupying the number 12 spot. The similarities between the two corpora become even more apparent if one takes into account the fact that the word '*is*' holds the number 11 spot in both. This means that the slight deviations in the tokens of each top ten are not too significant in the greater context.

One interesting observation is the greater prevalence of the words '*you'* and '*I*' in the World of Warcraft quest corpus as opposed to COPA, being even more frequent than the article '*a*' or the conjunction '*and*'. A possible logical explanation for this lies in the present study's corpus' *domain-specificity*. As has been touched upon in the sections, *The characteristics of a game* and *The anatomy of a quest in World of Warcraft,* the quest texts, while also transmitting stories, serve a clear gameplay purpose. They exist to aid *you*, the player in the completion of tasks, which yield in-game rewards. That is why they address *you* directly, because without *you* the game stops progressing.

The reason why the type '*I'* appears to a greater degree in the present study's corpus can also be explained by examining the general staging of quests. Quests, in general, as was the case with the example, *'Goldtooth*', are undertaken by talking to preprogrammed characters which inhabit the digital world of *Azeroth* in *World of Warcraft*. The use of the verb '*talking*' may be a bit misleading however, as in reality the player will simply read the *Description* of a quest and can then either choose to click an icon to either accept or decline the quest. In order better immerse players into the fantastical world, the narratives are framed through the eyes and senses of the preprogrammed characters. In the case of '*Goldtooth*', that preprogrammed character is the boy named Billy Maclure. The first two lines of the quest read, "*I was playing near the Fargodeep Mine, and I think I dropped, er...I mean I saw, the old lady's necklace. Don't ask me how it got there...it wasn't me!"*. This sample alone demonstrates how central the type '*I'* is to the first-person, personal narratives being told throughout the game's numerous different quests that constitute the present study's corpus. COCA, on the other hand, is made up of texts from wide selection of genres, all of which use a different concoction of language components that involve emphasizing some vocabulary over others. Its size and breadth of coverage unquestionably make its data more representative of vocabulary usage at a larger scale, but the parallels between the two corpora are too evident to be

brushed aside. The figures below, FIGURE 1.3 and FIGURE 1.4, show the token percentile distribution pattern of the top 50 most-frequent types in each corpus.

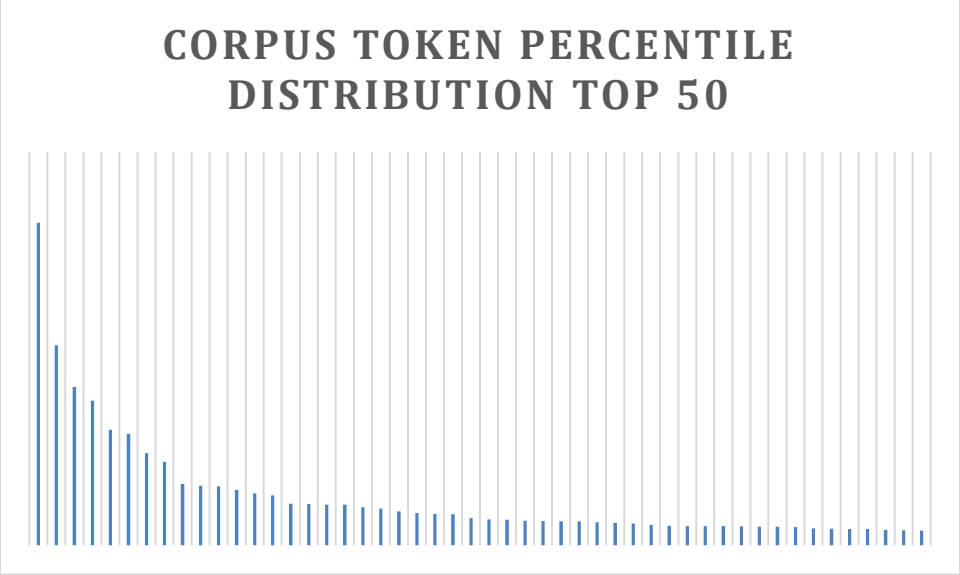

FIGURE 1.3 Corpus token percentile distribution of the present study for the top 50 words

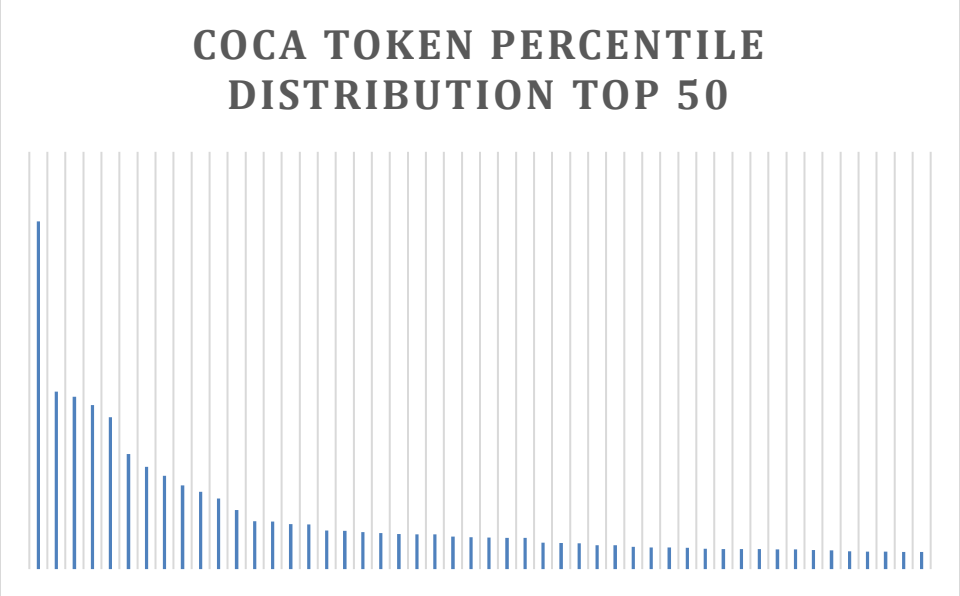

FIGURE 1.4 Corpus token percentile distribution of COCA for the top 50 words

The most salient observation that can be drawn from the two figures above, FIGURE 1.3 and FIGURE 1.4, is that, despite being orders of magnitude smaller in quantity, the present study's corpus is structurally remarkably similar to COCA, which is slightly over 3500 times its size. Such a finding speaks very much in favor of the notion that the quest texts in *World of Warcraft* represent general, and, by extension, natural language. Despite being created for a fantastical target setting, they are very much based on real, organic language and as such feature similar patterns in the distribution of their lexical elements. It would seem, therefore, that there is strong enough evidence to extrapolate from such a conclusion that this would still be the case even if the two corpora were equivalent in size. The two figures show a very clear trend where a comparatively small subset of types account for a significant percentage of all language use in their respective sets of data.

