Towards an Analysis of Gender in Video Game Culture: Exploring Gender-specific Vocabulary in Video Game Magazines

Thomas Schmidt, Isabella Engl, Juliane Herzog and Lisa Judisch

Media Informatics Group, University of Regensburg, Regensburg, Germany
thomas.schmidt@ur.de
isabella.engl@stud.uni-regensburg.de
juliane.herzog@stud.uni-regensburg.de
Lisa.judisch@stud.uni-regensburg.de

Abstract. We present preliminary results of a project examining the role and usage of gender specific vocabulary in a corpus of video game magazines. The corpus consists of three popular video game magazines with 634 issues from the 1980s until 2011 and was gathered via OCR-scans of the platform archive.org. We report on the distribution and progression of gender-specific words by using word lists of the LIWC for the categories "male" and "female". We can indeed show that words of the type male are considerably more frequent than words of the type female, with a slight increase of female words during 2006–2010. This is in line with the overall development of gaming culture throughout these years and previous research in the humanities. Furthermore, we analyzed how the usage of negatively connoted words for female depictions (e.g. chick, slut) has evolved and identified a constant increase throughout the years reaching the climax around 2001–2005, a timespan that coincides with the release and popularity of games encompassing rather sexist concepts. We discuss the limitations of our explorations and report on plans to further investigate the role of gender in gaming culture.

Keywords: Game Studies, Gender Studies, Corpus Linguistics, LIWC, Video Games, Video Game Magazines, Gender, Gaming

1 Introduction

Video games are one of the most important and financially successful media in the world today having significant influence on pop culture. According to statistics from 2019, around 2.5 billion of the human population play video games[1]. Due to this rise in popularity and the impact on our culture, it is not uncommon anymore for humanities researcher to analyze video games and gaming culture. This branch of humanities is also often referred to as *game studies* [9]. Although gaming is becoming more popular among females [5], the gaming culture was and still is strongly dominated by

[1] https://www.wepc.com/news/video-game-statistics/

Please cite as:
Schmidt, T., Engl, I., Herzog, J. & Judisch, L. (2020). Towards an Analysis of Gender in Video Game Culture: Exploring Gender-specific Vocabulary in Video Game Magazines. In *Proceedings of the Digital Humanities in the Nordic Countries 5th Conference* (DHN 2020) (pp. 333-341). Riga, Latvia. http://ceur-ws.org/Vol-2612/short20.pdf

males. For this reason, games often have been analyzed in the context of *gender studies*:

Dietz [2] used content analysis on 33 video games to identify an underrepresentation of female characters in those games. Furthermore, the women were often portrayed as “sex objects”. Dill and Thill [3] also used content analysis on the images of video game characters in video game magazines. They showed that video game characters were more often male than female but also that females are portrayed mostly as sexualized, scantily clad and as a mix of sex and aggression. They conclude that video games might play an important role in sexualizing women and overall sexism in culture. Downs and Smith [4] who analyzed 60 video games to again prove the underrepresentation of females but also show that females are more likely to be shown partially nude also confirm these results. Lynch et al. [7] conducted similar studies but with a higher number of games (n = 571). They showed that games depict females more often in rather secondary roles as well as sexualized with a peak of this depiction around 2005 and a decrease starting from there on. Kirkpatrick [6] performs discourse analysis with video game magazines of the UK between 1981 and 1995. Kirkpatrick argues that video game magazines started to become more and more focused on males and the depiction of male stereotypes after 1987 while the community was more inclusive before.

We want to further this research area, which is currently dominated by qualitative and hermeneutical research on few examples by applying computational text analysis on a large corpus of video game magazines to gather additional insights. While todays’ gaming culture is mostly situated on online platforms like *Twitch* or *Youtube*, in the 80s, 90s and the beginning of the 2000s video game magazines were the dominant media channel for video game culture and are thus often material for research in the area of gaming culture [3]. In the magazines, new games were introduced, courses of games were shown, problems were addressed, and new technology was presented. In the following, we present preliminary results analyzing the usage of gender-specific words in a corpus of video game magazines. We will also discuss how those results are in line with the aforementioned humanities research and outline how we want to continue our work.

