THE UNIVERSITY OF CHICAGO

Gender Representation of Psychosocial Factors in Role-Playing Video Game Dialogue

By

Weiwu Yan

May 2025

A paper submitted in partial fulfillment of the requirements for
the Master of Arts degree in the Master of Arts in
Computational Social Science

Faculty Advisor: Jean Clipperton
Preceptor: Sabrina Nardin

Acknowledgments

I would like to express my sincere gratitude to my advisor, Jean Clipperton, and my preceptor, Sabrina Nardin, for their invaluable guidance, insightful feedback, and unwavering support throughout this project. I am also grateful to my families and friends for their encouragement and feedback during the development of this work. This thesis would not have been possible without the continued inspiration and patience of everyone who supported me during this process.

Abstract

This study investigates gender differences in the linguistic construction of character dialogue within role-playing video games, focusing on psychosocial factors. A corpus of game dialogue was analyzed using both a lexicon-based dictionary and a transformer-based emotion classification model. Basic linguistic analysis revealed that player characters spoke significantly more lines, sentences, and words than non-player characters but used simpler sentence structures. However, no significant gender differences were observed in these linguistic metrics. Word frequency analysis showed that female characters used significantly more language associated with cognitive reflection and social connection, whereas male characters exhibited higher rates of swear word usage. Emotion classification revealed that female non-player characters expressed a broader range of emotions, including more disgust, sadness, surprise, joy, and fear, while male characters showed a stronger association with anger. Principal component analysis identified an affective positivity–negativity axis that significantly predicted gender, with male characters clustering toward greater negativity. Analysis of neutral characters indicated that their psychosocial profiles most closely resembled male characters, supporting the notion of masculine defaults. These findings highlight how gendered psychosocial patterns are embedded in video game dialogue and reveal both persistent stereotypes and emerging shifts toward more balanced gender representation in interactive narratives.

Keywords: gender representation, RPG, psychosocial factors, LIWC, DistilRoBERTa, PCA

1 Introduction

Video games have evolved from casual entertainment to complex cultural experiences, becoming an important subject of academic inquiry across disciplines such as psychology, sociology, and communication studies. As an interactive media, video games provide not only entertainment but also rich sites for investigating psychosocial factors through player and character behavior.

Among video game genres, role-playing games (RPGs) are uniquely suited for studying the linguistic construction of identity. Specifically, dialogue is a central mechanic in RPGs, producing expansive corpora of character speech that allow players to shape conversations and relationships. The volume and variability of dialogue in RPGs create valuable datasets for analyzing how language encodes psychosocial meanings.

Given the centrality of dialogue to RPGs, understanding how characters are linguistically constructed is especially important when examining issues of gender representation. Previous research on gender representation in video games demonstrates persistent stereotypes in character design and narrative roles, showing that male characters are often portrayed as rational and dominant while female characters are associated with emotionality and supportiveness (Behm-Morawitz & Mastro, 2009; Lynch et al., 2016).

While prior work has made important contributions to understanding gender differences in video games, most studies have focused on visual or narrative representations rather than systematically analyzing character dialogue (Behm-Morawitz & Mastro, 2009; Dietz, 1998; Jansz & Martis, 2007; Lynch et al., 2016). As a result, the linguistic construction of gendered identities in video games remains underexplored, particularly in the context of RPGs where dialogue plays a critical narrative role.

In addition to male and female characters, the rise of neutral or genderless characters in modern games prompts new questions about linguistic default norms. Whether neutral characters linguistically align more closely with male or female characters, or exhibit distinct patterns in scripted dialogue remains an open empirical question.

Advances in computational linguistics offer new opportunities to address these questions. Tools such as lexicon-based dictionaries and transformer-based models allow researchers to analyze psychosocial content at scale. Applying these methods to video game dialogue provides a powerful framework for understanding how gendered language is constructed within interactive narratives.

2 Literature Review

2.1 Video Game Industry

The video game industry has transformed over the past five decades from a niche entertainment to a dominant form of global culture (Kline et al., 2003). Its origins can be traced back to the early 1970s, with arcade titles such as *Pong* (1972) and *Space Invaders* (1978), which introduced interactive digital play to the public. These early games were mechanically simple and focused primarily on player reflexes and point-based objectives. However, they laid the foundation for a rapidly growing medium that would come to incorporate diverse gameplay mechanics, advanced audiovisual effects, and dynamic narratives.

Culturally, video games have grown from being associated with youth subcultures to occupying a central place in mainstream media consumption (Kent, 2001). With the emergence of home consoles in the 1980s, such as Nintendo Entertainment System and Sega Genesis, gaming became a household activity. By the 2000s, it had matured into a cross-generational form of entertainment, facilitated by the rise of online multiplayer systems and digital distribution platforms such as Steam and PlayStation Network.

Economically, the video game industry has outpaced both the film and music industries in annual revenue. According to Statista (2024), the global market of video games reached over $475 billion U.S. dollars in 2024, with projections indicating continued growth driven by mobile gaming, cloud gaming, and in-game purchases. The industry's economic influence has positioned it as a major player in technology sectors, attracting significant investment and driving innovation in artificial intelligence, graphics processing, and interactive design.

Technological advances in hardware and software have been central to the industry's evolution. Improvements in graphics processing units and memory capacity have enabled the rendering of realistic virtual environments and character models (Wolf, 2002). At the same time, the development of powerful and accessible game engines such as Unreal Engine and Unity has lowered the barriers to entry for developers while expanding the creative possibilities for game design (O'Donnell, 2014).

These technological innovations converged around the turn of the millennium, making the post-2000 era a pivotal period in the development of video games. With the release of sixth-generation consoles such as PlayStation 2, Xbox, and GameCube, developers could now implement voice-acted dialogue, expressive facial animations, and immersive 3D environments on a scale that was previously unattainable (Wolf, 2007). The impact of these

developments has been far-reaching. Games are no longer constrained to simple mechanics or superficial stories. Instead, they function as dynamic narratives where players form emotional connections with characters, engage in moral decision-making, and experience richly textured worlds.

This shift has solidified video games as a legitimate object of academic inquiry across a wide range of disciplines. In psychology, previous research has focused on the cognitive, emotional, and social impacts of gaming, highlighting how video games can foster attention, resilience, prosociality while being associated with risks such as addiction, depression, and violence (Granic et al., 2014). In communication studies, scholars argue that game studies are central to understanding broader patterns of media convergence, audience interaction, and digital culture, calling for video games to be fully integrated into critical media theory (Chess & Consalvo, 2022). From a sociological perspective, video games are recognized as key to understanding contemporary patterns of consumption, identity, and social interaction (Crawford, 2009). Together, these studies illustrate the growing recognition that video games offer a rich site for interdisciplinary research into psychosocial dynamics.

2.2 Role-Playing Games

Role-playing games are a genre characterized by narrative depth, character progression, and decision making. Unlike arcade-style games, which emphasize fast reflexes and linear gameplay, RPGs often unfold at a slower pace and prioritize story over action. In these games, players typically assume the role of a protagonist or lead a party of characters within a fictional world, engaging in conversations, transactions, and combat while developing their characters' abilities and relationships (Hitchens & Drachen, 2009). Core features of contemporary RPGs include character customization, branching storylines, and extensive dialogue systems that allow players to influence the direction and tone of the narrative.

The origins of digital RPGs can be traced back to the late 1970s and early 1980s with games such as *Akalabeth: World of Doom* (1980) and *Ultima* (1981), which were heavily influenced by tabletop RPGs like *Dungeons & Dragons* (1974) and *Call of Cthulhu* (1981). As hardware capabilities improved, RPGs evolved from text-based and top-down interfaces to complex three-dimensional environments, seen in genre-defining titles such as *Final Fantasy VII* (1997) and *The Elder Scrolls III: Morrowind* (2002).

A defining trait of RPGs is their emphasis on character development, not just through mechanical progression like skill acquisition and equipment upgrades, but also the growth of

personalities and relationships (Deterding & Zagal, 2018). Characters often express a wide range of emotions and feelings through conversations, facial expressions, and body postures, reflecting interpersonal dynamics and internal conflicts. Dialogue choices are particularly central to the genre, allowing players to guide conversations, determine character relationships, and shape narrative outcomes. This thesis focuses specifically on RPGs because these features produce expansive corpora of character dialogue, making RPGs a valuable source of linguistic data for psychosocial analysis.