Examining corpora this way reveals an interesting hypothetical opportunity to compare these corpora to ones that have been constructed using samples from English SLA learners. Not only would such a comparison help illustrate how learner language token distribution deviates from native language token distribution, but it would also function as a way to pinpoint any deficiencies in their lexical repertoire and how it is utilized when engaging in foreign language communication. This study posits that learner corpora is demonstrably different in its construction in comparison to English general corpora and other native-made language products.

The key observation drawn from the two figures evidences *World of Warcraft*'s potency as a language learning tool. It shows that the quest texts in the game are not made up of text which does not adhere to natural language patterns as far as the dissemination its lexicon is concerned. Perhaps unsurprisingly, it is not just the distribution of tokens in the corpus that act as a testament to its authenticity as a viable representation of contemporary language being converted to suit a fantasy setting. Looking at a rank-ordered listing of the actual types that make up the data on display in FIGURE 1.3 and FIGURE 1.4, shown in TABLE 1 below, reveals that they also have a significant amount of types in common.

COCA			Present study	
Rank	Type		Rank	Type
1	the		1	the
2	to		2	to
3	and		3	you
4	of		4	of
5	a		5	I
6	in		6	and
7	I		7	a
8	that		8	in
9	you		9	it
10	it		10	s
11	is		11	is
12	for		12	that
13	on		13	for
14	was		14	have
15	he		15	be
16	with		16	me
17	this		17	this
18	as		18	name
19	n't		19	with
20	we		20	will
21	be		21	my
22	have		22	we
23	are		23	your
24	not		24	are
25	but		25	from
26	at		26	but
27	they		27	can
28	do		28	at
29	what		29	on
30	his		30	bring
31	from		31	if
32	by		32	as
33	or		33	he
34	she		34	here
35	my		35	our
36	all		36	they
37	an		37	them
38	there		38	t
39	so		39	find
40	her		40	what
41	about		41	their
42	me		42	not
43	one		43	do
44	had		44	his
45	if		45	one
46	your		46	there
47	can		47	was
48	who		48	take
49	no		49	ll
50	out		50	some

TABLE 1 top 50 types. Unique tokens marked in yellow.

The table above is a list of all of the top 50 most frequent types for both the COCA (on the left) and the present study's corpus (on the right). Words unique to either of the corpora are marked in yellow. The two immediate observations that can be made are that over half of the 50 most frequent words are shared between the two corpora, 28 to be exact and that the 22 discrepancies are situated at the bottom halves of the lists. But perhaps even more noteworthy is the fact that the unique types, especially in the case of the present study's corpus, are not just some made-up, fictitious words. Types such as "*some*" and "*will*", while not found in the top 50 most frequent types in the COPA, are without question integral to the English language. Furthermore prevalence of the types "*bring*", "*find*" and *"take"* are common English imperatives that simply highlight the primary function of the texts in the current study's corpus; to instruct and guide the player in the completion of in-game tasks and objectives. They are also common building blocks in any sort of instructional texts, which are found in all kinds of text media from schoolbooks to instruction manuals. In that sense they are far from being fantastical like the setting that they are applied to; simply ordinary.

The fact that the token "*name*" is the 18th most common token in the *World of Warcraft* -quest text corpus may strike as being odd to some. However, there is a logical explanation for its placing beyond its utilization as both a verb and a noun. The quest texts, which, as has been mentioned, have been gathered from the *Wowhead* -website that hosts them as they appear in the game's code. In them, the character sequence "<*name*>" functions as an instruction for the program to replace that sequence of characters with the player character's name. As a result, players experience the game refer to them by their chosen character name, strengthening immersion. It, much like the token "*you*", clearly shows how the domain specificity of the study's corpus has an effect on the distribution of its lexical elements. Regardless, it is equally as important to highlight how much the two corpora actually have in common. It bears repeating that the most noteworthy observation to be made from examining the two corpora by looking at how types and tokens are distributed in each of them is that doing so illustrates how authentic texts, even ones that have been created for a product set in a fantastical, fictional universe, use language and language

components in very much the same way that they are used in real life. Even the names of places in that fictional setting are often derived from a union of two pre-existing English words. This is evidenced by the corpus' arguably-first made-up noun, *Plaguelands*, which is a proper noun that has been formed from the word *plague*, a virulent, difficult-to-rid disease that affects a significant portion of a population, and the plural form of the noun *land*, *lands*. It is the name of an in-game area that is being ravaged by a plague, hence why its called the *Plaguelands*. Its ranked as the 168th most frequent type in the corpus, meaning that the 167 types before it are all non-made-up types, which again further verifies the corpus data's texts as being representative of regular language. The noun *"Cenarion"*, which is located at the 186th spot based on frequency, appears to be the first truly made-up type within the study's corpus. It is an adjectival noun that is attached to beings and objects which relate to a demigod character called *Cenarius*. But even that word, especially given that the demigod is himself a half-man, half-horse hybrid, could possibly be based on the Latin word "*Centaurus"*, which is a half-human, half-horse creature from Greek mythology (Wowpedia, 2021). What this reveals is that even types in the study's corpus that do not see use outside the confines of the game and its fictional universe often originate from either combinations of pre-existing words, or even in the case that they are brand new, are created using pre-existing words as a starting-off point.

The next observation which should be addressed is the fact that the types "*she"* and *"her"*, which are absent from the study's top 50 most frequent types list, appear noticeably less in that corpus than their masculine counterparts. While there seems to be a similar trend overall in the English language based on their respective rankings in COCA, the phenomenon is far more pronounced in the study's data. Both the masculine tokens "*him"* and *"his"* can be found in the top 50 most frequent types list, at rank 33 and 44, respectively. A search of the entire corpus reveals the feminine forms *"she*" and *"her"* to be as far back the ranks as 114 and 95. What such observations hint at is that not only are female characters less likely to be included in the quest narratives, they also have less active roles than men, as the *"her*" token's higher ranking seems to suggest. This may or may not be intentional however, as it could be something that affects and has affected the fantasy - genre as a whole. The present study speculates that the writers who were tasked with

creating the in-game quest texts and their narratives were simply replicating gradually established, historically-evolved tendencies to cast women and men in certain roles. Roles that grant men the ability to neglect domestic responsibilities and perform heroic deeds while risking their lives.

