

---

## Representation of Gender-Based Violence Words in Genshin Impact

*I.G.N. Adhimas Maharditya*

Universitas Airlangga  
ign.adhimas.maharditya-2021@fib.unair.ac.id

Received: 3<sup>rd</sup> January 2023 | Revised: 17<sup>th</sup> May 2023 | Accepted: 23<sup>rd</sup> July 2023  
Email Correspondence: ign.adhimas.maharditya-2021@fib.unair.ac.id

### Abstract

This paper aims to examine a representation of gender-based violence words inside videogames. This study focused on processing textual data inside videogames, where Genshin Impact's in-game textual data is the chosen object of research. In previous researches regarding modern media, gender male was always related to violence words more than female. Uncovering the result from comparison of violence words used Genshin between male and female is the main goal of this study. Genshin was chosen because the dominant number of female characters (33) compared to male characters (17), which piqued a curiosity if this game is able to change the culture where female characters can be more related to violence words. Mixed-method analysis is used as the core method of this research, as aside of statistical analysis by corpus software, culture and social aspects are also analyzed based on statistical result. The result shown that despite big difference in character count, male characters still have more relation to violence words than female characters by around 2%.

**Keywords:** Corpus linguistic, violence words, gender-based violence words, Genshin impact

### 1. INTRODUCTION

Videogames nowadays are no longer about moving small number of pixels with little interactions, but also includes many rows of interactions between characters inside the game. This situation enables corpus-based researches, such as ludolinguistic, which is a term for corpus linguistic focused on videogames. Researches related in corpus videogames are still scarce, especially researches related to videogames' textual data (Heritage, 2021). Gendered violence has been included in a lot of issues related to social, political and cultural topic. The topic has spread into a lot of fields and the discussions are not only become a real-life conversation, but also online discussions on multiple platforms. The field of video games also cannot escape from discussion about violence related to gender, whether it is a conversation or chat between online game players, video game textbooks and textual data of video game (Baker, 2013), which is now easily obtained thanks to official and fan-made sites which provide the data. However,

---

statistical data and qualitative research of this circumstance are still scarce, let alone the combination of it. In video games, unfortunately there are still few researches regarding gender representation. Various researchers have investigated the representation of gender and sexuality via large scale content analyses. Several scholars have also examined how gender and sexuality are represented in texts around videogames, such as in magazine articles, player-to-player communication, websites, and interviews with players/content and many more (Heritage, 2021). Videogames are no longer limited to simple aspects like arcade games or simple pixel games like Pacman or some 1980s videogames. Modern videogames carry complex layers of meaning, which always reflect a certain set of ideologies about society and its power relationships (Ensslin, 2012).

Regarding this research, however, these studies mainly focus on visual aspects relating to gender, such as body size, the degree to which women are visually sexualized, or player's experiences. In this research the focus was on video games' texts. The linguistic analysis based on video game texts called Ludo linguistic. (Heritage, 2021) argues that examining on how gender is represented at a linguistic level within videogames as texts is still needed beside examining how gender is represented in videogames at a visual level. Corpus linguistic will be used as this research's method. (Heritage, 2021) continues that corpus softwares can reveal information which might not have been visible by manual analysis alone – such as by demonstrating what words are statistically likely to occur with what other words, or what words are statistically likely to appear in one videogame in comparison to language in general.

Based on previous ludo linguistic researches, (Heritage & Koller, 2020) argue that there are three types of video game data that can be examined: the first is the language used in videogame Para text, like tutorial books and magazines related to the video game. The next is player-interactions, which gathered through netno/ethno-graphic methods. Finally, the third and currently less frequent is analyses of videogames as textual sources of. The third type is what the focus of this research. Researching natural language processing (NLP) from videogames can also be done with corpus as well (Hämäläinen et al., 2022).

The object for this research is the textual data inside video games Genshin Impact. This game is released in September 2020 which appeared as one of the most popular video games in current years (2020-2022). The reason of this choice is not only because of the game's popularity, but also due to this Genshin's huge number of characters and texts from a lot of sources. The entire textual data from Genshin Impact will be examined via corpus linguistic softwares, LancsBox and AntConc to find the gender representations. This research is expected to increase the knowledge of corpus linguistics especially on word choice and reveal how particular words (including noun, verb, and adjective) are associated with a

---

particular gender. The data included is the video game textual data from their release data version until version 3.2 which is started at October 1st.