2 Corpus

We gathered the corpus via the platform *Internet Archive*[2], which is a platform that describes itself as a "non-profit library of millions of free books, movies, software, music, websites, and more”. Internet Archive also offers the OCR scans of various video game magazines in different languages since the 80s. For our preliminary analysis, we decided upon the magazines *Computer Gaming World* (1981–2006) with 268 issues, *PC Zone Magazines* (1993–2011) with 160 issues and *Computer and Video Games Magazines* (1981–2004) with 206 issues. We selected those magazines since they offered the most material for reasonable diachronic analysis and an acceptable

[2] https://archive.org/

OCR output compared to other magazines on Internet Archive. We developed Python-Scripts to crawl the OCR-scans, which are offered as txt-files. Furthermore, we applied several methods to improve the OCR like filtering non-alphanumeric tokens and spell correction with the library *symspellpy*[3]. Although we were able to create acceptable output via those methods, such heterogeneous material consisting of many pictures and advertisements is always prone to error concerning OCR optimization. For most analysis, we normalized the corpus by lowercasing the tokens. We used the *Natural Language Toolkit*[4] for general text processing steps like tokenization. Table 1 summarizes some of the main corpus statistics:

Table 1. General text properties of the video game magazines corpus

General text properties	n
Texts	634
Tokens	29,126,799
Average tokens	45,941.32
Average word length	4.07
Types	2,940,999
Average types	4,638.80

To perform diachronic analysis and explore developments over time we separated the corpus into six 5-year time spans beginning from 1981 and ending in 2011. Table 2 summarizes the main attributes of these sub-corpora.

Table 2. General text properties of the sub-corpora

	1981 – 1985	1986 – 1990	1991 – 1995	1996 – 2000	2001 – 2005	2006 – 2011
Issues	92	123	109	170	104	36
Tokens	1,813,073	2,406,343	7,076,137	10,664,775	5,474,272	1,692,109
Average number of tokens	19,707.32	19,563.76	64,918.69	62,733.97	52,637.23	47,003.03
Average word length	3.84	4.0	4.19	4.12	4.16	4.13
Types	253,251	356,746	631,676	956,606	551,472	191,248
Average number of types	2,752.72	2,000.37	5,795.2	5,627.09	5,302.62	5,312.44
Lexical diversity	0.1602	0.1894	0.1064	0.0972	0.1093	0.1143

Note that the six sub corpora are not equally distributed concerning the size. In regards to the number of issues and the size of those one can easily identify how video

[3] https://pypi.org/project/symspellpy/
[4] https://www.nltk.org/

game magazines were on the height of their popularity during the 90s and became less successful throughout the 2000s being replaced primarily by online content. Nevertheless, we regard all sub-corpora large enough for first explorations and we will focus on normalized results to enable valid comparisons.

3 Analysis

To explore gender aspects in this corpus of video game magazines we focused on the analysis of the usage and distribution of gender-specific vocabulary throughout the time spans. To determine such gender-specific vocabulary we refer to the 2015 version of the *Linguistic Inquiry and Word Count* (LIWC [8]), which is currently one of the most popular linguistic resources used in psychology research [10] but also Digital Humanities [1]. The LIWC can be regarded as large dictionary that is structured by several linguistic and semantic categories each consisting of words (or lemmas) that are associated with this specific category. The associations are based on research in the context of psychology (more information can be found in [8]). The LIWC also offers the categories "female" and "male". For example: the words in the female categories consist of some general synonyms for females e.g. *girl, women*, words describing relations like *sister, mother, wife*, general pronouns like *she*, some more special words like *princes and queen*, but also rather condescending words like *chick*. We furthermore extended the lists of the LIWC with other general synonyms of the words women and female. We were especially interested in more negatively connoted female terms that were not part of the LIWC and added them manually. Our final word lists consist of 128 words for the category male and 163 words for the category female. Table 3 gives some more examples for both lists:

Table 3. Examples of female and male words

Female words (examples)	Male words (examples)
female	boy
girl	bro
goddess	dad
her	father
lady	gentlemen
madam	he
maid	him
she	male
sister	mister
wife	prince
woman	uncle

First, we present the top five most frequent words of both gender categories throughout the five year time spans (see table 4).

Table 4. Most frequent male and female words per time span

Period	Top words male	Top words female
1981 - 1985	he; his; him; man; men	her; she; sis; girl; mother
1986 - 1990	he; his; him; men; king	her; she; sis; female; girl
1991 - 1995	he; his; him; king; man	her; she; sis; mother; queen
1996 - 2000	he; his; him; manager; man	her; she; girl, female; queen
2001 - 2005	he; his; him; man; manager	her; she; girl, woman; female
2006 - 2011	he; his; him; man; manager	her; she; girl; lady; woman

Not surprisingly, the pronouns are the most frequent per each category. However, it is notable that for the male category the personal pronoun "he" is the most frequent word while in the female category the possessive pronoun "her" is more frequent, possibly proving that female video game characters are rather rare and therefore the usage of "she" is not as necessary. Other than that, words describing gaming content are for males "king" and "manager" with the latter one becoming more frequent throughout the years. For females, frequent words besides the pronouns are in the early years words describing family relationships ("sis", "mother") and the word "queen". Another noticeable word is "lady" which becomes very frequent in 2006-2011. However, the word can be connoted positively as well as negatively; therefore, we cannot make any assumptions concerning the depiction of females.

