



Gender in Genshin Impact: A corpus-Assisted Discourse Analysis

Yifei Chen

Digital Humanities From A Computational Linguistic Perspective



Games and Gender Representation



Games have become a mainstream entertainment and a hobby of all ages and genders. Violence and sexist content in games have risen concerns about their potential (negative) influence in individual behaviours by corrupting their perspectives or cognitives.

Concerns on gender representation:

Females:

- Occurrence: Female characters appear in game less than male **chatterers**.
- Appearance: Over-sexualised with unrealistic body proportions and underwear-like costumes displaying their skins. (Objects of “male gaze”, or sexual attention from male characters.)
- Roles: secondary characters as damsels in distress who eventually become a prize presented to a male character for his period deed.



Lara Croft in Tomb Raider, [image source](#)

Games and Gender Representation



Games have become a mainstream entertainment and a hobby of all ages and genders. Violence and sexist content in games have risen concerns about their potential (negative) influence in individual behaviours by corrupting their perspectives or cognitives.



Concerns on gender representation:

Males (Additionally)

- Appearance: excessively masculine bodily features
- Behaviours: Tendency to violence, heroic deeds



Geralt of Rivia in *The Witcher 3*, [image source](#)

Research Gap

In-game text corpora are rare. This study contributes a new one.

Previous studies:

- Mostly content analysis with visual level data, very few in lexical level.
- One-shot games Mostly studied (which discarded after being played one) with one-time gameplay, and relatively old games.

This study:

- Lexical level;
- Studied a long-lasting game with daily gameplay, and updates with fresh contents are introduced regularly (approx. per 1.5 month): the newly released *Genshin Impact*.
- Why care? Long-term exposure.