Scholarly research on RPGs spans multiple disciplines but tends to cluster around three major areas. First, in educational psychology, RPGs have been studied for their ability to support skill acquisition and cognitive development. Extensive research shows that RPGs can foster collaboration, communication, and critical thinking skills, particularly when gameplay is structured around problem-solving scenarios and narrative challenges (Prager, 2019). Second, RPGs have been explored as therapeutic tools in mental health contexts. A scoping review highlights how RPGs are employed with adolescents and adults to promote self-efficacy, facilitate social skills, and address conditions such as depression and anxiety (Arenas et al., 2022). Third, scholars in interactive media have examined how the design values embedded in RPGs shape player experience. A content analysis of RPG postmortems reveals that successful games prioritize player agency, immersive storytelling, and meaningful choices, underscoring the central role of narrative design in crafting emotionally resonant experiences (Pulkkinen, 2014).

2.3 Gender Representation in Video Games

Gender representation in video games has been the subject of sustained academic attention, revealing persistent disparities and biases across multiple dimensions of the gaming ecosystem (Romrell, 2014). Scholars have examined not only how gender is depicted within games but also how it is shaped by the demographics and ideologies of game developers and interpreted by diverse player communities.

From the production side, the video game industry has historically been male-dominated. As of 2023, women represented only 31% of game developers globally (Weststar & Lentini, 2024). However, gender disparity is even more pronounced in leadership, with a report indicating that women comprised just 16% of executive team members at the top 14 gaming companies by revenue (20-first, 2020). Additionally, development studios often operate within masculinized design cultures in which certain values, themes, and player perspectives are privileged over others (Fron et al., 2007). As a result, male-centered design practices have

historically prioritized male protagonists, heterosexual romance narratives, and combat-oriented mechanics, frequently sidelining or simplifying female and non-binary characters.

On the consumer side, gendered assumptions about who plays games and what types of games they prefer have contributed to a cultural narrative that frames gaming as a predominantly male activity (Shaw, 2011). This perception persists despite research showing that nearly half of all players identify as women (Entertainment Software Association, 2023). Previous research shows that female players often report exclusionary or hostile experiences within gaming communities, which in turn may shape their engagement with certain genres or online play (Fox & Tang, 2014).

Within games themselves, gender representation is often skewed through visual and narrative dimensions. Content analysis finds that female characters are underrepresented in leading roles and frequently portrayed in hypersexualized ways, with exaggerated physical features, erotic movements, and revealing clothing (Lynch et al., 2016). In contrast, male characters are more commonly depicted as active agents who are strong, rational, and assertive, reinforcing normative gender binaries (Jansz & Martis, 2007).

Beyond physical traits, character personalities and capabilities are also shaped by gendered expectations. Female characters are more likely to express vulnerability, empathy, and nurturing behaviors, while male characters are more often written to exhibit stoicism, aggression, and emotional restraint (Behm-Morawitz & Mastro, 2009). Narrative roles often reflect these patterns, with women cast as caretakers, victims, or companions, and men as heroes, leaders, or saviors (Dietz, 1998).

Language use is another area where gendered scripting in video games becomes apparent. Recent research finds consistent gender asymmetries in both the quantity and content of speech (Rennick et al., 2023). Female characters, who make up 30% of all characters, are found to speak 35% of the dialogue. More importantly, female characters use more expressions of gratitude, hedging, and apologies, and swear significantly less, reflecting stereotypically feminine politeness and sensitivity.

Taken together, these studies reveal that gender representation in video games is shaped by multiple factors, including industry practices, player culture, character design, and patterns of language use.

2.4 Computational Linguistic Analysis

Even though previous work has examined gender differences in language use within video games, much of this work has focused on limited linguistic features (Heritage, 2021). Computational approaches to linguistic analysis have become increasingly sophisticated, drawing on both lexicon-based dictionaries and transformer-based models to capture nuanced psychosocial dimensions of language. These tools have been widely applied in social sciences to investigate how psychosocial factors vary by context and identity.

Lexicon-based dictionaries such as Linguistic Inquiry and Word Count (LIWC) have been used to study the relationship between language use, mental state, and social behavior. Specifically, LIWC operates using dictionaries in which words are matched to pre-defined categories that reflect psychological, emotional, cognitive, and social constructs (Boyd et al., 2022). When analyzing a text, LIWC scans each word and aggregates the frequency of words falling into each category, producing quantitative scores that reflect the linguistic and psychosocial profile of the text.

It has been proven effective in identifying patterns in language use across different social groups, including gendered differences in self-disclosure, affective intensity, and interpersonal style (Newman et al., 2008). One study using LIWC shows that female authors tend to use more emotional and interpersonal words, while male authors more frequently use language related to objectivity or authority (Kacewicz et al., 2014). Therefore, this tool is particularly suitable for large-scale analysis of expressive text, such as social media posts or video game dialogue, where subtle psychosocial signals may accumulate over time and across interactions.

In recent years, transformer-based language models have significantly advanced the field of textual analysis by leveraging self-attention mechanisms to learn deep contextual embeddings of language (Devlin et al., 2019). Models such as BERT and its successors are pre-trained on vast text corpora using masked language modeling and next sentence prediction. These pre-trained models can then be fine-tuned on downstream tasks like sentiment analysis, topic classification, and named entity recognition with relatively small labeled datasets. The BERT architecture's bidirectional nature allows it to capture semantic relationships from both left and right contexts, offering a major improvement over earlier word embedding methods like Word2Vec or GloVe, which rely on static representations.

Transformer-based models have also been used to study gendered language patterns, especially in online settings. For example, research leveraging unsupervised classifiers demon-

strates how gendered expression manifests in political discourse, customer reviews, and interpersonal communication on platforms like Reddit and Twitter (Field & Tsvetkov, 2020). These studies often combine deep learning with demographically labeled corpora to explore how language use differs between men and women in both content and style. Such methodologies are especially relevant for the analysis of video game dialogue, where character speech may be scripted in ways that reflect or challenge social norms about gender.

Both LIWC and BERT offer complementary strengths. LIWC provides interpretability through well-defined psychosocial categories, while BERT offers contextual flexibility and semantic depth. Together, they enable a multifaceted analysis of language use that is well-suited for investigating gender differences in dialogue-based media. In the context of RPGs where character dialogue is stylistically diverse and emotionally charged, these tools provide a rigorous methodological foundation for understanding how psychosocial factors are linguistically constructed and gendered.

2.5 The Intersection of Gender, Dialogue, and Psychosocial Factors

As the preceding literature demonstrates, substantial research examined gender representation in video games and computational methods for linguistic analysis of psychosocial factors. However, these areas of inquiry were largely studied in parallel rather than in intersection. While studies have analyzed gender stereotypes in character design and narrative roles (Behm-Morawitz & Mastro, 2009; Dietz, 1998; Jansz & Martis, 2007; Lynch et al., 2016), and others have used tools like LIWC and BERT to investigate gendered discourse (Field & Tsvetkov, 2020; Kacewicz et al., 2014; Newman et al., 2008), few studies have applied these methods to the gendered dialogue of fictional characters in interactive media, particularly in the context of RPGs (Rennick et al., 2023). This leaves open important questions about how gender is constructed and reinforced linguistically by game developers and how psychosocial factors are distributed across characters of different genders.

Moreover, while many studies have adopted a binary framework of male and female characters, fewer have considered how neutral or genderless characters are situated within this continuum (Rennick et al., 2023). As video games increasingly incorporate characters that resist traditional gender categorization, there is a growing need to understand whether these characters linguistically align more with masculine or feminine patterns, or whether they represent a distinct mode of expression.

This study addresses these gaps by applying linguistic psychosocial analysis to a large

dataset of character dialogue from RPGs. The analysis focuses on two central research questions:

1. **What differences exist in psychosocial factors between male and female characters in RPG dialogue?**

   I hypothesize that male and female characters will differ significantly in both psychological and social language patterns. Specifically, male characters are expected to use more language associated with assertiveness, causality, and certainty, while female characters are anticipated to use more language reflecting cognitive reflection, affective states, and social connection. Prior work finds that women use significantly more language related to psychological states and social affiliation, whereas men rely more heavily on concrete references like occupation and money (Newman et al., 2008). These findings suggest that female dialogue will emphasize internal states and interpersonal relationships, while male dialogue will focus more on externalized concerns and informational precision.

   In terms of emotional expression, male characters are hypothesized to express anger more frequently, while female characters are expected to display a broader range of emotions, including disgust, fear, joy, sadness, and surprise. This expectation aligns with a meta-analysis (Archer, 2004), which finds that males are more likely to engage in direct forms of aggression, a pattern that may translate into a greater frequency of anger in dialogue. In contrast, Allen and Haccoun (1976) identify higher emotional expressiveness among women, particularly in internalized emotions such as sadness and fear, suggesting that female characters may be portrayed with greater emotional diversity.

2. **Do neutral characters use language that more closely resembles male or female characters, or do they exhibit distinct linguistic patterns?**

   Neutral characters are expected to exhibit psychological and social patterns more closely aligned with male characters than female characters. This expectation is grounded in the documented phenomenon of masculine defaults, wherein behaviors and characteristics typically associated with the male gender role are treated as standard and normal within cultural settings (Cheryan & Markus, 2020). Research shows that masculine traits, such as assertiveness, independence, and emotional restraint, are often implicitly regarded as neutral, while feminine behaviors are seen as deviations from the norm. In the context of video game dialogue, this suggests that genderless characters are likely to mirror male linguistic styles.