5.2 Phase two

It is now time to move on to the second phase of the analysis; testing the coreness of the most token-rich verbs within the present study's corpus. An inspection of the gathered data reveals the following ten types as the most abundant verbs of the corpus, displayed in TABLE 2 below.

Rank	Type
14	have
15	be
24	are
30	bring
39	find
43	do
48	take
60	need
66	know
69	get

TABLE 2 the ten most token-rich verbs in the present study's corpus

5.2.1 Syntactic substitution

Testing the words' *syntactic substitutability*, i.e. can it be used as a substitute for other similar or less generic verbs, manifests a non-exhaustive 4-column table, TABLE 3, on the following page. The first column lists each of the ten types, although it should be noted that the type "*are"* has been grouped with the verbs' infinitive, "*be"*. The second column contains a number of different meanings that are associated with the type in question. The third and the fourth column house an example sentence and a version of that sentence where the type in question is used as a substitute for the predicate, respectively.

Type	Meaning	Example sentence	Substitution
have	1. Marks material ownership	”I own a car.”	”I have a car.”
	2. Expresses mental disposition	”I am doubtful.”	”I have my doubts.”
	3. Marks immaterial ownership	”You can choose between two options.”	”You have two options.”
	4. Expresses disapproval	”I won’t allow it.”	”I won’t have any of that.”
	5. To have sexual intercourse	”I have slept with many women.”	”I have had many women.”
	6. To use or exercise a quality	”Please be merciful.”	”Please have mercy.”
	7. To order someone to do something	”Paint the walls red.”	”Have the walls painted red.”
	8. To bear a child	”She bore a child during the war.”	”She had a child during the war.”
	9. To partake in something	”They played a game of tennis.”	”They had a tennis match.”
be/are	1. To symbolize	”God symbolizes love.”	”God is love.”
	2. To possess a quality	”It has green leaves.”	”Its leaves are green.”
	3. Marks existence	”I think, therefore I exist.”	”I think, therefore I am.”
	4. To be situated somewhere	”The book sits on the table.”	”The book is on the table.”
	5. To remain undisturbed	”Do not bother him.”	”Let him be.”
	6. To come and go	”I have visited many different places.”	”I have been to many different places.”

bring	1. To lead, carry something to come along	”She hauled a lot of stuff with her to the dormitory.”	”She brought a lot of stuff with her to the dormitory.”
	2. To cause something to happen	”His actions caused a great deal of pain to his loved ones.”	”His actions brought a great deal of pain to his loved ones.”
	3. To persuade someone	”I will try to convince them.”	”I will try to bring them around.”
	4. To force someone to do something	”He was forced to his knees.”	”He was brought to his knees.”
	5. To bear as an attribute or characteristic	”That man has years of experience in trading stock.”	”That man brings with him years of experience in trading stock.”
find	1. To come upon accidentally	”I encountered two errors in your code.”	”I found two errors in your code.”
	2. To come to possess something through effort	”I need to make the time to study this further.”	”I need to find the time to study this further.”
	3. To reach or connect to something	”The bullet hit the mark.”	”The bullet found its mark.”
	4. Expresses a certain sentiment	”I like painting miniature airplanes.”	”I find it fun to paint miniature airplanes.”
	5. To perceive oneself to be in a particular situation	”I caught myself reminiscing about the past.”	“I found myself reminiscing about the past.”
	6. To regain something that was lost momentarily	“She tried to regain her composure.”	“She tried to find her composure.”
	7. To render a verdict or make a statement	“They decreed him guilty of wire fraud.”	“They found him guilty of wire fraud.”
do	1. Bring to pass	“He was forced to carry out his master’s bidding.”	“He was forced to do his master’s bidding.”
	2. To bring into existence	“He has painted some beautiful landscape paintings.”	“He has done some beautiful landscape paintings.”

	3. To set or arrange	"She tied her hair in a bun."	"She did her hair in a bun."
	4. Get along, fare	"She seems to fare well against all of her opponents."	"She seems to do well against all of her opponents."
	5. To manage	"We will work with what we've got."	"We will make do with what we've got."
	6. To bring about an effect in someone/something	"You've made your parents proud."	"You've done your parents proud."
	7. To exert oneself	"You did what you could."	"You have done your best."
take	1. To capture	"The British stronghold was captured by enemy forces."	"The British stronghold was taken by the enemy forces"
	2. To grasp physically	"Grab the animal by its throat."	"Take the animal by its throat."
	3. To be afflicted by something in a sudden manner	"Within a few days, the whole town had become ill."	"Within a few days, the whole town had taken ill."
	4. To be captivated by someone	"I was quite charmed by her when we met in New York."	"I was quite taken with her when we met in New York."
	5. Move into a location	"The students filled the streets to vent their frustration."	"The students took to the streets to vent their frustration."
	6. Assume the duties related to something or someone	"He assumed the role of director for the upcoming film."	"He took on the mantle of director for the upcoming film."
	7. To choose	"Pick whatever role suits you."	"Take whatever role suits you."
	8. To require something	"You need a lot of time to finish what you started."	"It will take you a lot of time to finish what you started."

	9. To feel a certain way	"There is no need to be offended by what I just said."	"There is no need to take offense to what I just said."
	10. To adopt or make use of something	"They shelter from rain in a nearby church."	"They take shelter from rain in a nearby church."
Need	1. To require something	"I have to see my child right now."	"I need to see my child right now."
Know	1. Expresses familiarity	"Are you familiar with any of her literary works?"	"Do you know any of her litetary works?"
	2. To have an understanding of something	"Do you understand Russian?"	"Do you know Russian?"
	3. To have experience in something	"He has practiced Kung-Fu."	"He knows Kung-Fu."
Get	1. To gain a material possession	"I was given a medal for my bravery."	"I got a medal for my bravery."
	2. To successfully enact one's will	"He managed to make them clean their rooms."	"He got them to clean their rooms."
	3. To understand emotionally	"I feel he really understands me."	"I feel he really gets me."
	4. To be the recipient of an action	"He was run over by a car."	"He got run over by a car."
	5. To bring something somewhere	"Fetch me my rifle and bring it the garden."	"Get me my rifle and bring it to the garden."
	6. To seek something	"I wish to be accepted to the most prestigious university in the country."	"I hope to get accepted to the most prestigious university in the country."
	7. To elicit an emotional response	"What really annoys me is how he thinks he's better than all his peers."	"What really gets me is how he thinks he's better than all his peers."

	8. To understand something	"What I still don't understand is how he managed to escape using just a toothpick."	"What I still don't get is how he managed to escape using just a toothpick."