Genshin Impact

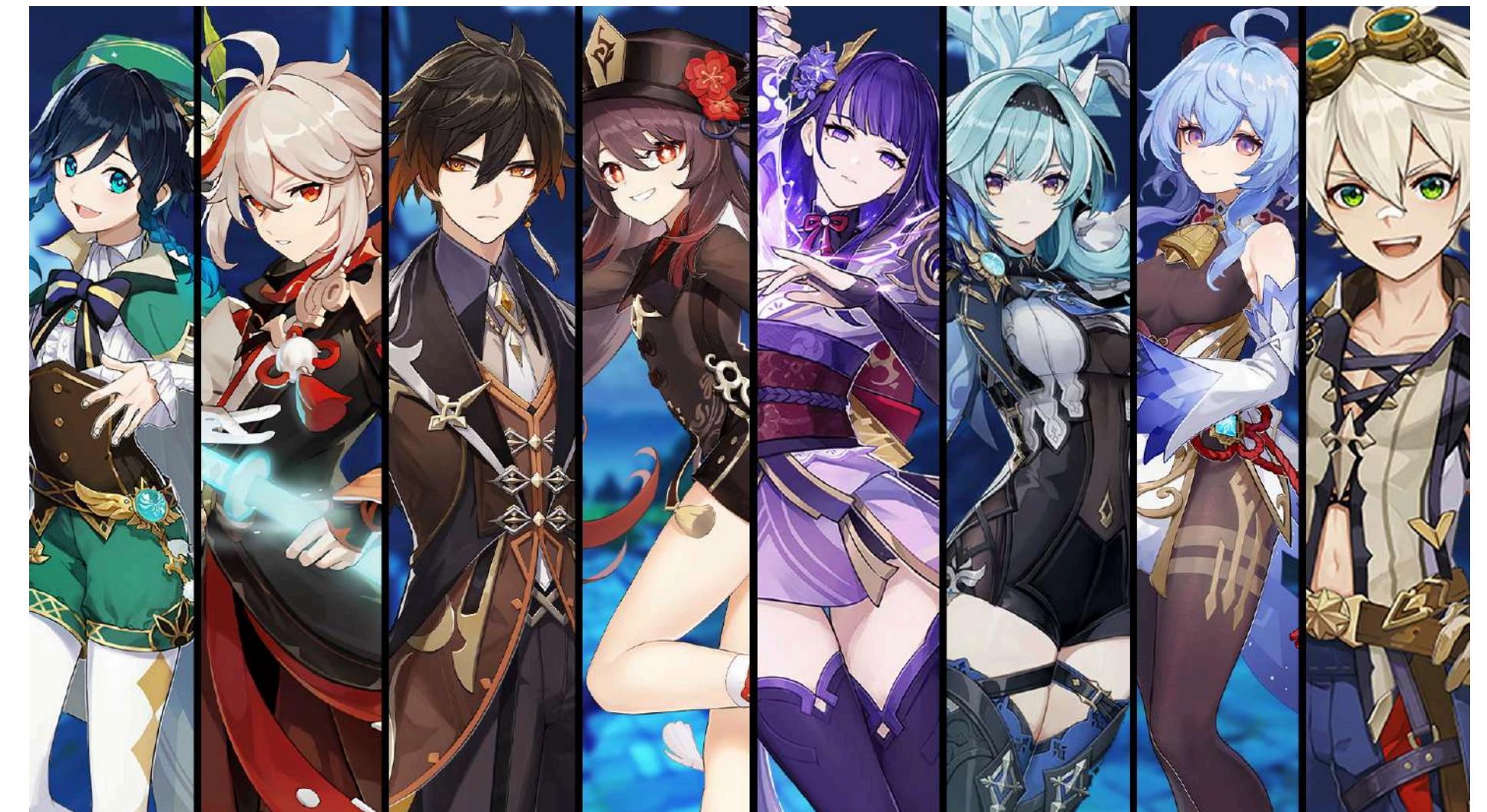
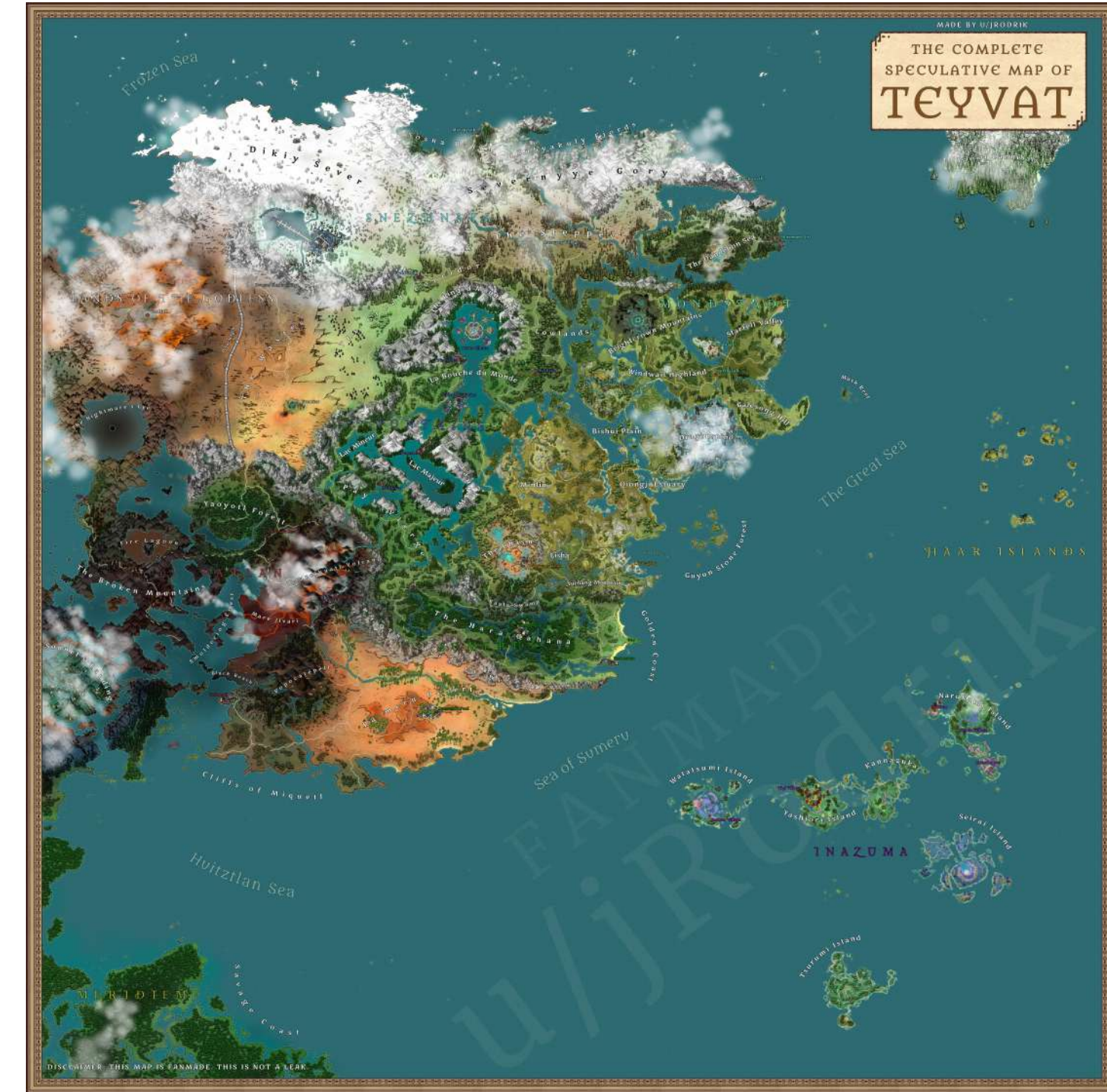
An action role-playing game released for multiple gaming platforms (incl. mobiles and consoles) from fall 2020.

