Journal
of Liberal Arts
Prince of Songkla University

Characterization and Lexical Construction of Heroism in the Resident Evil Game Series: A Corpus-Based Analysis

ลักษณะตัวละครและการสร้างวีรภาวะผ่านคำศัพท์ในเกมซีรีส์เรสซิเดนต์อีวิล: การวิเคราะห์เชิงคลังข้อมูลภาษา

Received: July 4, 2026 **Revised:** September 1, 2026 **Accepted:** September 25, 2026

Kriangkrai Vathanalaoha[1]
เกรียงไกร วรรธนะเลาหะ

Abstract

Heroism is defined as the courageous actions of heroes or heroines. However, it is not just about valiant deeds but it is also reflected in the rhetoric that presents their characters as idiosyncratic, which leads the audience to perceive how the speaker wields leadership. Using the Resident Evil game series, this case study examined how verbal choices by protagonists contributed to representations of heroism within the survival horror context. Only selected utterances of Claire Redfield, Leon S. Kennedy, Jill Valentine, and Carlos Oliveira were extracted from the scripts of the remakes of Resident Evil 2 (RE2) (2019) and Resident Evil 3 (RE3) (2020) to form a dedicated corpus. The reference corpus used to show differences in style consisted of the concatenated speech from all four characters. The results revealed that Leon's hedging (*ask, believe*) reflected his institutional self-consciousness, while Carlos's rescue lexicon (*clear, rescue, get you to a hospital*) foregrounded his bravery and protectiveness. Claire's affect markers and swearing (*care, asshole, bastard*) signaled her emotional openness, and Jill's moral evaluatives (*greed, get out of here*) underlined her interpersonal skill. Moreover, the shared collocates (*you, I, get,* and *do*) suggested that heroism was constructed as second-person and action-oriented discourse. Consequently, the results obtained from the corpus-based analysis not only highlight the value of digital humanities in research but also validate literary interpretations through empirical substantiation.

Keywords: characterization, corpus linguistics, heroism, lexical construction, Resident Evil

[1]***Affiliation:*** Faculty of Liberal Arts, Thammasat University, Thailand
หน่วยงาน: คณะศิลปศาสตร์ มหาวิทยาลัยธรรมศาสตร์ ประเทศไทย

Corresponding Author: Kriangkrai Vathanalaoha
Email: kriangkrai.v@arts.tu.ac.th

บทคัดย่อ

วีรภาวะ เป็นคำที่นิยามขึ้นเพื่อแสดงถึงคุณลักษณะของวีรบุรุษหรือวีรสตรีผู้กระทำการอันกล้าหาญในเหตุการณ์ต่าง ๆ ทั้งนี้ การเป็นวีรบุรุษไม่ได้ขึ้นอยู่กับพฤติกรรมความกล้าหาญเพียงอย่างเดียว แต่ยังสามารถพิจารณาได้จากวัจนลีลาที่แสดงถึงบุคลิกภาพที่เป็นเอกลักษณ์เฉพาะตัว ช่วยให้ผู้รับสารสามารถเข้าถึงภาวะผู้นำของผู้พูดได้ งานวิจัยนี้ใช้เกมซีรีส์เรสซิเดนต์อีวิลเป็นกรณีศึกษาเพื่อสำรวจว่าการเลือกใช้ถ้อยคำของตัวเอกมีส่วนในการนำเสนอภาพแทนของวีรภาวะในบริบทของเกมแนวสยองขวัญเอาชีวิตรอดอย่างไร โดยผู้วิจัยได้ถอดวัจนภาษาของตัวเอกทั้งสี่คือ แคลร์ เรดฟิลด์, ลีออน เอส เคนเนดี, จิล วาเลนไทน์ และ คาร์ลอส โอลิเวียรา จากบทสคริปต์เกมฉบับทำใหม่ ได้แก่ เรสซิเดนต์อีวิล 2 (พ.ศ. 2562) และ เรสซิเดนต์อีวิล 3 (พ.ศ. 2563) เพื่อสร้างเป็นคลังข้อมูลภาษาเฉพาะกิจ ข้อมูลการพูดของตัวละครทั้งสี่ถูกนำมาประกอบกันเป็นคลังข้อมูลอ้างอิงเพื่อแสดงให้เห็นถึงความแตกต่างทางลีลาภาษา ผลการวิจัยพบว่าลีออนมักใช้ถ้อยคำแสดงความไม่แน่ใจ (เช่น *ask*, *believe*) ซึ่งสะท้อนความตระหนักรู้ในบทบาทเชิงสถาบันของตน คาร์ลอสใช้ถ้อยคำเชิงภารกิจและการช่วยเหลือ (เช่น *clear, rescue, get you to a hospital*) อันแสดงถึงความกล้าหาญและการปกป้องผู้อื่น ส่วนแคลร์ใช้ถ้อยคำแสดงอารมณ์ความรู้สึกและคำหยาบ (เช่น *care*, *asshole*, *bastard*) ซึ่งสะท้อนการเปิดเผยทางอารมณ์ ขณะที่จิลใช้ถ้อยคำเชิงประเมินค่าทางศีลธรรม (เช่น *greed*, *get out of here*) ซึ่งแสดงถึงทักษะการปฏิสัมพันธ์ นอกจากนี้ คำปรากฏร่วมที่ตัวเอกทั้งสี่ใช้ร่วมกัน (ได้แก่ *you*, *I*, *get* และ *do*) แสดงให้เห็นว่าวีรภาวะถูกประกอบสร้างขึ้นผ่านวาทกรรมที่มุ่งสู่บุรุษที่สองและเน้นการกระทำ ดังนั้น ผลการวิจัยจากการวิเคราะห์ด้วยวิธีคลังข้อมูลภาษาไม่เพียงแต่เน้นย้ำถึงความสำคัญของการใช้เครื่องมือดิจิทัลในงานวิจัยด้านมนุษยศาสตร์เท่านั้น แต่ยังช่วยยืนยันการตีความทางวรรณกรรมด้วยหลักฐานเชิงประจักษ์อีกด้วย

คำสำคัญ: ลักษณะตัวละคร ภาษาศาสตร์คลังข้อมูล วีรภาวะ การสร้างคำศัพท์ เรสซิเดนต์อีวิล

Introduction

Launched in 1998, Resident Evil 2 (RE2) is a survival horror game from Capcom that gained praise for its graphic horror scenarios and likable protagonists — Claire Redfield and Leon S. Kennedy — who become involved in an obscure viral outbreak caused by the evil pharmaceutical company Umbrella in the fictional Raccoon City. This game became a great success, leading to the release of the sequel, Resident Evil 3: Nemesis (RE3), in 1999. The sequel introduced two more playable characters, Jill Valentine and Carlos Oliveira, who must search for all remaining vaccines and evacuate the city before its destruction. Since then, these characters have become archetypal heroes for international gamers.