Regarding the frequency of gender-specific words, we can show that male words are far more frequent throughout all the time spans compared to female words (see table 5) reaching almost 1% of all the words in certain time spans (table 5).

Table 5. Proportion of male and female words in the entire corpus

Period	# male tokens	Relative frequency male tokens (in %)	# female tokens	Relative frequency female tokens (in %)
1981 – 1985	14,533	0.802	2,212	0.122
1986 – 1990	20,153	0.838	3,198	0.133
1991 – 1995	54,459	0.770	9,951	0.141
1996 – 2000	73,429	0.689	13,711	0.129
2001 – 2005	36,922	0.675	6,230	0.114
2006 – 2011	14,327	0.850	2,689	0.159

Of course, this result does not come as surprise, since video games, especially in the 90s and early 2000s were primarily targeted towards males. Therefore, the majority of video game characters are men, which results in high levels of male pronouns. Over-

all, we can confirm the results of [2, 3, 4, 6] concerning higher amounts of male representation by using computational methods.

No trend is visible concerning the progression throughout the time spans. However, it is striking that the maximum proportion of female tokens can be found in the last time span beginning 2006, possibly mirroring a trend in the gaming industry to target women more strongly. This is also in line with research by [7] pointing towards 2005 as a year where the representation of females in video games improved.

For an in-depth analysis on how females were perceived in the magazines throughout the time, we selected nine tokens of our female list with rather strong negative connotations and examined their usage. Those words are: *chick, bitch, slut, pussy, harlot, strumpet, hooker, wench, whore*. Table 6 shows the proportion of the sum of those words per time span for all female words.

Table 6. Proportion of negative female words to all female words

	# female tokens	# negative female tokens	relative frequency to all female tokens
1981 – 1985	2,212	16	0.723 %
1986 – 1990	3,198	40	1.251 %
1991 – 1995	9,951	166	1.668 %
1996 - 2000	13,711	289	2.108 %
2001 - 2005	6,230	370	5.939 %
2006 - 2011	2,689	126	4.686 %

We identified a clear increase throughout the time spans reaching the climax in the time span of 2001 and 2005. The following line graph (Figure 1) illustrates this development for every single one of the words as proportion among all female words:

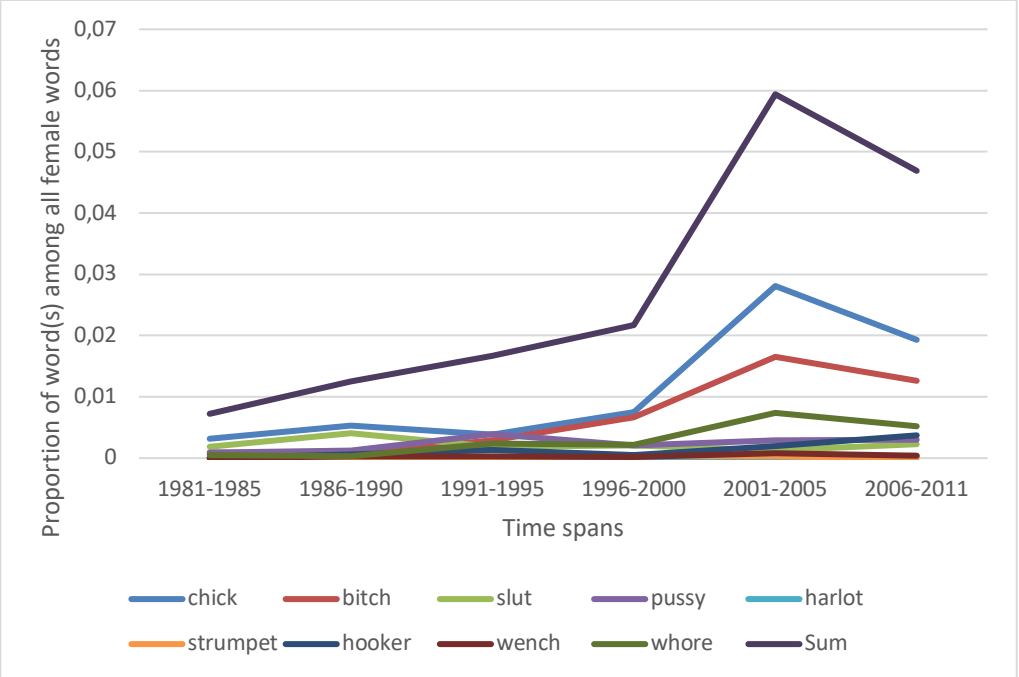

Fig. 1. Proportion of negative female words among all words based on time spans

The three words that have the most influence and show the strongest increase are indeed "chick", "bitch", "slut", "hooker" and "whore" which points to an increase concerning the sexualization of females in video games and video game magazines. Indeed, many video games sexualizing females were released in the time span of 2001–2005 (e.g. *Dead or Alive Xtreme Beach Volleyball*). The time span between 2001 and 2005 also coincides with the release of the game *Grand Theft Auto 3* in 2001, which encompasses, next to general female sexualization, also concepts like prostitution and which influenced many other games in this time span. This sexualization is in line with research by [2] and especially by [3] who identified via content analysis that 80% of female characters in video games are portrayed as sexualized. It is also in line with [7] who proved that the peak of sexualization was reached in 2005 and identified a decline shortly after.

Please note however, that, at the moment, our data is solely descriptive. We want to examine in future work if this trend is persistent with stronger statistical analysis. Furthermore, it is not sure, if this is a trend specific to video games, or if the data just mirrors a general pop cultural trend for this time. More research comparing similar analysis to other pop cultural artefacts is necessary to answer this question.

4 Future Work

We were able to extend existing rather qualitative and hermeneutical research in the context of gender studies, sociology and video game culture by applying computational text analysis on a corpus of video game magazines. By examining gender specific words, we identified the more frequent representation of males in gaming culture but also the oftentimes negative and sexualized depiction of females. Nevertheless, our corpus as well as our methodology is rather limited so we want to continue and improve our work:

The corpus is currently limited on three video game magazines. While the size was reasonable for first explorations, we plan on increasing the corpus. Furthermore, video game magazines offer many structural units like reviews, interviews, exhibition reports and it might be interesting to analyze and compare the content of this units instead of the entire magazine as one.

Concerning the applied methods, we focused on rather basic analysis with the LIWC. We want to explore more advanced methods to examine gender specific research questions. One method we want to pursue is the analysis of collocations concerning gender-specific words so we gain more insights about the context specific words are used in. Furthermore, we also want to apply sentiment analysis concerning the context to investigate which sentiments or emotions are connoted with gender-specific words. Other than that, we plan to also integrate qualitative work into our research. We propose that a mixed methods approach using qualitative and hermeneutical as well as computational quantitative approaches will lead to holistic insights of the role of gender in video games.

We currently focus for our research on the depiction of females, however there are many more gender-specific research possibilities in the context of gaming culture and video game magazines. In future work, we also want to analyze male stereotypes and depictions as well as the trend to more positive depictions of females in more detail. We also plan to investigate gender roles outside of the binary system in gaming culture. Nevertheless, it is very clear, that video game magazines are less important for current gaming culture. While we still think that video game magazines offer a unique and historical insight into gaming cultures, platforms that are more important for todays' community are online and should be investigated for a more holistic view on the topic.

References

1. Boyd, R.L.: Psychological text analysis in the digital humanities. In: Data analytics in digital humanities, pp. 161–189. Springer (2017)
2. Dietz, T.L.: An examination of violence and gender role portrayals in video games: Implications for gender socialization and aggressive behavior. Sex roles 38(5-6), 425–442 (1998)
3. Dill, K.E., Thill, K.P.: Video game characters and the socialization of gender roles: Young people's perceptions mirror sexist media depictions. Sex roles 57(11-12), 851–864 (2007)

4. Downs, E., Smith, S.L.: Keeping abreast of hypersexuality: A video game character content analysis. Sex roles 62(11-12), 721–733 (2010)
5. Hahn, S.: Gender und gaming: frauen im fokus der games-industrie, vol. 43. tran- script Verlag (2017)
6. Kirkpatrick, G.: How gaming became sexist: a study of uk gaming magazines 1981– 1995. Media, Culture & Society 39(4), 453–468 (2017)
7. Lynch, T., Tompkins, J.E., van Driel, I.I., Fritz, N.: Sexy, strong, and secondary: A content analysis of female characters in video games across 31 years. Journal of Communication 66(4), 564–584 (2016)
8. Pennebaker, J.W., Boyd, R.L., Jordan, K., Blackburn, K.: The development and psychometric properties of liwc2015. Tech. rep. (2015)
9. Raessens, J., Goldstein, J.: Handbook of computer game studies. The MIT Press (2011)
10. Tausczik, Y.R., Pennebaker, J.W.: The psychological meaning of words: Liwc and computerized text analysis methods. Journal of language and social psychology 29(1), 24–54 (2010)