The game features an anime-style open-world environment and an action-based battle system using elemental magic and character-switching.



Main Plot:

The player will be *The Traveler*, an interstellar adventurer who got separated from their twin sibling after landing in a fantasy world Called *Teyvat* with seven nations, each of which is tied to a different element and ruled by a different god. The Traveler journeys across the nations in search of his/her lost sibling, befriend numerous individuals, become involved in the affairs of nations, and begin to unravel the mysteries of the land.



[image source](#)

Genshin Impact

Reasons chosen:

1. A wide audience: millions of players cross multiple platforms with over 85 millions downloads (until 2022).
2. Heavily long-term daily exposure: The mechanism requires players to do daily missions to acquire necessary in-game resources, which estimably will consume over 1 hour per day.
3. Fresh contents continuation: new maps, characters, stories, NPCs with each patch update.



We want to know...

1. If the gender representation improves in this certain game (compared to previous studies);
2. If an analysis from linguistic perspective could also reveal the gender representation in games;

Data – Two Corpora

1. New Model Corpus (NMC) as a reference corpus. A general English corpus.
2. Specialised *Genshin Impact* corpus with 800,231 words from in-game quest dialogues.
 - All of the released quests from the game's first patch (1.0) until the end of 2.6 were collected from a fan website.
 - Including 1011 daily commissions and 84 main story, 98 character, 41 event and 1065 world quests (side quests) that contain dialogues were added to the database.
 - Junk words like dialogue markers were found and deleted by using Excel applications.

Methods

Two tools:

1. Sketch Engine: with simple clicks the corpus analysis could be done, such as keyword analysis, word lists, collocations, and concordance lines.
2. AntConc: similar to Sketch Engine.

Concordance

Word List

Word Sketch

Thesaurus

Find X

Sketch-Diff

Sketch-Eval

Corpus Info

Save

View options

KWIC

Sentence

Sort

Left

Right

Node

References

Shuffle

Sample

Filter

Overlaps

1st hit in doc

Frequency

Node tags

Node forms

Doc IDs

Query colour 16,486 (147.0 per million)

Page 1 of 825 Go Next Last

J2L

It would be tedious to list the types and colours of stone, ceramic etc. used at each site

J2L

types of stone used for various shades of colour are predictable and limited in number.

J2L

Birdcombe Avon. Here, sandstone furnished a buff colour , pennant stone a blue, liar the white for

J2L

most mosaics comprise three to six basic colours , a work of good quality will include many

J2L

therefore, to note ten or twelve different colours of tesserae in one pavement. In some, such

J2L

the Woodchester Orpheus mosaic. </p> 3.2 The colour of Tesserae <p> Sensitive use of shading

J2L

1976, 9). Elsewhere, intelligent use of colour is responsible for the blue shading which

J2L

are notable. </p><p> Whilst considering the colour of tesserae it is also pertinent to mention

J2L

: 0.5 cm. sq. and 1.5 cm. sq. </p><p> Like colour , the size of the tesserae affects the perspective

J2L

fairly dark tesserae (deep red is a favourite colour), so producing a stronger" proximity effect

J2L

panels (pl. 5b). At Leicester the rosettes - coloured (from the edges inwards) red, yellow and

J2L

be cramped (although" loose"). There are colour contrasts however: the simple guilloche

J2L

former. However, the-more subtle use of colour in the latter also produces a less contrived

J2L

angular appearance. An overall poverty of colour , and the use of slightly larger (but still

J2L

mosaic A). Although including the same basic colours , as well as tesserae of a similar size,

J2L

blending of many tones of five or six basic colours , is notable in both designs. It is a sensitivity

J2L

shows a generally consistent interlace of colour , one in every four tongues of the latter

J2L

Oceanus panel (contrast the confusion of colour around the heads of the lion and stag)

J2L

However, on balance, the use here of similar colours (red, yellow, grey, pale-blue, brown) and

J2L

Street mosaic, the presence there of a richly coloured figured panel (enclosed by a chain-guilloche