TABLE 3 a number of different meanings that are associated with the ten most token-rich verbs, along with sentences that showcase *syntactic substitutability*

The table demonstrates the high degree of *syntactic substitutability* that is a key characteristic of core vocabulary. Even this partial listing of meanings shows how these verbs can be- and are- used in a wealth of different scenarios as stand-ins for other verbs. This in part explains their high token counts within the study's corpus. However, verbs like *"need"* show that a type can have a very large amount of tokens in a corpus even without being polysemous. It would seem that the *need* to indicate a deficiency or a material shortage of something is such a fundamental aspect of language that it, along its usage as an auxiliary verb to mark an obligation, cause it to appear in abundance as part of the text data. Furthermore, it is reasonable to assume that the *domain-specificity* of the corpus, which has been discussed in the section named "*Corpus Linguistics"*, and also touched upon in the "*the anatomy of a quest in World of Warcraft"*, is another reason for why the type *"need"* has many tokens. The in-game characters *need* the player to do accomplish various tasks for them. They also *need* in-game items that the player has to then collect for them in order to complete a quest.

The fundamental goal of TABLE 3 is to demonstrate high-frequency verbs', as well as core vocabulary's, syntactic potential to be used in a variety of scenarios due to their often polysemous nature. This increases their likelihood to be encountered in language. However, the multiple different meanings which have been assigned to the types in the table are not necessarily encountered by players of *World of Warcraft*. Additionally, the token counts produced by *AntConc*, which was used to calculate the quantity of tokens that each of the verbs in TABLE 3 has, only include the single form listed in TABLE 3 under the type -column. Despite this, there appears to be a notable link between the high-frequency verbs in the corpus and a capacity for *syntactic substitutability*.

5.2.2 Antonymy

Trying to determine verbal antonyms, i.e., finding verbs with the exact opposite meanings to each the verbs in TABLE 2, is a rather involved process. A concession has to be made by excluding polysemy, which was demonstrated and discussed in relation to *syntactic substitutability*. Even still, attempting to nominate a precise antonym for the verbs *have, be/are, bring, find, do, take, need, know,* and *get* yields mixed results. The existence of grammatical negation in the form of the adverb *"not"*, which functions in much the same way an antonym would, complicates matters even further.

For the word *"have",* many of the possible candidates imply the relinquishment of ownership. In other words, they are not neutral in the same sense as how *having* or *not having something* describes a state. For example, in order *to abandon* something, one must have at one point possessed the item or thing in question. Similarly, *losing* something predicates prior obtainment and possession. In order for the word to be considered a proper antonym, it needs to be free from an association to the act of *losing*. The word "*lack"* is fairly close to being an antonym for the word *"have"*, but it distinguishes itself by expressing a non-binary state of possession. For example, saying that a painting lacks color does not mean it has no color whatsoever. It simply means that the person who has voiced such a judgment feels that there is not enough of it.

Finding an antonym for the word *"be"* is equally difficult. As a word that indicates a state of existing, it too suffers from the same dilemma as is the case with the word "*have"*. Words such as "*cease", "vanish"* and *"desist"* all imply cessation, rather than inexistence. The formation of a non-existent verb, for example, *"disexist"*, could be recognized as a pure antonym for the word *"be",* but such an assemblage is a mere fabrication and conjecture on behalf of the present study.

The third verb "*find"*, when thought of as the act of locating something, refers to a singular action, which ends when whatever is being looked for is located, i.e., *found*. Despite that, determining an antonym is once again exceedingly difficult. The verb "*overlook"* does partially imply the failure to notice or locate something, but it also carries with it the sense that someone is willingly

neglecting to take into account or reprimand an action or behavior that can be considered immoral or unjust.

As for the verb "*do"*, *"abstain"* appears to be a fairly-formidable antonym candidate, as doing so involves choosing not to do something. However, a noticeable difference between the two verbs emerges once again, just as it has when trying to determine an antonym for all of the already discussed verbs. In this case it happens to be volition. Individuals may involuntary do something, perhaps even remain unaware of it occurring. Whereas practicing abstainment requires that a person deliberately and consciously neglects engaging in some sort of, often unwanted, behavior.

Out of all of the ten most token-rich verbs in the present study's corpus, the verb *"take"* arguably possesses the strongest antonymical relation to another verb, which is *"give"*. Both involve the transfer something, either material or immaterial. The latter denotes relinquishment during such a transfer, while the former denotes acquirement. In the strictest sense, a party that receives something has to *take* what is being *given* in order for both actions to be completed. Because this strong interrelation that both verbs have with one another, this study judges them as each other's antonyms.

When examining the verb *"need"*, electing an unambiguous antonym seems once again difficult. However, a reasonable case could be made for the verb *"have"*. This is because *needing* or *having* both express on-going states that remain in-effect based on various conditions. A person may need various amounts of something or require something to varying degrees. Regardless, "*to need*" signals a deficit that remains in effect until the requisite conditions specified by that deficit have been met. Giving a person who needs six apples only two means that they still *need* four apples, and thus remain in a state of *need.* Once four more apples have been given to that person, that person no longer needs six apples, since they now have them. It should be stated that this line of thinking presumes that both states, *needing* or *having* something, stem from an identical set of conditions. In that instance, this study considers the verb "*have*" to be the antonym of *"need"*.

Finding an antonym for the second-to-last verb *"know"* is challenging. This time there are no obvious candidates. The existence of grammatical negation in the form of the adverb *"not"* makes

it exceedingly hard to nominate any verb. Additionally, depending on the context, the verb *"know"* is used to refer to either a cognitive capacity to perceive something, both for a long or a short time, or the mental retainment of information that has been accrued over a long period of time. Interestingly if something is not known to a person or a party, it is indicated by adjectives such as "*unknown"*. Vacuity, when in relation to the absence of perception or cognition, cannot therefore be expressed as an active state that a grammatical subject is in or is causing upon themselves. Based on the ruminations expressed above, this study fails to select an antonym for the verb *"know".*

The final verb's scenario is fairly similar to that of the verb *"take"*. The difference between the most basic meanings of the verbs *"take"* and *"get"* is that *taking* something requires proactivity while the process of *getting* something does not necessitate action from a recipient. Despite that dissimilarity, the verb *"give"* appears to work as an antonym for both, although such a relation is not as convincing in the case of the verb *"get"* as the one between the verb *"take"* and *"give"*. This is due to the impossibility of passively *giving* something, as any such action always requires effort from its performer.

