UNIVERSITY OF TARTU
DEPARTMENT OF ENGLISH STUDIES

Exploring Dialogue in RPG Video Games: An Additive Multi-Dimensional Approach

MA thesis

JANELY RÜDEIN
SUPERVISOR: Assoc. Prof. JANE PADRIK

TARTU
2025

ABSTRACT

Although video games increasingly capture the attention of global audiences, their function as input for second language acquisition remains unaddressed in current research. This thesis aims to provide a detailed description of dialogue in role-playing games (RPG) by conducting an additive multidimensional analysis on The Video Game Dialogue Corpus, focusing on the characteristic features of RPG dialogue, its comparison to written and oral registers, and the possible effects of subcategories, such as ratings, on dialogue. The thesis is divided into two chapters: the first chapter describes video games and previous research on video game dialogue, and the second chapter contains the empirical analysis. The results of the analysis establish the RPG dialogue as a complex register, combining features of spoken and written language.

Keywords: video games, role-playing games, video game dialogue, corpus linguistics, additive multidimensional analysis, multidimensional analysis.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ARPG – Action Role-Playing Game

CALL – Computer-Assisted Language Learning

DGBLL – Digital Game-Based Language Learning

JRPG – Japanese Role-Playing Game

MAT – Multidimensional Analysis Tagger

MMORPG – Massively Multiplayer Online Role-Playing Game

RPG – Role-playing Game

SLA – Second Language Acquisition

SRPG – Strategy Role-Playing Game

TTR – Type-token ratio

TTRPG – Table-top Role-playing Game

VGDC – Video Game Dialogue Corpus

WRPG – Western Role-Playing Game

INTRODUCTION

The interactive and entertaining medium of video games increasingly attracts the attention of players from all age groups across the world. The constantly growing influence and importance of video games as a form of entertainment is evidenced by the globally thriving video game industry and a rapidly growing player population (Newman 2004: 3; Wolf 2008: 1; Granic et al. 2014: 66). According to Kantar Emor (2022), it is estimated that in the age group of 11–16, about 76% of Estonian children have played video games and 58% play video games daily. Although gaming is frequently viewed as a hobby for children, research has shown that players come from various age groups, with a growing number of adults and older individuals participating in gaming. In their report, Entertainment Software Association (2024) mentions that the average age of a video game player in the United States has increased over the last 20 years. Furthermore, about 30% of video game players in the United States are over 50 years old (Entertainment Software Association 2024). Therefore, video games can be considered a popular form of media for all ages, and their impact and influence are comparable to film and television.

Following the rising popularity of the medium, the field of video game studies has become increasingly relevant. Still, disagreements appear over the object of the study and its characteristic features. Esposito (2005: 1) defines a video game as a game, possibly including a narrative, played with the assistance of audiovisual devices. In contrast, Granic et al. (2014: 67) emphasise the interactivity of video games in their definition, whereas Wolf (2008: 3; 5) requires the presence of a microprocessor, rules, conflict, interaction and reaction in a video game, also noting that the term is used loosely in society. Although the term “video game” is often used interchangeably with terms such as “computer game”, “console game”, etc, in this thesis, the term “video game” is chosen as an umbrella term for digital games regardless of their platforms. Video games are further classified into various genres and sub-genres. Burn

and Carr (2006: 20–21) and Zagal and Deterding (2018: 38–39) distinguish the genre of role-playing games (RPG) by the inclusion of specific elements such as a customisable protagonist, progression of skills through experience, multiple quests, and journey-based narratives. Research also emphasises the significance of dialogue and reading in the gameplay of RPGs (Burn and Carr 2006: 21; Rennick et al. 2023a: 2); importantly, these games tend to include significant amounts of text and dialogue.

Due to the popularity of video games, the language they contain constitutes a significant part of contemporary language use in digital domains. This popularity has also led to the use of video games in educational contexts, which in turn inspires game developers to create games targeted towards educational use, such as *Minecraft: Education Edition* (Kuhn 2017: 215). Although video games can be helpful for teaching different skills, they also emerge as a potential input for second language acquisition, as they are often played in foreign languages. Still, video games remain relatively unexplored in current linguistic research, despite their prevalence and significance as a cultural medium. As video games generally include either spoken dialogue and written text displayed on screens, they can provide a unique environment for experiencing and interacting with a foreign language. From a foreign language classroom perspective, video games can also provide opportunities for language acquisition in an enjoyable environment. Research addressing the use of video games in Second Language Acquisition (SLA) has reported positive effects. Cornillie (2022: 273) states that the engaging and interactive nature of games can promote the learning of foreign languages and claims that correlations have been found between gaming frequency and knowledge of lexis in a second language (Cornillie 2022: 278). Cornillie (2022: 278) also sees gaming as beneficial for acquiring new vocabulary in a second language. In their review, Li (2019: 484) also concludes that video games include multiple features, such as repetition, guessing, and collaboration, which can assist in the SLA process. Cornillie (2022: 276) further claims that the benefits of

SLA via gaming include accessibility, interactivity and informality. However, prior to the incorporation of video games into second language studies, their linguistic contents should be studied and described in detail.

Previous research at the English department of the University of Tartu regarding language use in video games has been limited, as the primary focus has been on either the language use of gamers (Lifländer 2020, Rätsep 2022) or the players' perception of the advantages computer games bring to English vocabulary acquisition and using video games to teach English vocabulary (Mehine 2021, Rodendau 2017, Mägi Ravn 2021). Whilst focusing on a single game, Saks (2017) analysed the use of verbs, and Jaagola (2016) focused on narrative. Meier (2022) also researched video games but focused on developing a new methodology for analysing gender in video games. Meier (2022: 77) developed the approach of *preferred playing,* which can be used to study the representation of gender in video games in a way that incorporates video-game-specific aspects which do not appear in other forms of media. Meier's approach, however, requires playing, which makes it unsuitable for analysing large collections of games. Thus, the current thesis aims to contribute to the research gap by conducting a linguistic analysis on an extensive video game dialogue corpus.

Although video games usually include large amounts of dialogue available in machine-readable format, the use of video game corpora in video game language research has only recently begun to be explored. Dixon (2024a), Heritage (2020), van Stegeren and Theune (2021), and Hämäläinen et al. (2022) mention the scarcity of corpus-based video game research but indicate a growing interest in the field. Furthermore, Rennick and Roberts (2024) and van Stegeren and Theune (2021) specify that current research adopting corpus linguistic approaches has primarily focused on the discourse surrounding video games rather than the text included directly within the games. The growing interest in video game language is also evident in the recent creation of several video game corpora. Currently existing video game

corpora include either a single game as in Hämäläinen et al. (2022), a few individual games such as Dixon's (2024a) Single Player Offline Game Corpus and Bianchi's (2024) ViPiCo, a game series, such as Lazzeretti and Gatti's (2023) corpus of *Animal Crossing* series. Recently, however, the call for a large-scale video game corpus has been answered by the Video Game Dialogue Corpus (VGDC), which Rennick and Roberts (2024) and Rennick et al. (2023a) claim as the first publicly available resource making large-scale empirical analysis of video game dialogue possible. Whilst the VGDC corpus can be considered large-scale in the context of game studies, as it includes data from 50 games, it only features games from the RPG genre. Therefore, an analysis of VGDC can provide insight into the characteristics of dialogue within the RPG genre.

In the thesis, the VGDC corpus will be analysed using the Multi-Dimensional Analysis (MDA) methodology developed by Biber (1988) to provide insight into the characteristics of the dialogue included in video games. Conducting MDA on video game dialogue data can provide new information describing English language use in video games. This approach has previously been used to study various multimedia texts, from song lyrics (Werner 2021), Internet language use (Biber and Egbert 2016), to tweets (Clarke 2022). Although Dixon (2022b) used MDA on video game dialogue data, he provided partial descriptions for a small dataset, meaning that the full scope of MDA has not been applied to VGDC or any other extensive video game dialogue corpora so far. The MDA methodology uses factor analysis to identify sets of co-occurring features within texts, which are grouped into specific registers and dimensions. Berber Sardinha and Veirano Pinto (2019: 1) emphasise the wide range of applications for the MDA method, its use in describing registers and their specific functions in communication. As MDA also incorporates register analysis, it can be used to compare texts within a framework of established registers and text types in written and spoken modes (Biber 1995: 19). Furthermore, conducting MDA on video game dialogue can establish the video

game dialogue register within Biber's register framework and highlight any differences and similarities with other registers included in the framework.

Having knowledge of the characteristics of video game language, such as its similarities to spoken or written language, can be beneficial for educators, who can recognise the impact of video game language on learners and therefore be more aware of aspects of language that are more familiar to students who play video games, or choose to incorporate video games for teaching specific linguistic features.

This thesis aims to answer the following questions: 1) To what extent is the dialogue within role-playing video games classified as one of the spoken or written registers in the framework of MDA? Although the dialogue included in video games aims to emulate spoken dialogue between characters, the language included is not spontaneous but scripted by the writers of the game as text. Therefore, the hypothesis for this thesis is that the dialogue within video games will not be classified as one of the spoken dialogue registers in the framework of MDA, but rather as a register of written language. Following that, the second research question asks: 2) Do different subgenres, age ratings and time periods influence the dialogue included in the RPGs? In light of the limited research on the effects of subgenre and age ratings on video game dialogue, this thesis will take an exploratory approach to gain deeper insights into the topic of RPG dialogue and seeks answers to the third research question: 3) Which linguistic features are characteristic of video game dialogue? Given that video game dialogue is an edited written text which aims to emulate spoken dialogue between characters, it could feature characteristic features of written or spoken registers.

The MDA is conducted on data from the *Video Game Dialogue Corpus* (Rennick and Roberts 2023; Rennick et al. 2023a), a publicly available corpus which contains data from 50 games classified as role-playing games (RPGs). The MDA is conducted using the Multi-dimensional Analysis Tagger (MAT) software (Nini 2019), which replicates Biber's (1988)

process for multi-dimensional analysis, tagging the text according to the MDA framework and providing each text with a specific text register and type.

The first chapter of the thesis focuses on establishing a definition for video games and role-playing video games within the thesis by examining video game research and previously proposed definitions for the concepts of game, video game, and video game genre. Additionally, the relevance of dialogue within video games is examined, alongside two approaches to studying video game language. The empirical analysis of the thesis appears in the second chapter of the thesis as the additive multidimensional analysis of RPG video game dialogue in VGDC. Within the second chapter, the method and methodology of MDA are described in detail, along with its central concepts such as dimension of variation and linguistic co-occurrence. Two different approaches to conducting MDA are also described. This is followed by the description of video game corpora and their specific creation process. A detailed description of the analysis process is provided in the third section of the second chapter, and the results of the analysis are also reported there. The final section of the second chapter is the discussion, where results for specific dimensions are examined qualitatively and interpreted with the goal of providing answers to the proposed research questions.

1. VIDEO GAME RESEARCH AND LANGUAGE IN VIDEO GAMES

The goal of the first chapter is to provide a detailed description and definition of the object of research, considering video games in general and, more specifically, role-playing video games (RPGs) and their dialogue. The first section establishes a definition for video games by exploring different definitions provided in previous research for video games and games in general. Following that, the currently prevalent categorisation system for video games, the genre system, is described, and the definition is further refined by focusing specifically on the genre of role-playing video games (RPGs). The chapter concludes with an examination of dialogue within video games and provides an overview of prior research focusing on the language included in video games, including the ludolinguistic and digital game-based language learning perspectives.

1.1. Role-playing Video Games: Definitions

1.1.1. Defining Games and Video Games

The term "video game" indicates that video games can be considered a specific type of game. Definitions for the act of game can be, therefore, incorporated as the starting point into the process of defining video games. Although a seemingly simple action, various definitions have been proposed for the act of gaming. Although Caillois initially defines a game as an activity following a set of rules (Perron 2023: 145), he also creates a systematic approach, where games are established simultaneously across a continuum and four categories. Caillois' (2001: 14–27) categorisation draws from the dominant element in the gameplay:

Agôn – games of skill and competition, such as football or chess

Alea – games of chance, such as the lottery

Ilinx – games characterised by rapid movement, disorder, and the pursuit of vertigo, such as spinning in circles, roller coasters

Mimicry – games involving roleplay and imitation, such as playing house

According to Caillois (2001: 13), further categorisation is possible as games are also located on a continuum, where on one end, *paidia* describes energetic and impulsive games which often feature improvisation, and on the other end *ludus,* applies to disciplined games with fixed rules, which often require skill and effort.

Crawford's (2003: 8) definition emphasises the involvement of multiple players, as he sees games as entertaining interactive conflicts where players have individual goals and interact with each other to disrupt each other's goals. Juul (2003: para.3) sets three requirements for the definition of game: (1) description of a rule system, (2) description of the interaction between the player and the game, and (3) description of the connection between playing and non-playing. Based on the examination of seven different definitions for the act of game, Juul (2003) concluded that most definitions include reference to some, if not all, previously mentioned requirements Additionally, Juul (2003) lists elements that reoccur across different definitions, such as rules, outcome, goals, interaction, contrast between game and non-game, and contrast between game and work. Establishing that games should be defined by specific shared features, Juul defines games as follows:

> A [classic] game is a rule-based formal system with a variable and quantifiable outcome, where different outcomes are assigned different values, the player exerts effort in order to influence the outcome, the player feels attached to the outcome, and the consequences of the activity are optional and negotiable. (Juul 2003: para. 15)

The definitions provided by Juul, Crawford and Caillois focus on similar aspects, such as interactivity, end goals, and gameplay elements, which vary depending on the type of game. Although originally created to define games without an electronic component, these definitions can also be used for describing video games, which combine a game with electronic components, with some modifications. Although video game genres are more numerous than the four categories established by Caillois (2001: 14–27), they also function as a mode of categorisation based on gameplay, among other features. Additionally, the descriptions for Caillois' games can apply to video games, as elements such as roleplay, competition, and

chance do appear in video games. The same applies to the *paidia-ludus* continuum, as different video game genres can be situated on that continuum. Crawford's (2003: 8) definition is mostly suitable for defining video games, except for his requirement for multiple players, which would exclude all single-player video games, unless the mentioned "player" could also be interpreted as digital, computerised players. Juul (2003: para. 46) recognises that video games push the boundaries of his game definition, and different genres of video games modify the game requirements to suit the characteristic style of gameplay for specific genres. Furthermore, Juul (2003: para. 46) argues that video games are distinct due to the computer's role as the enforcer of the rules, whilst also providing the player with the option to modify the rules.

As with games, the definitions of video games vary in required components and emphases. Esposito (2005:1) defines video games as games, possibly containing a narrative, played with the assistance of audiovisual devices, while Granic et al. (2014: 67) put emphasis on interactivity. Wolf's (2008: 3,5) definition combines the previous two, listing rules, conflict, interactivity, reaction, and microprocessors as obligatory elements for video games. Newman (2004: 27) suggests that video games include elements from all categories proposed by Caillois (2001: 14–27) with the addition of computer-generated environments and participation. Clearwater (2011: 36) also emphasises the equipment and context of playing as relevant characteristics. Perron (2023: 145) relates video games to games based on the necessity of rules, adding that rules become more prevalent for video games, where only specific actions, which are programmed into the game, are possible.

Based on the examination of several definitions of games and video games, video games are defined for the current thesis as a type of game, where characteristic game aspects, such as rules, interaction and outcome, are combined with obligatory electronic elements, which can differ for different platforms, such as personal computers, mobile phones, gaming

consoles. As with regular games, variation occurs within the concept of video games. Video games can be classified further according to genres and gaming platforms.

1.1.2. Defining Video Game Genres

Like other forms of popular media, video games can be categorised and described by their genres. Some examples of video game genres are action, adventure, first-person shooter, and strategy games, and further examples can be seen in Figure 1. Online game distribution platforms often rely on genre-based classifications for the categorisation of games, as seen in Figure 1. However, it is notable that the lists of genres on those platforms differ between different game distributors, as is evident in Figure 1: whilst both *Epic Games Store* and *Steam* list genres such as Action, Casual, Puzzle, RTS (Real Time Strategy), and Tower Defense, there are others that are featured on only one platform, such as Dungeon Crawler, MOBA (Multiplayer Online Battle Arena), and Sandbox and Physics. Furthermore, genres can be listed under different names and in different combinations, such as First Person and Shooter, which appear as two distinct categories on *Epic Games Store,* compared to First-Person Shooter and Third-Person Shooter in *Steam.* Different sub-genres are also featured in these categorisations. On *Steam,* genres such as Action RPG, Adventure RPG and Strategy RPG are listed alongside the genres Action, Role Playing, Adventure and Strategy. These distinct categorisation differences can be attributed to the fluid and constantly developing definition of genres within the field of video game studies and in the video game industry.

According to Arsenault (2009: 149), the genre-level classification has been used in video game studies as the middle ground to avoid overgeneralisations (considering video games in general) and hyper-specific approaches (focusing on specific games and game series) in video game research. It is important to mention that the characteristics used for determining the genre of video games are distinct from the elements used in literature and linguistics and

other media (Arsenault 2009: 149). A closer examination of this method of categorising video games is, therefore, required.

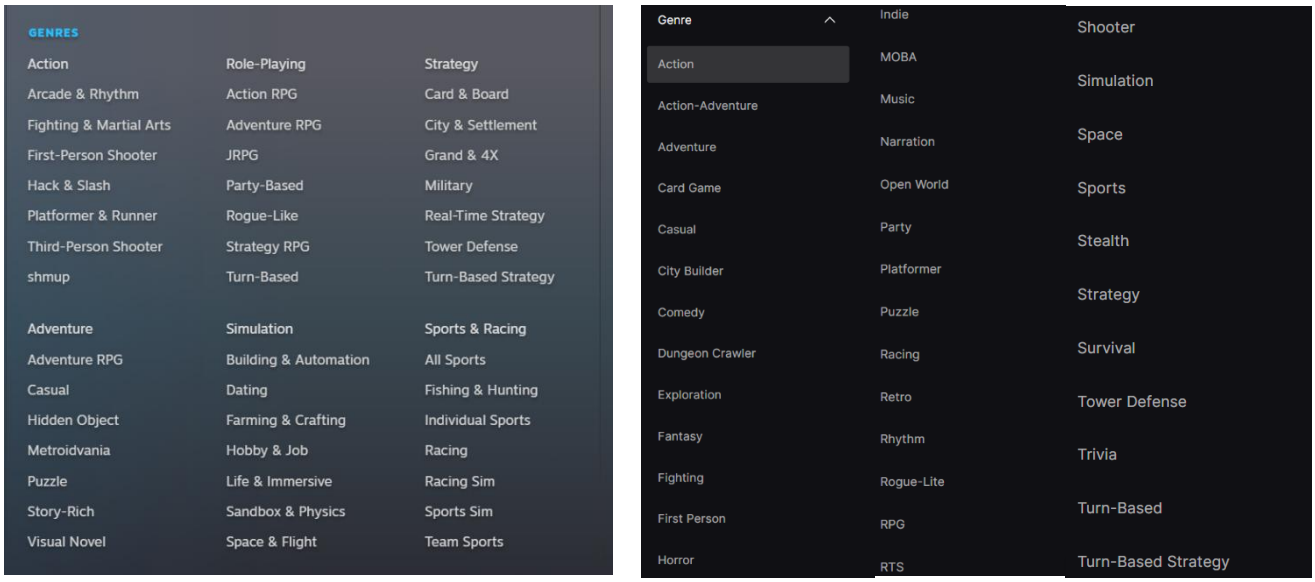


Figure 1. Video game genre categories as presented on game distribution platforms: *Steam* (left) vs. *Epic Games Store* (right)

Newman (2004: 12) notes that the terminology for video game genres originated from game reviews and magazines and was eventually adopted by the game industry. However, Newman (2004) still recognises the role of the gaming community and their expectations in the formation of genres. A similar emphasis on the players' role appears in Perron (2023: 150), who suggests that video game genres can be identified by their specific conventions, which players recognise and learn to expect in gameplay. Such conventions may include specific themes, game controls, plot elements and symbols among others (Perron 2023: 150). Taking a more general approach, Arsenault (2009: 150) considers genre as the combination of game mechanics, functions and aesthetics. Similarly to genre in other forms of media, Perron (2023: 151), Arsenault (2009: 149), Dixon (2022a: 166), and Clearwater (2011: 41) agree that video game genres and genre-defining conventions are not necessarily fixed but relatively mobile, as multiple genres can share conventions, and new games can define new genres and expand the existing ones.

Video game genres thus emerge as the combination of multiple aspects of video games. However, the ambiguity and complexity of the genre-based classification of video games have also been criticised. Arsenault (2009: 149) is critical of the genre classification of video games due to the lack of clarity and system for the conventions considered for determining genres. Arsenault (2009: 156–157) and Clarke et al. (2015: 15) find the current genre classification insufficient, claiming that the characteristics used for determining genres are dissimilar and therefore should not be grouped under a single concept. Similarly, Clearwater (2011: 35) finds genre labels problematic as they fail to address the variety included in a single video game. The diversity of characteristics included in a single genre label is also visible in Figure 1 in the names of different genres: the First Person Shooter genre combines the description of the primary gameplay activity (shooting) with a description of the camera perspective (first-person); Sports, Dating and Racing genre names refer to the primary theme of the gameplay; Turn-Based, MOBA, Open World, Platformer labels describe the prevalent mechanics of the gameplay; Indie describes the game developers rather than the game itself, and Fantasy and Horror, categories borrowed from literary genres, can also describe the themes found in the games.

Apperley's (2006: 7) criticism of the current genre system addresses its ties to previous media forms, which are considered dissimilar to video games. This perspective is supported by Clarke et al. (2015: 2), who claim that the genres established for other media forms are not suitable for describing the interactive medium of video games. Taking a less critical approach, Clearwater (2011: 43) concludes that genres should be viewed as an ongoing process, and therefore it is necessary to readdress and adapt the definition of genre constantly. This proposed adaptation of genres is, however, opposed by Clarke et al. (2015: 12) and Reinhardt (2019: 90), who claim that the constant development of genres and the addition of new genres add to the ambiguity of the system, and make it confusing for the users of these systems. Clarke et al.