Page 1 of 825 Go Next Last

Lexical Computing

AntConc 3.5.9 (Windows) 2020

File Global Settings Tool Preferences Help

Concordance

Concordance Plot

File View

Clusters/N-Grams

Collocates

Word List

Keyword List

Corpus Files

EN_1771_Mackenzie

EN_1771_Smollett,T

EN_1778_Burney,Fai

EN_1782_Burney,Fai

EN_1786_Beckford,W

EN_1788_Wollstone

EN_1790_Radcliffe,J

EN_1794_Godwin,W

EN_1794_Radcliffe,J

EN_1794_Rowson,S

EN_1795_Lewis,Mat

EN_1796_Bonhote,E

EN_1796_Burney,Fai

EN_1796_Hays,Mary

EN_1797_Foster,Hai

EN_1798_Brown,Chi

EN_1798_Wollstone

EN_1799_Brown,Chi

EN_1800_Edgewort

EN_1801_Edgewort

EN_1804_Opie,Ame

EN_1805_Lewis,Mat

EN_1806_Edgewort

EN_1809_More,Han

EN_1811_Austen,Jar

EN_1813_Austen,Jar

EN_1814_Austen,Jar

EN_1814_Scott,Walt

EN_1815_Peacock,T

EN_1817_Scott,Walt

EN_1818_Peacock,T

EN_1818_Shelley,Mi

EN_1819_Shelley,Mi

EN_1820_Scott,Walt

EN_1821_Galt,John

EN_1821_Peacock,T

EN_1822_Hogg,Jam

EN_1823_Cooper,Je

Concordance Hits 434

Hit KWIC

1 top of which were marked South-sea annuities, India-stock, and Three per cent. annuities consol. "Th

2 much as to rejoice at our conquests in India. You tell me of immense territories subject to

3 of another kingdom to establish an empire in India? to give laws to a country where the

4 those gentlemen do not deny their going to India. The fame of conquest, barbarous as that motive

5 . When shall I see a commander return from India in the pride of honourable poverty? You describe

6 a soldier in the service of the East-India company; that, in the course of duty, he

7 boundless, that the whole capital of the East India Company would not suffice to fulfil them." Cecili

8 am an Indian, but from a region of India which is wholly unknown." The Caliph, delighted

9 enthusiasts, who had travelled from the heart of India, and halted on their way with the emir.

10 emities of the valley with magnificent screens of India cloth, which were guarded by Ethiopian slaves wit

11 ventured to peep in; But the wealth of India would not have tempted her to cross the

12 was left in his bosom: The wealth of India would not have tempted him to a second

13 the command of a ship in the West India trade, had, during his voyages in fine weather,

14 she had married an officer in the East India service; that soon after their union he was

15 who brought her letters and intelligence from the India House. Mr Harley, rising at the same time

16 her, that the term of his residence in India was prolonged; pressing her to join him there,

17 offer of a nobleman to accompany him to India, as his confidential secretary. "A correspondence

18 of the sale of his father's West-India property, mentioned that the sum obtained for it

19 mentioned Ives's account of the elephants in India, who have been set to watch young children,

20 Irwin to Sir William Jones, from Spain to India, he passed with admirable celerity, and seized al

21 better connexion than the son of a West India planter, even though he may be a protégé

22 to go to Portsmouth, to join the West India fleet, which was to sail with the first

23 there, may easily be guessed. "Has the West India fleet sailed?" "No: it sails to-morrow morning,"

24 to settle with him. relative to his West India estates. He spent the remainder of the mornin

Search Term Words Case Regex

india

Advanced

Search Window Size 50

Start Stop Sort Show Every Nth Row 1

Kwic Sort

Level 1 1R Level 2 2R Level 3 3R

Clone Results

Methods

Corpus-assisted discourse study  (CADS), which includes:

- The quantitative Keyword Analysis (the most frequent 100 words);

More specifically: the specialised corpus was uploaded to SE's dataset to compare it with NMC, applying Keyword Analysis to identify salient features, in this case - if a specific gendered word was salient in the specialised corpus.

- The quantitative Collocation Analysis, combined with the qualitative Concordance Analysis:

More specifically: use collocational analysis of specific words to demonstrate how they co-occur with some words together more than others, in this case - to see how gendered pronouns collate with what words (adjs, verbs, nouns) and which frequency.

Simultaneously, use concordance analysis to comprehend in which way keywords appear in a corpus, in this case - to check whether those specific keywords were used in a gendered meaning.

Results - Top 50 Keywords

Top 50 frequent words:

- Even though gender is not a frequently observed notion in this corpus, it is still salient.
- The pronoun he appeared on the keywords lists (ranked 36) while the pronoun she did not make into the list.

The lack of any feminine pronouns or proper names shows that **female characters are linguistically underrepresented** in Genshin Impact.