These characters are celebrated not only because they boldly fought against zombies and protected the citizens but also because their characteristics stand out through the dialogue they produced throughout the series. In fact, the dialogue spoken among fictional characters help in distinguishing their traits, as the audience builds a mental representation of a character mainly from what that character says (Culpeper, 2001; Culpeper et al., 1998; Herman, 1991). In other words, characterization is predicated on the analysis of what they say and do in dialogue. The study of characterization has also been approached from a corpus linguistics perspective, as illustrated by Busse's (2006) research on Shakespeare, which subsequently fed

into Culpeper's (2009) investigation of keywords for characters in Shakespeare's plays. McIntyre (2010) elaborated on the importance of corpus linguistic methodology for distinguishing types of criminals in *Reservoir Dogs* by Tarantino.

Sommers (2014) argued that Sarkeesian's (2014) claim of sexist language based on close reading of snippets from video game dialogue and narratives was questionable, and this attracted controversy after its debut in 2014. Sarkeesian's qualitative analysis is thought-provoking but points to the lack of empirical investigation in game dialogue research. Corpus-based approaches to dramatic and game dialogue have emerged as a means of addressing precisely this shortcoming. These studies incorporate various analytical methods to identify keywords, lexical bundles, and collocations to support literary interpretation and provide empirical evidence.

Content analyses of sexuality and gender issues have also investigated ludological linguistics (Kirkland, 2009; Martins et al., 2011). Most recently, however, corpus linguistic methods have begun to be directly applied to well-defined large-scale video game dialogue datasets, demonstrating the viability and efficiency of treating game scripts as analyzable corpora (Dixon, 2024; Heritage, 2025; Rennick et al., 2023; Rennick & Roberts, 2024). Two gaps yet remain in this line of research. First, corpus-based game studies have only worked at the level of the franchise or the demographic aggregate, and none of them has taken each protagonist as the unit of analysis. Second, the heroism of the game protagonists has still not been explicated in terms of verbal choices. This study aims to fill these gaps by analyzing the direct speech of the four protagonists to substantiate what audiences have perhaps already inferred as heroic characterization.

Literature Review

1. Heroes as Deviance-Based Perspectives

Carlyle (1840) coined the term heroism to refer to a quality that any person could embody and that could be celebrated ritualistically through devotion and worship, like that given to Jesus Christ as "the greatest of all heroes" (p. 249). A hero is sincere, even in his faults, and must avoid being a "slave and coward" (p. 268), endowed with an awesome "clarity of vision" (p. 281). Nietzsche (2014) disputed this perspective, insisting that "there are no heroes as long as God has any place in life and thus, they created Him to teach others to leave their ways of living behind, elevating them into the status of a God" (p. 9). This notion suggested that transcendent beings pave the way for other human beings and that heroism had a baseline between the extraordinary and the ordinary, which could "inspire believers" (Carlyle, 1840, p. 383).

Heroism was not something that belonged solely to men. Scholarship regarding mythology supported by Campbell (1949) suggested that a hero is "the man or woman" who fights against the limits of life or one's personal discord (p. 30). Since mythology worships gods and goddesses alike, gender should not be a basis for discrimination. Heroes were psychologically different from ordinary people, and this difference made them act, speak, and behave differently; for instance, they appeared as models of great courage but were not wholeheartedly accepted by their followers. Gestalt psychologist Polster (1992) pointed out another view, according to which heroes are said to act out of a deep respect for human life.

Their determination pushes them toward unimaginable strength, not simply because of their physical prowess.

Heroism, in its present perception, is not only limited to physical acts but also includes deviance both in verbal and non-verbal activity. For example, heroism may be expressed through pro-social leadership (Allison & Goethals, 2016), altruism (Monin et al., 2008), improvising ability (Rand & Epstein, 2014) and time perspective (Zimbardo & Boyd, 2008). Deviance-based perspectives argue that heroes must be risk-taking (Franco & Zimbardo, 2016), impulsive (Eysenck & Eysenck, 1978), or maybe even psychopathic (Pallone & Hennessy, 1998).

2. Corpus Approaches to Characterization and Video Game Dialogue

Both top-down and bottom-up processing are necessary to construct a mental representation of characters from the text, which is called "characterization" (Culpeper, 2001); thus, dialogue is highly relevant in dramatic texts. It is reliance on schemas, previous knowledge of the world, and conventions learned through fiction that supports top-down processing. In contrast, bottom-up processing receives linguistic cues from the text to form an image of a character. In actuality, characterization involves the simultaneous operation of both forms of processing (McIntyre, 2010), and linguistic indicators are crucial for comprehending the explicit or implicit attributes of characters. The latter, however, necessitates "style markers" (Enkvist, 1973), which need to be identified in a corpus-based approach using "keywords."

Corpus linguistics has been widely applied to examine characterization in various kinds of dramatic texts. Busse (2006) focused on vocative constructions in Shakespeare, and Culpeper (2009) adopted this system for a keyword analysis of individual characters in Shakespeare. Similarly, McIntyre (2010) used a corpus stylistic methodology with *Reservoir Dogs* by Tarantino to show how linguistic evidence can be used to discriminate character types. In a corpus approach, Vathanalaoha (2022) focused on drama dialogue in more recent plays. These works together demonstrate the fruitful nexus of corpus linguistics and literary characterization, a tradition which the current study applies to the interactive form of video games.