One of recent studies about gender-based violence words is used to analyze Italian media (Busso et al., 2021). This research presented a cross-modal of newspaper and TV language regarding the sociological occurrence of gendered violence, through qualitative and quantitative analysis of an ad hoc compiled newspaper corpus and multimodal database. The corpus was analyzed to reveal the collocational behavior of related terms. The data confirm the results of their surveys that the gravity of violence seems to be somehow attenuated by external circumstances or by internal factors like temporary illness. The TV shows also reports on psychological violence and domestic abuse is related with femicide. While the study committed to research about gender-based violence words, there is still lack of comparison of similarities between violence words related to men and women. This research will give more insight about that comparison.

The other recent studies related to gender in video games are done in various video games. One of it is a popular criminal open world video game, Grand Theft Auto (Gutiérrez, 2014), though corpus is not the tool of the analysis. This research leans more on a social issue, not something linked to the private decisions of those individuals who "consume" video games or the people responsible for them (usually families). Responsibility comes before the market, before these products reach the market. The video game itself contains a lot of violence and harsh words directed towards women, such as Hit the bitch, and there are features which enabled the gamers to hit and even inflict a fatal damage to women characters in GTA.

The main topic of this research is discovering gender-based violence words in a video game Genshin Impact until at its lexical level. This research uses LncsBox to process the corpora from Genshin Impact textual data and to generate the visual graphic to depict whether gender difference really give impact on the result. The study of analyzing gender-based words at lexical level in video game has been done many times in the past, until this type analysis had a name, called ludo linguistic. The previous studies about ludo linguistic typically focus on visual aspects relating to gender, such as body size, the degree to which women are visually sexualized, or player's experiences (Baker, 2014). While in this research, corpus linguistic is implemented to discover the violence words within video games textual data. The endgame of this research is to determine the lexical items that related to violence in Genshin Impact and to know further about how the creator depicts how much men and women are using the violence words. In the end, the similarities and differences of how violence words related to both men and women will be compared based on statistic generated from LncsBox and theoretical analysis about gender

---

violence and gender representation. The endgame of this research is to determine the lexical items that related to violence in Genshin Impact and to know further about how the creator depicts how much men and women are using the violence words. In the end, the similarities and differences of how violence words related to both men and women will be compared based on statistic generated from LancsBox and theoretical analysis about gender violence and gender representation.

## **2. RESEARCH METHOD**

The approach used in this research is mixed-method analysis. The reason of using mixed-method analysis is because the data which gathered from corpus software are formed as numeric or statistical data, thus quantitative method is used here. But the use of qualitative method is still necessary to uncover the reason why a particular variable has certain number. According to (Dawadi et al., 2021), the qualitative aspect from mixed-method analysis provides a deeper understanding of the issue being investigated, which in this case is the statistical data processed from corpus software. Quantitative results can also be triangulated with qualitative research. Triangulation is the use of multiple methods or data sources to develop an overall understanding of a research problem or to check the validity of information. through the convergence of information from different sources (Carter et al., 2014). The research design used for this research is corpus linguistic. The materials for this research are a huge data of texts. The texts are processed by corpus linguistics' programs where the data collection and analysis will be done. Before the processing step, the texts will be filtered and cleaned from unnecessary and unrelated words, numbers and symbols as those things are able to hinder the true result.

The population of this research cover almost all of Genshin Impact's in-game components. Not only the details of characters, but also the description of weapons, artifacts, stories and even conversations with non-playable characters. At the current time, the content population will be from version 1.0 to version 2.7, where this research is made. The entire textual data is gathered from Genshin Impact's text files. Although accessing data via datamining or gathering it manually from inside game is difficult, I choose the website [www.genshin-impact.fandom.com](http://www.genshin-impact.fandom.com) as this site covered all in-game texts that required for this research. The data required consists of texts from main story quests, playable characters' story quests, hangout quests and artifact's descriptions.