The contemplations expressed above seem to suggest that finding antonyms for verbs is a difficult task. The existence of adjectives carrying a prefix denoting negation, such as "*inexistent"* and "*unknown"*, and the adverb "*not"* are likely to be some of the prime culprits for this particular verbal shortcoming. It would appear, then, that even words which could be considered as core vocabulary do not necessarily possess each of the ten characteristics outlined by Carter (1987). This observation is echoed by Carter when discussing *superordinateness*, which has been covered in *The nature of high-frequency words and core vocabulary* -section of the present study. In relation to *superordinateness*, Carter (1987, 183) makes a remark that words like *car* and *chair* are far more common words in the English language, unlike their superordinates *vehicle* and *furniture*. It can thus be argued that the ten ways to gauge a words coreness are not necessarily all applicable to each and every word or word class. They are more akin to a set of tools that may be utilized in order to inform the election of a set of words as possible core vocabulary candidates.

5.2.3 Extension

The quality of *collocability* is not examined because it deals with partnerships that word classes other than verbs have. *Extension* serves a very similar purpose as a way to investigate word partnerships which feature verbs, namely idioms, phrasal verbs and multiword verbs. The table below, TABLE 4, showcases a number of the phrasal verbs and idiomatic expressions which may be derived from each of the ten verbs, accompanied by as their respective meanings and example sentences.

Verb	Example phrasal verbs and idiomatic expressions	Meaning	Example sentences
have	have at	to deal with through confrontation	"Have at it, gentlemen!"
	have coming	to deserve a punishment or penalty	"You had that coming for you."
	have done with	to finish something	"I wish to have it done with by noon."
	have it in for	be determined to cause harm or suffering for someone	"The coach has always had it in for me ever since I became the new quarterback."
	have none of it	to refuse to involve oneself or entertain an idea	"We will have no such talk in this household!"
	have one's eye on	to observe constantly or have as an objective	"I have my eye on you, you little rascal."
	have it	to claim or allege	"Legend has it that our there's treasure buried deep beneath these ruins."
	have to do with	to be involved in something	"What does that have to with all of this?!"
	have an edge on	To have an advantage over someone or something	"I have an edge on him because of the way I was raised."
	have somebody over	to invite someone over for a particular activity	"We were delighted to have them over for dinner."
be/are	be crystal clear	information that is easy to understand	"The instructions were crystal clear."
	cannot be arsed	to be unwilling to put in the effort to achieve something	"I can't be arsed to call him after what he did."
	could be forgiven for	to be understandable that someone acts or thinks a certain way, even when	"You could be forgiven for thinking that the company has no executives based on how

		doing so is based on incorrect information	disorganized everything seems around here."
	would not want to be in someone's shoes	used to express that the person who is being referred to is in an undesirable or unenviable situation or position	"The amount of backlash he got for announcing that new web series was totally uncalled for. I really wouldn't want to be in his shoes trying to respond to all the hate."
	be all ears	to be excited and eager to hear what someone has to say	"I expected my team to be all ears when I told them about a great new idea I had while showering this morning."
	be up in arms	to voice disapproval or protest vigorously	"The coal miners were up in arms about yet another pay cut."
	be on the lookout for	to actively search for something	"I'm on the lookout for a pair of strong hands to fix my leaking roof."
	be on one's way	to be about to leave	"It's getting late, I must be on my way," she said as she grabbed her coat.
	be a thing of the past	to be obsolete	"Once we settle Mars, our lack of living space here on Earth will be a thing of the past."
	be past one's prime	to no longer be in as good of a condition as in the past.	"As a show host, the dude is way past his prime."
	be into something	to like something	"I'm not totally into the idea of us having to fix other people's mistakes."
bring	bring to mind	to remember	"Seeing our child frolicking in the fields brings to mind all of the times I did the same thing with my brothers and sisters."
	bring down to earth	to be brought back to reality	"I was envisioning myself as a captain in the British Royal Navy until the agitated yells of my mother calling me down for supper brought me back down to earth.
	bring out the worst in one	to evoke the worst traits of an individual	"The anonymity provided by the internet brings out the worst in people."
	bring to bear	to point or aim a weapon	"The USS Alabama brought all of its guns to bear on the stranded enemy vessel."
	bring forward	to present	"The evidence brought forward by the defendant was enough to convince the jury that no crime had been committed."

	bring up-to-date	to inform someone of the latest information about something	“Let me bring you up-to-date on what’s been happening around town while you’ve been gone.”
	bring the hammer down	to reprimand or condemn in a severe manner	“The league’s representatives brought the hammer down on Dwayne for a rule violation.”
	bring to light	to reveal something, often morally suspect	“A local newspaper story has brought to light a nationwide social problem.”
find	find one’s calling	to discover one’s passion in life	“I guess I still need to find my calling in life.”
	find a happy medium	to discover a healthy mixture or a suitable compromise between two extremes	“A film producer’s job is to find a happy medium between letting creative people run wild with their ideas and delivering a finished product on time.”
	find one’s way around	to navigate or move within a space without getting lost	“Finding your way around these labyrinthian halls can be difficult.”
	find it in oneself	to bring oneself to do something	“I hope they can find it in themselves to give the boy another chance.”
	find common ground	to find something that all parties involved agree upon	“Despite their differences, the two women found common ground on how to run a successful business.”
	find fault with	to find a problem with someone or judge someone or something harshly	“He could find fault with just about everything I did or said. There was no pleasing him.”
	find one’s match	to encounter one’s equal in ability, skill or superiority	“It seems that after all these years you’ve finally met your match.”
	find something wanting	to discover that something is lacking and does not meet certain requirements	“Even though the pitch he made was met with some initial enthusiasm, the plan he had outlined on paper was found wanting.”
	find favor with something or someone	to receive someone’s approval	“Albert found favor with the foreign king, who granted him a beautiful chateau on a hill with a view of the entire kingdom.”
do	do someone solid	to help someone	“Could you do me a solid and bring those boxes from the corner into the main hall?”
	do one’s bit	to contribute towards a collective goal	“Everyone needs to do their bit if we are to get this church built before the coming winter.”