Considerable advances have taken place in the last two years in terms of methodologies applied to corpus-based video game research. Rennick et al.'s (2023) study, considered the first large-scale analysis of gender bias in video games, examined over 6.2 million words used by 13,587 characters and found that only one-third of the dialogue was spoken by female characters, even when controlling for the word types spoken, with dialogue from male characters nearly double that from their female counterparts — and this effect was consistent even across games where a woman takes center stage as a protagonist. Motivated by this idea, Rennick and Roberts (2024) formally introduced the Video Game Dialogue Corpus (VGDC), the first open-source, fully coded corpus of in-game dialogue, with custom formats to represent the branching, multilinear structures found in interactive game narratives. At about the same time, Dixon (2024) released a corpus of seven registers from digital role-playing games known as the Single Player Offline Game Corpus (SPOC), creating an analysis framework for how language varies systematically with ludic function, ranging from quest dialogue to item descriptions to in-game books. These corpus resources have enriched the analytical tools in the field of video game language study and lent confidence to the more niche corpus-based approach taken. Heritage's (2025) pragmatic analysis of corpus data from the LGBTQ Video Game Archive showed that identities of transgender characters are implicitly constructed

through implicature and pronoun use. Heritage's work is an example in which character identity is coded not by formulaic dictums but by reiterated speech habits and implicit construction of language for the game protagonist. Additionally, Nae (2021) revealed how player immersion and gendered characterization form are central to an analysis of the survival horror genre, with examples from games such as those in the Resident Evil (RE) franchise. Though his work offered insights into survival horror games, it still lacked an explanation of the relationship between verbal behavior and the formation of identity represented by each protagonist.

3. Heroic Characters in Resident Evil Series

Resident Evil 2 (RE2) (2019) is a survival horror game developed and published by Capcom. It serves as a remake of the original RE2, released in 1998. In this remake, the protagonists Claire Redfield and Leon S. Kennedy accidentally encounter a zombie outbreak caused by a viral incident in Raccoon City, a fictional city in which the story takes place. Resident Evil 3 (RE3) (2020) is the sequel, centered around elite agent Jill Valentine and supporting mercenary Carlos Oliveira as they attempt to find a vaccine for Raccoon City before the city is incinerated in a nuclear strike.

Each protagonist has a unique psychological and behavioral profile. Claire is a strong-willed, tomboyish woman who, unlike her brother Chris (who trained Claire), was not formally trained in combat. In the RE2 remake, she is stoic and quick to anger, particularly in her swearing and shouting, but affectionate towards Sherry and forgiving towards Annette before her demise. At first, Leon is a naive rookie cop desperate to do the right thing, following his duty zealously. His struggle to come across as confident when navigating the deadly police station is as vocal in the remake as before, but there is a playfulness that clashes with his more serious tone with Marvin or Ada. Carlos has a reputation for violence, but he is also kind-hearted and humorous; he is knowledgeable about weaponry and driving. Cheerful yet determined, with a great sense of justice, Jill can cope well in stormy conditions.

Research Questions

The research questions for this study are as follows:

RQ1: What kind of keywords appear for each protagonist in RE2 and RE3?

RQ2: How are collocates (the words that frequently occur in proximity to named characters) distributed among all protagonists in RE2 and RE3?

RQ3: How is heroic characterization reflected through the keywords and collocates identified in RE2 and RE3?

Research Methodology

This corpus-based research compiled the direct speech (DS) of four protagonists from the game scripts of RE2 (2019) and RE3 (2020) and analyzed it with a corpus analysis tool as follows:

1. Corpus Selection

The RE series was utilized for a corpus-based investigation for several reasons. Unlike other action franchises whose protagonists are rarely voiced, this series is dialogue-driven, as the protagonists keep speaking in cutscenes, in-game commentary, and wireless exchanges. These dialogues provide a large body of DS, which reveals each protagonist's attributes.

This RE series also presents male and female protagonists in comparable narrative roles, so heroic speech can be compared across characters without confusing the roles with gender. RE2 (2019) and RE3 (2020) were selected rather than other entries in the series as both remakes were produced and directed by the same studio and share the same background, timeframe, and outbreak event. Moreover, these two remakes offer four contrasting individuals: a civilian, a rookie police officer, an elite officer, and a hired mercenary. The scripts of earlier entries, including RE1, are much shorter in both monologues and dialogues and the inclusion of other entries may add extraneous factors that potentially affect the findings.

2. Corpus Design and Compilation

The four sub-corpora were constructed from the complete scripts of RE2 (2019) and RE3 (2020). There were no official transcripts from Capcom; therefore, the transcripts were directly recorded from full walkthroughs without the caster's comments to extract every line of spoken dialogue from all characters. The isolation of the DS of the four playable protagonists was verified twice by the researcher to ensure that each line was accurately assigned to each character. Only utterances explicitly spoken by the protagonists, including interjections and contracted forms of spoken utterances (e.g., *I'm*, *gonna*), were used in this study. Each cleaned sub-corpus was kept as a plain-text file (UTF-8) and uploaded to the analysis software (LancsBox). The sub-corpora comprised Claire Redfield (1,649 tokens), Leon S. Kennedy (1,532 tokens), Carlos Oliveira (971 tokens), and Jill Valentine (854 tokens), for a combined total of 5,006 tokens. To foreground the verbal choices that distinguish one protagonist from another, an internal reference corpus was built upon the combined speech of all four characters. Each character's sub-corpus was consequently compared against the reference corpus to explore keywords derived from each character. This design controlled for the shared vocabulary of the horror genre and highlighted individual characterization.

3. Analytical Tool: LancsBox

LancsBox (Brezina et al., 2021), a corpus analysis program developed at Lancaster University, was used as the principal analytical tool. This study used three analytical functions in the program. The Words/NGrams function calculates frequency and dispersion indices for unigrams and n-gram types. This function was run using its keyness display, in which each n-gram in the specialized corpus was compared with the same item in the reference corpus by the log-likelihood statistic. The Keyword-in-Context (KWIC) concordance provided the concordance lines, which were then used to analyze what each keyword and its co-occurring lexical environment reveal about the behavior of individual terms within character utterances. GraphColl is a module that builds and visualizes networks of collocations based on a mutual information statistic (MI $\geq$ 3.0 with minimum collocation frequency set at 2). In this module, the node is the search word around which the collocates are plotted, and a closer node indicates a stronger collocational association.