The data from Genshin Impact texts are processed by two softwares: antconc and lancsbox. Antconc is used for searching the keynes. The keyword reference used to search the keyness is from 100.000 words of English news in [wortschatz.uni-leipzig.de](http://wortschatz.uni-leipzig.de). Lancsbox is used for processing the collocates. The type used is MI and the collocates within a 5 left – 5 right window spans of the node

words (“he” and “she”) were examined. Furthermore, a minimum occurrence of  $\geq 5$  was imposed for collocates. These two tables of collocate “he” and “she” are sorted by MI score.

### 3. FINDINGS AND DISCUSSION

#### 3.1 Representation of gender

Before analyzing the violence words based on gender, the first information to know is that as in Genshin Impact version 2.8, as comparison the total of male playable characters is 17, while the female one is 33. The amount of playable female characters are almost doubling male characters. But, in order to show the how Genshin Impact represents both genders inside the game, a keyword analysis is done with Antconc showing top 50 unique keywords in Genshin.

Table 1. Keyword List of Genshin Impact’s Wordlists

Rank	Freq	Score	Keyword	Rank	Freq	Score	Keyword
1	2584	7964.62	paimon	26	451	1244.63	ah
2	10337	6688.61	you	27	2512	1214.86	so
3	9737	6135.42	i	28	401	1210.39	ei
4	3839	3235.14	t	29	416	1205.32	hmm
5	7424	2961.24	s	30	1247	1181.99	don
6	2580	2685.15	her	31	2340	1178.74	what
7	2192	2640.9	me	32	1096	1126.05	him
8	2356	2403.79	she	33	1099	1115.4	ve
9	708	2180.19	liyue	34	361	1111.46	inazuma
10	1175	1982.54	let	35	1740	1087.47	like
11	806	1941.2	oh	36	2647	1086.11	he
12	611	1825.72	traveler	37	351	1080.66	abyss
13	7212	1760.33	it	38	7415	1071.48	that
14	571	1758.2	mondstadt	39	1836	1041.93	just
15	500	1539.52	shogun	40	355	1027.26	knights
16	1490	1535.45	here	41	382	985.3	jean
17	504	1526.68	huh	42	1866	944.09	no
18	1125	1486.8	ll	43	2026	918.12	my
19	480	1477.93	yae	44	306	909.77	uh
20	475	1462.53	dainsleif	45	572	836.88	talk
21	465	1431.73	zhongli	46	741	813.31	never
22	1490	1386.02	re	47	288	810.35	alright
23	448	1379.38	miko	48	261	803.53	favonius
24	3216	1282.97	but	49	438	802.73	vision
25	408	1256.19	venti	50	260	800.46	fatui

This data reveals that gendered lexical terms may not be the most frequent within the keywords, but that they are still a salient feature, with approximately 24% of the keyword list being referencing male/female social actors. Within these gendered terms, the data reveal some information about the representation of gender at a quantitative level. The keywords suggest that female characters and their pronouns are more frequently represented in this data set, as 8 of the keywords, “paimon”, “her”, “shogun”, “jean”, “she”, “yae”, “miko” and “ei”. Meanwhile, the male characters and their pronouns have

4 keywords, which are “dainsleif”, “venti”, “zhongli” and “he”. This shows that female characters are more centered in Genshin Impact’s storyline.

### 3.2 Collocates of Violence Words

After discovering the general gender representation, the next step is for uncovering the gender-based violence words. For this section, collocates within a 5 left – 5 right window spans of the node words (“he”, “she”,) were examined. Furthermore, a minimum occurrence of  $\geq 3$  was imposed for collocates. Finally, the verbs and nouns related to violence words was selected and represented in the following tables.

Table 2. Violence verb and noun collocates of “he”

Verb			Noun		
No	Collocate	Freq (coll.)	No	Collocate	Freq (coll.)
1	lose	21	1	end	18
2	disappear	16	2	blade	14
3	fight	15	3	death	7
4	fall	11	4	duel	6
5	defeat	11	5	pain	6
6	die	11	6	blood	6
7	face	9	7	battlefield	5
8	suffer	9			
9	finish	8			
10	encounter	8			
11	break	7			
12	draw	7			
13	steal	7			
14	pierce	7			
15	force	6			
16	cross	6			
17	collapse	5			
18	battle	5			

The data from Table 2 is the result of collocational analysis using lancsbox. From the verb collocates, the total of violence words is 169 from the total of 4241 verbs collocated with “he”. This means the number of verbs related to violence from male perspective is approximately 3,985% of total verbs collocated with “he”.