3 Data

3.1 Video Game Dialogue Corpus

The Video Game Dialogue Corpus (VGDC) is a comprehensive dataset that comprises over 6.2 million words of dialogue from more than 50 RPGs, with a diverse selection in terms of region and period (Rennick & Roberts, 2024). Available on a public GitHub repository, the VGDC provides programs that download and process each game's script on local machines. Eighteen RPGs released in the 21st century were selected to maintain a modern scope. This period reflects significant advancements in 3D graphics, interactive dialogue systems, and branching narratives that are extensively applied in contemporary RPGs. Some post-2000 games were excluded due to the unavailability of data, such as cases where fan transcripts were under copyright or game files were encrypted and inaccessible.

A data parser was developed to extract and clean dialogue and relevant metadata from game scripts. The parser filtered out non-dialogue elements and extracted character names alongside their corresponding lines of dialogue. Characters without unique names were excluded, as their gender is often unidentifiable, making it difficult to reliably associate language use with individual character identities. Additionally, characters were labeled as either male or female based on documentation provided by gaming community websites, such as Fandom (2025) and the Unofficial Elder Scrolls Pages (2025). In three special cases where players can select the gender of the main playable character, these characters were categorized as neutral. Furthermore, some characters may appear under aliases that are only later revealed in the game. In these instances, the dialogue attributed to aliases were merged with that of the character's canonical name to ensure consistency. Lastly, several metadata fields were added for each character, including playability, which refers to whether the character can be controlled by the player, as well as game title, release year, and country of development.

After applying the data parser to the selected games, the resulting data were concatenated, cleaned, and transformed for statistical analysis. Text enclosed in parentheses was excluded, as it generally represents internal thoughts or narrative asides rather than spoken dialogue. For example, in *Persona 3* (2006), the line "Anyway, school was school (boring as hell, in other words), but more stuff happened afterwards" contains a parenthetical that reflects the character's internal commentary rather than what is explicitly said aloud. Dialogue lines consisting solely of punctuation marks or special characters were also removed, since they sometimes do not carry meaningful linguistic content. In addition, the number of lines,

sentences, and words was calculated for each character. The number of lines corresponds to the number of parsed dialogue entries attributed to a character. Sentences were identified based on the presence of sentence-ending punctuation marks, including periods, exclamation points, and question marks, while words were defined as text segments separated by space.

3.2 Summary Statistics

The final dataset includes 3,043 game characters from eighteen RPGs developed in four countries—Japan, the United States, the Netherlands, and the United Kingdom—with dialogue comprising 247,644 lines, 512,601 sentences, and 3,653,484 words. Table 1 presents summary statistics on the volume of textual data in the dataset.

Summary Statistics

Title	Year	Country	Characters	Lines	Sentences	Words
Death Stranding	2019	Japan	23	1335	5178	39544
Disco Elysium	2019	UK	85	54172	87196	565346
Elder Scrolls Morrowind	2002	US	357	6594	18047	164517
Elder Scrolls Oblivion	2006	US	873	59633	108713	747469
Elder Scrolls Skyrim	2011	US	297	7175	20525	144468
Final Fantasy VII Remake	2020	Japan	71	7507	14004	69650
Final Fantasy X	2001	Japan	104	4925	9608	50710
Final Fantasy XII	2006	Japan	154	2340	6570	47785
Final Fantasy XIII	2009	Japan	21	946	2087	11306
Final Fantasy XIII-2	2011	Japan	54	3412	7910	51411
Final Fantasy XIV	2013	Japan	532	20640	74019	709366
Final Fantasy XV	2016	Japan	45	7317	11803	68781
Hades	2020	US	28	19718	35502	223440
Horizon Forbidden West	2022	Netherlands	121	6331	15989	111256
Horizon Zero Dawn	2017	Netherlands	61	2263	6299	40457
Persona 3	2006	Japan	45	5809	11896	94789
Persona 4	2008	Japan	63	18155	29718	192064
Persona 5	2016	Japan	109	19372	47537	321125

Table 1: Characters refer to fictional entities in video games.

Figure 1 visualizes the proportion of dialogue spoken by female characters across eighteen RPGs released between 2000 and 2022. The majority of games cluster below the 50% mark,

indicating that dialogue is predominantly male across most titles. Notably, *Horizon Forbidden West* (2022), featuring a female protagonist, stands out with a proportion of female dialogue close to 75%. The visualization illustrates both the temporal trends and regional variation in gender representation, highlighting a stagnant increase in female dialogue across regions over time.

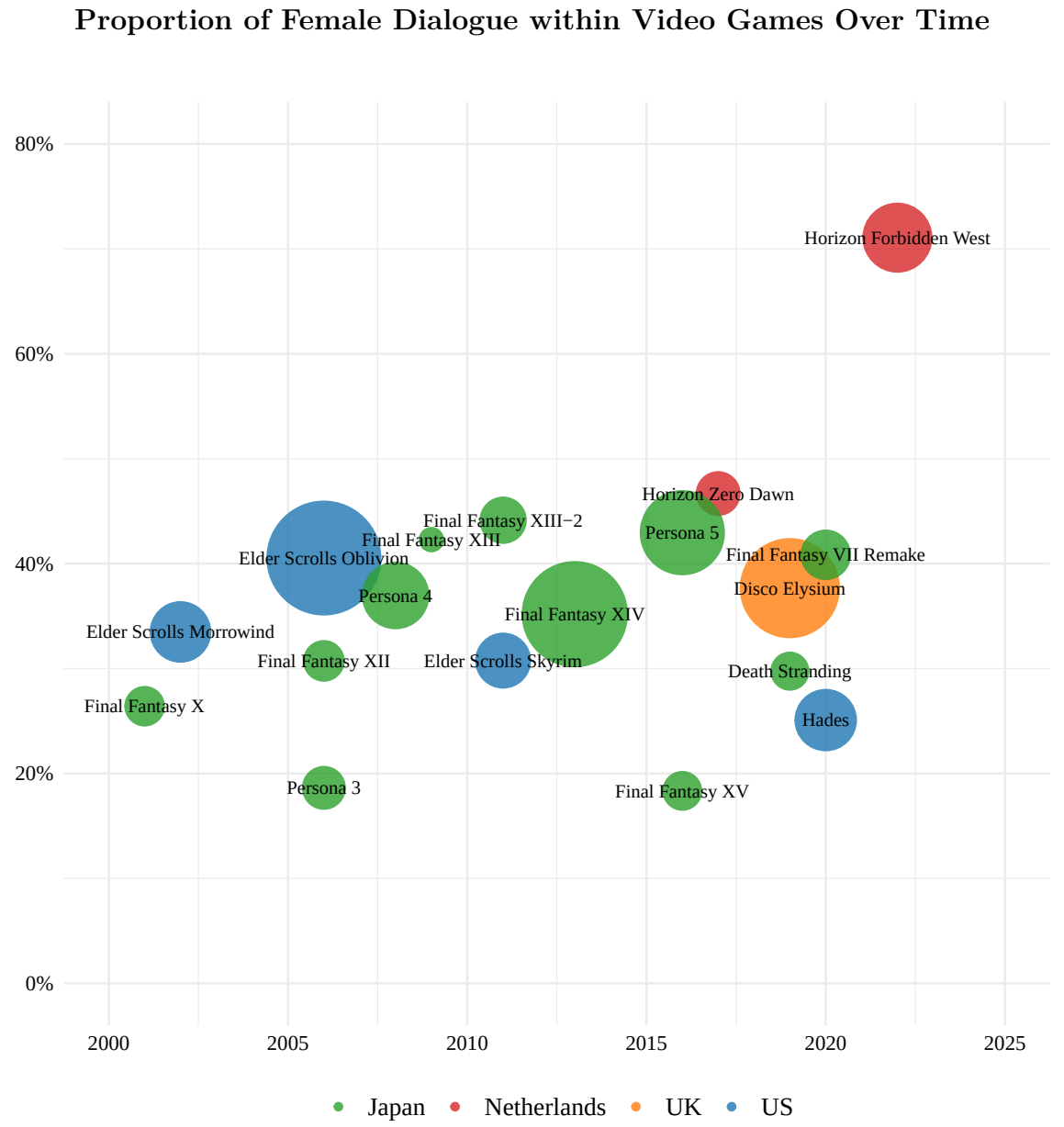

Figure 1: The color represents the country, and the size represents the volume of female-spoken words.