(2015: 12–13) also find the subjectivity and relativity of genres problematic, as many genres are defined by player communities rather than based on characteristics of games. Some examples of genres which emerged from the gamer communities include Metroidvania, 4X, and the Rogue-like and Rogue-lite game genres, which are also included in Figure 1. Clarke et al. (2015: 13) conclude that the instability of the genre system also affects the game industry and research negatively, hindering innovation and creating challenges for categorisation.

Alternative approaches have been proposed to combat the overlapping categories and ambiguity of the genres. Apperley (2006: 21) calls for the development of a system that would emphasise aspects of interactivity over other features, such as narrative and visual aesthetics, which are currently included in the concept of genre in some instances. In a similar vein, Dixon (2022a: 166), although not explicitly against genres, warns against generalising research results by game genres and recommends using game mechanics as the basis for categorisation instead. Newman (2004: 13) also advises against using genre as the main distinction in video game studies, seeing its potential to distract from the ludic aspects of video games, but suggests the use of different gaming platforms (such as console, personal computer, mobile) as an alternative basis for classification instead. Similarly, Wolf (2001: 114) proposes using interactivity as the primary basis for classifying video games, emphasising its necessity for all video games. Following this approach, Wolf (2001: 115–116) categorises 42 video game genres, drawing on aspects such as in-game interactions, objectives, character types and control mechanisms. Clearwater (2011: 34) remains critical of the genres proposed by Wolf (2011); he sees the proposed genres as suitable for earlier video games, but insufficient for describing the current state, thus further emphasising the constant development of genres. Clarke et al. (2015: 14-15) propose several alternatives for current genre considerations: establishing prototypical examples for genres or replacing specific genre labels with a combination of aspects that are relevant to a specific genre. Another alternative already used in *Steam* is the use of user-defined

tags for game identification; however, Clarke et al. (2015: 10) report that the tags often include information about specific details in games which are not related to genre, such as visual style, mood and setting.

Although several alternative classification options have been proposed in research, these options have not been widely adopted by the media, game industry or game studies, which still predominantly rely on the complex and ambiguous genre system to categorise different video games. Despite its several limitations, Clarke et al. (2015: 16) conclude that currently, genre functions as the primary classification system in media and marketing and for gaming communities. For developing alternative categorisation options, Clarke et al. (2015: 16) emphasise the necessity of researchers co-operating with gaming communities. The thesis will adopt the prevalent approach in the discipline and use the genre classification for describing video games, as this categorisation appears prevalently in video game language research and available datasets and allows for easier comparison with previous findings. Therefore, genre affiliation for the data within the thesis will follow the genres established in the marketing of video games.

1.1.3. Role-Play and Role-Playing Video Games

As the data for this thesis is a corpus of dialogue from video games in the role-playing games (RPG) genre (Rennick et al. 2023a: 2), a closer examination of the RPG genre and its characteristics is needed. The RPG genre could be considered a contemporary development of Caillois' game category of mimicry. Mimicry, as established by Caillois (2001: 20), relies on the player's imitation of another being or entity, animate or inanimate. Therefore, the act of mimicry can also be found in the gameplay of role-playing video games. Caillois (2001: 21) claims that mimicry appears in both animals and humans across all ages and illustrates that claim with examples of children mimicking adults, baby animals following the behaviour of

their parents, and adults acting in theatrical plays. For RPGs, similar behaviour occurs as the player acts a specific role of the playable character. Furthermore, Caillois (2001: 21-22) states that mimicry games do not have strict and fixed rules, and their pleasure is derived from the successful act of impersonation. Within those games, the player's goal is to convincingly act as the target character (Caillois 2001: 23).

Although the mimicry category can be used as the basis for defining RPGs, the gameplay in a contemporary video game goes beyond imitation. Burn (2023: 319) and Carr (2006: 21) generally describe RPG games as slower-paced and complex, and more oriented towards strategy rather than fast-paced action. Rennick et al. (2023a) and Domsch (2017: 266) distinguish RPGs from other genres based on their reliance on dialogue; a similar emphasis is noted in Mäyrä (2017: 271), who suggests that the RPG genre is distinct among video games due to gameplay, which is "typically highly language-oriented" and in Dixon (2023: 2), who emphasises that large amounts of both speech and texts are included in RPGs. Therefore, among video game genres, the RPG genre emerges as a valuable source for studying language use.

In his definition, Burn (2023: 319) claims that the mechanics featured in the RPG genre, which are mostly borrowed from TTRPGs, distinguish it from other genres. Bycer (2024: 5) and Burn (2023: 319) agree that these prevalent mechanics include detailed character customisation options, class systems, skill progression, and possible companion characters. Similarly, Reinhardt (2019: 92) defines RPGs as games featuring a playable character that progresses and develops throughout the game, large-scale, detailed settings, and progression through quests. Furthermore, both Bycer (2024: 198) and Burn (2023: 322) emphasise the importance of the player's interaction with the narrative of the game: the player's choices affect the gameplay.

Bycer (2024: 3), who describes RPGs as one of the oldest established genres for video games, recognises the constant change present within the genre, which complicates the formation of any specific definition. Despite these developments, Bycer (2024: 197) sees the current definitions of the RPG genre as problematic and calls for more specificity. Although Zagal & Deterding (2018) suggest that role-playing games have no universally established and recognised definition, their connection to tabletop role-playing games (TTRPGs) is widely acknowledged (Perron 2023: 151; Burn 2023: 319; Bycer 2024: 3). Perron (2023: 151) and Bycer (2024: 5) emphasise that the TTRPG *Dungeons and Dragons* has heavily influenced the development of the role-playing video game genre. Burn (2023: 319) and Apperley (2006: 17) also see fantasy fiction as influential for both TTRPGs and RPGs.

RPGs feature multiple characteristic aspects aside from their dependency on dialogue. In his definition, Burn (2023: 319) claims that the mechanics featured in the RPG genre, which are borrowed mainly from TTRPGs, distinguish it from other genres. Bycer (2024: 5) and Burn (2023: 319) agree that these prevalent mechanics include detailed character customisation options, class systems, skill progression, and possible companion characters. Similarly, Reinhardt (2019: 92) defines RPGs as games featuring a playable character that progresses and develops throughout the game, large-scale, detailed settings, and progression through quests. Furthermore, both Bycer (2024: 198) and Burn (2023: 322) emphasise the importance of the player's interaction with the narrative of the game: the player's choices affect the gameplay.

RPGs can be further classified into various sub-genres, which have more specific characteristics. Bycer (2024: 64–74) claims that subgenres emerge from the combination of RPG genre-specific mechanics and elements from other genres, mentioning the following subgenres: Action RPGs (ARPG), which additionally include characteristic elements of action games; Strategy RPGs (SRPG), which often focus on small-scale conflicts; and Massively

Multiplayer Online RPGs (MMORPG), which centre around a single controllable character in a multiplayer environment.

Another prevalent subdivision within the RPG genre is the dichotomy of Japanese RPGs (JRPG) and Western RPGs (WRPG). Burn (2023: 319) separates these subgenres based on the inclusion of specific game mechanics and themes. Additionally, Bycer (2024: 39–40) and Domsch (2017: 265) see the player's involvement in the narrative and the customisation of the protagonist as the main differences between the two: WRPGs tend to include highly customisable main characters and plotlines opposed to JRPGs, which feature already established characters and narratives. Therefore, it could be possible that this difference in the involvement in the narrative appears as a difference in some linguistic aspect in the dialogues of WRPGs and JRPGs. Although Bycer (2024: 196–198) argues that the RPG genre has evolved past the dichotomy of WPRGs and JRPGs due to the rising popularity of mixed-genre ARPG games and the evolution of game design, he still concludes that the distinction between these subgenres is necessary until alternative approaches to distinguishing different subgenres of RPGs are formed.

1.2. Video Games and Language Research

Video games provide a diverse range of data for research, as they often combine text, visuals, audio and player interactions with computers or other players. Furthermore, as video games include both written and spoken texts, which often co-occur, they can be analysed from a linguistic perspective. As a starting point, the following sections address the relevance of dialogue within video games, focusing more specifically on its significance within RPGs. Two different approaches to video game language can be found in current research, and the following sections will examine the specificities associated with the ludolinguistic and Second Language Acquisition (SLA) approaches to video game language.

1.2.1. Video Game Dialogue

Within video games, dialogue can appear in the format of written text on screen or as spoken dialogue in the form of audio within the game, or as a combination of the prior two. According to Domsch (2017: 251), dialogue, although not considered an obligatory aspect of a video game, has been present in video games since their creation in various capacities, but can nowadays be considered an important aspect of aesthetics in video games. This significance is also evident in the fact that video game scripts are written by teams of professionals, and famous actors, such as Keanu Reeves, Elliot Page, Angela Bassett, Matthew Perry, and Samuel L. Jackson, voice characters and act in video games. In addition to its aesthetic value, dialogue also functions as an important aspect of gameplay as it enables communication between the player and the game.

Domsch (2017: 254–256) divides in-game communication into two categories: ludic communication, which breaks the immersion of the game world to address the player and functions as a reminder of the rules, and diegetic communication, which occurs as dialogue spoken by in-game characters and directed at the player's character in the game, not the player, with the goal of immersing the player in the game world. Still, Domsch (2017: 264) confirms that combinations of ludic and diegetic dialogue can be used to convey instruction without breaking immersion: in those instances, the ludic content of the message is conveyed in a diegetic form. Domsch (2017: 263–264) describes that within video games, the diegetic communication, which can often be optional for the player and occur in a non-linear manner, can also be used to shape the narrative of the game.

As established, reliance on dialogue, or diegetic communication, is one of the characteristic aspects of RPGs. According to Mäyrä (2017: 283), the dialogue in RPGs is overall more refined compared to dialogue in TTRPGs, as it is not spontaneous and edited. Furthermore, Mäyra (2017: 288) describes the language in RPGs as including elements of

literary fiction and interpersonal communication. Mäyra highlights the complex nature of RPG dialogue as both a spoken and a written type of text, a written script emulating spoken dialogue.

Dixon's (2023: 5) analysis divided spoken dialogue in RPGs into two categories based on the necessity of the player's input. Dixon (2023: 111) identified an interactive register, which requires the player's response and an immersive register, which requires no input. Based on a situational analysis, Dixon (2023: 112) found that dialogue has various purposes within RPGs: narrative development, communication with non-playable characters, agency affirmation for the player, and providing direction for the player. Despite this variety of functions performed by the dialogue, Mäyra (2017: 282–283) mentions limited flexibility as a limitation of RPG dialogue, as players can only choose between a set of scripted options. Despite the insistence on interactivity that is present in the definitions of video games in general, Domsch (2017: 253) agrees that the communication within video games is always simulated, as the player's contribution is restricted by the options available to them within the game.

An example of how dialogue is presented to the player in *Disco Elysium* (ZA/UM 2019), an RPG, is provided in Figure 2. In the figure, the dialogue between the narrator and other non-playable characters is presented simultaneously as spoken dialogue (audio) and as white text in the black vertical panel on the right side. Below the written text, the numbered options in red illustrate the options available to the player to reply to the non-playable character, while the currently selected option is highlighted in white. The figure successfully conveys both freedom and limitations of choice apparent in RPG dialogue; although the player is provided with options which allow the customisation of the storyline and gameplay, the player must still choose between options which are pre-defined by the developers and cannot create their own unique options.

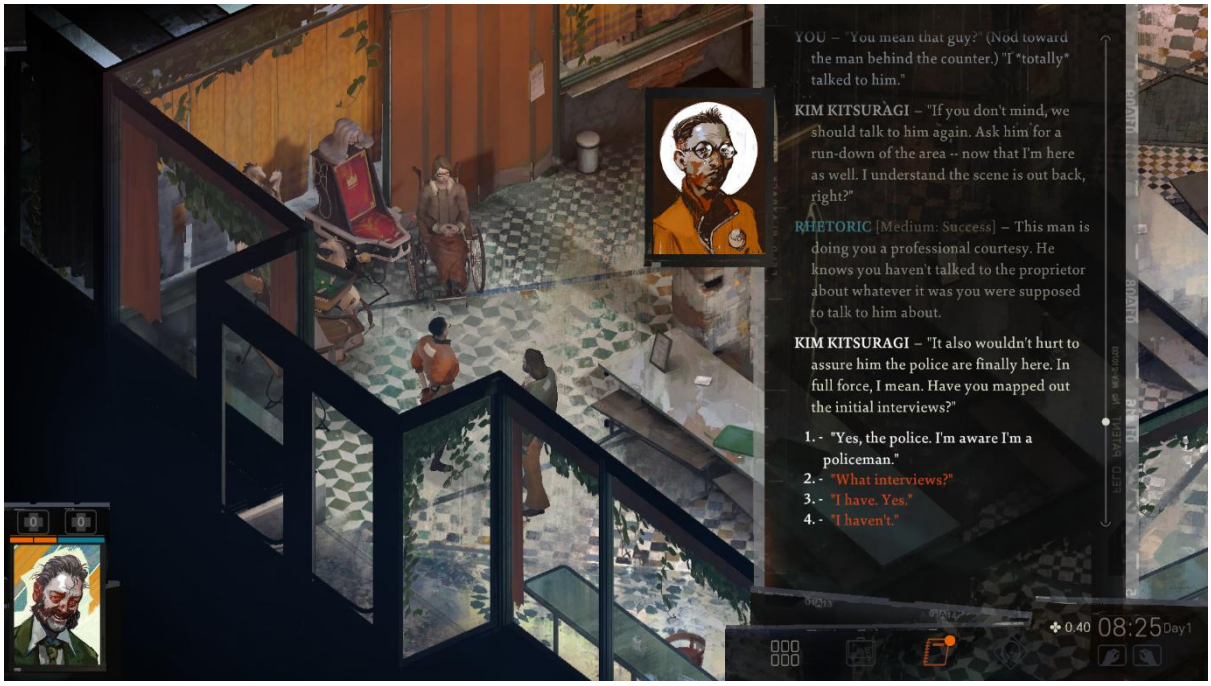


Figure 2. An example of dialogue in *Disco Elysium* (ZA/UM 2019), a role-playing video game.

Although dialogue occurs as a scripted text, contrary to communication between players within multiplayer games, which is spoken and spontaneous, it is a significant aspect of RPGs and video games in general, as scripted dialogue can occur across all game genres, whilst player-to-player communication is restricted to multiplayer games. Therefore, dialogue emerges as one of the most significant ways to experience language, both in written and spoken forms, in video games. Furthermore, the scripted and edited nature of dialogue also indicates its grammatical correctness (unless intended otherwise). Thus, players experience a significant amount of language in the form of dialogue during the playing process for most video games, necessitating the research of video game dialogue as linguistic data.

1.2.2. Ludolinguistics

According to Heritage (2025: 209), ludolinguistics, the study of language included in (video) games, is considered a sub-discipline of video game studies. The term ludolinguistics is connected to ludology, which is defined by Frasca (2003: 222) as the study of games and video games. The emergence of the field of video game studies and the further development of

the ludolinguistics field were shaped by a debate surrounding the most suitable approaches to researching video games. To understand the status of video game studies and ludolinguistics, a description of both perspectives and the outcome of the debate will be provided.

Pugh and Ramey (2022: 3) describe game studies as a relatively young discipline, with the first key texts published in the late 1990s by Chris Crawford, Janet Murray and Espen Aarseth. According to Arsenault (2023: 589), the debate first emerged in the foundational texts of the field, which had opposing viewpoints on the suitability of specific frameworks for analysing video games. Within the debate, there was a disagreement between the narratological side, which prioritised storytelling and narratives in video games, and the ludological side, which emphasised ludics, or gameplay, over other aspects (Pugh and Ramey 2022: 3).

Before delving into the specifics of the debate, it is necessary to describe both sides of the conflict. Taking a more ludological stance in the early 2000s, Frasca (2003: 221–222) claimed that existing research focused on similarities between games and narrative, and called for further approaches from other perspectives, as video games are distinct from other media due to their mechanics. According to Frasca (2003: 222), ludology as a discipline is a formalist approach examining the structure, elements and mechanics in games. Examining the debate in retrospect, Aarseth (2023: 255, 257) provides multiple definitions to the term ludology, stating that it can describe either the general discipline of game studies, a specific approach within game studies which questions or rejects the connection between games and narratives, or a movement during the late 1990s and early 2000s. Still, Aarseth (2023: 259) concludes that ludology cannot be considered a separate discipline, but rather a criticism of the incompetence in video game studies at the time of the debate or a reaction to the hybrid formats prevalent in the 1990s.

The other side of the debate promoted the narratological approach to video games. Arsenault (2023: 588) establishes the connection between narratives and games using Caillois'

game category of *agôn* (competitive games), claiming that conflict is an important aspect in both. Although structuralist narratological approaches focused on structural elements which helped in establishing narratives in games, and even functioned below the level of narrative, the post-structuralist narratological approaches in video game studies examine the ways in which narratives appear and function in video games (Arsenault 2023: 590–591). However, Arsenault (2023: 501–592) still emphasises that narrative is an optional element in video games and its significance differs in different games.

Despite the distinct differences in the viewpoints of the researchers, the results of the debate remain inconclusive. Pugh and Ramey (2022: 3) claim that the debate had no definitive outcome and that nowadays scholars combine approaches from both sides. In contrast, Pugh and Ramey (2022: 4) mention that the ludonarratological approach, which combines gaming and narrative theories, could be considered as one result of the debate. Taking a more critical stance in their descriptions, both Arsenault (2023: 589) and Aarseth (2023: 258) claim that the debate was distorted and amplified and has been given exaggerated importance in subsequent academic texts. Analysing the situation retrospectively, Aarseth (2023: 255) claims that the criticism was specifically towards inefficient use of narratological theory in the research of video games. In a similar vein, Arsenault (2023: 589) argues that the focus of the debate was on the suitability of using existing frameworks for the analysis of video games and the possibility of creating new approaches more suitable for the specific medium of video games. Furthermore, Aarseth (2023: 256) implies that the opposition was not facilitated by the contrasting approaches, but rather by the communication between researchers with different views on the matter. Ultimately, Aarseth (2023: 255) claims that the narratological approaches do not conflict with ludological ones, as the ludologists were ultimately still using narratology in their approaches to game research. Although it did not yield a concrete conclusion, the debate

clearly influenced the development of video game studies. In addition, both approaches are acceptable research approaches in video game studies today.

Before delving deeper into ludolinguistic studies, it is necessary to provide an explanation of the terms used to distinguish different forms of gameplay. Based on their analysis of the usage of the term "metagame", Carter et al. (2012) observed multiple contrasting meanings for the term. Although Carter et al. (2012: 1) define metagame as any activity which is relevant to the gaming experience whilst being external to the game itself, their analysis found that different game genres and gaming communities may use the term with different meanings. For example, one possible meaning denotes the use of strategy in competitive games that goes beyond the rules of the game, whilst the other, applied more specifically to role-playing games, describes the breaking of immersion during the process of role-play (Carter et al. 2012: 12–14). Furthermore, another meaning associated with the term metagame describes additional in-game activities which are not required for the completion of the game, but enhance the playing experience, such as achievements and campaign playing modes (Carter et al. 2012: 14). To alleviate the potential confusion arising from the use of the term metagame, Carter et al. (2012: 14–15) introduce the terms orthogame (describing the core gameplay), and paragame (activities distinct from but related to the orthogame) whilst also redefining the term metagame as gameplay using resources beyond the orthogame.

Despite its connections to ludology, ludolinguistic research has no specific connection with the ludological approach in video game studies. Research on orthogame data seems lacking; several collections of articles analysing the elements of video games separately, such as Lowood and Guins (2016), Newman (2004) and Wolf and Perron (2023), fail to address dialogue as a distinct element in video games. This research gap is confirmed by Heritage (2022: 2) and Bianchi (2024: 245), who agree that current ludolinguistic research has mainly utilised paragame data. Despite such focus, Heritage (2022: 3; 2025: 209) emphasises that

research on video game paratexts cannot be equated with research on data originating directly from video games, as genre conventions between different forms of media may affect their contents. Therefore, further research using orthogame data is crucial for establishing a better understanding of the textual contents of video games.

1.2.3. SLA Approaches to Video Game Research

According to Reinhardt (2019: 25), games are usually translated into languages that have the largest groups of speakers globally or that provide large markets for video games, such as English, Spanish, Chinese, and Japanese, among others. Still, for many players, like those in Estonia, video games remain an experience contingent on their understanding of another language. Reinhardt (2019: 1) speculates that games are effective in teaching because the learning occurs as a secondary process in an entertaining and motivating environment.

As such, video games have been researched from the perspective of second language acquisition (SLA). Peterson and Jabbari (2023: 1–2) specify Computer-Assisted Language Learning (CALL) and Digital Game-Based Language Learning (DGBLL) as the research fields which focus on the use of video games in SLA. According to Reinhardt (2019: 6), the field of CALL emerged alongside the first video games; however, as the availability of games was limited, the application of games in SLA was initially believed unlikely, but interest in CALL increased as games became more available to the public. In the 2000s, a significant interest within CALL was the use of Massively Multiplayer Online Games for language acquisition (Reinhardt 2019: 59). Reinhardt (2019: 59) also specifies that DGBLL emerged as a later development of CALL. Peterson and Jabbari (2023: 6) describe DGBLL research as increasingly qualitative, despite the popularity of quantitative approaches during the initial years.

Within DGBLL, Berns and Ruiz-Rube (2023: 35) categorise video games into serious games, which are created specifically for educational purposes, and commercial games, which

are created for entertainment purposes. However, Reinhardt (2019: 59) notes a decline in the use of serious/educational games due to their highly specific audience, claiming that educational video games have been replaced by educational online platforms. A similar classification system for game objectives is provided by Berns and Ruiz-Rube (2023: 35), who categorise two types of goals within games in DGBLL: affective goals, which are tied to engagement and immersion, and skill-based goals, which describe developing language skills. Reinhardt (2019: 30–31) also specifies that video games are mainly considered in relation to the concepts of incidental learning in an informal environment, but it is possible to use them intentionally in a classroom setting with the guidance of a teacher.