Top 50 keywords

No	Item	Frequency	Score	No	Item	Frequency	Score
1	be	46148	46500,434	26	go	4180	4071,959
2	the	34192	33582,082	27	get	3916	3946,814
3	to	23923	24106,176	28	just	3885	3915,578
4	i	23386	23565,088	29	there	3849	3879,303
5	you	23447	23423,969	30	here	3822	3852,098
6	it	14155	14263,788	31	as	3940	3838,218
7	of	14091	14199,301	32	if	3806	3835,976
8	that	13670	13775,095	33	your	3731	3760,405
9	have	13005	12882,208	34	on	4121	3733,606
10	not	12645	12742,289	35	they	3503	3530,668
11	do	11130	11119,584	36	he	3428	3455,097
12	and	10810	10893,315	37	would	3617	3437,437
13	this	8226	8289,637	38	like	3258	3283,802
14	in	7933	7994,405	39	no	3157	3182,033
15	we	7818	7878,529	40	one	2929	2952,297
16	for	7300	7356,584	41	time	2921	2944,236
17	will	6068	6011,227	42	about	2875	2897,886
18	me	5494	5536,831	43	well	2998	2803,601
19	so	5232	5272,835	44	know	2766	2788,056
20	what	5088	5127,739	45	at	2789	2787,126
21	but	5348	4594,937	46	let	2754	2775,964
22	with	4540	4536,333	47	now	2716	2737,675
23	can	4331	4364,974	48	from	2631	2652,028
24	my	4257	4253,623	49	come	2641	2639,278
25	all	4213	4173,88	50	look	2592	2568,307

Results – Verbs

Male characters mostly represented with verbs that have **bodily/physical meanings** (%73,91), while mental and cognitive verbs were few (%26,08).

*leave him be for a couple of days. **He'll recover** on his own soon enough.*
*He also said that **he** wouldn't let anyone **die** in vain*
*Suikou pleaded guilty. **He** admitted to **drinking** at the tavern the night*
*According to my investigation, **he** was **headed** southwest*
*Right, why don't we ask Taisuke? **He likes** cats a lot.*
*I've heard of any of the materials **he mentioned**, though.*

Women were represented **almost equally** by both bodily/physical (52,38%) and mental/cognitive verbs (47,61%).

*Do you have any idea when **she recovered** from her illness?*
*Has **she headed** off to work yet?*
*I was quite stunned when **she ran** past me.*
*Clearly, **she wished** for her people to be wise...*
***She loved** humanity very dearly.*
*Maybe **she thinks** humans are still too young to handle it.*



Verb collocates of *He*

Verbs Denoting Bodily/Physical Actions and Processes			Verbs Denoting Mental/Cognitive Actions and Processes		
Frequency	MI Score	Word	Frequency	MI Score	Word
5	8.23652	wakes	7	596.706	loves
5	7.23652	hides	10	5.56409	likes
10	6.77709	recover	11	5.30793	noticed
13	6.57063	joined	18	5.23902	mentioned
5	6.23652	crossed	44	5.23447	knows
7	6.19838	puts	18	5.23902	mentioned
9	6.13032	finds			
7	6.07809	dressed			
13	5.66083	disappeared			
7	5.59266	headed			
5	5.59266	ate			
8	5.48832	die			
5	5.48163	calls			
6	5.45516	drinking			
6	5.29010	showed			
5	5.28232	walking			
68	5.13787	left			
11	5.10723	works			

Verb collocates of *She*

Verbs Denoting Bodily/Physical Actions and Processes			Verbs Denoting Mental/Cognitive Actions and Processes		
Frequency	MI Score	Word	Frequency	MI Score	Word
10	713.322	returns	5	641.702	wished
8	671.176	recovered	6	634.563	loved
7	629.672	headed	33	625.365	wants
5	608.260	stayed	5	603.369	remembered
36	566.964	gone	21	592.351	needs
47	530.905	left	20	578.071	thinks
12	528.475	leaving	9	572.249	needed
7	525.608	ran	11	553.892	meant
10	519.911	running	34	543.933	wanted
7	515.922	wrote	30	538.599	knows
6	509.509	playing	8	522.633	forgotten
9	508.814	acting			

Results – Verbs

Appearance of verbs like *hides*, *recover*, *disappeared* and *die* shows that male characters were not represented as the sources of brutal physical actions, rather, they **were affected by** them, i.e. they are the receivers:

The only action that can be defined as negative was as *drinking* (excessive alcohol consumption was a specific trait of male characters who felt guilty over something or were just reckless):

*Suikou pleaded **guilty**. **He** admitted to drinking at the tavern the night
He's been here **drinking** for hours. Look, here's his tab.*

The use of the verb *playing* and *acting* show that female characters were represented as **unreliable and sneaky characters**.