	do one's damndest	to give the maximum amount of energy toward a goal	"I will do my damndest to make sure they will never step foot in our town ever again!"
	do somebody no harm	to recommend a course of action as worthwhile or helpful	"It would do him no harm to listen to and heed the advice of those around him.
	do one's worst	to challenge someone to perform something, often negative, to the best of their ability	"Come on! Do your worst! There's no way you could ever harm me."
	do something blindfolded	to do something with extreme ease; in a way that poses no challenge	"I have done this so many times I could do this blindfolded."
	do a number on	to hurt or harm someone or something	"Considering the amount of bruises he has received, last night in the ring must have really done a number on him."
take	be taken up with	to be occupied with something	"Here at work, most of my time is taken up with forwarding calls and arranging appointments."
	to take orders from somebody	to be expected to follow someone's commands	"Are you expecting me to drop everything I'm doing and help you? I don't take orders from you. You are going to have to talk to someone else if you want to solve that problem of yours."
	take offense	to become angry or agitated	"She took offense at the mere suggestion that she was wrong."
	take up with someone	to become friends with someone	"I don't want Jimmy to take up with the wrong crowd. It's tough enough to have to raise him as a single parent."
	take it	assume responsibility for a task	"We can assume that the police will take it from here. Let's leave."
	take notice	to become aware of something	"The department took no notice of the large amount complaints that had been filed against them."
need	in need of something	lacking something	"I'm in need of some advice. I cannot seem to come up with a compelling headline for this article that I am writing."
	need something so bad that one can taste it	to require or need something to almost an unbearable degree	"All of our funding has been pulled. We need new sponsors so bad that I can taste it!"

	need a fix of something	indicates an addictive need that must be satiated	“When I was addicted to smoking, I always needed a fix of nicotine to keep myself going.”
	need to let of some steam	to have a need to unleash pent-up emotions or energy	“Whenever she needs to let off some steam, she goes to gym.”
	one needs to get out more	used to express the idea that someone has been staying in a single location or focusing on a single activity for too long	“While we all respect your tenacity to keep hammering away at your assignment, you need to get out more to clear your head.”
	I don’t need this	an expression of frustration that is used especially when someone already has plenty of other things to deal with	“Please don’t tell me we’ve got no hot water. I really don’t need this right now. I have got to get to work!”
	something is just what someone needs	an ironic phrase used to indicate that something is not wanted.	“Great, a rainstorm in the middle of my morning commute—that is just what I need!”
know	what do you know	A phrase which indicates surprise	“Well, what do you know! If it isn’t my old pal Fred who returns to us after all these years!”
	Be in the know	To be well-informed about a topic, especially an uncommon one.	“You won’t find a better mechanic anywhere within a 300- mile radius. David is in the know when it comes to rust buckets like the one you are riding.”
	Know by heart	Used to refer to a person who has memorized or knows something really well	“We’ve read the passage in church so many times that everyone knows it by heart.”
	God only knows	Said in relation to things that no one, other than God, knows.	“God only knows how long I have to wait to hear back from the company.”
	For all someone knows	A phrase used to indicate uncertainty due to a lack of knowledge or information	“For all she knew, they could have already left the city.”
	To know someone to do something	To know from experience that someone will act a certain way or do something	“These birds are known to migrate in from the south during winter.”
get	Get a load of	Used to direct someone’s attention to something	“Get a load of those sweet rims on that four-by-four! The owner must have some serious cash!”

	Get busted	To be arrested or caught in the act of wrongdoing	“Hey, did you hear what happened to Tommy? He got busted trying to hot-wire a car next to a police station.”
	Get riled up	To become angry or agitated	“That man over there’s got the whole town riled up. They say he bought his way into the mayor’s office.”
	Get in on something	To become privy to secret information or to become involved with activity	“She likes to get in on the fun with her kids whenever she can make for it in her busy schedule.”
	Get the hang of something	to begin to understand or become familiar with how something works or should be operated	“I’m slowly starting to get a hang of how things work around here. The first couple of weeks have been a blur.”
	Get something out of something	To obtain information or other such benefits from an activity or exercise	“I hope you got at least something out of the experience. I’m sure it wasn’t a waste.”
	Get out of something	To no longer be in some state or condition	“His parents worried he would get out of shape lounging around the house all day.”

TABLE 4 a listing of example phrases and idiomatic expressions accompanied by a definition and an example sentence

The table above, which presents an incomplete listing of some of the phrases and idiomatic expressions that are associated with each of the ten most token-rich verbs in the present study’s corpus, highlights the high degree of *extension* that is commonly associated with core vocabulary. It needs to be reiterated that the table offers simply a sampling of some phrases and expressions, as a fully comprehensive variant would be too herculean a task to compile. Regardless, even this snapshot that TABLE 4 offers shows how a notable amount of the verbs’ idiomatic usage *extends* well beyond the boundaries of a singular word definition. It also shows how the core meaning of each verb appears to be extremely elementary, which in-turn facilitates their *extensive* usage within the English language.

5.2.4 The remaining factors

In regard to *superordinateness*, the hierarchical ordination of the verbs proves to be difficult. While arguably all actions that happen are predicated on *doing something,* such instances seem to invariably require some degree of volition. Hence, even though the verb "*do"* appears encompass all actions that can be taken by an individual, even the nine other most token-rich verbs that this study examines, it is difficult to establish an equally clear superordination between, for example, the verbs "*do"* and "*happen"*. One could argue that anything that is ever *done* also simultaneously *happens*, but the inverse of that is not necessarily always the case. The presence, or, in the case of something simply occurring, the absence of sentience could be reasoned to distinguish the two verbs from each other, but that does not solve the issue of which one is more fundamental, *i.e.*, *superordinate*. Due to the inconclusive nature of the considerations expressed above, the present study deems the process of identifying *superordinates* too unrealiable a factor to use to consider the *coreness* of the ten most token-rich verbs of the study's corpus.

The analysis that has been carried out conducted thus far provides a solid foundation to argue that each of the verbs are *free from culture-specific uses*. Their universal presence in the language and ubiquitous usage to describe a plethora different of actions and events reinforces the notion that the verbs are non-cultural and generic in nature, as none of them refer to any culture-specific behaviors. They function as vocabulary which is used to express universal human practices regardless of the socioeconomical environment where they occur. By contrast, an example of a far-less ubiquitous, culturally-delimited verb would be "*skirl"*, which is used to describe the act of playing a bagpipe and their high-pitched sound.

The present study lacks adequate data to determine whether or not the ten most token-rich verbs of its corpus are preferred by speakers when summarizing events or information. This is because the language data of the corpus reflects the computer game that it has been constructed for and does not therefore necessarily include utterances that have been transcribed from live speech. Testing whether speakers prefer to use the verbs when forming and relaying synopses would require a separate corpus of recorded speech by native speakers that would serve as the basis for a far-more factual analysis instead of conjecture.