Although DGBLL could be described as a developing field of study, research has already shown that using video games in language acquisition can have positive effects. According to Peterson and Jabbari (2023: 1) and Dixon (2023: 1), previous research has demonstrated that video games provide the player with an interactive and immersive environment for experiencing input in the target language. Dixon (2023: 54) agrees, adding that the possibility of practising target language output is beneficial to the players. Dixon et al. (2022: 1) note that previous meta-analyses in DGBLL, while finding positive correlations between SLA and video games, are problematic in their generalisations, as results are often conveyed as significant to specific game genres. Based on their meta-analysis of 26 DGBLL studies, Dixon et al. (2022: 17) conclude that the use video games can have a positive impact in SLA and specify that more benefit can be gained from using commercial games, multiplayer (but not Massively Multiplayer Online) games, and from the provision of additional study material to support the learning process.

The field of DGBLL has been productive regarding the study of RPGs. Reinhardt (2019: 58) approaches RPGs from Caillois' category of *mimicry* (simulation games) and sees it as an important aspect of language reproduction and transmission. Reinhardt (2019: 59) also

distinguishes *mimicry* from the other game categories due to its lack of a specific end goal and ties *mimicry* to SLA through the shared aspects of imitation, repetition and recognises the effectiveness of role-play as an SLA technique. Reinhardt (2021: 69), Peterson and Jabbari (2023: 2) agree that RPGs, specifically MMORPGs, have been the most significant genres featured in current DGBLL research. Peterson and Jabbari (2023: 5) frame their definition of the RPG genre through the lens of language acquisition possibilities: they describe RPGs as games that are designed to promote social interaction that often exceeds the boundaries of the games and results in the creation of online gaming communities. Peterson and Jabbari (2023: 9) claim that game-centred communities are also relevant for SLA, as they provide input to the learner whilst also providing opportunities for players to interact in an environment that is engaging their interests.

Despite the increasing interest in DGBLL, Peterson and Jabbari (2023: 6) pinpoint several limitations of the field, including excessive interest in vocabulary and factors affecting learners in DGBLL research. Furthermore, Peterson and Jabbari (2023: 6) address the issue of genres, emphasising that although current research focusing on RPGs and simulation games has shown positive results, it has left other game genres under-researched.

The specificity of video game data and methods used for researching it within DGBLL have been addressed by Reinhardt (2019: 94; 2021: 70, 84) and Dixon (2022a: 150; 2023: 52) who advocate for studying specific game mechanics separately, in contrast to previous studies which have framed their results around specific genres. Reinhardt (2019: 94) defines a game mechanic as a pre-programmed action within the game, that is made possible for the player and is used to progress within the game, e.g., racing, dialogue or building. Reinhardt (2019: 95) recognises that one game can utilise multiple different mechanics, but genres generally share similar mechanics. Game features, such as levels or quests, are also distinct from game mechanics, according to Reinhardt (2019: 95), as these shape the structure of the game and are

less related to the actions of the player. In Dixon's (2022: 150) perspective, game mechanics support the autonomy of the player within the game. Both Dixon (2023: 52) and Reinhardt (2019: 94) claim that researching game mechanics separately is more productive as it allows establishing more specific connections between game aspects and their possible effects on the learner.

Results from previous DGBLL research help to highlight several characteristics of language within video games. Conducting a lexical coverage analysis on a corpus including all language included in ten video games, Rodgers and Heidt (2020: 222–224) concluded that the lexical coverages of video game texts and more traditional forms of media, such as television and film, are similar; furthermore, they claim that the genre of the video game does not influence the linguistic content of the game. Bianchi (2024: 260), who focused her analysis on video game dialogue, also confirms the similarity between video games and spoken media on a lexical level and emphasises the internal variation of genres.

Dixon establishes that commercial games have been more conducive for SLA (Dixon 2022: 153–154) and that within video games, distinct differences can be observed between spoken language and written language (Dixon 2023: 53; Dixon 2022: 157). Overall, Dixon (2023: 53) emphasises that linguistic variety within video games should be recognised, as different types of text serve different purposes. To address this issue in DGBLL research, Dixon (2023: 53) establishes a method for analysing language in video games that incorporates the register analysis approach. Further discussion on Dixon's register analysis on video game dialogue and its results can be found in Section 2.1.1.

Previous theses at the English department at the University of Tartu have researched video games from the aspect of SLA. Mägi Ravn (2021) focused on the concept of gamification by creating a specific digital activity (resembling a video game) for teaching vocabulary. Rodendau (2017), who incorporated a video game into the classroom context, measured the

students' knowledge of game-specific vocabulary prior to and following multiple gaming sessions. Rodendau (2017: 52) concluded that students' knowledge of game-specific vocabulary did progress but notes that the selection of the game is important, as students' interest in the game can affect the gaming and learning processes and that the results of the study cannot be generalised to describe the general effects of using video games in SLA. Mehine (2021) researched the learners' perception of acquiring English and study skills via video games. Additionally, Mehine (2021: 30), Rodendau (2017: 44) and Mägi Ravn (2021: 36–38) all incorporated student feedback in their research, inquiring about the leaners' perception of using (video) games for SLA, concluding that students considered video games an enjoyable and motivating environment for vocabulary acquisition and practising already acquired skills. Although the topics researched at the University of Tartu are all related to video games, they do not include an analysis of the linguistic contents of commercial video games or the representations of specific video game mechanics but rather focus on the acquisition of specific selected vocabulary that is present in the game. Therefore, further studies are needed to provide an overview of the language included in the video games.

2. ADDITIVE MULTIDIMENSIONAL ANALYSIS OF ROLEPLAYING VIDEO GAME DIALOGUE IN VGDC

The goal of the thesis is to investigate the characteristic linguistic features of video game dialogue, more specifically RPG dialogue, and its relation to written and spoken modes of communication. To this end, an additive multidimensional analysis (MDA) was conducted on data from the Video Game Dialogue Corpus (Rennick et al. 2023a; Rennick and Roberts 2024). The MDA method allows the comparison of several linguistic features simultaneously across various established registers of texts. Although the additive MDA method has previously been used on video game dialogue data to some extent, a full additive analysis of a larger collection of data is absent from current video game language research. Therefore, the following analysis will offer new information on the characteristics of video game dialogue in the genre of role-playing games. The chapter starts with a description of the register analysis and multidimensional analysis methods, as they are closely related. Following that, the methodology of MDA is described in detail, distinguishing the strands of full and additive MDA approaches, the central concept of a dimension of variation. A thorough description of the method is then provided. Additionally, the Multidimensional Analysis Tagger (MAT) software is introduced as a suitable option for conducting MDA.

Prior to the description of the data used for the current analysis, the available video game corpora are described. These descriptions are followed by the examination of the video game corpus creation process and the proposed requirements for video game corpora. Subsequent sections describe the method used in the analysis, the MDA process, data collection and preparation and analysis itself. In the final section of the chapter, the results of the additive MDA are reported.

2.1. Register Analysis and Multidimensional Analysis

The following section will provide descriptions of two approaches to analysing variety in texts, register analysis and multidimensional analysis. Although MDA is the method used in this thesis, its close ties to the register analysis method require the description and examination of both methods. Following this comparison, the MDA method and methodology are described extensively, focusing on the central concept of dimension and linguistic co-occurrence. Two strands of MDA and their uses in linguistic research are then examined and contrasted. The suitability of the additive approach is confirmed based on the characteristics of the analysis process. In the final section, the MAT software is examined as one suitable and accessible option for conducting MDA.

2.1.1. Register Analysis and Its Application

Biber and Conrad (2019: 2) list genre, style, and register analyses as the three possible approaches for analysing variation in texts, but emphasise register as a relevant descriptor for all texts, over other approaches, which are reported to have a more specific purpose. Biber and Conrad (2019: 6) list context, linguistic components and the connections between the prior two as the necessary aspects of register analysis, whilst asserting that within register analysis, linguistic features prevalent to a particular register are considered functional. In other words, if specific linguistic characteristics frequently occur in a register, it is because they carry a relevant function within that register (Biber and Conrad 2019: 16). Biber and Egbert (2016: 97) list three stages for conducting register analysis: identification of registers, followed by the creation of a representative corpus of the observed registers and analysis of the corpus. A more in-depth, yet similar description appears in Biber and Conrad (2019: 7), starting from establishing a register via its distinct situational characteristics, which followed by creating a description of the characteristic linguistic features of the register, and the final step is analysis,

which focuses on describing and explaining the connection between the contextual aspects and the linguistic characteristics. Biber and Conrad (2019: 259) describe register analysis as a comparative approach, as contrasting different registers highlights the specificity of the register under review.

The register analysis method has been applied to the analysis of a wide variety of texts, including video game texts. Biber and Egbert (2016: 96) establish different Internet registers, including blogs, social media posts, and email messages and use the MDA approach to analyse the differences between different registers. Although Biber and Egbert (2016) include various registers and sub-registers in their research, the language included in video games is rightfully excluded from their data due to differences in the situational and communicative characteristics of the texts included in that register. Focusing on text within videogames, Dixon (2022a; 2022b, 2023, 2024b) has used the register analysis approach to establish video-game specific registers in the genre of single-player role-playing games. Following recommendations in DGBLL, Dixon (2022b: 13) considers game mechanics and registers as equivalent and suggests using register analysis for the identification of game mechanics. Dixon (2022b: 11) describes game mechanics as the options available to the player to use to progress within the game. Therefore, dialogue can be considered a game mechanic. Furthermore, Dixon (2023: 52) emphasises the similarity between game mechanics and registers, explaining that both can be distinguished on the basis of their specific purpose, which is realised via specific language use.

Dixon (2022b: 31) uses situational analysis, which is the analysis of non-linguistic context, to distinguish registers within video game texts. According to Dixon (2022a: 166), registers and mechanics are identical categories, as each is defined in terms of their specific functions. Dixon (2022a: 162) uses discriminant analysis to demonstrate that similarities do appear in the language used in a specific game mechanic despite the uniqueness of each game. Dixon (2022a: 168) concludes that specific game mechanics can be distinguished based on

their language use, and that the language found in the mechanics remains similar across different games. Furthermore, Dixon (2022a: 167) hypothesises that immersive game dialogue could be considered meaningful interaction, and if further research establishes game dialogue as similar to spoken dialogue, then knowledge gained from games could be used in real-life situations. Nevertheless, further research should analyse specific video game mechanics separately. Additional research should also be conducted on video game dialogue to establish the connection between real-life spoken dialogue and game dialogue.

Although Biber and Conrad (2019: 39) specify seven categories of situational characteristics for situational analysis, Dixon (2022b: 32) conducts his situational analysis considering three categories that were suitable for video game texts: participants, processing circumstances and communicative purposes. Based on these characteristics, Dixon (2022a: 152) initially establishes three registers within video games: the spoken register of dialogue trees, and two written registers relating to quests: quest objectives and quest stages. He later expands the register framework of video games, dividing the spoken dialogue register into interactive speech and immersive speech, and revising the written register category to include the following registers: character text, quest text, tutorial text and lore (Dixon 2022b: 31). In Dixon's later research, the written register of quest text is again split into quest objective and quest stages registers (Dixon 2024: 110).

Regarding the spoken registers, Dixon (2022b: 33–35) justifies the establishment of two distinct dialogue registers due to differences regarding the input of the player and passage of time during the dialogue, but notes that multiple differences also emerged in the communicative differences between the two dialogue registers. Dixon (2023: 58) establishes that immersive speech is used to develop narrative, whilst interactive speech allows the progression in the game, and is used to present the player with choices regarding quests, therefore echoing Domsch's (2017: 264) diegetic and ludic dialogue classification to some

extent. Dixon (2022b: 37) also specifies that while immersive speech does not require the player's participation, interactive speech cannot occur without the involvement of the player character. Dixon (2022b) follows the instructions set for register analysis by Biber and Douglas (2019): following the establishment of registers, he compiles a corpus including the registers and conducts multidimensional analysis to provide a detailed description of the registers under review.

In another article, Dixon (2022a: 159–170) uses the established registers to conduct discriminant analyses, focusing on complexity and measuring specific linguistic characteristics which are found to be common to spoken dialogue. Dixon (2022a: 167–168) concluded that dialogue in real life and in video games share specific characteristics, and similar conclusions appear in Dixon (2023:69), where Dixon specifies that immersive speech is more similar to spoken language than interactive speech. Additionally, Dixon (2022a; 2023) found that significant differences between video game spoken language and real-life spoken language appeared in the use of filled pauses, response forms and discourse markers. Furthermore, Dixon (2023: 71) hypothesises that the immersive speech register, if found to be significantly similar to real-life spoken dialogue in further research, could be beneficial for SLA.

In several instances, Dixon (2023: 72; 2022a 167–168) recommends further research to develop the method further: this includes recommendations on using a different, more diverse and expanded sample, identifying additional registers for video game texts, and investigating the similarities between spoken dialogue and game dialogue. These recommendations are significant because Dixon's research uses the Single Player Offline Game Corpus (SPOC), which, despite including 3.7 million words of text, represents only four commercial RPG games (Dixon 2024a: 108). Furthermore, two games are developed by the same game studio (Dixon 2024a: 108).

The current study can incorporate and develop Dixon's (2022a, 2022b, 2024b) results further by using data depicting already established registers, as Dixon (2024b: 28) recommends using the established video game registers in future research. Following this advice, this thesis will incorporate the definitions for the spoken RPG dialogue registers established by Dixon (2022a: 33; 2022b: 153; 2024b: 110): RPG dialogue contains two distinct registers, interactive speech and immersive speech, which differ in terms of the player's participation and the passage of time during dialogue. Both immersive speech and interactive speech registers are included in the sample that will be analysed to provide further descriptions of the spoken dialogue used in these video game registers. Following Dixon's recommendations, the sample analysed will contain different games and is significantly bigger than SPOC.

2.1.2. Multidimensional Analysis

Before delving into the process necessary for conducting multidimensional analysis (MDA), an overview of the method and its central concepts is required. Various definitions for the method have been offered in previous research, consistently emphasising similar core elements. Quaglio (2009: 58) and Berber Sardinha and Veirano Pinto (2019: 1) describe MDA as a quantitative corpus linguistic approach. Berber Sardinha and Veirano Pinto (2019: 1) specify that MDA was developed in the 1980s by Douglas Biber, with the goal of identifying patterns of variation in various registers. Similar descriptions are provided by Biber and Conrad (2019: 268), who specify that MDA allows the comparison of multiple registers by grouping multiple co-occurring variables into dimensions of variation.

The concept of co-occurrence appears significant to the method, as it emerges across several definitions of MDA. In their description of the MDA method, both Quaglio (2009: 58) and Biber and Conrad (2019: 268) emphasise the importance of linguistic co-occurrence, which has been established in prior linguistic research, in MDA. To summarise, linguistic co-occurrence refers to the concept that multiple linguistic variables occur simultaneously in a

specific context (Biber and Conrad 2019: 260). According to Quaglio (2009: 58), Biber developed the concept further, hypothesising that different text registers feature different patterns of co-occurring linguistic features due to their different functions in communication, and, therefore, linguistic co-occurrence of specific linguistic features indicates their shared function. Biber and Conrad (2019: 268–69) explain that MDA applies factor analysis to identify groups of co-occurring features from texts, which can then be used for describing the functions of specific text registers.

The methods of MDA and register analysis share multiple relevant aspects. Biber and Conrad (2019: 268) describe MDA as an alternative to register analysis, stating that its advantage is the possibility of comparing multiple registers by using groupings of linguistic characteristics rather than individual characteristics as the basis of comparison. The simultaneous comparison of multiple linguistic variables is considered significant by Biber and Conrad (2019: 260), who emphasise the increased accuracy and level of detail in register descriptions emerging from MDA compared to register analysis. Furthermore, MDA is considered a more efficient option as it allows the effective comparison of multiple registers simultaneously (Biber and Conrad 2019: 260). Biber and Conrad (2019: 274) observe the similarity between MDA and register analysis due to the shared assumption of functionality of linguistic features: register analysis considers specific linguistic features to be functional in a register, and this idea is amplified in MDA, where linguistic co-occurrence patterns are considered functional, meaning that a group of linguistic variables that occur together frequently are assumed to share a communicative function.

2.1.3. The Method of MDA

As described in the prior section, MDA focuses on the identification of patterns of co-occurring linguistic features. Within the MDA method, these patterns are known as dimensions

of variation, or dimensions. Biber and Conrad (2019: 269–270) and Berber Sardinha and Veirano Pinto (2019: 1–2) agree that a dimension represents a set of frequently co-occurring linguistic variables that serve shared functions in communication. Biber and Conrad (2019: 273) also emphasise that dimensions are not pre-determined but emerge via factor analysis, which highlights the value of MDA as a descriptive method. Each dimension includes specific linguistic features, which have negative and positive values based on their effect within that dimension (Biber and Conrad 2019: 273). Quaglio (2009: 59–60) and Biber and Conrad (2019: 273) note that linguistic features occur in a complementary distribution. A similar description is provided by Dixon (2022b: 85), who describes features as mutually exclusive, meaning that positive and negative features cannot occur together, and a high frequency of positive features also notes a lack of negative features on a dimension, and vice versa. Therefore, the dimension emerges as a continuum between the positive and negative features. Dimensional scores are used to situate a register on a dimension and describe the contents of a register within a specific dimension. Biber and Conrad (2019: 273) and Dixon (2022b: 85) explain that dimensional scores are calculated for each dimension by combining the values of the positively loaded and negatively loaded features identified in data that are characteristic of a specific dimension.

Prior to the calculation of dimensional scores, several steps are required to conduct MDA. Biber and Conrad (2019: 269) and Berber Sardinha et al (2019: 3) emphasise the importance of factor analysis in the method of MDA, as it is required for the identification of dimensions. Both Quaglio (2009: 58–59) and Biber and Conrad (2019: 269–270) provide specific, but similar, step-by-step instructions for conducting MDA, which are summarised below:

1) Selection of Linguistic features: Based on prior research, features regarded as functional for the register are selected for analysis.

2) Corpus Compilation: Considering prior research, a representative corpus is collected; registers included in the corpus are identified and described.
3) Identification of selected linguistic features in the corpus using automated tools.
4) Factor Analysis is used for the identification of co-occurring linguistic variable patterns and the reduction of multiple variables into dimensions of variation.
5) Quantitative analysis: Calculation of dimension scores based on the occurrence of features associated with a specific dimension, followed by the qualitative analysis of dimension scores.
6) Qualitative analysis is used to interpret the quantitative results with the aim of providing descriptions of communicative functions for each dimension.

Several important specifications should be considered. First, Biber and Conrad (2019: 260) emphasise that variety also occurs within a specific register; therefore, analysis should be conducted on a sample including a variety of representative texts from the register under review. Second, Biber and Conrad (2019: 269) emphasise that although factors/dimensions are identified using quantitative analysis, qualitative analysis is still required to connect the linguistic differences to functional differences between different registers. Following the discussion of the MDA method and the relevant recommendations, the current analysis uses the MDA approach as it is suitable for the analysis of an extensive corpus sample and can assist in creating a more detailed description of a specific register of texts.

2.1.4. Types of MDA

Although the MDA can be considered a method of research, specific approaches to MDA have emerged in linguistic research. According to Dixon (2022b: 75–76), Quaglio (2009), Berber Sardinha and Veirano Pinto (2019), and Berber Sardinha et al. (2019), the MDA method can be divided further into two specific approaches. One type, known as additive MDA, incorporates the dimension framework established by Biber, where the established

dimensions and dimensional scores for various text registers are compared and contrasted with new registers, establishing them by comparison and contrast to existing text registers. In the second type, which is known as novel MDA, new MDA, complete MDA or full MDA, new dimensions are established which are specific to the register under review. To determine their suitability for the current analysis, both approaches are described.

According to Dixon (2022b: 76), in the full MDA approach, researchers must choose the specific linguistic features to be used in the analysis prior to the identification of dimensions. Following that, factor analysis is used to reveal the patterns of co-occurring variables, which are also known as dimensions. Nini (2019: 70) claims that although the full MDA approach can be beneficial for describing the internal structures of specific registers, it does not include the option to demonstrate contrasts and/or resemblances to other registers, which is possible with the additive MDA approach. Still, in Nini's (2019: 70) evaluation, MDA research has tended to favour the full MDA method over Biber's framework. The full MDA process is described in detail by Biber and Egbert (2016: 96), who analyse a corpus of web registers and conclude that the MDA approach has been used for describing various Internet-specific registers, yet the overall variation in language use on the Internet is still relatively unknown.

In contrast to the full MDA approach, where register-specific dimensions are identified, Berber Sardinha et al. (2019: 165) describe additive MDA as the incorporation of new registers into an existing MDA framework, therefore increasing its range. Similarly, Nini (2019: 70) agrees that MDA using Biber's framework allows the evaluation of new text registers in comparison to established text registers. Berber Sardinha et al. (2019: 165) note that the dimensions described in Biber (1988) have become a common reference point in additive MDA studies, due to their stability and use in multiple tagging software. Based on his research results, Biber (1988) established a framework for describing register variation in most registers in

English. Biber's framework uses 67 linguistic features to establish six dimensions, which are summarised below. Names of the dimensions indicate the communicative functions associated with both sides of the continuum, or the presence/absence of a specific function in text.

Dimension 1: Involved vs. Informational Discourse (D1)

Nini (2019: 68) summarises D1 as the contrast between interactional and affective texts (such as spoken conversations) and informationally dense texts (e.g. academic texts). Dixon (2022b: 163) and Quaglio (2009: 58) claim that the first dimension in Biber's framework is the most significant. According to Biber (1988: 102), the positively loaded grammatical features on D1 include a high frequency of private verbs (e.g., *believe, decide, doubt, feel, fear, hope, imagine* (Biber 1988: 242)), contractions, present tense verbs, emphatics (e.g., *a lot, such, really, most, more* (Biber 1988: 241)), discourse particles (e.g., *anyway, anyhow, anyways, well* (Biber 1988: 241)), hedges (e.g., *something like, almost, maybe, kind of, sort of* (Biber 1988: 240)), amplifiers (e.g., *absolutely, completely, extremely, perfectly, totally, very* (Biber 1988: 240)) and stranded prepositions, among other features. In contrast, the negatively loaded features on D1 are a high frequency of nouns and prepositions, a longer word length, and a high type/token ratio (Biber 1988: 102). Texts with positive scores on D1 are characterised by the combination of frequent positive features and infrequent negative features. Biber (1988: 131) describes texts with high D1 scores as interactive and affective, presenting content in a more general and fragmentary way. Furthermore, these texts include shorter words and repetitive vocabulary (Biber 1988: 131). Texts with lower D1 scores include longer words and a varied vocabulary. Although Biber (1988: 132) recognises that spoken registers (with higher dimensional scores) tend to contrast with written registers on D1, he specifies that the dimension more specifically distinguishes edited informational and un-edited affective registers.