4. Analytical Procedure

LancsBox facilitated three procedures: keyword analysis, key n-gram analysis, and collocation network analysis. First, the keyword analysis was tabulated using the log-likelihood statistic at the threshold ≥ 3.84 ($p < 0.05$). Both positive keywords, over-represented in the character's speech, and negative keywords, which are under-represented in comparison with the other protagonists, were ranked by keyness value for interpretation. Both over- and under-represented keywords are essential not only to show what is foregrounded for each character but also what is uncommonly absent from their speech. Second, the key n-gram analysis was conducted using the same tool to investigate recurrent three- and four-word lexical bundles for each protagonist against the same reference corpus, and they were ranked in a similar fashion to the keyword analysis. Keyness for each n-gram was calculated by the same statistic used for the keyword analysis. Every ranked keyword and key n-gram was then examined through the Keyword-in-Context (KWIC) concordance to establish the semantic environment of each item through its left and right co-text. The concordances for both keywords and key n-grams were used to interpret the situations in which they are uttered, addressing RQ1. Third, the GraphColl tool created collocation networks to address RQ2. Collocates were tabulated for the four protagonists' names (*Claire*, *Leon*, *Jill* and *Carlos*), which served as the nodes for drawing the words occurring around each name. The mutual information (MI) statistic was applied, with a threshold ≥ 3.0, a minimum collocation frequency of 2, and a collocation span of five words to the left and right of the node (5L–5R). The closer nodes suggest stronger associations, while the farther ones suggest weaker associations. All the node distances were visualized to interpret the linguistic "aboutness" of heroism in the RE corpora.

5. Contextual Analysis and Analytic Consistency

Beyond frequency counts and keyword lists, in-game situational factors were used for contextual analysis. Every keyword and key n-gram found in the ranked lists was investigated in its KWIC concordance, which was used to trace back to the scenes in the recorded walkthrough. Each line was assigned to one of four situations: combat and threat, rescue and assistance, exploration and navigation, and non-interactive cutscene dialogue. For analytic consistency, no inter-coder reliability check was performed; however, two measures were adopted to maintain it. First, speaker attribution was conducted twice, in two passes separated by a one-week interval to ensure that only utterances spoken by a protagonist were included, that interjections and contracted forms were retained and that non-verbal vocalizations were excluded. Second, each interpretive claim was grounded in concordances and could be traced to a retrievable item in the corpus.

Findings

The results tabulated by LancsBox are presented in three subsections: keyword analysis per protagonist, key n-gram analysis per protagonist, and collocation network analysis for all protagonists. Positive keywords (+) are those items that appear unusually frequently in the specialized sub-corpus compared with the reference corpus, while negative keywords (–) are unusually under-represented in comparison.

1. Keyword Analysis

1.1 Carlos Oliveira

Table 1

Keywords (+ and –) for Carlos Oliveira

No.	Keywords (+)	No.	Keywords (–)
1	*Jill*	1	*what*
2	*bard*	2	*Sherry*
3	*clear*	3	*I*
4	*she's*	4	*oh*
5	*Subway's*	5	*yeah*
6	*Carlos-less*	6	*you*
7	*comin'*	7	*Marvin*
8	*cruel*	8	*I'm*
9	*head*	9	*hey*
10	*killed*	10	*no*
11	*he's*		
12	*nah*		
13	*rescue*		
14	*shelter*		
15	*street*		
16	*substation*		
17	*thought*		
18	*we'll*		
19	*ahead*		
20	*whoa*		

Despite Carlos's main mission to find Dr. Bard, the scientist who invented the vaccine, the positive keywords, as shown in Table 1, were mostly Jill-oriented, with *Jill* being the highest item on the list, followed by action-oriented items like *rescue*, *clear*, *shelter*, *street* and *substation*. The colloquial markers *comin'*, *nah*, and *whoa* correspond to an informal sociolect and a youthful persona. In addition, words like *cruel* and *killed* highlight the moral dimension of his character's voice. His negative keywords — most interestingly *I*, *hey*, *no*, and the contracted form *I'm* — signal a speech style that is less self-referential and less interactionally casual than the reference corpus baseline. These negative keywords suggest that he rarely focuses on his well-being and that he infrequently rejects victims' requests.

1.2 Claire Redfield

Table 2

Keywords (+ and –) for Claire Redfield

No.	Keywords (+)	No.	Keywords (–)
1	*Sherry*	1	*down*
2	*care*	2	*leave*
3	*Chris*	3	*stay*
4	*mom*	4	*this*
5	*it's*	5	*it's*
6	*why*	6	*good*
7	*asshole*	7	*out*
8	*both*	8	*me*
9	*check*	9	*going*
10	*anybody*	10	*hey*
11	*keep*		
12	*Kennedy*		
13	*long*		
14	*looking*		
15	*looks*		
16	*more*		
17	*run*		
18	*save*		
19	*orphanage*		
20	*Sherry's*		

Claire's positive keywords, as shown in Table 2, leaned heavily on proper nouns: Sherry, Chris, mom, and Kennedy, a cue that her world revolves around key relationships. Lexical items such as *care*, *save*, *run* and *check* manifest her protective behavioral profile. Her lexical register demonstrates informal and expressive choices, such as *asshole*, among her high-frequency keywords. These words further suggest Claire's evaluative and interrogative stance, as seen in *it's*, *why*, *both* and *looking*. Her negative keywords (all less frequent than in the reference corpus) were *down*, *leave*, *stay*, *good* and, finally, her most negative word is *out*.

1.3 Jill Valentine

Table 3

Keywords (+ and –) for Jill Valentine

No.	Keywords (+)	No.	Keywords (–)
1	*Carlos*	1	*Jill*
2	*Brad*	2	*Sherry*
3	*am*	3	*Leon*
4	*greed*	4	*Claire*
5	*Nicholai*	5	*what*
6	*after*	6	*this*
7	*out*	7	*going*
8	*here*	8	*me*
9	*get*	9	*good*
10	*you*	10	*ok*

In the case of Jill, her keyword profile, as shown in Table 3, highlights the social and ethical aspects through which she might be characterized. In short, her top keyword, *Carlos*, demonstrates her fixation on her partner in RE3. *Brad* and *Nicholai*, which name villains and helpers, place Jill more clearly in a network of interpersonal relationships closely connected to her professional ones. The keyword greed also appeared here, indicating that Jill is expressing moral judgement — a typical behavior of the character-driven agent with a sense of justice. The word *am* may represent an assertive, self-identifying Jill in direct speech, while the word *after* may index pursuit-oriented narrative sequences.