Meanwhile, for the noun, the total of violence words is 62 from the total of 1117 nouns collocated with “he”. This means the number of nouns related to violence from male perspective is approximately 5,551% of total nouns collocated with “he”.

Table 3. Violence Verb and Noun Collocates of “She”

Verb			Noun		
No	Collocate	Freq (coll.)	No	Collocate	Freq (coll.)
1	finish	10	1	monster	7
2	fall	10	2	loss	6
3	disappear	9	3	blood	6
4	strike	8	4	flame	5
5	fear	8	5	death	5
6	burn	8	6	enemy	5
7	force	6			

The data from Table 3 is the result of collocational analysis using lancsbox. From the verb collocates, the total of violence words is 59 from the total of 3940 verbs collocated with “he”. This means the number of verbs related to violence from female perspective is approximately 1,497% of total verbs collocated with “she”.

Meanwhile, for the noun, the total of violence words is 34 from the total of 1287 nouns collocated with “he”. This means the number of nouns related to violence from female perspective is approximately 2,642% of total nouns collocated with “she”.

#### 4. CONCLUSION

The corpus collected from Genshin Impact textual data served a lot of required data for research. According to the data presented in the previous section, it can be concluded that by using collocates “he” and “she”, male characters has related more into violence words than female characters. From the verb collocates, male characters have 2,488% more than female character, meanwhile for noun collocates, male characters have 2,909%. The ratio of violence words with total of words and difference between “he” and “she” violence word collocates in Genshin Impact data is really small. This is understandable despite Genshin Impact is a Role-Playing Game which emphasis on combat style, the game itself has Teen rating (> 12 years old) according to ESRB, which means the developers intentionally toned down the violence and harsh words inside the game.

---

## REFERENCES

- Baker, P. (2013). Sexed Texts: Language, Gender, and Sexuality. *Equinox*.  
<https://doi.org/10.1558/genl.v7i2.265>
- Baker, P. (2014). Using Corpora to Analyze Gender. In *International Journal of Corpus Linguistics* (Vol. 19, Issue 4). <https://doi.org/10.1075/ijcl.19.4.06kre>
- Busso, L., Combei, C. R., & Tordini, O. (2021). Narrating Gender Violence A Corpus-Based Study on the Representation of Gender-Based Violence in Italian Media. In *Language, Gender and Hate Speech A Multidisciplinary Approach*. Fondazione Università Ca' Foscari.  
<https://doi.org/10.30687/978-88-6969-478-3/002>
- Carter, N., Bryant-Lukosius, D., Dicenso, A., Blythe, J., & Neville, A. J. (2014). The use of triangulation in qualitative research. In *Oncology Nursing Forum* (Vol. 41, Issue 5, pp. 545–547). Oncology Nursing Society. <https://doi.org/10.1188/14.ONF.545-547>
- Dawadi, S., Shrestha, S., & Giri, R. A. (2021). Mixed-Methods Research: A Discussion on its Types, Challenges, and Criticisms. *Journal of Practical Studies in Education*, 2(2), 25–36.  
<https://doi.org/10.46809/jpse.v2i2.20>
- Ensslin, A. (2012). *The Language of Gaming*. Palgrave.
- Gutiérrez, E. J. D. (2014). Video Games and Gender-based Violence. *Procedia - Social and Behavioral Sciences*, 132, 58–64. <https://doi.org/10.1016/j.sbspro.2014.04.278>
- Hämäläinen, M., Alnajjar, K., & Thierry, P. (2022). Video Games as a Corpus: Sentiment Analysis Using Fallout New Vegas Dialog. *HAL Open Science*, 664.
- Heritage, F. (2021). Language, Gender and Videogames. In *Language, Gender and Videogames*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-74398-7>
- Heritage, F., & Koller, V. (2020). Incels, in-groups, and ideologies. *Journal of Language and Sexuality*, 9(2), 152–178. <https://doi.org/10.1075/jls.19014.her>