Dimension 2: Narrative vs. Non-narrative Concerns (D2)

Nini (2019: 68) associates D2 with the presence of narrative in texts, where fictional texts receive higher dimensional scores. The positively loaded grammatical features for D2 include past tense verbs, perfect tense verbs, public verbs (e.g., *acknowledge, admit, claim, declare, deny, explain* (Biber 1988: 242)) and third person pronouns (Biber 1988: 102). According to Biber (1988: 137), low dimensional scores for D2 reflect a lack of narrative concern in text, which can be associated with the lack of a narrative purpose, but also the presence of some other specific function. Biber (1988: 141–142) also notes that many registers are unmarked on D2, meaning that a combination of narrative and non-narrative discourse can occur in registers, as in personal letters and spoken conversations.

Dimension 3: Situation-Dependent vs. Explicit Reference (D3)

Nini (2019: 68) summarises D3 as the measure of the text's dependence on context, contrasting dependent registers with low dimensional scores (such as broadcasts) with independent texts with high dimensional scores (such as academic texts). According to Biber (1988: 102) features contributing to high dimensional scores on D3 are *wh*-relative clauses on both object (e.g., *the man who Sally likes* (Biber 1988: 235)) and subject (e.g., *the man who likes popcorn* (Biber 1988: 235)) positions, phrasal co-ordination, and pied-piping constructions (e.g., *the manner in which he was told* (Biber 1988: 235)), whereas negatively loaded features on D3 are time/space adverbials (e.g., *above, around, behind, beside, afterwards, earlier, later, now, previously* (Biber 1988: 224)) and a general high frequency of adverbs. Biber (1988: 145) specifies that in registers characterised by low dimensional scores (broadcasts, for example), reference to the context of the text is possible and necessary for comprehension, whilst in registers with high scores (such as documents), explicit reference to context occurs within the text.

Dimension 4: Overt Expression of Persuasion (D4)

According to Biber (1988: 151), D4 reflects the degree of persuasion found in the text. Positively loaded features for D4 are associated with expressions of the author's stance and persuasion, and the negative side of D4 marks the absence of the positive features (Biber 1988: 148). Positive features include infinitives, predictive modal verbs (e.g., *will, shall* (Biber 1988: 242)), necessity modal verbs (e.g., *should, must* (Biber 1988: 242)), suasive verbs (e.g., *agree, ask, command, demand, insist, recommend, suggest* (Biber 1988: 242)), and conditional subordination (e.g., *if, unless* (Biber 1988: 236)). Biber (1988: 148) ties high scores on D4 to expressions of opinion and argumentation, as visible in professional letters and editorials, whilst low scores are characteristic of texts which have a descriptive function. Biber (1988: 151) also specifies that most registers are unmarked on D4.

Dimension 5: Abstract vs. Non-Abstract Information (D5)

Nini (2019: 68) summarises D5 as the measure of abstract information included in the texts and uses scientific discourse as an example of a register including abstract information and thus having a high D5 score. The abstract information label is associated with a high frequency of conjuncts, agentless passives, *by*-passives, and past participial clauses (Biber 1988: 103). According to Biber (1988: 151–153), academic writing and documents tend to feature high D5 scores, while fictional and conversational registers have low dimensional scores, indicating their lack of technical content.

Dimension 6: On-line Informational Elaboration (D6)

According to Nini (2019: 68), high scores on the sixth dimension can be associated with time constraints during the production of the text. This manifests in the texts as the high frequency of *that*-clauses as verb complements (e.g., *I said that he went* (Biber 1988: 231)), demonstratives, *that*-relative clauses on object position (e.g., *the dog that I saw* (Biber 1988: 234)), and *that*-clauses as adjective complements (e.g., *I'm glad that you like it* (Biber 1988:

231)). Biber (1988: 156) highlights interviews, prepared and spontaneous speeches as the primary registers with high D6 scores, which all share informational goals but time-critical production.

Despite the prevalence of the full MDA approach, which requires the establishment of register-specific dimensions, similar trends have emerged across different studies. Berber Sardinha and Veirano Pinto (2019: 20); Biber and Egbert (2016: 130), and Biber and Conrad (2019: 290) confirm that research findings indicate that two specific dimensions tend to emerge in full MDA studies for the majority of registers: a dimension contrasting spoken/written texts and a dimension contrasting texts with a narrative and non-narrative texts. Berber Sardinha and Veirano Pinto (2019: 20) even suggest that some dimensions of variation could be universal, whilst others apply to specific registers. As these dimensions are included in Biber's (1988) framework for additive MDA as D1 and D2, conducting additive MDA allows the identification of these dimensions in new registers, thus expanding their universal reach.

Having summarised the dimensions for additive MDA, the steps involved in the analysis process must also be considered. Berber Sardinha et al. (2019: 169) include specific instructions for conducting additive MDA, which are summarised below:

1. Selecting an MDA study as a basis for comparison.
2. Choosing the dimensions for analysis.
3. Identifying features relevant for the chosen dimensions and identifying them in a corpus.
4. Calculating the dimension scores based on the normalised frequency of linguistic features.
5. Comparison between computed dimension scores and those provided in the base study.

Additive MDA is reported to have several advantages over full MDA, although existing research shows a preference for the full MDA approach, according to Nini (2019: 70). Berber Sardinha and Veirano Pinto (2019: 3) and Berber Sardinha et al. (2019: 182) consider the additive approach less challenging than full MDA as it does not involve extracting register-specific factors, adding that factor analysis presents a significant challenge in the context of full MDA, as it requires experience and knowledge in advanced statistical techniques. Furthermore, Berber Sardinha et al. (2019: 182) emphasise the value of additive MDA studies for the base studies as they extend the scope of their results. Berber Sardinha and Veirano Pinto (2019: 3) recommend additive MDA for providing descriptions of specific text registers, and Berber Sardinha et al. (2019: 182) even advocate for the broader use of additive MDA among researchers. An example of thorough additive MDA can be found in Quaglio (2009: 58), who analyses a corpus of scripts from a television show, focusing on the first dimension. Quaglio (2009: 69) concludes that the corpus of television scripts is less diverse, but still similar to spoken conversations on the first dimension.

Previous research indicates the versatility of MDA for analysing text from different discourse domains. Berber Sardinha and Veirano Pinto (2019: 2) mention that MDA has been used for describing a wide variety of registers, including online texts, academic writing, popular media texts, fictional texts and texts depicting different varieties of English. Furthermore, Dixon (2022b: 74) claims that MDA has been extensively applied in the analysis of academic texts. Nini (2019: 67) also recognises the flexibility of MDA as a method, claiming that it can be used for the analysis of a wide range of texts and combined with other methods. Similarly, Berber Sardinha and Veirano Pinto (2019: 1) see it as a productive approach that can be used with various languages and domains.

The MDA method has also been successfully used to describe video game registers. In his dissertation, Dixon (2022b: 75–76), who claims to be the first researcher to conduct MDA

on a corpus of video game texts, combines both MDA approaches to provide a comprehensive description of text registers in RPG video games, but uses them to achieve different goals. In the full MDA approach, Dixon (2022b: 86) establishes four dimensions of variation distinctive to video game registers, including Interactive Discourse vs. Information on Specific Events; Interpretation of Past Events; Concise Technical Descriptions, and Procedures and Directives. In his additive MDA approach, Dixon (2022b: 162) investigated the similarities between spoken dialogue in real life and in RPG dialogue using the first dimension in Biber's framework. The results of MDA across the first dimension indicate that video game dialogue registers are similar to real-life spoken dialogue and distinct from written texts (Dixon 2022b: 166).

Dixon reaches several significant conclusions in his additive MDA research. Dixon's (2024b: 26) results indicate that video game registers are linguistically distinct, which also signifies their specific communicative purpose. Based on the dimensional scores, Dixon (2022b: 166; 2024b: 26) concludes that spoken game registers are similar to spoken communication registers in Biber's framework but specifies that the interactive speech register resembles real-life dialogue more closely than the immersive speech register. Dixon (2024b: 27) found that immersive and interactive speech registers had different dimensional scores, further confirming their differences. Additionally, the high frequency of nominal features in game registers indicates their resemblance to written scripts (Dixon 2024b: 27). However, it is necessary to recognise that Dixon's (2024b: 27) results only reflect the linguistic characteristics of four games that were included in the sample, and Dixon advises analysing other dimensions and increasing the size of the corpus. Connecting his results to DGBLL, Dixon (2024b: 27) suggests that MDA analysis results indicate that games expose players to real-life language.

For the current analysis, additive MDA emerges as the most suitable approach due to its clear and straightforward method and valuable results, which can be used to establish RPG

dialogue in comparison to several other text registers, highlighting the similarities and differences between video game dialogue and other well-established text registers. Conducting MDA also enables the description of the use of several linguistic features, compared to the narrower focus of register analysis. Furthermore, the additive approach benefits from the possibility of incorporating Dixon's (2022b) results and using them for comparison on the first dimension, whilst also providing further information about RPG dialogue on other dimensions included in Biber's framework.

2.1.5. Multidimensional Analysis Tagger

The Multidimensional Analysis Tagger (MAT) (Nini 2019) software, which is publicly available online, is used for conducting the MDA in this thesis. According to Nini (2019: 70), the software replicates Biber's analysis and allows users to conduct MDA on freely chosen datasets. Nini (2019: 90–91) encourages conducting MDA on specific registers to create models and descriptions of registers. The software calculates dimensional scores for the input using the dimensions established by Biber (1988) and assigns the most similar established register to the input on all dimensions. Additionally, MAT assigns the input a text type, which is as categorisation established by Biber (1989) (Nini 2019: 72). Nini (2019: 67) describes text type analysis as a later development in MDA, as the types are assigned on the basis of linguistic similarity across the six dimensions. The output of the program includes grammatically tagged input files, several files with statistics regarding the dimensional scores and graphs depicting visualisations of the dimensions. Nini (2019: 71) also confirms the reliability of the tagger results.

In Nini's (2019: 70) evaluation of previous MDA research, she finds that research using the MDA approach mainly uses factor analysis and cluster analysis to establish new registers and dimensions for specific datasets; in contrast, the original model of MDA is used less. Still,

Nini (2019: 70) supports using the original model to describe and establish new registers in contrast to the existing ones, which puts emphasis on the unique characteristics of new registers. Nini (2019: 70) recognises that the method should be easily accessible and presents the MAT software to encourage further analyses using MDA.

Nini (2019: 71) mentions that the main difference between the MAT and the original is the use of a different part-of-speech tagger, as the original option used by Biber is not publicly available. MAT uses the Stanford part-of-speech tagger (Toutanova et al. 2003) to tag the input with part-of-speech tags. Nini (2019: 71) acknowledges that this difference indicates that the results obtained using the MAT will not be entirely identical to the original and confirms that the output of the MAT is still reliable and comparable with the original results. The reliability of the tagger was confirmed by the successful replication of Biber's results for the LOB corpus; additionally, the reliability of the model was confirmed by the replication of Biber's results using an identically constructed sample (Nini 2019: 73). Nini (2019: 82) concludes that Biber's model is suitable for linguistic analysis and can be performed successfully using MAT. A manual is provided alongside MAT, which provides detailed instructions for working with the software (Nini 2019: 72). The full description of the software is provided in Nini (2019).

The MAT has been used for achieving different research goals and for the analysis of a diverse range of texts. Grieve and Woodfield (2023) use MAT to analyse a corpus of news articles written by one person, contrasting fake news with real articles. Similarly, Grieve (2023) uses MAT for contrasting opinion articles written by different authors in the same newspaper. Brugman et al. (2020) use MAT combined with other approaches on a corpus of satirical news articles. Regarding the use of MAT on popular media texts, Werner (2021) compiles a corpus of lyrics from pop music and uses MAT to establish the register of pop lyrics. Although video game dialogue has not yet been analysed using the MAT, other internet texts, such as scripts depicting spoken texts from *YouTube* videos, have been analysed using the MAT by Cooper

(2023). Berber Sardinha and Veirano Pinto (2019: 3) describe MAT as a user-friendly option for conducting MDA and recommend it for both beginners and advanced users, emphasising its accessibility in comparison to the Biber tagger. In light of this discussion, MAT can be considered an accessible and reliable option for conducting additive MDA and will be used in the current analysis.

2.2. Data

This section introduces currently available video game corpora and explains problems relevant to the creation of such corpora and their use in current video game language research. Following that, the corpus used as the data sample in the analysis is described in detail, focusing on its suitability for the current study.

2.2.1. Video Game Corpora

Despite the increasing interest in video game research, corpora including video game texts remain limited. Although the availability of corpora including video game texts is increasing, Rennick and Roberts (2024: 95) report specific challenges related to data collection, representation and distribution that may affect the creation of such corpora and thus account for their current scarcity. Similar concerns are raised by Heritage (2021: 93), who emphasises the difficulties of data collection from video games. Furthermore, as the data included in the games can often be subject to copyright, making the data of video game corpora publicly accessible may be complicated.

Despite the complex process of corpus creation, several corpora of video game texts have been created. The sizes of video game corpora differ drastically, ranging from including single games, as in Hämäläinen et al. (2022) and Erdur (2022), a few games, such as Dixon's (2024a) SPOC and Bianchi's (2024) ViPiCo, to a series of games, as in Heritage (2022) and

Lazzeretti and Gatti (2021). Several video game corpora include paratexts, meaning texts which are related to the video game domain but not included in the games themselves, such as Burghardt and Tiepmar's (2021) GWTC, depicting walkthroughs of video games and Juraska et al.'s (2019) corpus, which contains data depicting discussions on the topic of video games.

Although smaller video game corpora are becoming increasingly available, large-scale video game corpora, such as Video Game Dialogue Corpus (VGDC) (Rennick and Roberts 2024; Rennick et al. 2023a), which includes data from 50 video games, are not commonplace in current research. The current state of video game corpora research has been addressed by Dixon (2024a), Heritage (2020), van Stegeren and Theune (2021) and Hämäläinen et al. (2022), who agree that the current research including video game corpora is insufficient, whilst noting a growing interest in the topic, which is confirmed by the recent creation of several video game corpora. Furthermore, Heritage (2021: 86), Rennick and Roberts (2024: 93) and van Stegeren and Theune (2021) specify that previous studies have focused on the discourse around video games (paratexts) rather than the text included in the games. Rennick and Roberts (2024: 93) explain that issues related to data accessibility may explain the preference for analysing paratexts over in-game texts. Still, Heritage (2021:107) argues that paratexts cannot give an accurate overview of the language included in the games and therefore cannot be considered a sufficient substitute for data depicting in-game language.

Therefore, accessibility emerges as an important aspect in the creation of video game corpora. This issue has been addressed by Heritage (2021) and van Stegeren and Theune (2021), who analyse several methods for obtaining dialogue data from video games. Collecting data directly from the game using automated scripts or specific software, whilst technically challenging, is reported as yielding the most authentic and representative data (Heritage 2021: 99-102). Whilst van Stegeren and Theune (2021: 3) agree on the high quality of game-extracted data, they express concern over its reliance on many external factors such as the technical

proficiency of the user and the presence of official or fan-made software or tools for extracting data. Transcribing from game playthroughs, which benefits from the possibility to include highly specific tags in the data, suffers from its extensive duration, which makes obtaining all dialogue data from one game or multiple games a time-consuming process (Heritage 2021: 102–104). This approach can therefore be considered unsuitable for large-scale video game corpora.

Another option involves using fan-made transcripts of video game dialogue. While these sources are easily accessible, this type of data could include transcription and other errors and thus require further inspection and processing (Heritage 2021: 104–106). Van Stegeren and Theune (2021: 4) are also critical of the accuracy and reliability of crowd-sourced data. Still, fan-made data can be considered more efficient than manual transcriptions of data. Although van Stegeren and Theune (2021: 4) suggest that fan-made transcripts may be less accurate than texts extracted from games, Rennick and Roberts (2024: 95) argue that video game corpora need not rely solely upon the games themselves for high-quality data, as fan transcripts can also be considered accurate and representative sources. Furthermore, Heritage (2021: 106–107) and Rennick and Roberts (2024: 97) emphasise the difficulties of accessing video game dialogue data, pointing out that in some instances, crowdsourced data may be the only available or easily accessible option for collecting data. Therefore, existing video game corpora often incorporate data from multiple sources, as illustrated by the Video Game Dialogue Corpus.

2.2.2. Video Game Dialogue Corpus

As established previously, the collection of video game dialogue data can be a laborious and time-consuming process, requiring specific technical skills. Therefore, the current analysis uses an existing video game corpus, the Video Game Dialogue Corpus (VGDC) (Rennick and Roberts 2024; Rennick et al. 2023). In their study, Rennick et al. (2023a) compiled VGDC to

analyse dialogue differences between female and male characters in role-playing video games. By analysing the dialogue of 50 role-playing video games, Rennick et al. (2023a: 9) demonstrate the existence of a gender bias in the domain of video games, visible in the under-representation of female and non-male characters overall, the imbalance in the amount of dialogue between different genders, the use of gendered stereotypes in the content of the dialogue, and the use of gendered roles for characters. The VGDC is innovative among video game corpora due to its size and ease of access (Rennick et al. 2023a: 2).

Although VGDC is available in an online repository, the creators of the corpus agreed to provide direct access to the data upon personal request via email to ease the process of analysis for the current study. Currently, the corpus includes dialogue data from 50 video games from the genre of role-playing games (Rennick and Roberts 2024: 98). Despite slight imbalances in data as mentioned in Rennick et al. (2023b: 8–9), Rennick and Roberts (2024: 98-99) claim that the sample including the data from 50 games is representative of RPG genre due to the inclusion of games from different sub-genres (W-RPG and J-RPG), time periods (from 1985–2020) and with different age ratings (according to the Entertainment Software Ratings Board rating system and ranging from child to adult). The data included in the VGDC ranges from 400 to 780,000 words per game (Rennick et al. 2023b: 7–8). According to Rennick et al. (2023b: 9) and Rennick and Roberts (2024: 95), the data for the corpus was collected from various sources: game files, fan transcripts, and fan websites. The list of games included in the VGDC, along with their descriptive categories, such as subgenre, age rating, time period, and size, is provided in Appendix 1.

The corpus data is formatted as .json files, which the authors claim assists in the preservation of the structure of dialogue (Rennick and Roberts 2024: 101). Besides the dialogue data, VGDC also includes significant metadata for each game, including information about each speaking character, details about actions, locations and possible choices for the player.

Further detailed descriptions of the corpus creation process and the games included in the corpus can be found in Rennick et al. (2023a), Rennick et al. (2023b), and Rennick and Roberts (2024).

The sample used by Rennick et al. (2023a) and Rennick and Roberts (2024) is relevant to the current study as it contains a significant amount of video game dialogue. Furthermore, the sample meets the requirements for video game corpora set by van Stegeren and Theune (2021: 6): the data benefits from the inclusion of contextual tags, such as the choices available to the player, descriptions of actions, names of characters, locations of dialogue and other relevant information. Furthermore, the data is representative of popular games and includes the original dialogue written for the games. Although the data depicts games from the role-playing genre, it can still be considered diverse, as the corpus includes games from several subgenres, age ratings and time periods, and different game series (Roberts and Rennick 2024: 99-100). Finally, the corpus conforms to the requirement of accessibility, as it is available online in an easily accessible format. Therefore, VGDC emerges among video game corpora as a suitable and accessible source of video game dialogue data.

2.3. Data Analysis and Results

The current analysis was conducted using the MAT software, but the data of VGDC required pre-processing before it could be analysed using the MAT. The following section features the description of data preparation for the additive MDA analysis and a description of the analysis process itself. These descriptions are followed by the results of the additive MDA analysis, which are reported in multiple sections, including general results, results for the subcategories, and results across Biber's dimensions for additive MDA.

2.3.1. Data Preparation

Before the multi-dimensional analysis of linguistic features could be conducted with the MAT software, several steps were taken to prepare the data for analysis. The goal of the preparation was to preserve the text depicting the dialogue and remove all other elements in the file, such as the names of speakers and comments. The removal of the unrelated elements was required as the MAT software includes type-token ratio (TTR) as one of its measures; therefore, any additional text could affect the results of the analysis.

For each game, a dialogue data file and a metadata file were downloaded. To process the data, Python scripts were written and executed in the code editor application Visual Studio Code (Microsoft 2023). Within the scripts, regular expressions were utilised to seek specific text patterns within the files, which were then either modified or removed, as necessary.

The process of data cleanup was conducted in several stages. In the first stage, JSON structural elements (curly braces and square brackets) were removed from the dialogue files. After that, other non-dialogue elements were removed. The removed elements included action, location, context, choice and other descriptions which were tagged with specific keywords (Rennick and Roberts 2024: 101). Following the second step, data was examined for remaining game-specific non-dialogue elements (which were indicated with an underscore), and scripts were adjusted and executed again. The final step in the data processing was the removal of the names of the speakers. The names of speaking characters were provided in the metadata files accompanying the dialogue data.

After automatic processing using Python scripts, all text files were manually examined using the text editor software Notepad++ (Ho 2025) for any remaining non-dialogue data to ensure that all irrelevant elements were removed. During the examination, some errors were found relating to text encoding, meaning that certain characters were replaced by strings of Unicode hex codes. As these codes share a specific structure, the Replace tool in Notepad++

(Ho 2025) was used in the regular expressions search mode to replace the Unicode hex codes with the characters they represent.