1.4 Leon S. Kennedy

Table 4

Keywords (+ and –) for Leon S. Kennedy

No.	Keywords (+)	No.	Keywords (–)
1	*Ada*	1	*Sherry*
2	*gun*	2	*Carlos*
3	*lieutenant*	3	*Jill*
4	*FBI*	4	*ok*
5	*open*	5	*stay*
6	*ask*	6	*leave*
7	*believe*	7	*this*
8	*Claire*	8	*yeah*
9	*Marvin*	9	*hey*
10	*raccoon*	10	*out*

Table 4 reveals that Leon's positive keyword profile includes institutional vocabulary (*lieutenant*, *FBI*) and relational proper nouns (*Ada*, *Claire*, *Marvin*), demonstrating his fixation on both authority and personal connection. *Gun* and *raccoon* are contextually pinned to the game's immediate scene. Most significantly, *ask* and *believe both rose* to the surface as important keywords highlighting an epistemic verbal style in which Leon asks for confirmation and expresses doubt. It is these traits that corroborate his characterization as a self-conscious rookie officer having to negotiate institutional authority and to face exceptional circumstances.

2. Key N-Gram Analysis

2.1 Carlos Oliveira

Table 5

Key 3-Grams and 4-Grams for Carlos Oliveira

No.	Key 3-Grams	No.	Key 4-Grams
1	*oh my god*	1	*get you to a hospital*
2	*where are you*	2	*gotta be kidding me*
3	*what the hell*	3	*I'll be right there*
4	*be right there*	4	*oh my god Jesus*
5	*gotta be kidding*	5	*what the hell is*
6	*get you to*	6	*you gotta be kidding*
7	*I'll be right*	7	you go on ahead
8	*the hell is*	8	*never seen anything like*
9	*to a hospital*	9	*you to a hospital*
10	*you gotta be*	10	*be right there Jill*

As Table 5 shows, Carlos's lexical bundles highlight exclamatory and interrogative language with dominant entries *oh my god*, *what the hell* and *gotta be kidding me*, which show how he encounters unexpected predicaments during the zombie attack. In addition, his verbal behavior includes a motif of rescue-related phrases, *get you to a hospital*, *I'll be right there* and *be right there*, underlining his status as a protective call-to-action hero.

2.2 Claire Redfield

Table 6

Key 3-Grams and 4-Grams for Claire Redfield

No.	Key 3-Grams	No.	Key 4-Grams
1	*hold on Sherry*	1	*what the hell is*
2	*you hurt her*	2	*bastard, if you hurt*
3	*a way out*	3	*found a way out*
4	*are you ok*	4	*get back to Sherry*
5	*back to sherry*	5	*I'll be right there*

No.	Key 3-Grams	No.	Key 4-Grams
6	*bastard if you*	6	*you gotta be kidding*
7	*do you need*	7	*alright I swear you*
8	*alright i swear*	8	*do you need help*
9	*to a hospital*	9	*are you ok sherry*
10	*I swear you*	10	*If you hurt her*

Claire's verbal sequences in Table 6 revolve around Sherry: *hold on, Sherry*, *back to Sherry* and *are you ok, Sherry*, which establish the child's safety as the most important concern of her dialogue. The 3-gram *you hurt her* and its associated 4-gram *bastard, if you hurt* reflect Claire's determination to protect the child and demonstrate her ad hoc use of aggressive language.

2.3 Jill Valentine

Table 7

Key 3-Grams and 4-Grams for Jill Valentine

No.	Key 3-Grams	No.	Key 4-Grams
1	*oh my god*	1	*get you to a*
2	*where are you*	2	*gotta be kidding me*
3	*what the hell*	3	*I'll be right there*
4	*out of here*	4	*alright, I swear you*
5	*be right there*	5	*you to a hospital*
6	*get out of*	6	*what the hell is*
7	*get you out*	7	*get out of here*

As can be seen in Table 7, Jill has several verbal sequences similar to those of Carlos, especially *oh my god*, *where are you*, *what the hell*, and *be right there*, which imply that these two characters share a repertoire of formulaic expressions to some extent. This suggests that they have the strongest narrative relationship in RE3, as they are part of a cooperative journey. Escape and evacuation are among the dominant features establishing Jill's verbal action in both 3-gram and 4-gram patterns, as seen in the foregrounding phrases *out of here*, *get out of* and *get you out*.

2.4 Leon S. Kennedy

Table 8

Key 3-Grams and 4-Grams for Leon S. Kennedy

No.	Key 3-Grams	No.	Key 4-Grams
1	*about me just*	1	*gonna leave you here*
2	*don't even know*	2	*just gonna leave you*
3	*gonna leave you*	3	*not just gonna leave*
4	*forget about me*	4	*forget about me just*
5	*what the hell*	5	*what the hell was*
6	*leave you here*	6	*the hell was that*
7	*not just gonna*	7	*this is not how*

As Table 8 shows, Leon's lexical bundles are quite different from those of the other three protagonists. The high frequency of *gonna leave you (here)* and of its key 4-grams, *just gonna leave you* and *not just gonna leave*, indicates a consoling speech style in which Leon expresses his intention not to abandon others. The 4-gram *forget about me just* demonstrates the presence of self-deprecating or self-sacrificial language as he is obliged to respond as a policeman. As Leon is a newcomer to the Raccoon City Police Department (RPD), his speech also contains frequent expletives in ominous situations.

3. Collocation Network Analysis (GraphColl)

Table 9

Shared Collocates Across All Four Protagonists (MI ≥ 3.0, Min. Frequency = 2)