4 Methods

4.1 Linguistic Inquiry and Word Count

Linguistic Inquiry and Word Count is a widely used text analysis program designed to quantify linguistic, psychological, and social properties of language. Originally developed in the mid-1990s (Pennebaker & Francis, 1999), LIWC has undergone several major revisions—LIWC1999, LIWC2001, LIWC2007, LIWC2015—and most recently, LIWC-22. Each iteration has expanded the internal dictionary and improved software functionality to accommodate advances in linguistic theory and computational methods. LIWC-22 represents the most comprehensive version to date, incorporating over 12,000 words, stems, phrases, and emoticons, organized into a large set of hierarchical categories that tap into linguistic structure, cognitive processes, social themes, and emotional expression (Boyd et al., 2022).

In this study, LIWC-22 software was purchased from its official website and used to provide insights into specific word use and its psychosocial implications. The analysis focused on core dictionaries, including broad categories such as *cognitive processes*, *affect*, and *social behavior*, as well as their more specific subcategories. These three main categories were selected due to their theoretical relevance to how gender is expressed and perceived in language. Cognitive processes capture reasoning and reflection, which are often associated with internal thought and problem-solving styles. Affect reflects emotional tone and intensity, making them central to analyzing gendered emotional expression in character dialogue. Social behavior relates to interpersonal orientation and communicative intent, offering insight into how characters engage with others. Each category reports the proportion of words in a text segment that match entries in its predefined lexicon, enabling a nuanced quantification of psychosocial dimensions of character dialogue. These metrics offer a robust and interpretable framework for investigating patterns of psychosocial factors across gender groups in video game characters.

To conduct this analysis, character dialogue was imported into the LIWC-22 application, and the LIWC-22 English Dictionary was selected. The analysis was performed with each game character's complete dialogue treated as a single text entry. Upon completion, generated scores were exported and merged back into the original dataset for further statistical analysis.

4.2 DistilRoBERTa

Transformer-based language models such as RoBERTa represent a major advancement in natural language processing by moving beyond traditional bag-of-words or n-gram approaches. Rather than treating words independently, these models generate dense, context-sensitive vector representations, known as embeddings, that capture the meaning of a word based on its surrounding text (Liu et al., 2019). This allows them to model complex linguistic phenomena such as polysemy and syntax. Built upon the transformer architecture, RoBERTa and its variants leverage self-attention mechanisms to dynamically weigh the importance of each word in a sentence relative to the others, enabling more nuanced and flexible understanding of language.

DistilRoBERTa is a distilled version of RoBERTa that offers a lighter and faster alternative to large-scale transformer models while maintaining strong linguistic performance. Using the technique of knowledge distillation, DistilRoBERTa retains approximately 97% performance on benchmark tasks while reducing model size by 40% and improving inference speed by 60% (Sanh et al., 2019). Knowledge distillation involves training a smaller student model to reproduce the behavior of a larger teacher model, compressing knowledge by minimizing the differences between their predictions. In addition to using fewer parameters, DistilRoBERTa simplifies certain components of the transformer architecture, such as by reducing the number of layers, while preserving critical attention mechanisms that allow it to capture contextual relationships effectively.

The DistilRoBERTa model used in this study is publicly available via Hugging Face. This `emotion-english-distilroberta-base` was fine-tuned on `distilrobeta-base` specifically for emotion classification tasks using a combination of emotion-labeled datasets drawn from diverse sources, including Twitter, Reddit, and television shows (Hartmann, 2022). These sources contain a mix of spontaneous and scripted language, offering a wide range of emotional expressions and linguistic styles. This diversity makes the model particularly well-suited for analyzing video game dialogue, which often blends realistic conversation with narrative-driven scripting. The model is trained to predict seven basic emotion categories: anger, disgust, fear, joy, neutral, sadness, and surprise. Due to its easy accessibility, reliable performance, and computational efficiency, it offers a practical solution for large-scale emotion classification in dialogue-based corpora such as video game scripts.

When employed, the classifier preprocesses the input text to prepare it for emotion classification. This involves tokenization which converts each line of dialogue into tokenized representations. The model then transforms these tokens into contextual embeddings to

capture nuanced semantic and syntactic patterns. Additionally, only the first 514 token embeddings of each input are processed, as this is the maximum sequence length supported by DistilRoBERTa. Any input exceeding this limit is truncated, which may result in the loss of information from longer lines of dialogue. During inference, the model computes weighted embeddings to assign a probability distribution across seven emotion categories and selects the label with the highest probability as the predicted emotion.

In this study, the DistilRoBERTa model was initialized with default parameters and applied to each individual line of dialogue using the transformers pipeline provided by Hugging Face. For each line, the model returned a set of confidence scores for all seven emotion labels. These scores were aggregated and stored for each character to capture the distribution of emotional expression across their dialogue. For the purposes of analysis, the neutral category was excluded, as it does not reflect an active emotional state and could obscure patterns in expressive emotional content. For each dialogue line, only the emotion label with the highest predicted probability was retained, a decision made to simplify interpretation and ensure that each line contributed clearly to the dominant emotional profile of the character.

Figure 2 displays the overall frequency of each predicted emotion across all lines of video game dialogue in the dataset. The most frequently expressed emotion is surprise, followed by anger, disgust, and joy, while sadness and fear are the least common. These results suggest that video game dialogue tends to favor emotionally intense or reactive expressions over more subdued or negative states.

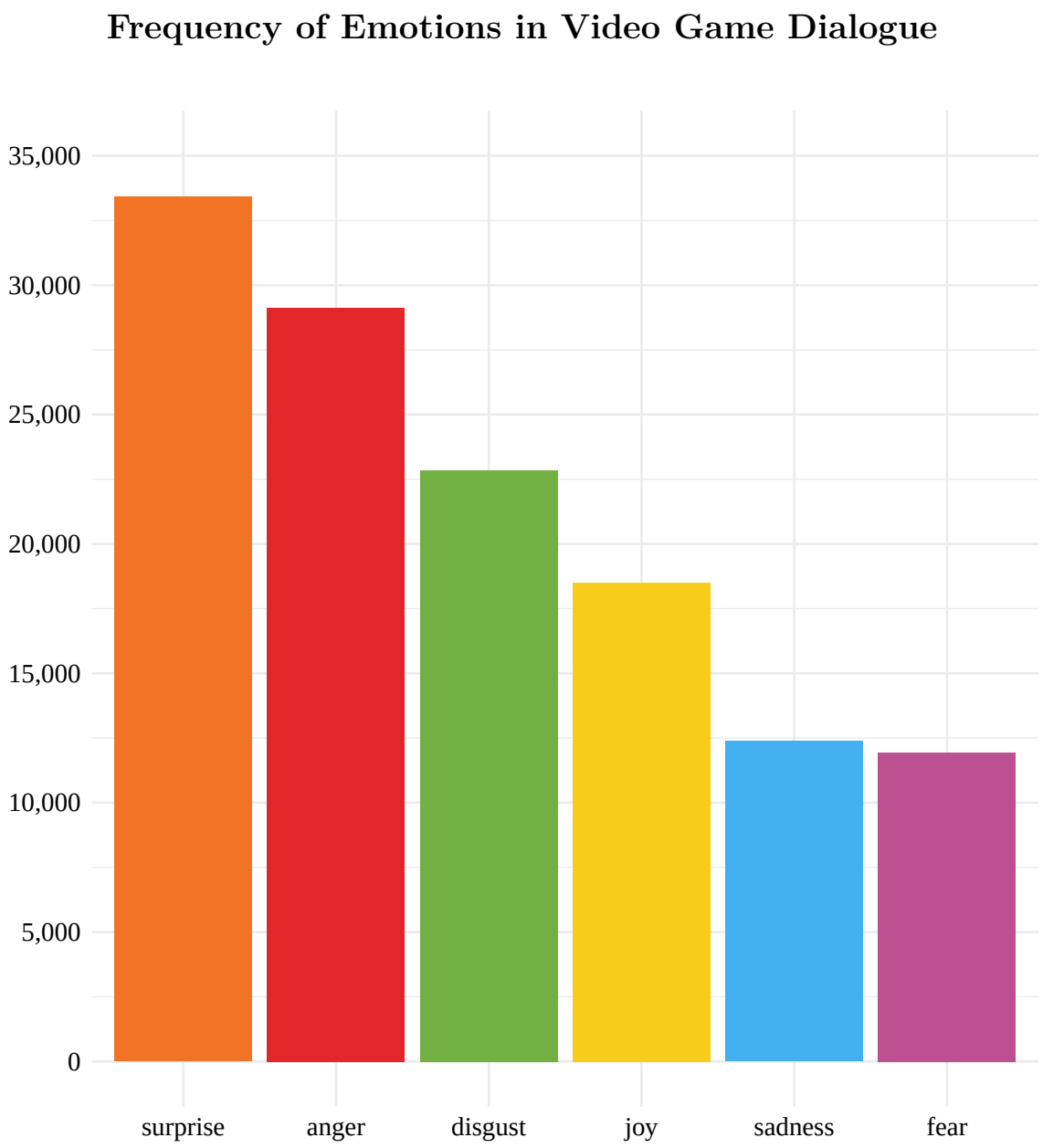

Figure 2: The color scheme is inspired by *Inside Out* (2015).