Finally, the present study is incapable of rendering a decisive verdict on the tonal neutrality of the ten most token-rich verbs of its corpus. Much like the factor of *summary*, testing *associationism, neutral field of discourse* and *neutral tenor of discourse* requires for each of the ten verbs to be examined as part of a separate study, as the present study's corpus is unable to provide a large enough amount of textual data from various different fields of discourse, which could be used to test each of the verbs' atonality by proving that they are present in any and all forms of discourse. The *World of* Warcraft's in-game quest texts, while varied in their execution of dramatical, interactive narratives, have been created for a specific purpose of providing fictional narratives and instructions.

Trying to analyze the coreness of the ten most token-rich verbs of the present study's corpus by using *Syntactic substitution, antonymy, collocability, extension, superordinateness, freedom from culture-specific uses, summary, associationism, neutral field of discourse* and *neutral tenor of discourse* reveals that the study's data is ill-equipped to completely account for all the ten factors. What is needed are different kinds of corpora with which to test the factors of *summary, associationism, neutral field of discourse* and *neutral tenor of discourse*.

6 RESULTS

The two-phased analysis, which sought to identify and examine the exact kind of vocabulary that players of the MMORPG *World of Warcraft* encounter in in-game quest texts, has uncovered a number of noteworthy findings. First of these is that the lexical composition of those texts very closely mirrors the lexical composition of the English language as a whole. Despite being roughly 3500 times smaller in size, the distribution of the types in the present study's corpus very closely resembles the allocation ratios of types in the one-billion-word COCA (Corpus of Contemporary American English) corpus. This is significant, because it gives the lexical structuring within the quest texts irrefutable authenticity as representations of natural language. The token distribution patterns identified in the present study further cement the already-acknowledged pattern that sees a small subset of very highly-frequent words account for an extremely high percentage of all language use (Lüdeling & Kytö 2006, 818; Milton 2009, 46). It bolsters their candidacy as components of vocabulary instruction in language classrooms.

As the present study has found, the similarities between the two corpora are not limited to the numerical distribution of their respective tokens. There also exists a nonnegligible portion of types that the two corpora have in common, which accounts for over half of all the types in their respective listings of the 50 most token-rich types. It means that the most often encountered vocabulary within *World of Warcraft*'s quest texts is not only non-fictitious, but also a fundamental part of the English language. Despite engaging with a fictional universe that contains all kinds of different, fabulous creatures, such as centaurs, the players of *World of Warcraft* are actually coming in contact with language that is demonstrably applicable to real world linguistic interaction. The placement of the shared tokens is also significant, as the discrepancies between the two corpora are distinctly located at the bottom-half of the top 50 listings. It means that the shared tokens are by-and-large equally as prevalent in both the study's *domain-specific* corpus and its larger COCA counterpart, which once more speaks to the quest texts' potency as great sources for language learning. Even the discrepancies between the top 50 tokens of each corpus that the analysis identified can be explained by the *domain specificity* (defined in the *Corpus linguistics* - section of the present study) of the present study's corpus, which causes tokens such as "*I*", *"you", "bring", "find" and "take"* to occur more frequently than in a corpus that factors in different

fields of discourse and mediums. It also should be pointed out once more that even those differing tokens are common, non-game-invented vocabulary that are an integral part of everyday language exchanges that occur between native speakers of English. This demonstrates that invented, fictitious vocabulary plays a much smaller role within the corpus than what one might expect based on the game's fantasy setting.

The uneven and slanted representation of males and females within the text which the study has uncovered seems to be a language-wide matter, and could be investigated further by examining the roles between the two genders that the quests establish as well as by a more thorough analysis of the context that the tokens *"she", "he", "her"* and *"his"* appear in.

The second phase of the analysis has revealed that the ten most token-rich verbs being examined unequivocably display a number of the characteristics that are commonly ascribed to core vocabulary. They all, with the exception of the verb "*need*", possess a high degree of *syntactic substitutability*. That is, their polysemic nature allows them to act as substitutes for numerous other less common vocabulary. The analysis shows that the most common verbs in the study's corpus have so many different use cases in the English language that listing them all is a task that is way too involved for the scope of the present study. Despite the verbs' great capacity to act as verbal substitutes, the difficulty of trying to elect clear-cut antonyms for them proves that types which exhibit some of the characteristics of core vocabulary cannot necessarily be examined using all of the ten factors: *syntactic substitution*, *antonymy*, *collocability*, *extension*, *superordinateness*, *freedom from culture-specific uses*, *summary*, *associationism*, *neutral field of discourse* and *neutral tenor of discourse.* Some of the factors have been found to be inapplicable to verbs or would require the utilization of different kinds of corpora as source data to draw conclusions from. The ten verbs, which have been shown to possess a high degree of *syntactic substitutability*, display a similarly strong aptitude for *extension*. This means that their usage extends beyond the boundaries of a single word definition, especially as constituents of phrasal verbs and idiomatic expressions. However, it bears mentioning that the present study has not verified whether the phrasal verbs and idioms presented are found within its corpus and at what frequency. They simply serve to demonstrate the capacity to which each of them may be *extended.* The fact that they are

abundant can be seen as another key characteristic that further strengthens the sense that the verbs are core vocabulary.

7 CONCLUSION

Based on the evidence presented in the previous section, it is now time to answer the two research questions that were first introduced the in *the research question* -section of the present study.

What is the kind of vocabulary that players most often encounter as part of in-game quest texts in World of Warcraft?

The kind of vocabulary that players of *World of Warcraft* most often encounter within in-game quest texts is by-and-large equivalent to the most-often occurring vocabulary in the English language. The most prevalent vocabulary in the study's corpus is non-fictitious, common English vocabulary that is approximately equally as prevalent in all forms of English discourse, both written and spoken. Fabulous words have a very minor role in the overall lexicon.

Does the vocabulary that is encountered possess characteristics of core vocabulary?

The ten most-token rich verbs in the analyzed sample possess characteristics of core vocabulary, the clearest two being *syntactic substitutability* and *extension*. While the present study has not fully been able to examine each of Carter's (1987, 180-185) ten factors used to determine how core to a language a word is and each of the ten verb's relationship to those factors, it argues that the verbs are core vocabulary. The reason for this is two-fold. Firstly, their high frequency of occurrence, verified in phase one of the analysis, means that they are more likely to be encountered by players than other, less common verbs. That frequency is a key metric in determining a word's coreness (Carter 1987, 187). Secondly, the verbs' extreme *syntactic substitutability* and *extension* capability showcase their relevancy in all forms of English discourse. This means that their pertinence extends well beyond the scope of the game that the present study draws its data from. They can therefore be argued to be core components of English: its core vocabulary.