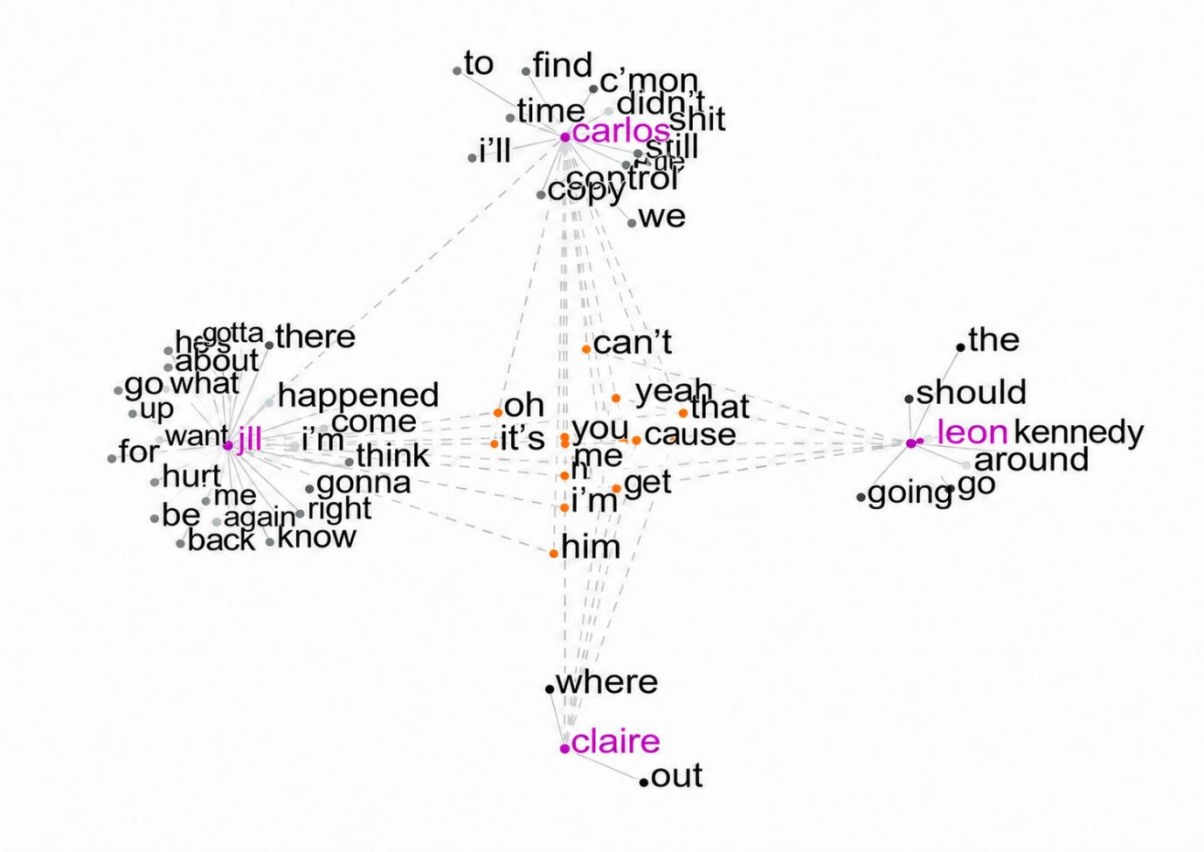
No.	Collocate (Frequency)	No.	Collocate (Frequency)
1	*you (228)*	10	*do (32)*
2	*i (114)*	11	*can (24)*
3	*that (71)*	12	*you're (24)*
4	*it (65)*	13	*oh (22)*
5	*me (64)*	14	*yeah (22)*
6	*get (51)*	15	*Jill (21)*
7	*it's (43)*	16	*can't (20)*
8	*I'm (42)*	17	*him (17)*
9	*are (40)*	18	*Carlos (16)*

As shown in Table 9 and Figure 1, the GraphColl analysis identified the collocates shared across all four protagonists, representing the collective linguistic "aboutness" of heroism in the RE corpus. The term *you* was the most frequent shared collocate (228 occurrences), followed by *I* (114), *that* (71), *it* (65), *me* (64), *get* (51), *it's* (43), *I'm* (42) and *are* (40). The remaining shared items, also in descending order, were do (32), *can* (24), *you're* (24), *oh* (22), *yeah* (22), *Jill* (21), *can't* (20), *him* (17) and *Carlos* (16). These collocates suggest heroism expressed through

modalities of certainty (*can*, *can't*), action verbs (*get*, *do*), and colloquial words (*yeah*, *oh*). Moreover, the use of *I* and *you* throughout RE2 and RE3 represents friendship or intimacy revolving around the protagonists.

Figure 1

Shared Collocates Among Four Protagonists



Discussion

The corpus-based analysis presents empirical evidence of how the four protagonists embody different images of heroism while exhibiting shared linguistic features representative of the narrative conventions of the survival horror genre. Each of the three research questions is answered in turn in the following subsections.

1. Individual Keyword and N-Gram Characterization (RQ1)

The keywords of the four protagonists show how heroic features are lexically constructed across character types, substantiating earlier findings in corpus-based characterization studies (Culpeper, 2009; McIntyre, 2010).

On the whole, the keyword profile for Jill Valentine suggests that she is an interpersonally oriented character who focuses on her relationships with other people. Given her sense of morality, it is also notable that *greed* appears as one of the key findings, which indicates Jill's concerns about moral principles, and this aligns with her status as an elite agent with a strong sense of justice (Allison & Goethals, 2016). This is underscored by the

evacuation-themed lexical bundles (*out of here, get out*), which contribute to a female heroic quality of liberation and protection.

As for Carlos Oliveira, the keywords mostly concern Jill, liberation, and human morality. This pattern fits accounts of heroic action (Franco & Zimbardo, 2016), in which acts of risk-taking are verbalized in danger-laden and commitment-laden language. His lexical bundles establish what could be called a "rescue lexicon", a set of multi-word phrases that accomplish the daring work of reassurance and action.

The keywords of Leon S. Kennedy characterize him as a figure defined by institutional authority because he needs to collaborate with a mysterious woman, Ada, who is monitored by the Umbrella organization. The institutional terminology highlights Leon's self-awareness of being a rookie, while *ask* and *believe* mirror an interpersonal processing style in which he tries to make sense of unfamiliar supernatural events. This concurs with Herman's (1991) concept of verbal cues that can shape fictional characters, as the way Leon speaks clearly exposes his uncertainty and self-guilt. In addition, his repeated phrases with the internal rhyme *gonna leave you (here)* encode a moral implication of refusal to desert others, affirming a heroic agency.

For Claire Redfield, who has the most relationally and emotionally rich keyword profile of the four protagonists, the visibility of *Sherry*, *Chris*, *mom*, *orphanage*, and *Sherry's* establishes her as the most socially grounded figure. Her heroism is expressed not through fighting monsters or risking personal harm, but through her caring for others in a protective fashion. The combination of affective terms with expressive swearing (*asshole*, *bastard*, and *the hell*) shows that there is a tension between her rough, tomboyish appearance and an inner nurturing spirit; a dichotomy in characterization that is consistent with Campbell's (1949) thesis about heroism being rooted in deep respect for all human life regardless of its outer manifestations. Her 3-gram *you hurt her*, and the 4-gram *bastard, if you hurt*, substantiate this dichotomy in a combatant statement combining protective love with violent means.