4.3 Principal Component Analysis

Principal component analysis (PCA) is a widely used dimensionality reduction technique that transforms a set of possibly correlated variables into a smaller number of uncorrelated components while retaining as much variance in the data as possible (Jolliffe & Cadima, 2016). PCA is a deterministic method that identifies the directions along which the variance of the data is maximized. Technically, it operates by computing the covariance matrix of the standardized input variables, followed by decomposition to extract eigenvectors and eigenvalues. The eigenvectors represent the principal components, while the eigenvalues indicate the amount of variance captured by each component. By projecting the data onto the leading principal components, PCA reduces dimensionality while preserving the most informative structure of the dataset.

This process enables the uncovering of latent patterns and relationships that may not be apparent in the original high-dimensional space. PCA is particularly valuable for exploratory data analysis and visualization, as it reduces noise and redundancy, facilitates the detection of clustering or outlier structures, and provides interpretable two-dimensional mappings of the relationships between observations and features (Hotelling, 1933).

In the context of video game dialogue, PCA offers a useful method for synthesizing multiple psychosocial features into a lower-dimensional representation. Rather than analyzing each linguistic variable in isolation, PCA enables researchers to identify broader psychosocial dimensions along which characters may differ, such as emotional intensity, cognitive complexity, or social engagement. By applying PCA to variables drawn from both LIWC-22 and `emotion-english-distilroberta-base`, this study seeks to identify principal dimensions of variation in character dialogue. The resulting components offer insight into psychosocial factors across characters, providing a holistic view of how language reflects gendered identities in game narratives.

In this study, PCA was conducted on a subset of psychosocial features from LIWC-22 and `emotion-english-distilroberta-base`. Although the initial dataset included 18 core categories, only six of them, including cognitive processes, social behavior, politeness, conflict, anger, and joy, were selected based on their interpretability and prior validation in gendered language research (Archer, 2004; Newman et al., 2008). These variables were standardized using z-score scaling prior to PCA to ensure comparability across features. PCA was then performed, and the first two principal components were retained for visualization and interpretation. Furthermore, to assess the relationship between gender and the derived psychosocial dimensions, a binomial logistic regression was conducted with gender

as the dependent variable and first two components as independent variables.

5 Results

5.1 Descriptive Statistics

Two-sample t-tests were conducted to assess whether gender and playability were associated with statistically significant differences across five basic linguistic variables, including the number of lines, the number of sentences, the number of words, the number of words per line, and the number of words per sentence. Results indicate that player characters (PCs) speak significantly more lines ($t = -6.304$, $p < .001$), sentences ($t = -6.092$, $p < .001$), and words ($t = -5.253$, $p < .001$) compared to non-player characters (NPCs). However, PCs use significantly fewer words per line ($t = 16.753$, $p < .001$) and words per sentence ($t = 7.446$, $p < .001$). In contrast, no significant gender differences are found within either PCs or NPCs across any of the five linguistic variables. However, descriptive patterns reveal that male PCs speak more than female PCs on average, with neutral PCs falling between the two. Conversely, among NPCs, female characters speak more on average than their male counterparts. Table 2 shows the statistics of linguistic variables across character groups.

Descriptive Statistics

	Male PC	**Female PC**	**Neutral PC**	**Male NPC**	**Female NPC**
	Mean (SD)	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*
Lines	1544.25 (3035.85)	960.08 (734.16)	1268.00 (913.19)	36.63 (119.37)	50.65 (300.28)
Sentences	2603.89 (4416.21)	1847.24 (1525.37)	1501.00 (1138.90)	91.91 (253.64)	115.01 (513.06)
Words	17300.62 (27297.86)	12037.71 (11618.95)	8028.67 (6278.12)	695.35 (1793.65)	861.28 (3639.62)
Words per Line	13.01 (8.43)	12.41 (4.86)	6.28 (1.16)	27.56 (26.39)	26.30 (24.28)
Words per Sentence	6.32 (2.14)	6.15 (1.23)	5.34 (0.67)	7.63 (2.57)	7.63 (2.25)

Table 2: There are no neutral NPCs.

Two-sample proportion tests were conducted to compare the proportion of male and female characters within both PC and NPC groups. When calculating gender proportions, neutral PCs, whose gender can be selected by the player, were included in both male and female counts to reflect their dual-gender representation. Results show that male characters appear at significantly higher proportions than female characters in both categories: among PCs, 58.6% are male and 41.4% are female ($\chi^2 = 5.172$, $p = .023$); among NPCs, 65.0% are male and 35.0% are female ($\chi^2 = 527.943$, $p < .001$).

5.2 Linguistic Inquiry and Word Count

To examine gendered patterns in psychosocial factors, LIWC features were analyzed separately for PCs and NPCs, with two-sample t-tests comparing male and female characters. Among PCs, female characters use significantly more discrepancy-related language ($t = 2.330$, $p = 0.022$), while male characters use significantly more swear words ($t = -2.850$, $p = 0.006$). Table 3 presents the statistics of LIWC features between male and female PCs.

LIWC Statistics (PC)

	Male Mean	Female Mean	t	df	p
Cognitive Processes	12.70	13.60	1.91	90.9	0.059
Insight	3.07	3.17	0.59	90.9	0.556
Causation	1.35	1.41	0.80	90.1	0.427
Discrepancy	2.60	2.87	2.34	79.8	0.022*
Tentativeness	2.75	2.93	1.30	91.0	0.196
Certitude	0.86	0.92	0.69	90.8	0.491
Differentiation	3.42	3.69	1.96	88.1	0.054
Affect	5.83	6.08	0.75	81.8	0.455
Positive Tone	3.44	3.67	0.97	80.7	0.333
Negative Tone	2.03	2.24	1.67	81.6	0.099
Positive Emotion	0.82	1.00	1.86	76.3	0.067
Negative Emotion	0.71	0.77	1.02	76.0	0.313
Swear	0.29	0.08	-2.86	60.3	0.006**
Social Behavior	4.47	4.49	0.07	90.9	0.944
Prosocial Behavior	0.69	0.79	1.37	90.7	0.173
Politeness	0.64	0.60	-0.23	61.7	0.823
Conflict	0.49	0.49	-0.04	81.5	0.970
Moralization	0.42	0.39	-0.54	88.2	0.591
Communication	1.68	1.72	0.37	85.6	0.710

Table 3: A positive t-statistic indicates that female characters use that psychosocial feature more frequently than male characters, while a negative t-statistic indicates the opposite.

For NPCs, several gender differences emerges. Female NPCs use significantly more language related to cognitive processes ($t = 2.250$, $p = 0.025$), discrepancy ($t = 2.500$, $p = 0.012$), negative emotion ($t = 2.390$, $p = 0.017$), and politeness ($t = 2.960$, $p = 0.003$). Compared to male NPCs, they also use significantly less swear words ($t = -3.780$, $p < .001$) and language related to moralization ($t = -2.430$, $p = 0.015$). Table 4 presents the statistics of LIWC features between male and female NPCs.

LIWC Statistics (NPC)

	Male Mean	**Female Mean**	**t**	**df**	**p**
Cognitive Processes	10.90	11.30	2.23	2104	0.026*
Insight	2.28	2.31	0.55	2071	0.581
Causation	1.04	0.98	-1.51	2363	0.131
Discrepancy	2.41	2.60	2.49	1849	0.013*
Tentativeness	2.32	2.46	1.89	1990	0.059
Certitude	0.67	0.70	0.73	1954	0.464
Differentiation	3.23	3.35	1.51	2333	0.130
Affect	6.36	6.35	-0.10	2665	0.920
Positive Tone	3.93	4.00	0.52	2791	0.604
Negative Tone	2.19	2.20	0.05	2415	0.962
Positive Emotion	0.89	0.96	0.75	2787	0.4510
Negative Emotion	0.62	0.71	2.39	2123	0.017*
Swear	0.20	0.10	-3.78	2055	$< .001$***
Social Behavior	4.81	5.01	1.56	2272	0.120
Prosocial Behavior	0.97	1.04	1.15	2547	0.249
Politeness	0.65	0.84	2.99	2305	0.003**
Conflict	0.55	0.51	-0.85	2022	0.395
Moralization	0.54	0.46	-2.43	2755	0.015*
Communication	1.54	1.66	1.63	1780	0.103

Table 4: A positive t-statistic indicates that female characters use that psychosocial feature more frequently than male characters, while a negative t-statistic indicates the opposite.