*We caught a suspicious person... **She** was **acting** very furtively...
Ningguang, has been **acting** strangely in the shadows, **she** wants the rocks created by Geo Vision
holders...
She's acting strange, huh?
Is she playing dumb?
I have heard of this device before. **She's playing** games with us...*



Results - Adjectives

Also shows **male characters affected by external forces**. They could be victims of some events or they were in bad psychological states:

*He's a **sensitive** kid and he cares a lot about your feelings.*
*We both think that **he's** been **possessed** by an evil spirit*
*I can't help but feel that **he's** still **alive** down here somewhere.*
*He didn't have enough to eat.**He** won't be so **sad** once he's full, I thought.*

They were also represented positively as self-confident, and negatively as guilty and impatient:

*Teucer's a good kid when **he's** not being so **stubborn**.*
*Clear sign of a **guilty** conscience. **He** does a lot behind the scenes...*
*Wow! **He** sure sounds **confident** in his cooking! I like that.*

Females were frequently represented with their psychological conditions. They would get angry quickly and do things that were not appropriate, or they would be happy out of simple things. Women were more frequently described with unstable psychological states or happy moods:
unfortunately, no clear evidence shown from the paper.

Adjective collocates of *He*

Frequency	MI Score	Word
6	717.762	awake
6	641.209	sensitive
8	627.073	upset
5	591.459	stubborn
6	591.459	possessed
6	585.570	guilty
7	567.755	confident
11	556.667	alive
6	554.536	sad
5	528.232	injured

Adjective collocates of *She*

Frequency	MI Score	Word
7	621.655	angry
7	572.557	mad
5	538.599	tired
19	537.473	alone
5	524.014	famous
17	509.878	happy
14	505.968	busy

*Bet Lisa does some crazy things when **she** gets **mad**.*
*Without desserts and books to console herself with **she** was so **angry** she quite simply lost it.*
*Little Liu was so **happy**. **She** said she wanted to prepare a gift for her father.*
*Barbara is always **happy** when **she** sees such beautiful scenery.*

Results - Adjectives (Continues)

They were also hard workers, high ranking individuals (Kamisato Ayaka, 2022) and famous persons of the world they live in:

*She's in one piece and then some. **She's** just **busy** with her research topic.
So Alice went to all these lengths because **she** was too **busy** to be with Klee?
...Xiangling, the famous chef. The spicy dishes **she** makes are **famous**...
Kamisato Ayaka, also known as the Shirasagi Himegimi... **She** must be super-**famous** in Inazuma...*

They were also frequently described as being alone, yet not in a needy way. Female characters wanted to accomplish things by themselves, without the help of others:

***She** just wants to train **alone**.
She went to Khaenri'ah **alone**, without telling me
She plans to head up **alone** first, camp out for a bit, and observe the conditions*

Results – Nouns (Roles)

Both female and male characters were presented within familial roles.

Males: son, dad.

*“Uncle” was used to address the elderly in a respectful manner. There was no mention of a familial role.

*It's a new story told by **Uncle** Ramsay. **He** said he's been to many countries...
Uncle He, **he**'s our senior here.*

Noun collocates of *He*

Frequency	MI Score	Word
7	619.838	son
29	5.88993	uncle
18	567.512	dad
5	532.963	fighter

Noun collocates of *She*

Nouns Related to Roles			Nouns Related to Professions		
Frequency	MI Score	Word	Frequency	MI Score	Word
25	636.769	girl	7	661.865	consciousness
5	613.322	guardian	5	635.562	astrology
14	593.415	sister	8	6.00394	company
20	568.319	child	5	565.518	shop
7	554.336	mother	5	558.303	gnosis
9	552.179	shrine maiden	13	553.773	eternity
8	519.239	granny	5	529.672	disaster
5	515.922	deity	5	513.322	magic
6	501.379	daughter	6	507.433	medicine

Results - Nouns (Roles)

no clue why males are not analysed in 2 groups. 🗨️

(Two groups of results were obtained when noun collocates of she were examined.)

Females: *mothers* and *daughters*.

*From the moment my **mother** married my father, **she** became the **mother** of the Kamisato household*
*My **mother** loved him dearly. **She** passed soon after.*
*Could I ask you to bring these toys to my **daughter**... **She** stays at Qingce Village.*
*I'm always telling my **daughter** that if **she** gets the chance to learn from Her Excellency directly.*

Similar to the word *uncle*, *child*, *girl* and *granny* did not show any familial bond or role. They were mostly used as an honorific expression.