2. Common Collocates and the Narrative Architecture of Heroism (RQ2)

The results revealed that *you* (228) was the most prominent shared collocate among these four protagonists, as shown by the full collocation network analysis. This finding is indicative of the second-person narrative structure common to role-playing games, where character speech often orients itself externally towards other characters. This pervasive second-person address can underpin the game's "choose-your-own-adventure" quality in its narrative, immersing the player directly in the diegetic world and framing each protagonist as a guide or partner within the interactional architecture of each game. This reflects the most recent observations of Rennick and Roberts (2024), who state that these dialogue patterns appear as genre-specific interactional norms in role-playing games (RPGs).

Concurrently, the shared collocate profile contained a very high frequency of *are*, mirroring the genre's focus on exploration and navigation. These linguistic realizations usually correspond to the relational attributive processes described in Halliday's (1994) Systemic Functional Grammar that unite places and people (*it is the subway*, *that's where they are*), which are typified by this genre's concern with mapping menacing locations and marking assailants. This locative function of the verb *be* further agrees with Nae's (2021) account of

survival horror games in terms of their "spatial narrative logic" (p. 235), where movement in dangerous space becomes a dominant feature of the ludic and diegetic content.

The most frequent action verbs, such as *do* and *get*, in the shared collocates instantiate material processes, which further represent an action-oriented linguistic register that replicates the character movements of the game during play. They also encode a modal obligation and directional imperative (*you can go, get out, do this now*), which reinforce the protagonists' agency as survivalist agents of rescue and directive action. The fact that all four characters share these items highlights that material agency is a fundamental linguistic characteristic of RE heroic discourse, recurring from one character to another. In particular, this genre's reliance on material processes is consistent with Dixon's (2024) register analysis of RPG language which found action-oriented utterances to be among the most distinctive registers in single-player game discourse.

3. Keywords, Collocates and Heroic Characterization (RQ3)

Linguistic evidence from keywords and collocates can be considered evidence of the discursive construction of heroism in the RE game series. It reveals two broad linguistic trends as follows:

First, heroic action is lexically realized as the movement of another person out of danger. The item *out* and the sequences *get you out*, *out of here*, and *get out of* express a literal spatial relation between a person and a hazardous location. These sequences are mostly other-directed: the person who is moved out is the addressee, not the speaker. This other-directedness can be related to the altruistic component of heroism, which is considered a sacrificial action taken at personal cost for the benefit of another (Monin et al., 2008; Rand & Epstein, 2014). This spatial signal of heroic action is consistent with Nae's (2021) account of the narrative logic of survival horror as expressed through the diegesis of gameplay.

Second, character-specific and shared lexical features are parallel to Culpeper's (2001) categorization of explicit and implicit characterization. Explicit characterization can be shown through named references, such as *Sherry*, *Ada*, and *Jill*, because they directly refer to the characters' physical appearances and vocal varieties. On the other hand, implicit characterization arises from stylistic regularities, as seen in Leon's hedging, Carlos's mercenary lexicon, Claire's affect and colloquial swearing, and Jill's unapologetic moral phrases. This finding aligns with the suggestion that identity in video game discourse is often implicitly represented through words that one opts to use rather than being overtly demonstrated (Heritage, 2025). The lexical patterns analyzed by corpus-based methodology can help render these implicit characteristics visible and quantifiable.

4. Contributions and Implications to the Field

The findings offer several contributions to the study of characterization and video-game language. Firstly, the study reaffirms that keyword, n-gram, and collocation analysis, traditionally applied to drama and prose by Culpeper (2009) and McIntyre (2010), can be applied to the interactive, multimodal text of video games. Moreover, as opposed to qualitative interpretation through verbal cues, quantifying characterization against a reference corpus, as addressed by Sarkeesian (2014) and Sommers (2014), makes the analysis empirical and replicable. Secondly, this study can enhance Culpeper's (2001) distinction between explicit

and implicit characterization, as the findings can uncover aspects that might have been overlooked by qualitative linguistic scholars. In other words, the findings regarding process types, modal verbs, and orientational metaphors that are vital for implicit characterization could only be made visible through corpus-based analysis, showing recurrent and stronger associations with each protagonist. Thirdly, the findings can contribute to the psychology of heroism in game studies by demonstrating abstract attributes through observable language. For instance, the prominent use of material and relational processes, alongside spatialized metaphors of escape, collectively describes how heroism is perceived through direct speech acts involving the protection and movement of others through dangerous spaces. These findings are underpinned by linguistic evidence in survival-horror heroic accounts (Franco & Zimbardo, 2016). However, they also yield a nuance showing that not all heroes are purely action-oriented but can also be caring and humanistic, as seen in Claire's verbal cues.

Conclusion

The application of LancsBox to the analysis of the characterization of heroes has been demonstrated through the compilation of direct speech from the four protagonists of the RE series. The analysis based on keywords, key n-grams, and collocation network suggested that heroism was a foregrounding linguistic structure produced by individual verbal input as well as discourse properties shared among the protagonists. The keyword analysis pointed to individual personality traits; for instance, Jill is characterized as reasonable and interpersonally skilled; Carlos as protective and brave; Leon as having institutional self-consciousness and coping with guilt and emotional burden; and Claire as courageous and empathetic towards victims. The collocate network illuminated the shared narrative architecture of survival horror texts through second-person narrative (*you*), material processes, relational processes, and orientational spatial metaphors.

The findings provide empirical support for previous literary readings of the RE series while addressing the methodological criticism of game dialogue scholarship concerning the balance between qualitative and quantitative approaches. Additionally, the research highlights the importance of digital humanities methodologies for researching corpus stylistics, signaling a new branch of studies extending from traditionally dramatic text-based corpus studies (Culpeper, 2009; McIntyre, 2010) to video games as an interactive medium. The paper contributes to corpus research on game dialogue (Dixon, 2024; Heritage, 2025; Rennick et al., 2023; Rennick & Roberts, 2024), showing that video game scripts are a promising and accessible source of linguistic information that is important for corpus-based methodology and discussion in the field of game studies.