To explore how neutral characters align linguistically with gendered character types, a distance matrix was constructed using average pairwise LIWC differences across all character groups. Neutral characters are most similar to male PCs ($d = 0.310$), followed by female PCs ($d = 0.350$), female NPCs ($d = 0.585$), and male NPCs ($d = 0.715$).

5.3 DistilRoBERTa

Poisson regressions were conducted to examine differences in emotional expression between male and female characters. Poisson regression was selected because the dependent variable, the number of lines expressing a particular emotion, is a count variable, and Poisson models are specifically suited for modeling the distribution of non-negative integer outcomes (Hayat & Higgins, 2014). This approach accounts for the fact that emotional expressions are discrete events that can vary across characters with different total dialogue volumes.

Results indicate that female NPCs express significantly less anger ($\beta = -0.297$, $z = -17.333$, $p < .001$). The coefficient suggests that female NPCs express approximately 74.3% as much anger as male NPCs. In contrast, female NPCs express significantly more sadness than their male counterparts ($\beta = 0.120$, $z = 7.872$, $p < .001$), corresponding to an increase of approximately 22.1%. Female NPCs also show significantly higher expression of disgust, fear, joy, and surprise, with increases ranging from 5% to 10%. Table 5 presents the statistics of emotion labels between male and female NPCs.

Emotion Statistics (NPC)

	Estimate	**Standard Error**	**z**	**p**
Anger	-0.297	0.017	-17.33	$< .001^{***}$
Disgust	0.095	0.019	5.12	$< .001^{***}$
Fear	0.051	0.025	2.02	0.043^{*}
Joy	0.062	0.021	2.90	0.004^{**}
Sadness	0.200	0.025	7.87	$< .001^{***}$
Surprise	0.073	0.017	4.37	$< .001^{***}$

Table 5: Male characters serve as the reference group. A positive estimate indicates that female characters express that emotion more frequently than male characters; a negative estimate indicates the opposite.

When comparing male and female PCs, gender differences in emotional expression become less pronounced. Specifically, there are no statistically significant differences in the expression of anger, surprise, sadness, or fear. Female PCs express slightly more joy ($\beta = 0.114$, $z = 5.131$, $p < .001$) and less disgust ($\beta = -0.156$, $z = -7.172$, $p < .001$), though both

differences fall within a 15% margin. Table 6 shows the statistics of emotion labels between male and female PCs. Notably, these distinctions are further diminished in games featuring mixed-gender PC groups, with expression of disgust decreasing by 6% in particular. As shown in Figure 3, the overlap between male and female emotional profiles increases progressively from NPCs to PCs, and is most substantial among PCs in diverse lineups.

Emotion Statistics (PC)

	Estimate	**Standard Error**	**z**	**p**
Anger	-0.006	0.019	-0.32	0.750
Disgust	-0.156	0.022	-7.17	$< .001^{***}$
Fear	0.049	0.030	1.65	0.099
Joy	0.114	0.022	5.13	$< .001^{***}$
Sadness	0.018	0.028	0.65	0.516
Surprise	0.014	0.016	0.88	0.381

Table 6: Male characters serve as the reference group. A positive estimate indicates that female characters express that emotion more frequently than male characters, while a negative estimate indicates the opposite.

Moreover, neutral characters exhibit a notably similar level of anger expression to that of male NPCs. However, for other emotions, the similarity varies across character groups, making it difficult to draw a direct connection between neutral characters and any single group. To provide a more comprehensive comparison, a distance matrix was constructed to quantify emotional similarity between neutral characters and other character groups. The analysis reveals that neutral characters are most similar to male PCs ($d = 0.088$), followed by male NPCs ($d = 0.090$), female PCs ($d = 0.093$), and female NPCs ($d = 0.110$).

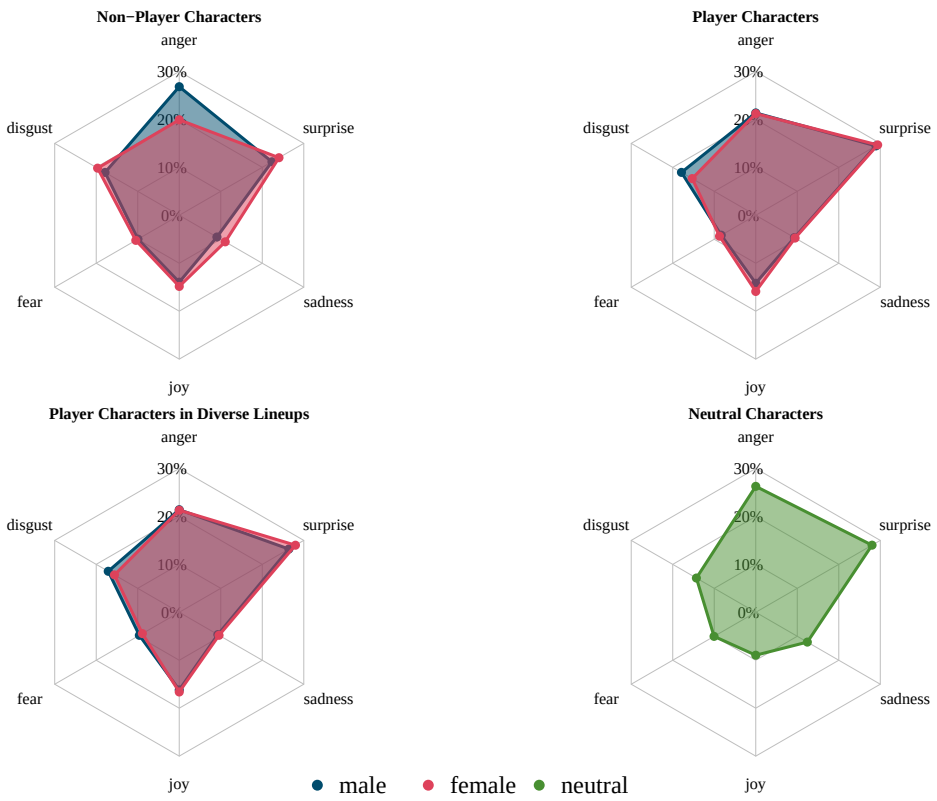

Figure 3: The color represents the gender. Diverse lineups refer to mix-gender main character groups.

5.4 Principal Component Analysis

As shown in Table 7, PC1 has a standard deviation of 1.45 and accounts for 34.8% of the variance, while PC2 has a standard deviation of 1.28 and explains an additional 27.3% of the variance. Together, the first two components capture approximately 62.1% of the overall variance in the dataset.

PCA Statistics

	PC1	PC2	PC3	PC4	PC5	PC6
Standard Deviation	1.4456	1.2809	1.0234	0.9390	0.5835	$< .0001$
Proportion of Variance	0.3483	0.2734	0.1745	0.1470	0.0568	$< .0001$

Table 7

Figure 4 visualizes the distribution of six psychosocial features across four major RPG franchises—*Final Fantasy*, *Horizon*, *Persona*, and *The Elder Scrolls*—based on scores from the first two principal components. Each point in the figure represents the mean score for a specific gender group within a given franchise. PC1 reflects a affective negativity–positivity axis, where higher values correspond to greater usage of anger-related language, while lower values indicate greater use of joy-related language. PC2 differentiates internally versus externally oriented discourse, with cognitive processes loading more strongly at one end and social behavior at the other. Additionally, conflict and politeness capture distinct social behaviors associated with negative and positive emotional states.

From a gender perspective, a clear pattern emerges. On average, female characters are positioned toward the positive side of the emotional spectrum, while male characters are positioned toward the negative side. A noteworthy exception is found in the Horizon franchise, where female characters exhibit language most strongly associated with anger compared to any other gender group across the games analyzed. However, no substantial gender differences are observed along the axis distinguishing internal cognitive processes from external social behaviors.

In contrast, while emotional expression does not vary dramatically across franchises, distinct differences in the balance between cognitive and social language are evident. Specifically, characters within the same franchise, regardless of gender, exhibit similar patterns in their use of cognitive versus social language.