*Hahaha, don't mind Changsheng, **she's** a good **girl**, really.*
*Klee needs to be a good **girl** and think about what **she** did...*
*The old lady was **Granny** Oni. **She's** the one that took me in.*
*How do you know **Granny** Ruoxin? **She** was a famous opera singer in her youth, too.*

Noun collocates of *He*

Frequency	MI Score	Word
7	619.838	son
29	5.88993	uncle
18	567.512	dad
5	532.963	fighter

Noun collocates of *She*

Nouns Related to Roles			Nouns Related to Professions		
Frequency	MI Score	Word	Frequency	MI Score	Word
25	636.769	girl	7	661.865	consciousness
5	613.322	guardian	5	635.562	astrology
14	593.415	sister	8	6.00394	company
20	568.319	child	5	565.518	shop
7	554.336	mother	5	558.303	gnosis
9	552.179	shrine maiden	13	553.773	eternity
8	519.239	granny	5	529.672	disaster
5	515.922	deity	5	513.322	magic
6	501.379	daughter	6	507.433	medicine

Results – Nouns (Roles, more specifically...)

Males were identified as fathers' sons that would  remind their wealth or fame to the others.

Dads on the other hand were strong, had wits and wealth.

*His ... name is Date, **son** of Eurydates. **He** said that he wanted to build a new generation...*

*But the editors said that **he** is actually the second **son** of the Feiyun Commerce Guild...*

*You can go ask my **dad**. **He** knows a lot of things.*

***Dad** said that **he** will hire a few people to protect us while we're traveling.*

Mother on the other hand, seems to be secondary to the father figure, mostly used to describe the close relationship between parents. And *daughters* do not share the same function as *sons*.

*From the moment my **mother** married my father, **she** became the **mother** of the Kamisato household*

*My **mother** loved him dearly. **She** passed soon after.*

*Could I ask you to bring these toys to my **daughter**... **She** stays at Qingce Village.*

*I'm always telling my **daughter** that if **she** gets the chance to learn from Her Excellency directly.*

Results – Nouns (Professions)

Male characters were represented as fighters which shows masculine physical powers:

Mr. Guy is a pretty mega awesome **fighter**. **He** beat up a whole camp of hilichurls...
*I wonder how strong of a **fighter he** is. I'd like to go a few rounds with him.*

Women has a prominent association with supernatural powers. In the use of *guardian*, *shrine maiden*, and *deity nouns*, or were associated with powers in the consciousness, mystical, and metaphysical realms.

*Shogun is the most exalted and revered... **She** is a **deity** who reigns on high...*
*Madame Ping is the **guardian** of the Cleansing Bell. **She** lives near Yujing Terrace.*
*Lady Yae, the head **shrine maiden**, **she** has direct and close contact with the Almighty Shogun*
*She said that **she** used **magic** she secretly learned from the bake-danuki.*
*As well-versed in **astrology** as **she** was, my master only knew astrology.*
*... **she** placed her **consciousness** inside her sword, where she remains in a meditative state*

The use of *Gnosis* proper noun also confirm this trend (Gnosis is a key held by the gods in Genshin). It is always associated with the *she* pronoun.

*She no longer needs the power of the **Gnosis**. **She** tells me she has severed ties with Celestia.*
***She** just grabbed your **Gnosis** and left!*
Female characters were also associated with commercial and medical professions.
*Marjorie? Oh, right, **she** runs that souvenir **shop**, doesn't she?*
*I asked Jilliana for some cough **medicine**, but **she** was out of ingredients.*

Back to the questions.

We want to know...

1. If the gender representation improves in this certain game (compared to previous studies);
2. If an analysis from linguistic perspective could also reveal the gender representation in games;

Summary – If the gender representation improves

- Women in *Genshin Impact* are linguistically overrepresented, emotionally prominent, and still defined by some stereotypical negative traits and roles;
- Positive changes:
 - No longer victims of male violence;
 - More independent social roles, coming to fore with their supernatural powers and achievements in their professions;
 - The dominant, abusive and violent male figure was not found in *Genshin Impact*.

Summary – If linguistic perspective helps

The study largely based on two standards: frequency and human judgements.