Following the preparation process, the final sample consisted of 50 text files, including the dialogue of 50 role-playing games, a total of 6.1 million words. The sample includes games from eight different game series. The full composition of the sample is provided in Table 1 below.

Table 1. Composition of the VGDC sample by game series

Title	**Size (games)**	**Size (words)**	**Proportion of sample (%)**
Elder Scrolls	4	1 144 739	18.56%
Dragon Age	2	982 374	15.92%
Mass Effect	3	1 017 311	16.49%
Persona	3	587 017	9.52%
Kingdom Hearts	4	126 982	2.06%
King's Quest	9	126 866	2.06%
Monkey Island	3	54 481	0.88%
Final Fantasy	17	1 532 447	24.85%
Other (no series)	5	596 270	9.66%
Total	50	6 168 487	100%

2.3.2. Data Analysis

The additive MDA analysis presented in this section is guided by the research questions established in the introduction. To reiterate, the current analysis seeks to establish RPG video game dialogue in comparison to spoken and written registers and provide descriptions of the language included in RPG dialogue. An additional goal is to determine the influence of

subcategories, such as sub-genre, time period and age rating, on the language included in RPG dialogue.

The MAT software (Nini 2019; Toutanova et al. 2003) was used for the analysis of data. For each round of analysis, the option “Tag and Analyze” was chosen in the MAT software, with the specifications not to correct Z-scores, to count all tags, and to use 400 as the token count for calculating type-token ratio, since this is the default option in the software, and compatible with Biber’s analysis, according to Nini (2020). For each round of analysis, the option to create graphical representations of the dimensional scores was chosen for all six dimensions. Following the analysis process, the option Inspect was used to examine the occurrence of specific grammatical features and to find examples.

Following the initial analysis of the full sample of VGDC, the analysis process was conducted multiple times, with each round focusing on a different categorisation of the data, to reveal potential differences between different subcategories as proposed in the second research question. The categorisations mostly followed those used for the sampling frame in Rennick and Roberts (2024: 99). The data was categorised as follows: age rating (subcategories: adult, teen and child); subgenre (subcategories: Western RPG and Japanese RPG); and time period (subcategories:1985–1999 and 2000–2020). An additional round of analysis was conducted on each game series separately to check whether series that represent a greater proportion of the sample influence the results overall and to find whether significant differences emerge between the different game series.

The output files with dimensional score statistics were used to calculate the mean, minimum, and maximum dimensional scores, as well as the range of dimensional scores and standard deviation of dimensional scores. The dimensional score results for the full VGDC sample are provided in Table 2, and the dimensional score results for all subcategories are included in Appendix 2.

Although the MAT output also includes comparison to text types, which are established on the basis of the dimensional scores across six dimensions, the analysis and examination of those results were excluded from the current thesis to maintain focus on the descriptions of the dimensions separately. Still, as these types are relevant to MDA, they could be addressed in future research.

2.3.3. MDA Results for the VGDC Sample

The MDA results for the VGDC sample, including the mean scores for all six dimensions, the internal variation of the sample (expressed as the range of each dimensional score), the minimum and maximum dimensional scores, and the assigned closest registers, are included in Table 2. Overall, the sample was assigned five different registers based on their similarity to Biber's results across six dimensions: personal letters, academic prose, general fiction, broadcasts and official documents. The closest register label is assigned to the sample on all dimensions; it is assigned on the basis of the dimensional score for each text included in the sample, in comparison to Biber's (1988) results. The most common register assigned to the texts on a specific dimension is considered the closest register.

Regarding the assigned closest registers, it is significant that only one assigned register out of six – broadcasts on D5 – describes spoken communication. The first dimension (D1) includes the most significant internal variation, as the range for dimensional scores for that dimension is largest at 25.12 compared to the remaining dimensions, where the range is below ten. This relatively narrow range across most dimensions could reflect the internal uniformity of the dialogue, further validating its status as a distinct register. In contrast, the wide range on D1 also highlights the range of variation on the first dimension and a possible similarity to multiple registers on D1. The results for the first dimension merit further inspection to provide explanations for the variation occurring on this dimension.

Table 2. Dimensional score results for the full VGDC sample

Dimension	Mean Dimensional Score	Minimum Dimensional Score	Maximum Dimensional Score	Range	Standard Deviation	Closest register
D1	15.30	-1.16	23.96	25.12	6.02	Personal Letters
D2	-1.63	-3.95	0.01	3.96	0.82	Academic Prose
D3	-2.24	-4.73	0.75	5.48	0.94	General Fiction
D4	2.64	-1.05	6.78	7.83	1.86	Personal Letters
D5	-1.42	-2.54	0.08	2.62	0.47	Broadcasts
D6	-0.42	-1.81	0.88	2.69	0.50	Official Documents

2.3.4. MDA Results for the Subcategories

Tables 3–6 provide results of the MDA of the samples by different subcategories and the closest registers assigned by the MAT to each subcategory sample; specific dimensional scores for the subcategories are included in Appendix 2. Although the differences between the sample sizes (in words) for some subcategories seem significant, it is important to remember that with the progress of technology, game sizes, and therefore also game text sizes, have increased significantly.

Age ratings

The results of the MAT for the age rating subcategories, presented in Table 3, closely resemble the results obtained from the analysis of the full VDGC sample, with minor exceptions. Whilst some deviations emerged in D2 and D3, the assigned closest registers are identical for the other dimensions. For the second dimension, both Teen and Adult categories were assigned Conversations as the closest text register, whilst texts in the Child category were assigned the category of Academic prose, which is identical to the results of the VGDC sample.

Results for the second dimension are therefore unexpected, as the general sample is more similar to a written register, whilst teen and adult games include language resembling spoken language. Similarly, the adult age rating category emerges as the exception on the third dimension, where it is assigned the register of press reportage.

Table 3. Closest assigned text registers across six dimensions and assigned text type for age ratings. The most notable results are indicated with shading for emphasis.

Category	Adult	Teen	Child	VGDC
D1	Personal Letters	Personal Letters	Personal Letters	Personal Letters
D2	Conversations	Conversations	Academic Prose	Academic Prose
D3	Press Reportage	General Fiction	General Fiction	General fiction
D4	Personal Letters	Personal Letters	Personal Letters	Personal Letters
D5	Broadcasts	Broadcasts	Broadcasts	Broadcasts
D6	Official Documents	Official Documents	Official Documents	Official Documents
Sample size (games)	12	20	18	50
Sample size (words)	3 731 441	2 061 092	489 007	6 168 487

Time periods

Regarding the subcategorisation of time periods, presented in Table 4, the results also mirror the results of the entire VGDC sample, except for the period of 2000-2020 on the second dimension, where it is assigned Conversations as the closest register. Although this result diverges from the general results for the second dimension, it aligns with the age rating subcategory results, indicating a possible variation within the sample regarding that dimension.

Table 4. Closest assigned text registers across six dimensions and assigned text type for time periods. The most notable results are indicated with shading for emphasis.

Category	1985-1999	2000-2020	VGDC
D1	Personal Letters	Personal Letters	Personal Letters
D2	Academic prose	Conversations	Academic Prose
D3	General Fiction	General Fiction	General fiction
D4	Personal Letters	Personal Letters	Personal Letters
D5	Broadcasts	Broadcasts	Broadcasts
D6	Official Documents	Official Documents	Official Documents
Sample size (games)	21	29	50
Sample size (words)	485 323	5 796 217	6 168 487

Sub-genres

The results for the JRPG and WRPG subgenres, presented in Table 5, show near-identical patterns in their assigned registers when compared to the results of VGDC. The only exception, the conversations register, which was assigned on the second dimension in the JRPG sub-genre, is consistent with results in other sub-categories, and requires closer examination, as the registers of conversation and academic prose contrast in their production modes.

Table 5. Closest assigned text registers across six dimensions and assigned text type for sub-genres. The most notable results are indicated with shading for emphasis.

Category	WRPG	JRPG	VGDC
D1	Personal Letters	Personal Letters	Personal Letters
D2	Academic Prose	Conversations	Academic Prose
D3	General Fiction	General Fiction	General fiction
D4	Personal Letters	Personal Letters	Personal Letters

D5	Broadcasts	Broadcasts	Broadcasts
D6	Official Documents	Official Documents	Official Documents
Sample size (games)	24	26	50
Sample size (words)	3 983 357	2 298 183	6 168 487

Game Series

Regarding the composition of the VGDC sample (Table 1), it is reasonable to compare the assigned registers by game series to prevent a disproportionate game series from distorting the results. From the MDA results for the game series, presented in Table 6, the second dimension emerges as anomalous, reflecting the results for other subcategories. Whilst the MAT results for other dimensions (1, 3, 4, 5, and 6) are generally similar to the results of VGDC in their assigned registers, the second dimension appears again as the exception. Most of the series were assigned the conversations register on the second dimension, which does not conform to the general results. Additionally, the series *Monkey Island* was assigned the Official documents register, which is comparable to the general academic prose register. However, if five series out of eight feature a distinctly different categorisation, it is necessary to check whether the contents of the rest of the sample, which were categorised as similar to academic prose, skew the overall results for that dimension.

For the second dimension, the label Academic Prose was assigned to the *Kingdom Hearts* series and the Other category, which combines five games that do not belong to any series included in the VGDC. Based on data from Table 1, only 11.72% of the sample was labelled as Academic Prose, which is interesting as it is the register assigned for the VGDC sample. In comparison, the conversations register appears for 87.4% of the sample on the second dimension, and the official documents register appears for 0.88% of the sample. Although the difference between these registers may appear great, further examination in the following section will reveal similar tendencies for these registers on the second dimension.

Therefore, the second dimension and the assigned registers on D2 will be examined further qualitatively to determine their characteristics and functions on that dimension.

Table 6. Closest assigned text registers across six dimensions for the game series. Games included in the Other category: *Stardew Valley, Super Mario RPG: Legend of the Seven Stars, Chrono Trigger, Horizon Zero Dawn, Star Wars: Knights of the Old Republic.* The most notable results are indicated with shading for emphasis.

Series	D1	D2	D3	D4	D5	D6
Elder Scrolls	Personal letters	Conversations	Press reportage	Personal Letters	Broadcasts	Official Documents
Dragon Age	Personal letters	Conversations	General Fiction	Personal Letters	Broadcasts	Official Documents
Mass Effect	Personal letters	Conversations	Press reportage	Personal Letters	Broadcasts	Official Documents
Persona	Personal letters	Conversations	General Fiction	Personal Letters	Broadcasts	Conversations
Kingdom Hearts	Personal letters	Academic Prose	General Fiction	Personal Letters	Broadcasts	Official Documents
King's Quest	Prepared Speeches	Conversations	General Fiction	General Fiction	Broadcasts	Official Documents
Monkey Island	Personal letters	Official Documents	General Fiction	Personal Letters	General Fiction	Conversations
Final Fantasy	Personal letters	Conversations	General Fiction	Personal Letters	Broadcasts	Official Documents
Other	Personal letters	Academic Prose	General Fiction	Personal Letters	Broadcasts	Official Documents
VGDC	Personal letters	Academic Prose	General Fiction	Personal Letters	Broadcasts	Official Documents

Although the data was analysed in different sub-categories, the register of RPG video game dialogue seems to be quite consistent across different subcategories. Across the six dimensions, the registers assigned to the subcategories mostly follow the closest registers of the full sample, but more minor variations tend to appear in the assigned registers on Dimension 2 and Dimension 3.

The results from the analysis of subcategories indicate that RPG video game dialogue combines features from both spoken (*conversations, broadcasts*) and written (*personal letters, general fiction, official documents*) registers. This is consistent with the overall characteristics of RPG dialogue, as it is a text written with the goal of resembling spoken conversation. As the results were mostly uniform across different subcategories, focusing on the comparison of the subcategories would be ineffective and subsequent analysis will focus on providing a description of the full VGDC sample across the six dimensions established by Biber (1988).

2.3.5. MDA Results across Six Dimensions Established by Biber (1988)

Biber (1988: 122–125) provides dimensional score ranges for different text registers based on his analysis. These scores, along with the data of VGDC, were used to create box plots, which would aid in the comparison of the sample to other established registers across six dimensions. The plots were created using a Python script.

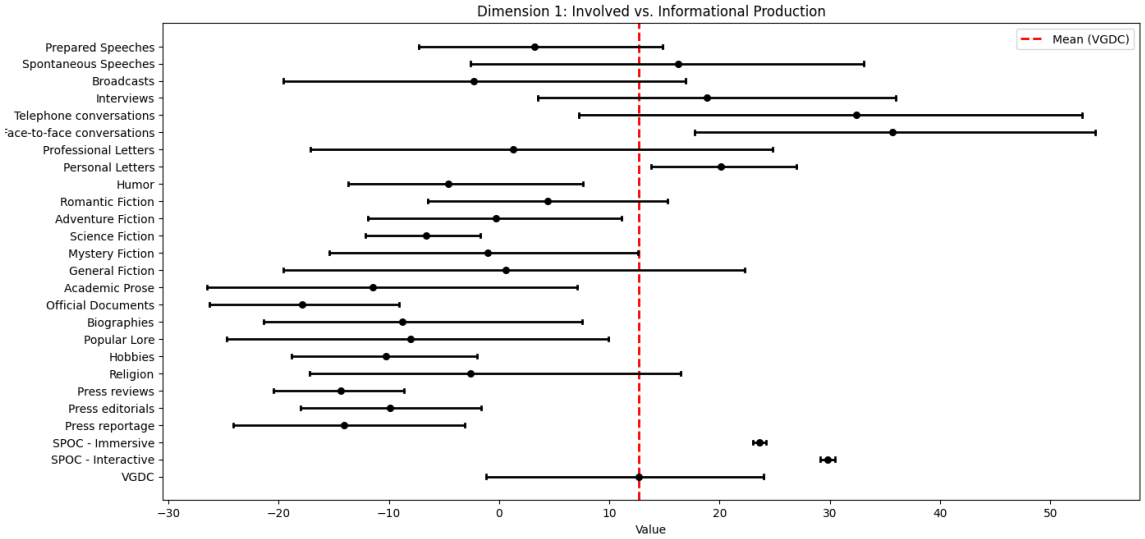
The dimensional score range comparison for the first dimension is provided in Figure 3. The plot for the first dimension also includes Dixon's (2024b: 16) additive MDA results for spoken RPG dialogue registers, which are marked as SPOC – Immersive and SPOC – Interactive, for comparison. From the examination of Figure 3, it appears that the range of the VGDC sample is smaller compared to the registers established by Biber (1988), yet more extensive than the range of spoken registers in SPOC. This variation could be explained by the size differences between the samples: Biber (1988: 67) used about 500 texts from two different

corpora in total, whilst Dixon's sample (2024b) contains text from 4 separate games. As variation can occur within a sample, it is reasonable to assume that a more extensive sample would include more variation. However, it is unexpected that the VGDC and SPOC results appear different and without extensive overlap, as the VGDC sample includes both spoken registers included in the SPOC. Still, it is significant that the positive end of the range for VGDC overlaps the immersive speech register in SPOC, indicating some similarity between the samples. One possible explanation for the different dimensional scores could be the limited scope of the SPOC samples, which further suggests a great range of variation found within the register of RPG dialogue, appearing as distinct differences in dialogue across different games.

Although the VGDC sample was assigned the personal letters register (Table 2), the mean dimensional score for VGDC seems to be closest to the spontaneous speeches and interviews registers, which are both spoken registers. This resemblance to involved production and spoken registers suggests the frequent occurrence of private verbs, contractions, present tense and emphatics in the dialogue of VGDC. Regarding the personal letters register, the mean dimensional score of VGDC seems close to the negative edge of its range, and the range of the VGDC sample is more extensive than the range of the personal letters register. This indicates a similarity between spontaneous speech, personal letters, interviews, and RPG dialogue registers on the first dimension, which will be examined further qualitatively in the following section. Overall, the positive mean dimensional score for D1 (Table 2) and comparison to other registers on that dimension suggest a closer resemblance to involved production, which is characteristic of spoken registers, such as conversations. A similar result emerged in Dixon's (2024b: 18) MDA study, where he found that the dialogue registers of SPOC featured the highest D1 scores. Still, the extensive range for dimensional scores on the first dimension and the negative minimum dimensional score indicate a great range of variation occurring in the sample and a possible similarity to informational production for specific games, which is more

characteristic of written registers, like fiction. This will be examined further qualitatively in the next section.

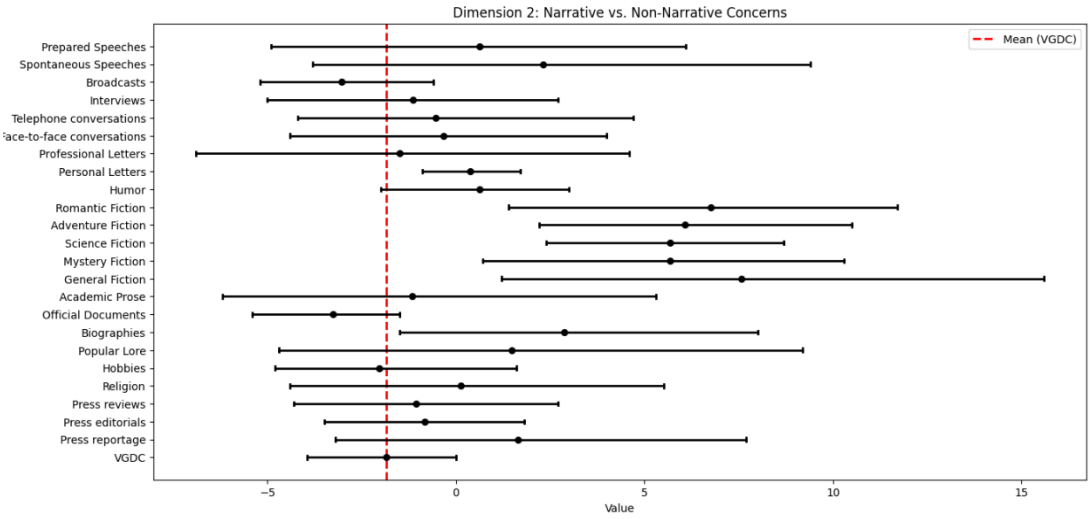
Figure 3. The range of dimensional scores for the VGDC across the first dimension and comparison to Dixon's (2024b: 16) results for SPOC. The smaller black points mark the mean dimensional score values for the register, the end caps for ranges appear as small black lines at the minimum and maximum dimensional score values, and the red dashed line marks the mean dimensional score for the VGDC sample.



The plot for the second dimension is available in Figure 4. The second dimension was analysed as closest to the academic prose register, although analysis of the sub-categories revealed the significance of the conversations register for that dimension. The range of the dimensional scores is significantly narrower than it was on the first dimension. Although the mean dimensional score for D2 is similar to that of the academic prose register, it is necessary to note that the mean dimensional score of VGDC fits in the ranges for most of the registers included in the plot, except for registers describing different genres of fiction and a non-fictional literary genre, which are known to include narration. Therefore, RPG dialogue can be contrasted with narrative registers for the second dimension, indicating a lower frequency of

past tense verbs, public verbs, and perfect aspect verbs, which are associated with conveying narrative concern. The register of conversations, which was prevalent in the result of the subcategories, has a wide range of variation within the second dimension on the positive end of the scale, and therefore, it is understandable that RPG dialogue would fit under that label. Similarly, the academic prose register features a wide range but occurs more on the negative side of the dimension. The range of the VGDC sample overlaps with the positive end of the academic prose register and the negative end of conversational registers. Therefore, the academic prose and conversational registers can be considered similar on the second dimension, and these similarities will be examined further qualitatively in the following section. Overall, the negative mean dimensional score and scores occurring on the negative end of the scale characterise RPG dialogue as text including non-narrative concerns.

Figure 4. The range of dimensional scores for the VGDC across the second dimension



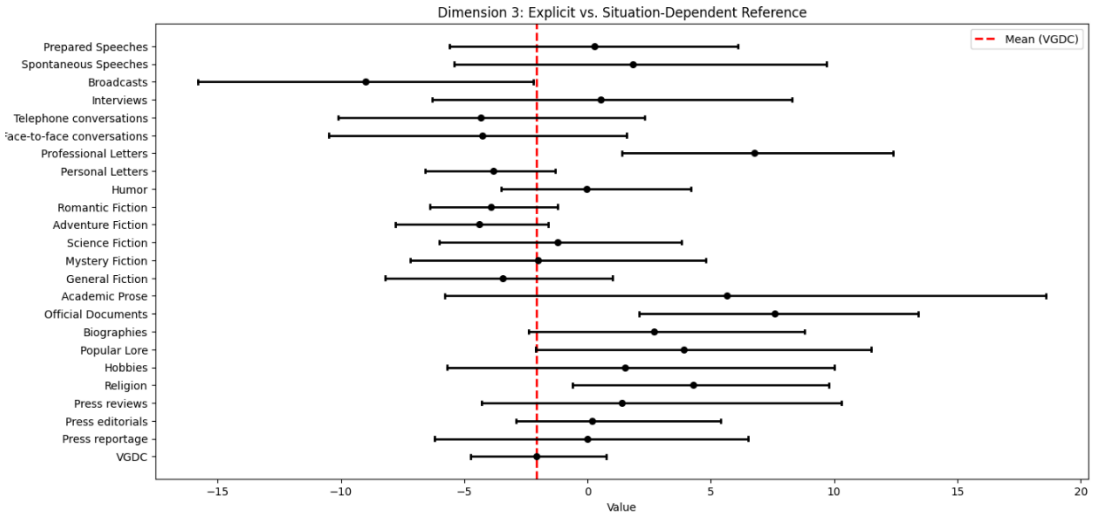
The plot for the third dimension is available in Figure 5. The MAT tagger assigned the VGDC sample as closest to the general fiction register on the third dimension. The sample features a narrow range on this dimension. In contrast with the previous dimension, the mean dimensional score is only in the vicinity of two additional registers, mystery fiction and science

fiction. The negative mean dimensional score and a range occurring on the negative side of the dimension describe RPG dialogue in VGDC as including explicit reference, similarly to fictional texts. Within dialogue, this appears as the high frequency of time and place adverbials and adverbs in general, combined with the low frequency of *wh*-relative clauses, phrasal coordination, and pied-piping constructions (Biber 1988: 102).