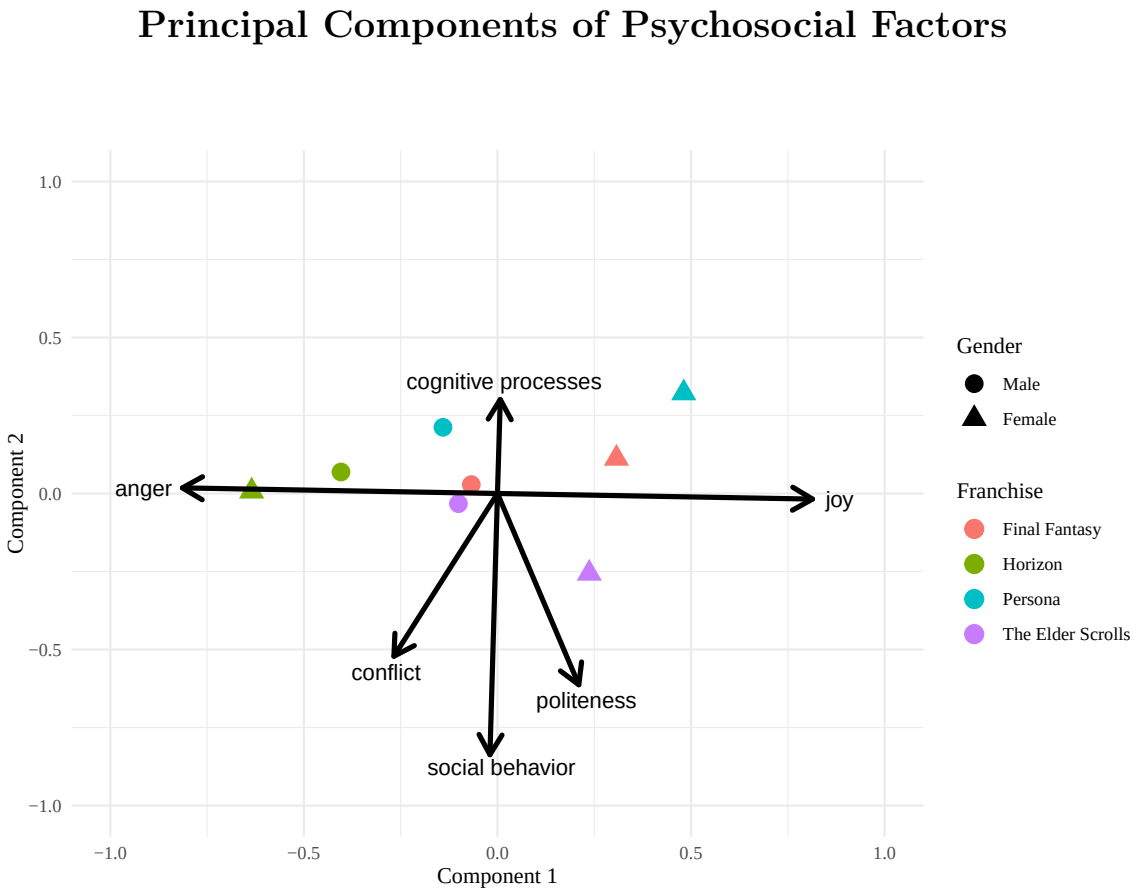

Figure 4: The color represents the franchise, and the shape represents the gender.

Furthermore, the logistic regression reveals that PC1 is a significant predictor of gender ($\beta = 0.155$, $z = 4.981$, $p < .001$), while PC2 is not ($\beta = -0.054$, $z = -1.515$, $p = .130$). Specifically, higher scores on PC1, which reflect greater use of language associated with anger relative to joy, are associated with a greater likelihood of a character being male.

6 Discussion

The initial analysis of basic linguistic features reveals important structural patterns in character dialogue across player and non-player roles. PCs speak significantly more lines, sentences, and words than NPCs, but their dialogue tends to be more concise, using fewer words per line and per sentence on average. This suggests that while PCs are more verbally central to the narrative, their speech is designed to be efficient and accessible for gameplay purposes. In terms of gender distribution, no statistically significant differences are found in basic linguistic metrics within either PC or NPC groups. However, Figure 1 indicates that male player characters speak more than female player characters. Additionally, proportion tests confirm a persistent gender imbalance. Male characters are significantly more numerous than female characters in both PC and NPC categories. This pattern is consistent with prior studies documenting the overrepresentation of male characters and the marginalization of female characters in video game narratives (Lynch et al., 2016; Rennick et al., 2023). These results highlight that gender asymmetries are embedded not only in character representation but also in the volume of linguistic participation, setting the foundation for more detailed psychosocial analyses.

Building on the fundamental analysis, LIWC analysis reveals significant gender differences in the psychosocial dimensions of character dialogue. Specifically, female characters, across both PC and NPC groups, use significantly more language related to cognitive reflection and social connection. In contrast, male characters, particularly NPCs, exhibit higher rates of swear word usage, indicating a stronger association with assertiveness, confrontation, and aggression. These findings align with prior research on gendered communication patterns, suggesting that male dialogue continues to emphasize dominance and externalized action, while female dialogue prioritizes internal reflection and relational communication (Behm-Morawitz & Mastro, 2009; Rennick et al., 2023).

Analysis of emotional expressions reinforces the gendered dynamics observed in psychosocial language use. Among NPCs, female characters express significantly less anger and significantly more sadness, disgust, fear, joy, and surprise compared to male characters. These

results are consistent with traditional gender stereotypes, where men are associated with externalized emotions such as anger, while women are depicted with a broader range of internalizing and relational emotions (Allen & Haccoun, 1976; Archer, 2004). However, when examining PCs, gender differences in emotional expression are much less pronounced. Moreover, in games featuring mixed-gender PC groups, emotional differences are further diminished, suggesting a deliberate narrative balancing of emotional expression among playable protagonists. These findings suggest that while background characters continue to reinforce conventional emotional gender norms, player-facing characters are increasingly designed to exhibit more balanced and inclusive emotional profiles, reflecting broader shifts toward more equitable character representation in contemporary game narratives.

PCA provides further insights into how psychosocial language patterns vary across gender and franchise. From a gender perspective, female characters are generally positioned toward the positive side of the emotional spectrum, while male characters are positioned toward the negative side. This pattern is further supported by logistic regression analysis, which reveals a significant association between the use of anger-related language and the likelihood of a character being male. This finding aligns with previous analyses showing that male characters are more associated with anger-related language, while female characters use more affectively positive expressions (Allen & Haccoun, 1976; Archer, 2004). A noteworthy exception is observed in the Horizon franchise, where female characters exhibit the highest association with anger across all gender groups and franchises analyzed. This pattern likely reflects the unique narrative positioning of Horizon's female protagonist, Aloy, who is designed to embody traits traditionally associated with male heroic figures. This phenomenon can be interpreted in two contrasting ways. On one hand, it may signal progress toward gender equality by presenting male and female protagonists with comparable personalities and capabilities. On the other hand, it may also reinforce the enduring norm that protagonists must conform to a masculine standard, sidelining qualities more often associated with femininity, such as compassion or patience, as narratively secondary or less heroic. In this view, gender parity is achieved not through a diversification of narrative traits but through the replication of existing masculine archetypes.

PCA also reveals distinct franchise-level differences in the balance between cognitive and social language features. *Persona* characters exhibit the highest emphasis on cognitive processes across all franchises. This is consistent with the its gameplay revolving around managing dual lives, confronting inner fears, and resolving deeply personal and social dilemmas (Williams, 2025a). In contrast, characters from *The Elder Scrolls* series are most strongly associated with social behavior language. *The Elder Scrolls* games are structured around open-world exploration where players interact with diverse factions, guilds, and political

powers (Williams, 2025b). Meanwhile, *Horizon* and *Final Fantasy* characters occupy a more balanced position between cognitive and social processes, reflecting narrative structures that emphasize both personal growth and relationship development (Alexander, 2025; Naletilic, 2023). These franchise differences indicate that the narrative architecture and thematic focus of a game influence how language is scripted along cognitive and social dimensions. Games that center identity exploration and internal conflict foster cognitively focused dialogue, while games emphasizing world-building and teamwork produce more socially oriented character interactions.

Beyond binary gender comparisons, the analysis also examines how neutral characters linguistically relate to male and female character groups. Using distance matrices based on LIWC features and emotional expression profiles, neutral characters are found to align most closely with male PCs across psychosocial dimensions. The consistent proximity to male characters, particularly male protagonists, supports the broader finding that linguistic neutrality often defaults to masculine norms in narrative design. Rather than establishing a distinct or balanced linguistic identity, neutral characters seem to inherit many of the linguistic patterns typically associated with male roles, reinforcing the phenomenon of masculine defaults observed in prior research (Cheryan & Markus, 2020). This suggests that even attempts at gender neutrality in game design may unintentionally reproduce male-coded norms, limiting the potential for genuinely inclusive character representation.

The findings of this study carry several important implications for both game design and broader conversations about gender representation in digital media. First, the persistence of gendered language patterns indicates that traditional gender norms continue to shape how characters are scripted. However, the reduction of these differences among player characters suggests a shift toward more balanced and inclusive design in protagonist roles. This trend may reflect an emerging recognition among developers that players expect nuanced, multidimensional characters regardless of gender. At the same time, the fact that neutral characters linguistically align more closely with male protagonists highlights the resilience of masculine defaults in game narratives. If left unexamined, these defaults risk reinforcing narrow standards for what constitutes a universal character. For researchers and developers alike, this underscores the importance of not only diversifying character design but also critically evaluating the language patterns embedded in dialogue, which shape players' perceptions of gender in subtle yet powerful ways.