The findings of the present study are complementary to earlier research which has demonstrated the beneficial impact that playing online video games has on vocabulary learning and retention

(Behbahani & Vahdat 2013; Chen & Yang 2012; Ovalle & Vásquez 2019). By revealing the ordinarity, and the semantically-multiapplicative nature of the vocabulary used in *World of Warcraft*'s in-game quest texts, they establish a further rationale for why playing online games strengthens foreign language learners' vocabulary skills. That vocabulary, which players are exposed to as they complete quests in *World of Warcraft*, can demonstrably be used in contexts that are wholly separate from the context that it was acquired from. Its versatility and universality appear to benefit players in all English language interaction.

One of the main limitations that the present study and its findings suffer from has to do with the extreme amount of data that it deals with. Using a corpus of 282,958 words, it simply cannot examine all of the available data and each individual type in detail. For example, with the omission of all other parts of speech save for verbs, the study is unable to examine the relative coreness of other word classes or semantic components. It is thus unable to decisively confirm the exact prevalence of core vocabulary in its data. It cannot do so even in the case of verbs, as its analytical sample size of ten is as an amount insufficient to draw accurate generalizations from. In order to better gauge the coreness of the vocabulary, the present study would benefit from corpora that would feature a number of the forms of discourse that are absent from its own corpus. An example of such a corpus would be one which consists of various types of written and spoken summaries that have been generated by native speakers. Such a corpus could feasibly facilitate the evaluation of the present study's sample verbs' prevalence in summation.

The research and results that have been presented in this study naturally leave behind several untapped targets for future academic investigation. Among these is a fully-fledged examination of the gender roles present in the game's quests. Such an examination could utilize the token amount discrepancy between the male and female genders as a basis for a thorough examination of not just the quest texts, but also, for example, how different graphical elements in the game portray gender. It may find that the lexical underrepresentation of females in *World of Warcraft* is the cumulative by-product of multiple factors, such as what roles they occupy within the in-game universe.

As the data of the present study only consists of the in-game quest texts of *World of Warcraft*'s base game, future studies could supplement its corpus to include all the expansions that have been

released for the 16-year-old game. Such a corpus could then reflect the entirety of the game's quest-related lexicon, and thus give an even more accurate representation of the words used to guide and entertain players as they engage with the game's world. That corpus could then be compared and contrasted with corpora created from other massively multiplayer online roleplaying games in an attempt to find similarities in the games' lexicons and the distribution of tokens in them.

BIBLIOGRAPHY

Adams, E., & Dormans, J. (2012). *Game mechanics : Advanced game design*. Berkeley, Calif.: New Riders.

Behbahani, A., & Vahdat, S. (2013). The Effect of Video Games on Iranian EFL Learners' Vocabulary Learning. *The Reading Matrix,* 13(1), 61-71. Retrieved from https://www.researchgate.net/publication/301340668_The_effect_of_video_games_on_Iranian_EFL_learners'_vocabulary_learning

Caillois, R. (1961). *Man, play, and games*. New York: Free Press of Glencoe.

Chen, H., & Yang, T. (2012). 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

Chotipaktanasook, N., & Reinders, H. (2018). A massively multiplayer online role-playing game and its effects on interaction in the second language: Play, interact, and learn. In B. Zou, & M. Thomas (Ed.), *Handbook of Research on Integrating Technology Into Contemporary Language Learning and Teaching* (pp. 367-389). IGI Global. http://doi:10.4018/978-1-5225-5140-9.ch018

De Aguilera, M., & Méndiz, A. (2003). Video games and education: (Education in the face of a "parallel school"). *Computers in Entertainment*, 1(1), 2-14. Retrieved from https://www.researchgate.net/publication/220686511_Video_games_and_education_Education_in_the_face_of_a_parallel_school

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York (N.Y.): Palgrave Macmillan.

Gee, J. P. (2008). *Cats and Portals: Video* Games, Learning, and Play. *American Journal of Play,* 1(2). Retrieved from https://www.journalofplay.org/sites/www.journalofplay.org/files/pdf-articles/1-2-article-cats-and-portals.pdf

Juul, J. (2005). *Half-real : Video games between real rules and fictional worlds*. Cambridge, Mass: MIT Press.

Lüdeling, A., & Kytö, M. (Eds.). (2009). *Corpus linguistics : An international handbook. volume 2*. Berlin ; New York: Walter de Gruyter.

Marzano, R. J., & Simms, J. A. (2013). *Vocabulary for the common core*. Bloomington, Ind.: Marzano Research Laboratory.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol ; Buffalo: Multiligual Matters.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Newgarden, K., Young, M. F., & Zheng, D. (2012). Multimodal analysis of language learning in World of Warcraft play: Languaging as Values-realizing. *ReCALL*, 24(3), 339-360. doi:10.1017/S0958344012000183

Newgarden, K., & Zheng, D. (2016). Recurrent languaging activities in World of Warcraft: Skilled linguistic action meets the Common European Framework of Reference. *ReCALL*, 28(3), 274-304. doi:10.1017/S0958344016000112

Ovalle, J., & Vásquez, G. (2019). The Influence of Video Games on Vocabulary Acquisition in a Group of Students from the BA in English Teaching. *GIST Education and Learning Research Journal*, 19(1), 172-192. doi:10.1017/S0958344016000112

Peterson, M. (2013). *Computer games and language learning* (First edition ed.). New York, NY: Palgrave Macmillan.

Piirainen-Marsh, A., & Tainio, T. (2009). Other-Repetition as a Resource for Participation in the Activity of Playing a Video Game. *Modern Language Journal,* 93(2), 153-169. doi:10.1111/j.1540-4781.2009.00853.x

Rankin, Y., Gold, R., & Gooch, B. (2006). 3D Role-Playing Games as Language Learning Tools. *EG Education Papers*. The Eurographics Association. doi: 10.2312/eged.20061005

Reinhardt, J. (2018). *Gameful second and foreign language teaching and learning : Theory, research, and practice.* Cham: Palgrave Mcmillan.

Scott, M., & Tribble, C. (2006). *Textual patterns key words and corpus analysis in language education*. Philadelphia: J. Benjamins.

Statista. (2015). Number of active video gamers worldwide from 2015 to 2023. Retrieved from https://www.statista.com/statistics/748044/number-video-gamers-world/

Statista. (2020). Number of video gamers worldwide in 2020, by region. Retrieved from https://www.statista.com/statistics/293304/number-video-gamers/#statisticContainer

Weisser, M. (2016). *Practical corpus linguistics : An introduction to corpus-based language analysis* (First edition ed.). Chichester, England: Wiley Blackwell.

Wowpedia. (17.4.2021). Cenarius. Retrieved from https://wowpedia.fandom.com/wiki/Cenarius