In comparison with other registers on D3, the dimensional scores occur at the centre of the dimension, but the negative mean dimensional score and a range occurring on the negative side of the dimension describe RPG dialogue in VGDC as leaning towards situation-dependent reference but possibly including both explicit and situation-dependent reference.

According to Biber (1988: 142), the lower scores for fiction on D3 indicate the use of situation-dependent reference in those registers. The lowest scores for D3 appear for broadcasts, where the listener and speaker do not share production circumstances, and this requires the text to be highly descriptive (Biber 1988: 147). Situation-dependent reference, although to a lesser extent, can also occur for RPG dialogue, as the producer of the text (the game) and the listener (the player) do not share the production circumstances; however, as the text must be understood by the player, it is reasonable to make it as detailed and descriptive as possible to ensure comprehension. Biber (1988: 147–148) specifies that fictional genres on D3 feature reference to the text-internal context, which also applies for RPGs, where game-internal context is referred to; however, the player has direct access to the production circumstances in RPGs.

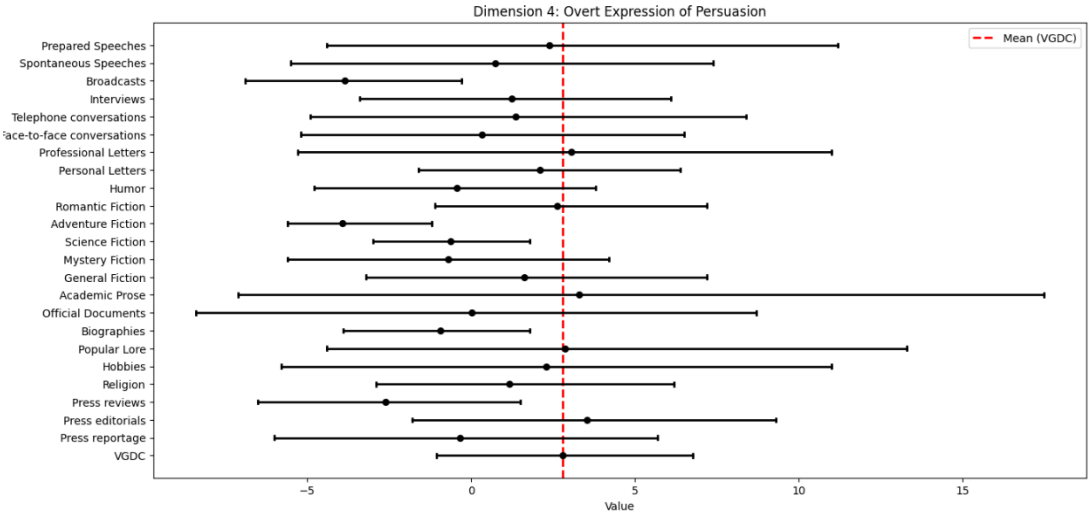
Figure 5. The range of dimensional scores for the VGDC across the third dimension



The plot for the fourth dimension is available in Figure 6. On the fourth dimension, the VGDC is assigned closest to the register of personal letters, as in the first dimension. Similarly, the range of the dimensional scores is narrower than for the first dimension, but more extensive than for the other dimensions. The mean dimensional score for the VGDC sample fits in the range for most registers in this dimension, but the mean dimensional score is also adjacent to the popular lore, romantic fiction, and professional letters registers. Compared to other registers, the dimensional scores for VGDC are high, but the minimal score has a negative value, indicating that the RPG dialogue can include features used to convey persuasion.

Biber (1988: 148) interprets personal letters as having intermediate scores on the fourth dimension, concluding that they are moderately persuasive. In dialogue, this can be seen in the moderate frequency of infinitives, different types of modals, and suasive verbs (Biber 1988: 103). In the context of RPG dialogue, the use of persuasive features can be interpreted as related to the necessity of having the player follow the direction provided by the dialogue when it functions as a ludic text.

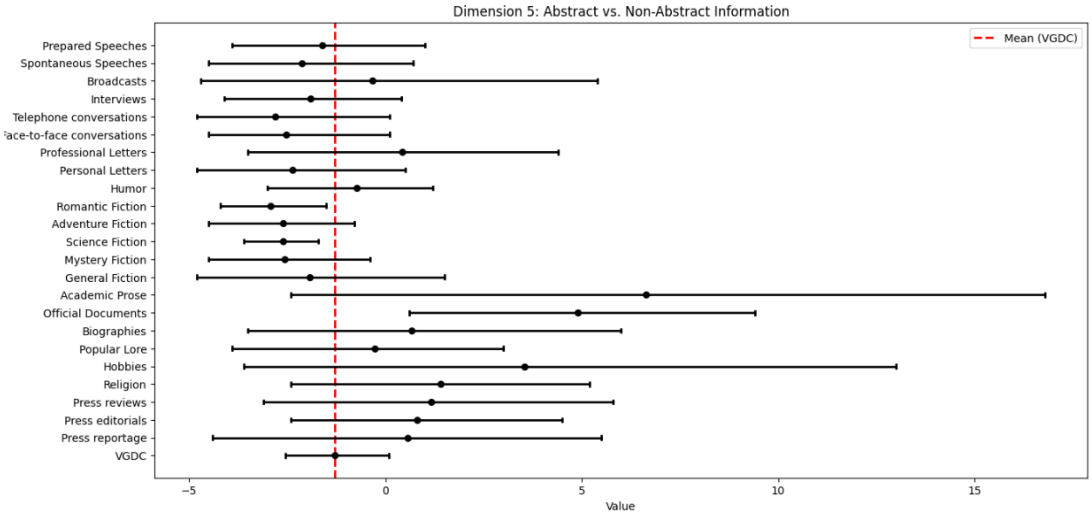
Figure 6. The range of dimensional scores for the VGDC across the fourth dimension



The plot for the fifth dimension is presented in Figure 7. The VGDC sample was assigned as closest to the broadcasts register, which was the only oral register assigned by the MAT across the six dimensions. The VGDC sample has the narrowest range and a relatively intermediate dimensional score, close to fictional and oral registers on this dimension, despite the wide variation found in most registers on this dimension. The intermediate dimensional score appears within the texts as the intermediate frequencies of conjuncts, agentless passives, past participial clauses and *by*-passives (Biber 1988: 103). Overall, the VGDC sample seems to include a mixture of abstract and non-abstract information, with a tendency to lean towards non-abstract information as present in fiction and oral registers and characterised by a lower type/token ratio, indicating a less diverse vocabulary.

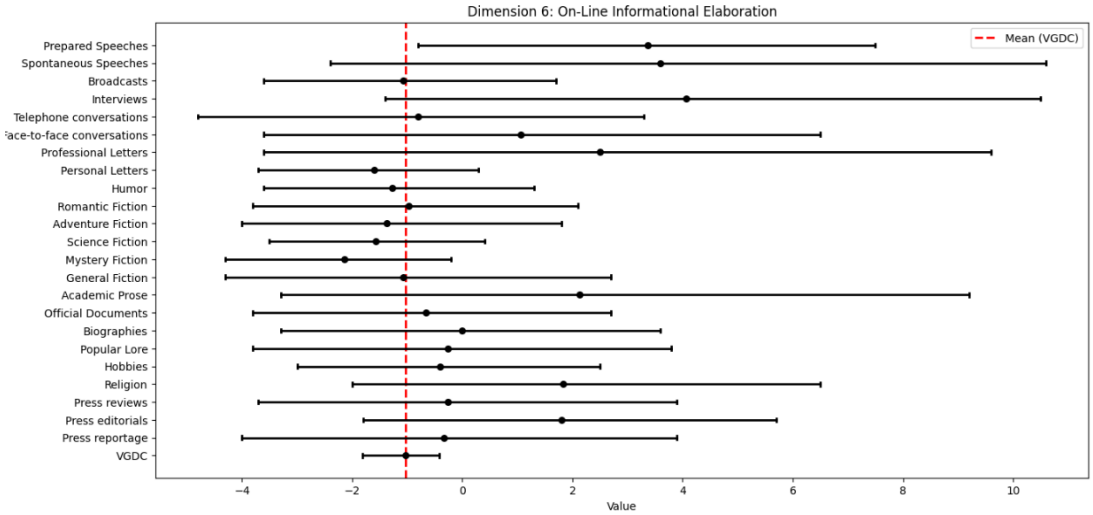
According to Biber (1988: 154), registers depicting active interaction between participants tend to have lower scores on the fifth dimension, which indicates their less abstract content and non-technical focus. Such interpretation can also describe RPG dialogue well, as these are active conversations between characters.

Figure 7. The range of dimensional scores for the VGDC across the fifth dimension.



The plot for the sixth dimension is available in Figure 8. The results of the MAT indicate that the VGDC sample resembles the official documents register on this dimension. In a stark contrast with the other registers, the VGDC sample features a narrow range of variation. The mean dimensional score of the sample, occurring on the negative end of the dimension, seems to indicate a similarity to the written registers. This indicates that the texts in VGDC include a higher frequency of phrasal coordination, together with a lower frequency of several types of *that*-clauses and demonstratives. The lower dimensional score describes RPG dialogue as including text not produced under time constraints, which is characteristic of written text, but not spoken real-life dialogue. Thus, the sixth dimension seemingly confirms the resemblance between written text and RPG dialogue. Biber (1988: 154) interprets negative scores on D6 as texts with informational focus, which have no production constraints. Regarding RPG dialogue, the informational focus could apply to the ludic use of diegetic text (instructions conveyed via immersive dialogue), where relevant information is incorporated into diegetic texts, such as dialogue. Furthermore, RPG dialogue does not have production constraints and is carefully edited prior to game release; therefore, RPG dialogue resembles fictional, written texts on D6.

Figure 8. The range of dimensional scores for the VGDC across the sixth dimension



From the plots (Figure 3–8), it is apparent that VGDC contains texts that are quite similar, as the ranges for the dimensional scores are quite narrow across most dimensions, except for Dimension 1. This could indicate that RPG dialogue in VGDC emerges as a distinct register, with its own specific characteristic language that serves specific functions within RPG games. The results of the additive MDA analysis provide answers to research questions established for the analysis. The results indicate that RPG dialogue within VGDC can be described as including involved production, indicating some similarity to spoken texts (D1), a lack of narrative (D2), situation-dependent reference (D3); persuasive language (D4), a lack of abstract/technical information (D5), and text not produced under time constraints, which indicates resemblance to written texts (D6). Overall, this suggests that the language of RPG dialogue in VGDC can be characterised as resembling both written and spoken texts, which accurately describes the production process of video game dialogue. Descriptions of specific linguistic features pertaining to specific dimensions were also provided in the analysis. Furthermore, a lack of influence from the sub-categories was observed in the analysis.

2.4. Discussion

The exploratory approach of the current analysis prioritises the first and second dimensions, as initial results suggested significant internal variation occurring on those dimensions. Therefore, the following sections will include a qualitative analysis of the first two dimensions, illustrated by examples from the VGDC dataset. Following the qualitative analyses, limitations of the current analysis and recommendations for future research are conveyed.

2.4.1. Qualitative Analysis of D1 Results

Overall, the high dimensional score of VGDC on D1 indicates that RPG dialogue could be classified as involved production, indicating its similarity to spoken registers. The high dimensional score is associated with the high occurrence of private verbs (e.g. *believe, decide, feel*), *that*-deletion, contractions, present tense, second person pronouns, demonstrative pronouns coupled with the low occurrence of prepositions, lower frequency of nouns, lower word length, higher TTR, higher frequency of attributive adjectives, among other grammatical features (Biber 1988: 102). According to Biber (1988: 104), these features associated with higher dimensional scores on D1 appear in non-informational texts, which are produced in real time without the possibility of modification after production. These contrast with the low-scoring texts, which are produced with the goal of conveying significant amounts of information and are therefore carefully constructed and edited (Biber 1988: 105). According to the results on D1, RPG dialogue in VGDC resembles spontaneous, un-edited texts more than edited texts, meaning that the writers of the dialogue succeed in emulating spontaneous spoken conversations in a written text, to some extent.

Biber (1988: 197) establishes that the first dimension describes the purpose of the text and the production circumstances. Biber (1988: 105) characterises positive-scoring D1 texts as interactive, but also more fragmented and general in their content. This effect is achieved by

using various co-occurring grammatical features. Biber (1988: 106) lists private verbs, present tense and second-person pronouns as features used for marking the interaction between the participants of a conversation. Additionally, Biber (1988: 132) mentions lower TTR in interactive texts as an indicator of a more general vocabulary in spoken, spontaneous texts. Other positively loaded features on D1 reflect the generalisation found in involved production, as information is presented in a non-specific manner (Biber 1988: 107). Biber (1988: 108) thus concludes that the time limits set on production affect the precision of the resulting text.

The involved production label emphasises the interactive goals of dialogue over the informational goals, relating more to Domsch's (2017: 254–256) category of diegetic dialogue, and Dixon's (2024b: 6) register of immersive speech, which are both used with the goal of creating immersion. However, the extensive range for the dimensional scores on the first dimension indicates that dialogue in specific games could be viewed as more informational and closer to written texts, reflecting the variability present in the register of RPG dialogue. Some possible interactive purposes of RPG dialogue could be the establishment of the context of the game via interactions with non-playable characters in the game.

The VGDC sample was assigned the register of personal letters as the closest register in Biber's analysis. Biber (1988: 108) mentions specifically the register of personal letters as anomalous on the first dimension, as it contains written texts that are markedly interactive and produced without time limits affecting their content. According to Biber (1988: 131), the high D1 score for personal letters reflects its linguistic similarity to spoken conversations. The personal letter register assignment to RPG dialogue further confirms that written dialogue within RPGs appears similar to spoken dialogue, but not completely, as the sample was not assigned labels related to spoken dialogue. Still, the personal letters register successfully captures the combination of written and spoken modes of production for the RPG dialogue.

A wide range of variation occurs regarding the linguistic features included in D1. *Final Fantasy I* had the minimal dimensional score (-1.16) on D1 and should therefore include dialogue that is more similar to written text produced without time limits and with possibilities for editing. In contrast, *Persona 3*, with the greatest dimensional score of 23.96, should include text resembling real-life spoken dialogue. In the following, examples from the dialogue will be provided, depicting the use of specific linguistic characteristics from multiple games, to highlight their difference on D1.

Positive loading features on D1, which are related to spoken dialogue produced with time constraints, occur more frequently in *Persona 3* and less frequently in *Final Fantasy I.* As in spoken conversations, private verbs (*I guess we'll see; I feel kinda bad though, spying on him like this…; I believe I mentioned him earlier, but this is Akihiko*) are used to convey the opinions and thoughts of the speaker. Furthermore, the use of present tense verbs (*Is it okay for him to be here; It's getting late, so are you ready to go?; But first, we need to go to the auditorium*) reflects the speakers' interactions with the surrounding environment and highlights the interactivity of the text further. Several grammatical features also help in reducing the amount of text and generalising it: contractions (*He'll eventually be moved to a room in the boys' dorm; I'm Ms. Toriumi; I've got something to take care of for the archery team*) and analytic negation (*It doesn't matter anymore; I can't blame you, though; Well, she won't be gone long*) often appear in spoken texts. Similarly, *that*-deletion (*I thought you might know, but I guess not; They say it's due to stress but …; I bet you're stoked too, right?!*) is characteristic of spoken texts and could be seen as a contraction. The use of *do* as a pro-verb (*He does his work well; The team did a great job; I'll do my best*) and using demonstrative pronouns (*I'm aware of that; That is why you must be made aware of the nature of your power; So this is it, huh?*) successfully generalises the content of the text, and is also characteristic to spoken language.

In contrast, *Final Fantasy I* features several linguistic features associated with written texts, which appear less frequently in *Persona 3.* The higher average word length in *Final Fantasy I* was greater than the word length in *Persona 3* (4.13 > 3.85), which contrasts with spoken dialogue, where words tend to be shorter. Furthermore, *Final Fantasy I* featured a higher count of prepositions (*In sadness, the Queen locked herself inside; Just as in Lukahn's prophecy; Thank you for saving the princess*), place adverbs (*The Elves live across the sea; I saw it north, near the waterfall; He floated far to the west.*), attributive adjectives (*Garland used to be a good knight until…; Matoya has poor eyesight;*) and agentless passives (*Our Prince was laid under a curse and our treasury ransacked; The Clinic was destroyed and the town was cursed; However, my supply of adamant is exhausted.*) which reflect a text that is constructed with care, and therefore tends to include more specific information.

On the first dimension, some similarity to Dixon's (2024b) results was noted, as his immersive speech register was included in the range of VGDC results. This similarity was expected as dialogue classified as immersive speech was included in the VGDC sample. Conversely, Dixon's interactive speech register, which did not overlap with VGDC on D1, suggests that language occurring in RPG dialogue can be presented in different ways within the genre across different games. Furthermore, the connection between sample sizes and the ranges for dimensional scores across D1 confirms that variety occurs even within the same register, thus necessitating the use of extended sample sizes in future research to create a more accurate description of the language included in the RPG dialogue register.

Regardless of the differences occurring on the dimensional scale for the VGDC sample, the overall results describe RPG dialogue on the first dimension as a written text that is very similar to spoken dialogue, reflecting the production circumstances for the texts in this register. Additionally, the texts lean more towards involved production, indicating the successful emulation of spoken registers.

2.4.2. Qualitative Analysis of D2 Results

The negative dimensional score for the VGDC sample on the second dimension indicates that dialogue within RPG video games can be classified as non-narrative, which also characterises spoken registers. According to Biber (1988: 108), positive features on D2 indicate the use of narratives within texts. These features include past tense and perfect aspect verbs, which are used for describing past events, third person pronouns, which mark the participants of the narratives excluding the speaker and listener, public verbs, which are used for reported speech, and present participial clauses which are used for descriptive purposes (Biber 1988: 137). Furthermore, Biber (1988: 109) speculates that synthetic negation occurs in narratives due to its emphatic value.

In contrast to narrative texts, negatively scoring texts on D2 are non-narrative; these texts lack features used for constructing a narrative, instead, present tense verbs and attributive adjectives, which are used to describe current events, occur frequently (Biber 1988: 109). Therefore, dialogue in RPGs can be classified as a non-narrative text, further indicating its resemblance to spoken registers. The similarity between dialogue and non-narrative texts seems reasonable, as dialogue between characters could focus on the immediate surroundings of the speakers and current events, resembling the involved production on D1. Biber (1988: 109) additionally mentions that D2 can also be interpreted as marking contrast between active, event-oriented and motionless, descriptive texts. This links the functions of the first and second dimensions and further confirms that active involvement, which is an important characteristic of spoken dialogue in real life, is also necessary in RPG dialogue, as it occurs between multiple characters and requires input from multiple parties.

The results of MAT classified the VGDC sample as the academic prose register on the second dimension. Biber (1988: 135) notes that registers on D2 may receive high scores, such as those found in fictional texts; low scores, characteristic of official documents; and

intermediate scores, commonly associated with spoken conversational registers. Figure 4 reveals that academic prose has a wide range on D2, which also overlaps with the narrower official documents range, indicating the similarities between these registers on D2. In comparison, the registers depicting spoken conversations feature quite similar ranges and mean dimensional scores, indicating their general similarity. Both conversations and academic prose registers are generally marked as non-narrative, but their wide ranges demonstrate the possibility of narration occurring within these registers. This observation is confirmed by Biber (1988: 141), who describes the conversations register with an intermediate value on D2, as it can combine narrative and non-narrative concerns, for example, when telling stories during conversations. In these instances, Biber (1988: 142) describes the narrative as subordinate to a larger purpose. Regarding RPG dialogue, it is possible that storytelling occasionally occurs within the dialogue, as it can aid in the creation of immersion. Similarly to D1, this strengthens the argument for the immersive and diegetic functions of the RPG dialogue register.

For the second dimension, *King's Quest III* had the highest dimensional score (0.01), indicating a combination of narrative and non-narrative constructions, whilst *Stardew Valley,* with the minimum dimensional score of -3.95, should be more non-narrative in comparison. Whilst the primary indicator of non-narrative features is the lack of features associated with narrative concerns, *Stardew Valley* also features a higher count of present tense verbs (*It's good to see you again; Do you have scarecrows on your farm?; On days like this, I like to help Evelyn with the public gardens*) and attributive adjectives (*I always enjoyed exploring those overgrown fields by myself; It was a good dream; Pretty cool strategy, huh?*), which are used to address and describe the current situation. Furthermore, past participial WHIZ deletion relatives (the omission of the relative pronoun and the verb *be* in a reduced relative clause) (*You're not a spy sent by the Shadow People, are you?; No responsibilities, Tunnelers game tonight, fridge stocked with all the goodies a guy could need; You found a golden scroll written*

in an unknown language?) are used to describe objects in the immediate context. *Stardew Valley* also features a lower word length (3.8) compared to 4.1 for *King's Quest III,* which is associated with non-narrative texts on the second dimension.

In contrast, *King's Quest III* features several characteristics associated with narrative construction, including past tense verbs (*Obviously, you were raised by a naughty wizard; It was horrible!; You ate it all*) and perfect aspect (*The impulse to buzz around has vanished; You have always admired this lovely tapestry; The soft grass of the forest has been replaced by the coarse grass of the seashore*), which are used to relate past occurrences to the listeners/readers in narratives. For RPG games, these narrative features can be used to explicitly describe the background information for specific locations and characters without using dialogue.

Third-person pronouns (*They are resting on the oak table; Evidently, they don't hear you; She growls menacingly*) have similar functions in narratives and in RPG dialogue, as they are used to include third persons, excluding the speaker and the listener, in the narrative. Similarly, public verbs (*That boy ought to swim back to shore, you overhear one fish say to another; I'll explain it all later!; She replies, Hard times hit Daventry right after you were kidnapped.*) are used to include reported speech in a narrative text. Synthetic negation (*Funny, no response; Otherwise, you have no interest in it; There are no climbable trees here*) and present participial clauses (*Looking at the butter churn makes you ache as you remember the long hours…; The trapdoor groans shut, disappearing into the floorboards; The burning torches throw flickering shadows across the walks, making the room even more ominous.*) also appear in *King's Quest III* with a descriptive function.