7 Conclusion

This study contributes to the understanding of how gender is linguistically constructed in role-playing video games by computationally examining character dialogue through the lens of psychosocial factors. However, this study has several limitations that should be acknowledged. First, the dataset was restricted to a limited number of RPG titles, primarily drawn from a few countries and spanning a narrow range of time periods. As a result, the findings may not generalize to other game types, cultural contexts, or historical eras in game design. Second, while LIWC categories and emotion labels provide structured, interpretable features, they may not fully capture the complexity and nuance of in-game language, such as sarcasm, implications, or shifting tones. Third, the emotion classifier was trained on social media and television dialogue, which may not perfectly align with the linguistic style and structure of video game scripts. Finally, the gender labels used in this analysis were based on design documentation, which may not reflect how characters are actually perceived by players, particularly in cases involving ambiguous, customizable, or nonbinary identities.

Future research can address these limitations in several ways. Expanding the dataset to include a wider range of games across genres, cultural contexts, and historical periods would improve the generalizability of findings. Incorporating additional layers of character metadata, such as race and age, could offer deeper insights into how gender interacts with other identity dimensions. Developing or fine-tuning classification models on video game dialogue specifically would also enhance the validity of emotion detection. Furthermore, analyzing emotional trajectories across different points in the game narrative could reveal how character expression evolves with plot development. Finally, complementing computational approaches with qualitative methods, such as close readings or player interviews, would provide a richer, more context-sensitive understanding of how gendered language is constructed and perceived in interactive narratives.

Availability Statement

The data collected by Seán G. Roberts is available at GitHub,
https://github.com/seannyD/VideoGameDialogueCorpusPublic.git.
The code for project reproduction is available at GitHub,
https://github.com/naivetoad/MACSS-Thesis.git.
The LIWC-22 is commercially available at its official website,
https://www.liwc.app/.
The DistilRoBERTa model trained by Jochen Hartmann is available at Hugging Face,
https://huggingface.co/j-hartmann/emotion-english-distilroberta-base.

References

20-first. (2020). Top 14 Global Gaming Companies. *20-first*.

Alexander, C. (2025). How to Play the Final Fantasy Games in Order. *IGN*.

Allen, J. G., & Haccoun, D. M. (1976). Sex Differences in Emotionality: A Multidimensional Approach. *Human Relations*.

Archer, J. (2004). Sex Differences in Aggression in Real-World Settings: A Meta-Analytic Review. *Review of General Psychology*.

Arenas, D. L., Viduani, A., & Araujo, R. B. (2022). Therapeutic Use of Role-Playing Game (RPG) in Mental Health: A Scoping Review. *Simulation & Gaming*.

Atari. (1972). Pong. *Atari*.

Atlus. (2006). Persona 3. *Atlus*.

Behm-Morawitz, E., & Mastro, D. (2009). The Effects of the Sexualization of Female Video Game Characters on Gender Stereotyping and Female Self-Concept. *Sex Roles*.

Bethesda Game Studios. (2002). The Elder Scrolls III: Morrowind. *Bethesda Softworks*.

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). The Development and Psychometric Properties of LIWC-22. *University of Texas at Austin*.

Cheryan, S., & Markus, H. R. (2020). Masculine Defaults: Identifying and Mitigating Hidden Cultural Biases. *Psychological Review*.

Chess, S., & Consalvo, M. (2022). The Future of Media Studies is Game Studies. *Critical Studies in Media Communication*.

Crawford, G. (2009). Forget the Magic Circle (or Towards a Sociology of Video Games). *Under the Mask 2*.

Deterding, S., & Zagal, J. (2018). *Role-Playing Game Studies: Transmedia Foundations*. Routledge.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Human Language Technologies*.

Dietz, T. L. (1998). An Examination of Violence and Gender Role Portrayals in Video Games: Implications for Gender Socialization and Aggressive Behavior. *Sex Roles*.

Entertainment Software Association. (2023). 2023 Essential Facts About the U.S. Video Game Industry. *Entertainment Software Association*.

Fandom. (2025). *https://www.fandom.com/*.

Field, A., & Tsvetkov, Y. (2020). Unsupervised Discovery of Implicit Gender Bias. *ArXiv*.

Fox, J., & Tang, W. Y. (2014). Sexism in Online Video Games: The Role of Conformity to Masculine Norms and Social Dominance Orientation. *Computers in Human Behavior*.

Fron, J., Fullerton, T., & Morie, J. F. (2007). The Hegemony of Play. *Situated Play*.

Garriott, R. (1980). Akalabeth: World of Doom. *California Pacific Computer Company*.

Garriott, R. (1981). Ultima. *California Pacific Computer Company*.

Granic, I., Lobel, A., & Engels, R. C. M. E. (2014). The Benefits of Playing Video Games. *American Psychologist*.

Guerrilla Games. (2022). Horizon Forbidden West. *Sony Interactive Entertainment*.

Gygax, G., & Arneson, D. (1974). Dungeons & Dragons. *Tactical Studies Rules*.

Hartmann, J. (2022). Emotion English DistilRoBERTa-base. *Hugging Face*.

Hayat, M. J., & Higgins, M. (2014). Understanding Poisson Regression. *Journal of Nursing Education*.

Heritage, F. (2021). *Language, Gender and Videogames*. Springer International Publishing.

Hitchens, M., & Drachen, A. (2009). The Personal Experience of Narratives in Role-Playing Games. *Intelligent Narrative Technologies II*.

Hotelling, H. (1933). Analysis of A Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology*.

Jansz, J., & Martis, R. G. (2007). The Lara Phenomenon: Powerful Female Characters in Video Games. *Sex Roles*.

Jolliffe, I. T., & Cadima, J. (2016). Principal Component Analysis: A Review and Recent Developments. *The Royal Society*.

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology*.

Kent, S. L. (2001). *The Ultimate History of Video Games*. Three Rivers Press.

Kline, S., Dyer-Witheford, N., & Peuter, G. D. (2003). *Digital Play: The Interaction of Technology, Culture, and Marketing*. McGill-Queen's University Press.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *ArXiv*.

Lynch, T., Tompkins, J. E., van Driel, I. I., & Fritz, N. (2016). Sexy, Strong, and Secondary: A Content Analysis of Female Characters in Video Games Across 31 Years. *Journal of Communication*.

Naletilic, D. (2023). All 3 'Horizon' Games in Order. *Fiction Horizon*.

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender Differences in Language Use: An Analysis of 14,000 Text Samples. *Discourse Processes*.

O'Donnell, C. (2014). *Developer's Dilemma: The Secret World of Videogame Creators*. MIT Press.

Pennebaker, J. W., & Francis, M. E. (1999). Linguistic Inquiry and Word Count. *University of Texas at Austin.*

Petersen, S. (1981). Call of Cthulhu. *Chaosium.*

Pixar Animation Studios. (2015). Inside Out. *Walt Disney Studios Motion Pictures.*

Prager, R. H. P. (2019). Exploring the Use of Role-playing Games in Education. *The MT Review.*

Pulkkinen, J.-M. (2014). Design Values of Digital Role-Playing Games. *University of Tampere.*

Rennick, S., Clinton, M., Ioannidou, E., Oh, L., Clooney, C., T., E., Healy, E., & Roberts, S. G. (2023). Gender Bias in Video Game Dialogue. *Royal Society Open Science.*

Rennick, S., & Roberts, S. (2024). The Video Game Dialogue Corpus. *Corpora.*

Romrell, D. (2014). Gender and Gaming: A Literature Review. *AECT International Convention.*

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, A Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter. *ArXiv.*

Shaw, A. (2011). Do You Identify as A Gamer? Gender, Race, Sexuality, and Gamer Identity. *New Media & Society.*

Square. (1997). Final Fantasy VII. *Square.*

Statista. (2024). Video Game Industry Worldwide 2024. *Statista.*

Taito. (1978). Space Invaders. *Taito.*

The Unofficial Elder Scrolls Pages. (2025). *https://en.uesp.net/wiki/Main_Page.*

Weststar, J., & Lentini, A. (2024). Developer Satisfaction Survey 2023. *International Game Developers Association.*

Williams, C. (2025a). Every Persona Game and Spin-Off in Order. *IGN.*

Williams, C. (2025b). How to Play the Elder Scrolls Games in Chronological Order. *IGN.*

Wolf, M. J. P. (2002). *The Medium of the Video Game.* University of Texas Press.

Wolf, M. J. P. (2007). *The Video Game Explosion: A History from Pong to PlayStation and Beyond.* Greenwood Publishing Group.