Several limitations should be acknowledged. The corpus focuses on two game remakes, restricting the generalizability of the findings to other entries in the franchise known to RE fans. Besides, the sub-corpora of Jill Valentine and Carlos Oliveira have smaller token counts, which might weaken the keyword statistics. Future studies may widen the scope to include the entire RE game series and non-protagonist characters, apply a semantic domain analysis using tools such as Wmatrix (Rayson, 2008), or exploit larger resources like the Video Game Dialogue Corpus (Rennick & Roberts, 2024) or SPOC (Dixon, 2024) as comparable reference corpora. Future work involving this methodology across different game genres or franchises would strengthen the ecological validity of corpus-based game dialogue analysis. Ultimately,

the framework defined here provides a basis for repeatable and scalable integration of corpus linguistics into ludological and digital humanities research.

References

Allison, S. T., & Goethals, G. R. (2016). Hero worship: The elevation of the human spirit. *Journal for the Theory of Social Behaviour*, *46*(2), 187-210. https://doi.org/10.1111/jtsb.12094

Brezina, V., Weill-Tessier, P., & McEnery, A. (2021). *#LancsBox v. 6.x* [Software package]. http://corpora.lancs.ac.uk/lancsbox

Busse, B. (2006). *Vocative constructions in the language of Shakespeare*. John Benjamins. https://doi.org/10.1075/pbns.150

Campbell, J. (1949). *The hero with a thousand faces*. Pantheon Books.

Carlyle, T. (1840). *On heroes, hero-worship, and the heroic in history*. James Fraser.

Culpeper, J. (2001). *Language and characterisation: People in plays and other texts*. Longman.

Culpeper, J. (2009). Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's Romeo and Juliet. *International Journal of Corpus Linguistics*, *14*(1), 29-59. https://doi.org/10.1075/ijcl.14.1.03cul

Culpeper, J., Short, M., & Verdonk, P. (Eds.). (1998). *Exploring the language of drama: From text to context.* Routledge.

Dixon, D. H. (2024). Introducing the single player offline game corpus (SPOC): A corpus of seven registers from digital role-playing games. *Corpora*, *19*(1), 107-122. https://doi.org/10.3366/cor.2024.0300

Enkvist, N. E. (1973). *Linguistic stylistics*. Mouton.

Eysenck, H. J., & Eysenck, S. B. G. (1978). Impulsiveness and venturesomeness: Their position in a dimensional system of personality description. *Psychological Reports*, *43*(3), 1247-1255. https://doi.org/10.2466/pr0.1978.43.3f.1247

Franco, Z., & Zimbardo, P. G. (2016). The psychology of heroism: Extraordinary champions of humanity in an unforgiving world. In A. G. Miller (Ed.), *The social psychology of good and evil* (pp. 494-523). Guilford Press.

Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). Arnold.

Heritage, F. (2025). Social identity and implicature: Exploring the pragmatics of transgender coming out narratives in videogames. *Journal of Pragmatics*, *246*, 158-169. https://doi.org/10.1016/j.pragma.2025.07.007

Herman, V. (1991). Dramatic dialogue and the systematics of turn-taking. *Semiotica*, *83*(1-2), 97-121. https://doi.org/10.1515/semi.1991.83.1-2.97

Kirkland, E. (2009). Masculinity in video games: The gendered gameplay of Silent Hill. *Camera Obscura, 24*(2), 161-183. https://doi.org/10.1215/02705346-2009-006

Martins, N., Williams, D. C., Ratan, R. A., & Harrison, K. (2011). Virtual muscularity: A content analysis of male video game characters. *Body Image*, *8*(1), 43-51. https://doi.org/10.1016/j.bodyim.2010.10.002

McIntyre, D. (2010). Dialogue and characterisation in Quentin Tarantino's Reservoir Dogs: A corpus stylistic analysis. In D. McIntyre & B. Busse (Eds.), *Language and style* (pp. 221-243). Palgrave Macmillan.

Monin, B., Sawyer, P. J., & Marquez, M. J. (2008). The rejection of moral rebels: Resenting those who do the right thing. *Journal of Personality and Social Psychology*, *95*(1), 76-93. https://doi.org/10.1037/0022-3514.95.1.76

Nae, A. (2021). *Immersion, narrative, and gender crisis in survival horror video games*. Routledge.

Nietzsche, F. (2014). *Also sprach Zarathustra (Auswahl)*. (in German) [Thus spoke Zarathustra (selections): A dual-language book] (S. Appelbaum, Ed. & Trans.). Dover Publications.

Pallone, N. J., & Hennessy, J. J. (1998). Brain dysfunction and criminal violence. *Society*, *35*(6), 21-27. https://doi.org/10.1007/BF02686049

Polster, E. (1992). *A population of selves: A therapeutic exploration of personal diversity*. Jossey-Bass.

Rand, D. G., & Epstein, Z. G. (2014). Risking your life without a second thought: Intuitive decision-making and extreme altruism. *PLOS ONE, 9*(10), Article e109687. https://doi.org/10.1371/journal.pone.0109687

Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, *13*(4), 519-549. https://doi.org/10.1075/ijcl.13.4.06ray

Rennick, S., Clinton, M., Ioannidou, E., Oh, L., Clooney, C., E. T., Healy, E., & Roberts, S. G. (2023). Gender bias in video game dialogue. *Royal Society Open Science*, *10*(5), Article e221095. https://doi.org/10.1098/rsos.221095

Rennick, S., & Roberts, S. (2024). The video game dialogue corpus. *Corpora, 19*(1), 93-106. https://doi.org/10.3366/cor.2024.0299

Sarkeesian, A. (2014, August 25). *Women as background decoration: Part 2* [Video]. Feminist Frequency. https://feministfrequency.com/video/women-as-background-decoration-part-2/

Sommers, C. H. (2014, September 17). *Are video games sexist?* [Video]. YouTube. https://www.youtube.com/watch?v=9MxqSwzFy5w

Vathanalaoha, K. (2022). Corpus stylistics in contemporary English dramas: Keywords and semantic fields of delusions. *GEMA Online Journal of Language Studies*, *22*(2), 43-62. https://doi.org/10.17576/gema-2022-2202-03

Zimbardo, P. G., & Boyd, J. N. (2008). *The time paradox: The new psychology of time that will change your life*. Free Press.