A possible reason for the high dimensional score for *King's Quest III* on the second dimension is related to its time of release: in 1986, the interactive aspects of video games were still limited, and therefore several aspects, such as descriptions of surroundings, sounds and speech, expressions of the player character's emotions and thoughts, are conveyed to the player

via dialogue, whereas nowadays, they could be conveyed in other ways of interacting with the game, such as the player character's movement, or changing camera viewpoint.

2.4.3. Discussion of the Analysis Results

This thesis highlights the underexplored area of RPG dialogue corpus studies and provides initial information to help bridge that gap. By conducting an additive MDA analysis on the largest currently available video game corpus, the thesis provides data-based descriptions of the language included in RPGs, expanding the results of prior research, which has primarily focused on describing game-related texts and smaller datasets.

As established in the previous chapter, the problematic nature of genre classifications for video games has emerged as a significant issue in video game studies. As Rennick et al. (2023a), Domsch (2017: 266), Mäyra (2017: 271), and Dixon (2023: 2) emphasise the importance of dialogue within the genre of RPG, language-oriented classifications could be established for RPGs to acknowledge the variation within the genre. The findings of the current study could be incorporated into a language-oriented classification of RPGs in connection with Biber's (1988) dimensions of variation, distinguishing more interactive and affective games from static ones, and more-narrative games from less narrative ones. Establishing such classifications addresses the criticism of Clearwater (2011: 35) and Wolf (2001: 114) by highlighting the variety present within a single genre and incorporating interactivity into the classification of video games. Nevertheless, further research on a more extensive sample of RPG dialogue could provide a more detailed examination of the variety within the genre, which could be used for establishing different sub-categorisations for RPGs.

The findings of the analysis can also be interpreted in relation to the different descriptions of video game dialogue in prior research. Mäyra (2017: 283) described RPG dialogue as refined in comparison to spoken language in TTRPGs, emphasising the lack of

spontaneous production for RPG dialogue. The findings of the current analysis provide a similar description, whilst noting the mutual influence of spoken and written registers. Such description contrasts with Dixon's (2024b: 18), who emphasises the similarity between oral registers and RPG dialogue. Although the origin of the RPG genre is in fantasy fiction, as established by Burn (2023: 319) and Apperley (2006: 17), its influence on the dialogue could only be noted on the third dimension, where the sample was assigned the label of "General Fiction", further indicating that video games include a register that is modified to the goals and functions of the medium. Mäyra's (2017: 288) description of RPG language in relation to both fictional and interactional registers, however, can be confirmed by the results of the analysis, the personal letters register, which is interpreted as a combination of oral and written registers, appeared on two dimensions.

Overall, previous descriptions of video games emphasise the interactivity of the medium (Peterson and Jabbari 2003: 1; Dixon 2023: 1). Although this interactivity not associated with the dialogue, results on the first dimension indicate that some level of interactive communication can occur in RPG dialogue, but differences appear between individual games. Although Bianchi (2024: 260) noted the similarity between spoken media and video game dialogue on the lexical level, similar results appeared only on the fifth dimension, which described the use of abstract and non-abstract information.

Regarding the differences between the subgenres of RPGs, Bycer (2024: 39–40) and Domsch (2017: 265) noted that the player's involvement in the narrative is an indicator used to distinguish JRPGs and WRPGs. As significant differences did not emerge between these subcategories, this description cannot be confirmed in the current analysis. Additionally, Dixon (2023: 112) and Domsch (2017: 263–264) list narrative development as a significant function for RPG dialogue and general video game dialogue. This narrative function of the dialogue did not appear in the findings, as the data was categorised as mostly non-narrative on D2. This

indicates that narrative establishment occurs in dialogue without the use of explicit narratives and linguistic features associated with it.

The additive MDA results provide answers to the research questions established in the thesis. In relation to the first research question, which focused on RPG dialogue and its similarity to written and spoken text, the results on the first, second and fifth dimensions seem to indicate a resemblance between spoken dialogue and RPG dialogue. However, as the sample was not assigned any oral registers on the first two dimensions, text in the RPG dialogue cannot be considered entirely identical to spoken dialogue, which is expected as it is not spontaneously produced. Furthermore, the results for the third, fourth, and sixth dimensions also establish that RPG dialogue is similar to written texts, indicating the combined influence of both written and spoken texts on RPG dialogue. Based on the qualitative analysis, results on the first dimension classify RPG dialogue register as texts, which are written but share characteristics with many spoken texts. Results for the second dimension further confirm this similarity, as RPG dialogue was classified as non-narrative text, which also characterises spoken conversational texts. Still, Bianchi's (2024: 260) observation on the genre-internal variation was confirmed by the wide range of dimensional scores on the first dimension.

The results also provide insight into the second research question regarding the influence of different subgenres, such as age ratings and time periods in the dialogue of RPGs, highlighting that subgenres do not have a significant effect on the linguistic characteristics of dialogue in RPGs. The final research question, which focused on providing a description of RPG dialogue, was answered with the results of the analysis, where several conversational linguistic features were observed in RPG dialogue, which indicate its lack of a narrative and involved and affective focus. The analysis also confirmed the first hypothesis, as the dialogue was generally classified as written language.

The results of the analysis can be compared to findings in previous research. Despite the lack of similar studies that could be used for comparison, as the application of the additive MDA approach to RPG data remains relatively unexplored, results for the first dimension can be compared to Dixon's (2024b) results and, therefore, develop Dixon's description of the RPG dialogue register further, highlighting the impact of the written registers on RPG dialogue. Overall, the findings show a general similarity to the results observed in Dixon (2024b), as RPG dialogue text was described as similar to conversational texts on D1. However, the current analysis highlights the significance of the personal letters register on D1, which indicates the mutual influences of the written and spoken modes of production in the RPG dialogue. Furthermore, in contrast to Dixon's (2024b) results, VGDC includes a significant variety appearing across the first dimension for RPG dialogue, suggesting some difference in the language use of dialogue across different games. This result is consistent with Bianchi (2024: 260), whose lexical coverage analysis results suggest that significant differences appear in the language use of different video games, which may not be related to game genre.

The analysis also provided new information about the contents of RPG dialogue for the remaining dimensions of Biber's MDA framework. Although Dixon's (2024b) results establish RPG dialogue as similar to spoken conversation, analysis on six dimensions indicates that RPG dialogue includes linguistic characteristics from both written and spoken registers. Contrary to expectations, the VGDC sample was assigned mostly written registers in the analysis process, although the influence of spoken modes of communication also appeared in the written register of personal letters. The results can also be interpreted from the perspective of SLA and used to give an overview of the language included in video games, and more specifically, RPG video games. Based on the results, the players of RPG games will likely experience dialogue that has mostly conversational characteristics as L2 input.

2.4.4. Limitations and Recommendations for Future Research

While the MDA analysis of RPG dialogue in VGDC provides valuable information about RPG dialogue, it is important to acknowledge several limitations. Although the sample size for the current analysis is more extensive than those used in previous similar studies and claimed as representative of the RPG genre by the authors of the corpus, it cannot be considered representative of all RPG games, as the genre is constantly developing and new games are released. Furthermore, the results for the current analysis confirm the suggestions from previous research, indicating that significant variety occurs within the RPG genre. Therefore, it is recommended to increase the sample size in the following studies to remain aligned with the innovation within the genre and to provide a better overview of the variation within the RPG dialogue register. Furthermore, more specific qualitative analyses should be conducted for the remaining four dimensions, which were not addressed in the current analysis, to understand how RPG dialogue is situated on those dimensions and how their linguistic content affects the function of the register.

Despite these limitations, the findings of this thesis contribute to the currently developing field of video game language research. As the current state of research concerning the texts included in video games is scarce, the current findings can be used as a basis for conducting comparative research on other registers related to video games or video game dialogue from other game genres. The analysis conducted in this thesis is significant, as previous studies with a similar focus have used substantially narrower samples of data; therefore, the current study expands the findings of previous research. The analysis also provides new information concerning the RPG dialogue register and Biber's (1988) framework of MDA, as previous studies have only described this register on the first dimension. Furthermore, as the additive MDA process has not been previously applied to the VGDC, the current findings can serve as a basis for describing its linguistic content. Regarding SLA, the

current thesis has highlighted the use of specific linguistic features in RPG dialogue, the knowledge of which is essential for incorporating video games into SLA studies.

Further research can expand on the results of the current analysis by conducting additive MDA on other texts included within RPG video games, such as fictional narrative lore texts in RPG video games. The differences between sub-categories could be examined in greater detail, and an analysis could be conducted on how the characteristics of video game dialogue differ for different sub-categories and how that affects the function of dialogue in categories related to age rating, for example. Further MDA research could also be conducted on video game data from other genres to encourage their further incorporation into DGBLL and ludolinguistic research. Additionally, analysis and interpretation of text types (based on the dimensional scores across the six dimensions) are recommended to understand further how the register is situated regarding the established text registers in English. It would be beneficial for subsequent MDA studies on video game data to conduct an additive MDA on the immersive and interactive dialogue in VGDC separately or conduct a full MDA using the same VGDC sample, comparing the results to Dixon's (2022b) and extending them, if possible. Compiling further video game orthogame data corpora and using them in linguistic research is also recommended for establishing a better understanding of language use in video games.

CONCLUSION

The globally rising player counts and a player population including people from all ages and nationalities confirm the increasing significance of video games as a form of entertainment. However, with the continuing advancement of technology, video games have evolved from simple forms of entertainment to complex, immersive experiences that engage players on multiple levels. Language is a significant aspect within video games, where it functions as one way of communication between players and the computer, or between several players. Additionally, language can also be used for storytelling, character development and world-building, functions which are realised through video game dialogue. For many players across the world, video games are experienced in a second language, making them sources of L2 input. However, to realise the potential of video games as tools for language education, the educators must be familiar with the linguistic content of various video games. Therefore, the aim of the thesis was to analyse the language of video game dialogue, more specifically RPG dialogue, where dialogue serves an important role, to understand the linguistic characteristics of video game dialogue. To fulfil that aim, the method of additive multidimensional analysis was used, which provides a detailed description of the linguistic characteristics of RPG dialogue and highlights the functionality of RPG dialogue.

Three research questions were established for the thesis, which aimed to establish a detailed description of RPG dialogue. The first research question focused on describing RPG dialogue in relation to spoken and written registers, hypothesising that the written mode of production for the dialogue will affect its goals of appearing conversational and spontaneous. The second research question addressed the possible influence of several sub-categories (age rating, time period, sub-genre) on the RPG dialogue, and the final research question concerned the description of prevalent linguistic characteristics in RPG dialogue.

The first chapter of the thesis focuses on providing a description of the object of the analysis and an overview of video game research. In the first section, the definition for RPGs is established by examining proposed definitions for games, video games and role-playing video games. Video games are established for the thesis as a type of game with technological aspects for receiving the player's input and outputting information to the player. More specifically, role-playing games are defined as video games in which dialogue and text play a significant role in gameplay. The second section of the first chapter focuses on linguistic research conducted on video game data. The significance of dialogue within video games is established as it can be considered a central game mechanic in video games and particularly in RPGs. Following that, two approaches to video game language, ludolinguistics and Digital Game-Based Language Learning, are examined, concluding that the current state of video game language research is characterised by a lack of orthogame data-based research.

The second chapter of the thesis focuses on the imperial analysis of video game data in VGDC using the additive MDA approach. The first section of the second chapter provides a description of the methodology and method of MDA used in the analysis. The similarities between register analysis and MDA methods are also introduced. Following the description of the method, the central concepts within MDA, such as dimension and linguistic co-occurrence, are examined. The process of conducting MDA is described in detail, including the differences between the two approaches to MDA: the full and the additive MDA. Overall, additive MDA is chosen as a suitable method for conducting a linguistic analysis of VGDC and MAT software is introduced as an accessible and reliable option for conducting additive MDA. The following section describes the existing video game corpora and their compilation process, concluding that the interest in video game corpora is increasing, but is inhibited by their technically complex creation process. Additionally, the VGDC corpora and their suitability for analysis are examined. The VGDC is the largest publicly available video game dialogue corpus, which

features 50 role-playing video games from two sub-genres, different age ratings and time periods.

The data analysis process and results are provided in the third section of the second chapter. The section begins with an overview of the data preparation process, where the VGDC data was processed to remove features irrelevant for the current analysis. This is followed by a description of the analysis process, which was conducted using the MAT software. The section concludes with the results of the analysis, which are reported in multiple stages: first, general results for the full sample, then results for the different subcategories of data, and finally, the results for Biber's six dimensions of variation. The second chapter concludes with the discussion section, which features the qualitative analysis of the first two dimensions in Biber's framework, the discussion of the results, the limitations of the analysis and recommendations for future research.

The method used for the analysis of RPG dialogue in the current thesis was additive multidimensional analysis. The method of MDA allows examining the use of multiple co-occurring linguistic features simultaneously as they are interpreted as indicators of shared functions of language in a specific register. The multiple linguistic features observed are grouped into dimensions of variation using factor analysis, which finds frequently co-occurring linguistic features. These dimensions are then interpreted as indicators of specific communicative functions. MDA can be further divided into full and additive MDA approaches, which differ in the extent to which conducting factor analysis independently is required. Additive MDA was determined to be a suitable method for providing descriptions of RPG dialogue in the current thesis. In the thesis, the additive MDA process was conducted on data from the Video Game Dialogue Corpus (Rennick et al. 2023a; Rennick and Roberts 2024), which features the dialogue of 50 role-playing games. Prior to the analysis, the data was downloaded from an online repository and processed to remove inconsistencies and irrelevant

features. Additive MDA was conducted using the MAT software, which is publicly available and suitable for MDA.

The findings of the analysis were reported in several stages. First, the general results for the VGDC sample were reported. The VGDC sample was assigned mostly written registers as the closest registers across the six dimensions. Results for the first dimension characterise data in VGDC as involved production, indicating a resemblance to spoken language. For the second dimension, the dimensional scores indicated a lack of narrative production in VGDC, which further confirms its connection to spoken registers. However, dimensional scores on the third dimension signify the use of explicit reference in the data of VGDC, which highlights the connection between RPG dialogue and written registers. For the fourth dimension, VGDC was marked as moderately persuasive, and the dimensional score for the fifth dimension describes VGDC as including a mixture of non-abstract and abstract information, and tendencies towards non-abstract information, which is similar to spoken registers. The sixth dimension marks the VGDC as text produced without time constraints, indicating a resemblance to written registers. Based on these results, the first research question can be answered. The dialogue in the VGDC includes language that can be characterised as a mixture of spoken and written registers. Such results also reflect the production methods of video game dialogue and its goals within the game.

To answer the second research question, the MDA was conducted on several categorisations of VGDC, including time period, age rating, sub-genre, and game series. The results indicate that although occasional variations were observed in the dimensional scores and assigned closest registers for the different sub-categories, no significant differences appeared in the MDA analyses of different data categories. This indicates that RPG dialogue functions as a distinct register, with its own characteristics and specific functions.

Characterised as a mixture of spoken and written registers, the dialogue of VGDC also includes linguistic features, which are significant to both modes of production. The analysis thus also provided an answer to the third research question. Regarding the resemblance to spoken registers, VGDC includes a higher frequency of present-tense verbs, attributive adjectives, private verbs, contractions, and a limited lexical diversity compared to academic texts. In contrast, characteristics connecting VGDC to written registers include a higher frequency of phrasal coordination, passive voice, and time and place adverbials.

Further qualitative analysis was also conducted on the results of the first and second dimensions, as these dimensions had unexpected results. For the first dimension, the range of the dimensional scores for VGDC was quite extensive compared to the other dimensions. This indicates a variation occurring within the dataset, which can be interpreted as variety occurring within the register of RPG dialogue. For the second dimension, several categorisations of the data were assigned conversational registers, contrasting with the academic prose register assigned to the full sample. Closer examination revealed similar functions of the conversational and academic registers on the second dimension, as both are marked as non-narrative registers. Differences in these registers are therefore only related to the degrees of non-narrative concern featured in the data.

Based on the analysis of RPG dialogue in VGDC, several conclusions can be drawn. It appears that RPG dialogue cannot be concretely characterised as either spoken or written register, but rather a combination of the two. Both the production circumstances and functional goals of the RPG dialogue emerged in the description of the register. While the subcategories appeared to have limited impact on the language use in RPG dialogue, some variation is still evident within the sample, suggesting that linguistic diversity may still occur within a highly specific register.

While the current analysis of RPG dialogue in VGDC offers valuable insights, it has limitations, specifically regarding the representativeness of the data due to the constantly evolving nature of the RPG genre and the high degree of variety present within the genre. Additionally, the absence of qualitative analysis for several dimensions limits the precision of the description provided for the RPG dialogue data in VGDC. Although subject to these limitations, the findings of this thesis contribute to the linguistic research of video game language by providing a description of RPG dialogue based on the analysis of an extensive video game corpus. The findings of the thesis can contribute to future research as a basis of comparison for analysing different RPG dialogue datasets or for analysing dialogue in other game genres. Additionally, the current analysis develops the findings of previous MDA research conducted by Dixon (2022b) by providing additional information concerning the RPG dialogue on Biber's (1988) MDA framework. Considering these limitations, future research would benefit from further expansion of the RPG dialogue dataset, analysis of other specific registers in video games, and further analysis of differences between sub-categories. Furthermore, additional research should incorporate other video game genres, which currently remain unaddressed. Overall, these limitations highlight the necessity of creating further accessible and large-scale orthogame corpora, which could be used in linguistic research.

Although video games are primarily used for entertainment purposes, it is necessary to acknowledge their hidden educational potential. The immense joy of the playing process, intricate adventures in immersive game worlds and the supportive and collaborative social networks of players across the world can be combined with other benefits, such as foreign language acquisition. The dialogue of a video game, whilst functioning as a window into the mechanics of the game, simultaneously immerses the player into the narrative world of each video game. Thus, the dialogue fosters a deep connection between the player and the game. As the medium of video games evolves, the linguistic analysis of game texts and language will

remain essential for understanding how players understand and interact with the digital world and what further benefits could be gained from these interactions.

REFERENCES

Aarseth, Espen. 2023. Ludology. In Mark J.P. Wolf and Bernard Perron (eds.) *The Routledge Companion to Video Game Studies.* 255-260, New York and London: Routledge.

Apperley, Thomas H. 2006. Genre and game studies: Toward a critical Approach to video game genres. *Simulation & Gaming,* 37: 1, 6-23.

Arsenault, Dominic. 2009. Video Game Genre, Evolution, and Innovation. *Eludamos. Journal for Computer Game Culture,* 3: 2, 149-176.

Arsenault, Dominic. 2023. Narratology. In Mark J.P Wolf and Bernard Perron (eds.) *The Routledge Companion to Video Game Studies.* 588-597, New York and London: Routledge.

Berber Sardinha, Tony and Marcia Veirano Pinto. 2019. Introduction. In Tony Berber Sardinha and Marcia Veirano Pinto (eds.) *Multi-Dimensional Analysis: Research Methods and Current Issues.* 1-11, London and New York: Bloomsbury Academic.

Berber Sardinha, Tony, Veirano Pinto, Marcia, Mayer, Cristina, Zuppardi, Maria Carolina and Carlos Henrique Kauffmann. 2019. Adding Registers to a Previous Multi-dimensional Analysis. In Tony Berber Sardinha and Marcia Veirano Pinto (eds.) *Multi-Dimensional Analysis: Research Methods and Current Issues.* 165-189, London and New York: Bloomsbury Academic.

Berns, Anke and Iván Ruiz-Rube. 2023. Use of Digital Game Creation Tools in Computer Assisted Language Learning: A Systematic Literature Review. In Mark Peterson and Nasser Jabbari (eds.) *Digital Games in Language Learning: Case Studies and Applications.* 35-52, Abington, Oxon; New York, NY: Routledge.

Bianchi, Francesca. 2024. Video Game In-Game Dialogue. An Analysis of Lexical Coverage. *Lingue Linguaggi,* 66, 241-266.

Biber, Douglas. 1988. *Variation Across Speech and Writing.* Cambridge: Cambridge University Press.

Biber, Douglas. 1989. A Typology of English Texts. *Linguistics,* 27: 1, 3-43.

Biber, Douglas. 1995. *Dimensions of register variation: A cross-linguistic comparison.* Cambridge: Cambridge University Press.

Biber, Douglas and Susan Conrad. 2019. *Register, Genre, and Style*. 2nd ed. Cambridge, NY, VIC, New Delhi: Cambridge University Press.

Biber, Douglas and Jesse Egbert. 2016. Register Variation on the Searchable Web: A Multi-Dimensional Analysis. *Journal of English Linguistics,* 44(2), 95-137.

Burghardt, Manuel, and Jochen Tiepmar. 2021. The Game Walkthrough Corpus (GWTC) – A Resource for the Analysis of Textual Game Descriptions. *Journal of Open Humanities Data*, 7, 14.

Burn, Andrew. 2023. Role-Playing. In Mark J.P. Wolf and Bernard Perron (eds.) *The Routledge Companion to Video Game Studies*. 318-329, New York and London: Routledge.

Burn, Andrew and Diane Carr. 2006. Defining Game Genres. In Diane Carr, David Buckingham, Andrew Burn, and Gareth Scott (eds) *Computer Games: text, narrative and Play,* Malden and Cambridge: Polity Press, 14-29.

Brugman, Britta C., Burgers, Christian, Beukeboom Camiel J., and Elly A Konjin. 2020. Satirical news from left to right: Discursive integration in written online satire. *Journalism,* 23:8, 1626-1644.

Bycer, Joshua. 2024. *Game design deep dive: Role-playing games.* Boca Raton, FL: CRC Press.

Caillois, Roger, and Meyer Barash. 2001. *Man, play, and games*. Urbana: University of Illinois Press.

Carr, Diane. 2006. Games and Narrative. In Diane Carr, David Buckingham, Andrew Burn, and Gareth Scott (eds) *Computer Games: text, narrative and play,* Malden and Cambridge: Polity Press, 30-45.