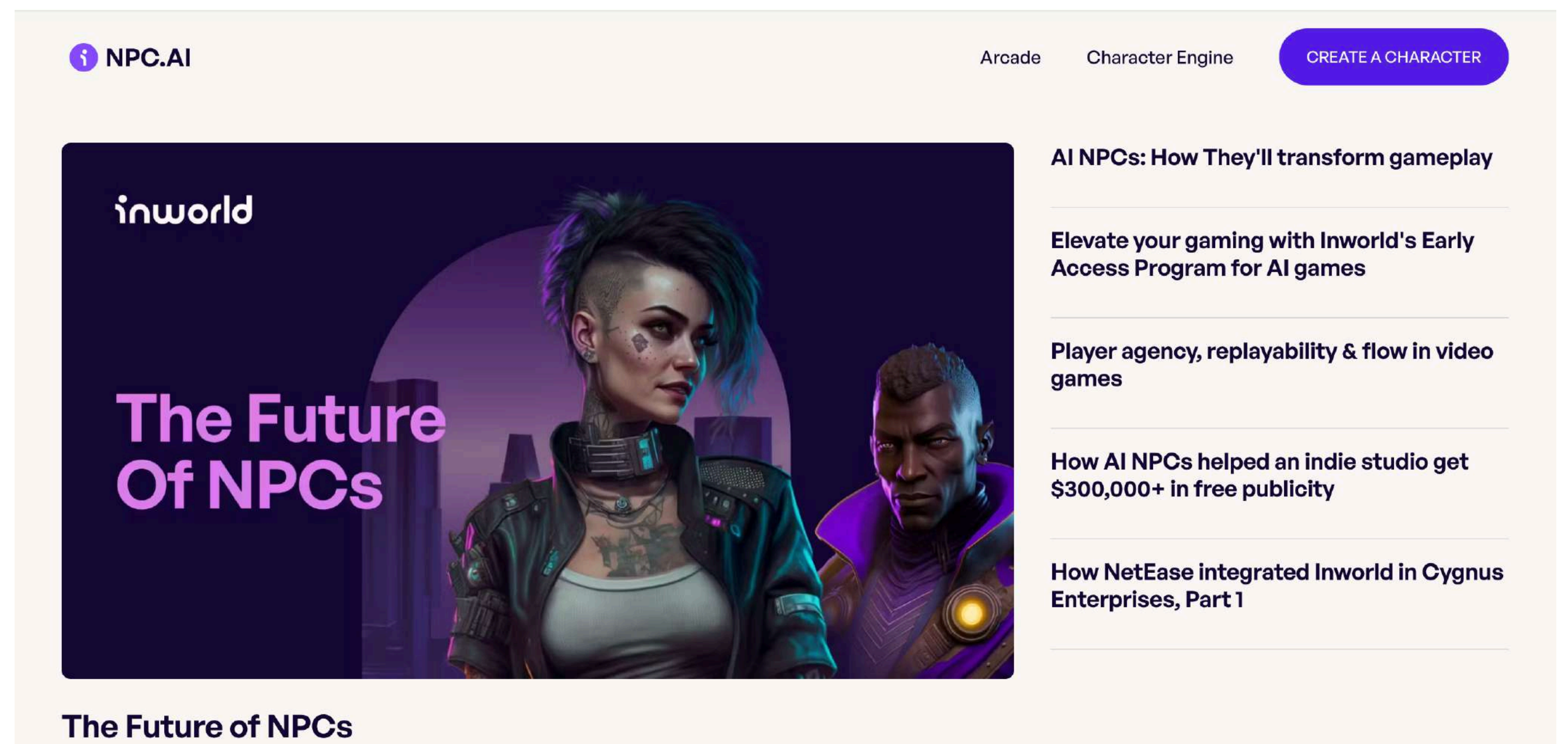
- Frequency: to rank the repetitive linguistic patterns, identify the salient ones;
- Human judgement: to check if these patterns convey gender representation.



But the reliability of human judgements is not validated. Some conclusions drawing from the comparisons between males and females are lack of supporting evidence.

Nevertheless, it still proves that lexical level analysis have an equal efficacy compared content analysis in studying gender representations, or potentially, into other fields using game-related corpora.

Significance in reality: AI NPC.



Further Reading And Other Links (If Interested)



- Heritage, F. (2020). Applying corpus linguistics to videogame data: Exploring the representation of gender in videogames at a lexical level. *Game Studies*, 20(3). http://gamestudies.org/2003/articles/heritage_frazer.
- Heritage, F. (2021a). Language, Gender and Videogames: Using Corpora to Analyse the Representation of Gender in Fantasy Videogames (1st ed.). *Palgrave Macmillan*. <https://doi.org/10.1007/978-3-030-74398-7>.
- Heritage, F. (2021b). Maidens and Monsters: A corpus assisted critical discourse analysis of the representation of gender in The Witcher videogame series (Doctoral Dissertation, Lancaster University).). <https://doi.org/10.17635/lancaster/thesis/1333>.
- Genshin Impact's official website, where you could explore the art design: <https://genshin.hoyoverse.com/en/>.

GAME
OVER