Carter, Marcus, Gibbs, Martin, and Mitchell Harrop. 2012. Metagames, paragames and orthogames: a new vocabulary. In *Proceedings of the International Conference on the Foundations of Digital Games (FDG'12).* Association for Computing Machinery, New York, NY, USA, 11-17.

Clarke, Isobelle. 2022. A Multi-dimensional analysis of English tweets. *Language and Literature,* 31: 2, 124-149.

Clarke, Rachel Ivy, Ha Lee, Jin and Neils Clark. 2015. Why Video Game Genres Fail: A Classificatory Analysis, *Games and Culture,*12:5, 1-21.

Clearwater, David. 2011. What Defines Video Game Genre? Thinking about Genre Study after the Great Divide? *Loading… The Journal of the Canadian Game Studies Association,* 5:8, 29-49.

Cooper, Christopher R. 2023. The identification of YouTube videos that feature the linguistic features of English informal speech. *Applied Corpus Linguistics,* 3:3.

Cornillie Frederik. 2022. Digital Games and Technology-Mediated Gameful Environments for L2 Learning and Instruction. In Nicole Ziegler and Marta González-Lloret (eds) *The Routledge Handbook of Second Language Acquisition and Technology,* 272-285. New York: Routledge.

Crawford, Chris. 2003. *Chris Crawford on game design*. Indianapolis, Ind: New Riders.

Dixon, Daniel H. 2022a. The Linguistic Environments of Digital Games: A Discriminant Analysis of Language Use in Game Mechanics. *CALICO Journal*, 39:2, 150-171.

Dixon, Daniel H. 2022b. *The Language in Digital Games: Register Variation in Virtual and Real-World Contexts.* Doctoral Dissertation, Northern Arizona University.

Dixon, Daniel H, Dixon Tülay, and Eric Jordan. 2022. Second Language (L2) gains through digital game-based language learning (DGBLL): A Meta-analysis. *Language Learning & Technology,* 26:1, 1-25.

Dixon, Daniel H. 2023. A Methodological Framework for Analysing the Language in Digital Games. In Mark Peterson and Nasser Jabbari (eds.) *Digital Games in Language Learning: Case Studies and Applications.* 52-74, Abington, Oxon; New York, NY: Routledge.

Dixon, Daniel H. 2024a. Introducing the Single Player Offline Game Corpus (SPOC): A Corpus of seven registers from digital role-playing games. *Corpora*, 19:1, 107-122.

Dixon, Daniel H. 2024b. Measuring the linguistic similarity of discourse from open-world role-playing games to the real world through an additive multidimensional analysis. *Register studies*, 6:1, 1-30.

Domsch, Sebastian. 2017. Dialogue in Video Games. *Dialogue Studies,* 28, 251-271.

Entertainment Software Association. 2024. *2024 essential facts about the U.S. video game industry*. Available at https://www.theesa.com/wp-content/uploads/2024/05/Essential-Facts-2024-FINAL.pdf, accessed March 2025.

Erdur, Nilüfer. 2022. Gender in Genshin Impact: A Corpus-Assisted Discourse Analysis. *Language Education and Technology*, 2(1), 74-93.

Esposito, Nicolas. 2005. A short and Simple Definition of What a Videogame Is. *Proceedings of DiGRA 2005 Conference: Changing Views: Worlds in Play,* https://doi.org/10.26503/dl.v2005i1.177, accessed March 2025.

Frasca, Gonzalo. 2003. Simulation versus Narrative: Introduction to Ludology. In Mark J.P. Wolf and Bernard Perron (eds.) *The Video Game Theory Reader.* 211-237, London and New York: Routledge.

Granic, Isabela, Adam Lobel, and Rutget C.M.E. Engels. 2014. The Benefits of Playing Video Games. *American Psychologist*, 69:1, 66-78.

Grieve, Jack and Helena Woodfield. 2023. *The Language of Fake News.* Cambridge: Cambridge University Press.

Grieve, Jack. 2023. Register variation explains stylometric authorship analysis. *Corpus Linguistics and Linguistic Theory,* 19: 1, 47-77.

Heritage, Frazer. 2020. Applying corpus linguistics to videogame data: Exploring the representation of gender in videogames at a lexical level. *The International Journal of Computer Game Research*, 20: 3.

Heritage, Frazer. 2021. Language, Gender, and Videogames. Switzerland: Palgrave-Macmillan.

Heritage, Frazer. 2022. Magical women: Representations of female characters in the Witcher video game series. *Discourse, Context & Media*, 49: 100627.

Heritage, Frazer. 2025. "Plow Her Well. Show Her You're a Man": Language, Sex, and Heteronormativity from a Diachronic Perspective in *The Witcher* Video Game Series. In Matthew Wysocki and Steffi Shook (eds) *The Bloomsbury Handbook of Sex and Sexuality in Game Studies.* New York, London, Dublin: Bloomsbury Academic, 207-220.

Ho, Don. 2025. *Notepad++*, version 8.7.5. [Computer Software].

Hämäläinen, Mika, Khalid Alnajjar, and Thierry Poibeau. 2022. Video Games as a Corpus: Sentiment Analysis using Fallout New Vegas Dialog. In Proceedings of *the 17th International Conference on the Foundations of Digital Games (FDG'22).* Association for Computing Machinery, New York, USA, Article 56, 1–4.

Jaagola, Karl. 2016. Narrative in Mass Effect 3. MA thesis. University of Tartu, Department of English Studies.

Juraska, Juraj, Kevin Bowden, and Marilyn Walker. 2019. Viggo: A Video Game Corpus for Data-to-Text Generation in Open-Domain Conversation. *Proceedings of the 12th International Conference on Natural Language* https://doi.org/10.18653/v1/w19-8623.

Juul, Jesper. 2003. The Game, the Player, the World: Looking for a Heart of Gameness. In *Level Up: Digital Games Research Conference Proceedings*, edited by Marinka Copier and Joost Raessens, 30-45. Utrecht: Utrecht University, 2003. Available at https://www.jesperjuul.net/text/gameplayerworld/, accessed March 2025.

Lazzeretti, Cecilia and Maria Cristina Gatti. 2023. Dialogic interaction between player and non-player characters in Animal Crossing: a corpus-based study. *Language and Dialogue*, 13: 1, 81-102.

Li, Juan. 2019. *The Role of Massively multiplayer role-playing games in facilitating vocabulary acquisition for English language learners: A mixed-methods study.* Doctoral Dissertation, Clemson University.

Lifländer, Loore. 2020. An analysis of Estonian gaming streamers' English-Estonian code switching. MA thesis. University of Tartu, Department of English Studies.

Lowood, Henry and Raiford Guins (eds). 2016. *Debugging Game History: A Critical Lexicon*, Cambridge, London: The MIT Press.

Kantar Emor. 2022. Eesti alaealiste elanike kokkupuuted hasart- ja videomängudega 2022. Aruanne. Available at https://15410.ee/wp-content/uploads/2023/03/Eesti-alaealiste-elanike-kokkupuuted-hasart-ja-videomangudega-2022.pdf, accessed March 2025.

Kuhn, Jeff. 2017. Minecraft: Education Edition. *CALICO Journal,* 35: 2, 214-223.

Mehine, Indra. 2021. Motivation and study skills in learning English vocabulary through computer games in Estonia. MA thesis. University of Tartu, Department of English Studies.

Meier, Marie-Luise. 2022. Gender in Games – methodology and analysis. Doctoral Dissertation. University of Tartu, Department of English Studies. Tartu: University of Tartu Press.

Microsoft. 2023. *Visual Studio Code,* version 1.92.0.[Computer software].

Mägi Ravn, Morten. 2021. Gamification in education: practising English phrasal verbs using a gamified activity. MA thesis. University of Tartu, Department of English Studies.

Mäyrä, Frans. 2017. Dialogue and Interaction in Role-Playing Games. *Dialogue Studies*, 28: 271.

Newman, James. 2004. *Videogames*. London: Routledge.

Nini, Andrea. 2019. The Multi-Dimensional Analysis Tagger. In Berber Sardinha, Tony and Veirano Pinto, Marcia (eds.), *Multi-dimensional Analysis: Research Methods and Current Issues,* 67-94, London; New York: Bloomsbury Academic.

Nini, Andrea. 2020. *Multidimensional Analysis Tagger (v. 1.3) – Manual.* Available at https://sites.google.com/site/multidimensionaltagger/home, accessed February 28, 2025.

Perron, Bernard. 2023. Conventions. In Mark J.P. Wolf and Bernard Perron (eds.) *The Routledge Companion to Video Game Studies*. 145-154, New York and London: Routledge.

Peterson, Mark and Nasser Jabbari. 2023. Digital Games and Foreign Language Learning: Context and Future Development. In Mark Peterson and Nasser Jabbari (eds.) *Digital Games in Language Learning: Case Studies and Applications.* 1-14, Abington, Oxon; New York, NY: Routledge.

Pugh, Tison and Ramey, Lynn. 2022. Introduction: Ludology, Narratology, and Teaching Literary Games. In Tison Pugh and Lynn Ramey (eds.) *Teaching Games and Game Studies in the Literature Classroom.* 1-25, London, New York, Dublin: Bloomsbury Academic.

Quaglio, Paulo. 2009. Television Dialogue: The sitcom *Friends* vs. natural conversation. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Reinhardt, Jonathon. 2019. *Gameful Second and Foreign Language Teaching and Learning: Theory, Research and Practice.* Cham: Palgrave Macmillan.

Reinhardt, Jonathon. 2021. Not all MMOGs are created equal: A design-informed approach to the study of L2 learning in multiplayer online games. In Mark Peterson, Kasumi Yamazaki and Michael Thomas (eds.) *Digital Games and Language Learning: Theory, Development and Implementation.* 69-89, London, New York, Oxford, New Delhi, Sydney: Bloomsbury Academic.

Rennick, Stephanie, Clinton, Melanie, Ioannidou, Elena, Oh, Liana, Clooney, Charlotte, T.E., Healy, Edward, & Roberts, Seán. G. 2023a. Gender bias in video game dialogue. *Royal Society Open Science*, 10:5.

Rennick, Stephanie, Clinton, Melanie, Ioannidou, Elena, Oh, Liana, Clooney, Charlotte, T.E., Healy, Edward, & Roberts, Seán. G. 2023b. Gender bias in video game dialogue: Supplementary materials. *Royal Society Open Science*, 10:5.

Rennick, Stephanie and Roberts, Seán G. 2024. The video game dialogue corpus. *Corpora*, 19:1, 93-106.

Rodendau, Elis. 2017. Using RuneScape for language acquisition purposes in an upper secondary EFL classroom. MA thesis. University of Tartu, Department of English Studies.

Rodgers, Michael P. H. and Julian Heidt. 2021. Levelling up comprehensible input and vocabulary learning: the lexical profile of videogames. In Valentin Werner and

Friederike Tagge (eds) *Pop Culture in Language Education. Theory, Research, Practice,* 215-227. Oxon and New York: Routledge.

Rätsep, Hannaleena. 2022. The analysis of clippings in Twitch live stream chat rooms. BA thesis. University of Tartu, Department of English Studies.

Saks, Karl Erik. 2017. An analysis of verbs within video game structures based on a video game verb theory and The Secret of Monkey Island. MA thesis, University of Tartu, Department of English Studies.

Toutanova, Kristina, Klein, Dan, Manning, Christopher and Yoram Singer. 2003. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In *Proceedings of HLT-NAACL 2003,* 252-259.

van Stegeren, Judith and Mariët Theune. 2021. Fantastic Strings and Where to Find Them: The Quest for High-Quality Video Game Text Corpora. In J.C. Osborn (ed.) *AIIDE 2020 Workshops: Joint Proceedings of the AIIDE 2020 Workshops co-located with 16th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE 2020)*. CEUR Workshop Proceedings, vol. 2862.

Werner, Valentin. 2021. Catchy and conversational? A register analysis of pop lyrics. *Corpora,* 16:2, 237-270.

Wolf, Mark J.P. 2001. Genre and the Video Game. In Mark J. P. Wolf (ed) *The Medium of the Video Game.* 113-135. Austin: University of Texas Press.

Wolf, Mark. J.P. 2008. The Study of Video Games. In Mark J.P. Wolf (ed) *The Video Game Explosion: A History from PONG to PlayStation and Beyond*. 21-28. Westport, Connecticut, London: Greenwood Press.

Wolf, Mark J.P and Bernard Perron (eds). 2023. *The Routledge Companion to Video Game Studies*, New York and London: Routledge.

Zagal, José P. and Sebastian Deterding. 2018. Definitions of "Role-Playing Games". In José P. Zagal and Sebastian Deterding (eds) *Role-Playing Game Studies: Transmedia Foundations*. 19-53. New York, Oxfordshire: Routledge.

ZA/UM. 2019. *Disco Elysium.* [Computer Game].

Appendix 1: The List of Role-Playing Games Included in VGDC

The List of Role-Playing Games Included in VGDC (adapted from Rennick and Roberts 2024: 100-101; Rennick et al. 2023b: 7)

Title	Subgenre	Age Rating	Time period	Length (words)
Stardew Valley	WRPG	Child	2000-2020	53870
King's Quest I: Quest for the Crown	WRPG	Child	1985-1999	1396
King's Quest II: Romancing the Throne	WRPG	Child	1985-1999	402
King's Quest III: To Heir is Human	WRPG	Child	1985-1999	3151
King's Quest IV: The Perils of Rosella	WRPG	Child	1985-1999	1772
King's Quest V: Absence Makes the Heart Go Yonder!	WRPG	Child	1985-1999	10833
King's Quest VI: Heir Today, Gone Tomorrow	WRPG	Child	1985-1999	72074
King's Quest VII: The Princeless Bride	WRPG	Child	1985-1999	21358
King's Quest VIII: Mask of Eternity	WRPG	Child	1985-1999	15880
King's Quest Chapters	WRPG	Child	2000-2020	113 053
Monkey Island 2: LeChuck's Revenge	WRPG	Child	1985-1999	9285

The Curse of Monkey Island	WPRG	Child	1985-1999	30577
The Secret of Monkey Island	WRPG	Child	1985-1999	14619
Super Mario RPG: Legend of the Seven Stars	JRPG	Child	1985-1999	13755
Kingdom Hearts	JRPG	Child	2000-2020	19847
Kingdom Hearts II	JRPG	Child	2000-2020	47843
Kingdom Hearts III	JRPG	Child	2000-2020	40808
Kingdom Hearts 3D: Dream Drop Distance	JRPG	Child	2000-2020	18484
Horizon Zero Dawn	WRPG	Teen	2000-2020	50996
Star Wars: Knights of the Old Republic	WRPG	Teen	2000-2020	439667
Chrono Trigger	JRPG	Teen	1985-1999	37982
Final Fantasy	JRPG	Teen	1985-1999	2763
Final Fantasy II	JRPG	Teen	1985-1999	8689
Final Fantasy IV	JRPG	Teen	2000-2020	17822
Final Fantasy V	JRPG	Teen	1985-1999	12408
Final Fantasy VI	JRPG	Teen	1985-1999	14750
Final Fantasy VII	JRPG	Teen	1985-1999	100584
Final Fantasy VII Remake	JRPG	Teen	2000-2020	86487
Final Fantasy VIII	JRPG	Teen	1985-1999	51368
Final Fantasy IX	JRPG	Teen	2000-2020	99245
Final Fantasy X	JRPG	Teen	2000-2020	84290

Final Fantasy X-2	JRPG	Teen	2000-2020	31635
Final Fantasy XII	JRPG	Teen	2000-2020	128742
Final Fantasy XIII	JRPG	Teen	2000-2020	12674
Final Fantasy XIII-2	JRPG	Teen	2000-2020	68922
Lightning Returns: Final Fantasy XIII	JRPG	Teen	2000-2020	47916
Final Fantasy XIV	JRPG	Teen	2000-2020	689360
Final Fantasy XV	JRPG	Teen	2000-2020	74792
Mass Effect	WRPG	Adult	2000-2020	385744
Mass Effect 2	WRPG	Adult	2000-2020	270209
Mass Effect 3	WRPG	Adult	2000-2020	361358
Elder Scrolls: Daggerfall	WRPG	Adult	1985-1999	43855
Elder Scrolls: Morrowind	WRPG	Adult	2000-2020	173666
Elder Scrolls: Oblivion	WRPG	Adult	2000-2020	777177
Elder Scrolls: Skyrim	WRPG	Adult	2000-2020	150041
Dragon Age 2	WRPG	Adult	2000-2020	281116
Dragon Age: Origins	WRPG	Adult	2000-2020	701258
Persona 3	JRPG	Adult	2000-2020	44447
Persona 4	JRPG	Adult	2000-2020	156753
Persona 5	JRPG	Adult	2000-2020	385817

Appendix 2: Dimensional Scores for VGDC across Six Dimensions

Subcategory	D1	D2	D3	D4	D5	D6	Closest Text Type
VGDC	15.3	-1.63	-2.24	2.64	-1.42	-0.42	Involved Persuasion
JRPG	16.86	-1.58	-2.59	2.68	-1.27	-0.33	Involved persuasion
WRPG	14.61	-1.69	-1.85	2.58	-1.58	-0.53	Involved persuasion
Adult	18.28	-1.27	-1.69	4.17	-1.25	-0.5	Informational interaction
Child	15.02	-2.08	-2.45	1.31	-1.72	-0.41	Involved persuasion
Teen	14.97	-1.45	-2.38	2.9	-1.26	-0.39	Involved persuasion
1985-1999	12.28	-1.88	-2.39	1.64	-1.51	-0.38	Involved persuasion
2000-2020	18.31	-1.45	-2.13	3.36	-1.36	-0.46	Informational interaction
Closest Register	Personal Letters	Conversa-tions	General Fiction	Personal Letters	Broad-Casts	Official Documents	NA
Mean	15.3	-1.63	-2.24	2.64	1.42	-0.42	Involved persuasion
Max	23.96	0.01	0.75	6.78	0.08	0.88	NA
Min	-1.16	-1.63	-2.24	2.64	-1.42	-0.42	NA

RESÜMEE

TARTU ÜLIKOOL
ANGLISTIKA OSAKOND

Janely Rüdein

Exploring Dialogue in RPG Video Games: An Additive Multi-Dimensional Approach
Dialoogi uurimine rollimängude videomängudes: aditiivne mitmemõõtmeline lähenemine

Magistritöö
2025
Lehekülgede arv: 98

Annotatsioon:

Kuigi videomängud on mängijate meelt lahutanud juba pikemat aega, on videomängude keele uuringud alles arenev uurimisvaldkond. Käesolev magistritöö uurib keelekasutust rollimängužanri videomängude (RPG) dialoogis kasutades aditiivse mitmemõõtelise analüüsi meetodit. Magistritöö eesmärkideks on kirjeldada RPGde dialoogi keelt: tuua välja videomängudialoogile omaseid keelelisi tunnuseid, mängukeele sarnasust kirjalikule ja suulisele keelele ning videomängude erinevate aspektide (ajaperiood, alamžanr, vanusepiirang) võimalikku mõju mängudialoogile.

Püstitatud eesmärkide saavutamiseks kasutati aditiivse mitmemõõtmelise analüüsi meetodit Video Game Dialogue Corpuse (VGDC) andmetel. Antud korpus on töö kirjutamise hetkel suurim avalikult ligipääsetav videomängudialoogi korpus, kuhu on kogutud 50 populaarse videomängužanri mängu tegelastevaheline dialoog. Töö meetodivalik lähtus mitmemõõtmelise analüüsi võimekusest analüüsida mitmete keeleliste tunnuste koosesinemist. Analüüsi läbiviimiseks kasutati Multidimensional Analysis Tagger tarkvara.

Analüüsi tulemused kirjeldasid valimi keelelist sarnasust kirjalikule keelele viiel dimensioonil kuuest. Siiski selgus, et VGDCs sisalduvas keeles esineb nii kirjalikule kui ka suulisele keelele omaseid tunnuseid, mis seostub videomängudialoogi loomise kontekstiga. Lisaks näitas VGDC analüüs, et erinevate videomängudele omaste aspektide mõju on tekstile pigem väike. Kahe esimese dimensiooni täiendav kvalitatiivne analüüs kirjeldas rollimängude dialoogi registri sisemist varieeruvust ning sarnasust mittenarratiivsetele tekstidele. Antud magistritöö tulemused panustavad mängukeele uuringute valdkonda, kirjeldades põhjalikult mängusisest keelt suurema valimi põhjal ning laiendades eelmiste uuringute tulemusi.

Magistritöö on jaotatud kahte peatükki, millest esimene annab ülevaate varasematest videomängude uurimustest, defineerides videomängud ja rollimängude žanri videomängud antud töö raames ning keskendudes videomängude dialoogi tähtsusele. Sellele järgneb teine peatükk empiirilise analüüsiga. Teises peatükis kirjeldatakse mitmemõõtelise analüüsi meetodi keskseid mõisted ja metoodikat, kahte erinevat lähenemist antud meetodile ning aditiivse meetodi sobivust antud analüüsiks. Samuti kirjeldatakse teises peatükis olemasolevaid videomängukorpuseid ja valitud andmestiku sobivust antud analüüsiks. Sellele järgnevad analüüsiprotsessi kirjeldus, analüüsitulemused ja nende arutelu.

Märksõnad: Aditiivne mitmemõõtmeline analüüs, mitmemõõtmeline analüüs, videomängud, rollimängud, korpuslingvistika.

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Janely Rüdein,

1. .annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose
   Exploring Dialogue in RPG Video Games: An Additive Multi-Dimensional Approach,
   mille juhendaja on Jane Padrik,

1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace´i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Janely Rüdein

Tartus, 13.05.2025

Autorsuse kinnitus

Kinnitan, et olen koostanud käesoleva magistritöö ise ning toonud korrektselt välja teiste autorite panuse. Töö on koostatud lähtudes Tartu Ülikooli maailma keelte ja kultuuride instituudi anglistika osakonna oma õppekava magistritöö juhendist ning on kooskõlas heade akadeemiliste tavadega.

Janely Rüdein

Tartus, 13.05.2025

Lõputöö on lubatud kaitsmisele.

Jane Padrik

Tartus, 13.05.2025