

# Corpus analysis on violence in video games



Introduction:-

Through corpus analysis, being a semantic approach to analysing a corpus which is a set of methodically or arbitrary collected and electronically stored real-life language samples such as speeches, magazine articles and/or texting messages.

The main goal of corpus analysis is to detect certain decree of language use, grammatical or lexical patterns, for example, that are pertinent to a particular genre or type of text. This serves as a valuable source for dialectology, sociolinguistics and other related fields.

In this essay, I aim to exhibit through the application of corpus analysis along with the help of discourse analysis in knowing whether violence in video games can or may lead to being violent and/or aggressive in real life while in a way maybe showcasing the same violent/aggressive nature while playing video games.

In order to this, corpus analysis will be carried out on a set of at least 20 files being proposed as our use case, in this case, our main corpus in knowing and researching about violence in video games.

Corpus analysis comprises of terms such as collocation, concordance, tokens, stoplists, lemmatization, clusters/n-grams, keyword lists and wordlists.

I shall dig deep into these concepts and apply them individually to our corpus and investigate how violence maybe shaped across video games and to also check for any similarities or differences if any.

<https://assignbuster.com/corpus-analysis-on-violence-in-video-games/>

Also, what I might have achieved from the results, I shall delve into the utilization of the procedures to the corpus that we need to inspect. For my study of corpus analysis, I shall use Antconc software to help me in drawing conclusions about the aforementioned terms.

Corpus Methodology:-

Wordlist:-

After attaching the files in Antconc and hitting the Word List tool, the number of word types are 7645. Word types refer to the count of each word form. We also see the number of word tokens turn out to be 62146. Word tokens refer to the total count of each instance of each word. On checking the results, we see that different forms of the word are counted as different word types, eg ‘ game’ and ‘ games’ are different word types and are counted separately in the Word List. Attaching a small extract of 20 lines showing the list of words and their frequency.

Word	Frequenc y
the	2650
of	1994
and	1638
to	1540
a	1248

in 1165

that 930

video 849

games 843

violent 649

on 538

is 530

for 523

s 414

as 405

game 403

violenc  
e 385

be 352

are 350

or 349

Tokenization:-

Next, we shall dive into the concept of Tokenization. Tokenization is considered to be a process of delimiting and conceivably dividing sections of a text into words. After applying the Punctuation setting in the Token Definition option in Antconc to test for tokenization and thereafter running the search again, we see that the number of word types and word tokens have changed. Word types being 11330 and word tokens being 63791. Attaching a small extract of 20 lines showing the new list of words and their frequency.

Word	Frequenc y
the	2621
of	1990
and	1607
to	1529
in	1150
a	1144
that	903

video 831

. 656

game  
s 632

violen  
t 605

on 524

is 520

for 516

( 456

as 401

are 349

be 344

game 344

or 337

From the above table and after applying the process of tokenization, we see that the punctuations like periods (.) and parentheses [()] are considered as separate words.

<https://assignbuster.com/corpus-analysis-on-violence-in-video-games/>

Stoplists:-

Next, we shall check how to use stoplists in order to remove certain words. For this process, we shall use the Natural Language Tool Kit Python library text file to help with the same.

After applying the stoplist to our corpus, the number of word types is 11200 and the number of word tokens is 39467. In the stoplist we have also added punctuation marks and special characters that needs to be removed when we run the analysis on the given corpus. In the result, we only see a list of grammatical words unlike when we saw words like ' the', ' and' in when no filter was applied. Also, we have added punctuation marks to the stoplist file in order to remove them while running the analysis on the corpus. Showing the list below of 20 line results after attaching the stoplist text file.

Word	Frequenc y
video	831
games	632
violent	605
game	344
violence	243
may	181

play 159

aggressi  
on 148

media 141

aggressi  
ve 132

studies 131

moral 128

new 128

playing 122

research 121

people 117

found 113

study 109

social 108

also 103

Lemmatization:-



The next concept that I would be going over involved in corpus analysis is called lemmatization.

Combining together different forms of the same word is called lemmatization. To query any version of a base word and get relevant results; lemmatization helps us to do this. For eg, 'turn' is the lemma for turns, turning, turned. These words are inflections for the word 'turn'.

Normally, search engines would prefer using lemmatization as any articulations forms of any words would give us admissible results.

Attaching a 20 lines extract of the result below showing various lemma forms.

Word	Lemma	Frequency
game		344
games	game	976
632		
video		831
videos	video	838
7		
violent		605
605	violent	605

play 159

played 59

playing play 344

122 plays

4

studied 3

studies

131 study

study 246

109

studying

3

violence

violence 243

243

may 181 may 181

aggressio aggressi

n 148 on 148

148 on

media

141

medium 147

medium

6

effect 71 effect 146

effected

1 effects

74

aggressiv aggressive 132  
e 132 ve

use 58

used 33 use 131  
uses 3

using 37

new 128 new 130  
newer 2

moral moral 128  
128

research

121

researche research 123  
d 1

researche

s 1

child 30

children child 117  
87

people people 117

117

found

113

found 114  
founded

1

link 56

linked 18

link 113  
linking 10

links 29

social

social 108  
108

While taking out grammatical or function words like ' to', ' from', ' my', ' they', essentially the sentences do make complete sense even after taking out words like prepositions, pronouns, etc. However, when we do remove them, some of the sentences might become meaningless or pointless.

For eg, ' I want to play video games with Simon.'

By taking out the prepositional words from the above sentence, the sentence now becomes:

' I want.'

This is still grammatically sound, but pretty much pointless or meaningless.

Concordances:-

<https://assignbuster.com/corpus-analysis-on-violence-in-video-games/>

The concordance is a way to view words in context. Word frequencies are useful, but without context we risk drawing false conclusions (<https://pdfs.semanticscholar.org/fa9a/84ae6215251b879569c4f902e14d71fb49f9.pdf>).

Lines of text of a given sentence featuring the word/words under investigation are in the centre of the page, namely in this case checking for concordances for words like “ violence”, “ video”, “ games” and the neighbouring context known as collocates in those same sentences from which we could draw about arrangements of links between the stated words.

When searching through the results for ‘ game’, we get 344 hits.

On applying wildcard character ‘\*’ along with ‘ game’ such as ‘ game\*’, the number of hits are 1307.

On searching for ‘ violence’ in the concordances tool, and while adjusting the sorting levels such as First (1L), Second (1R), Third (2R), we see that most of the words 1L of ‘ violence’ appear to be mostly prepositions, conjunctions and determiners.

For, eg - words like ‘ about’, ‘ and’, ‘ between’.

	Grammatical Part of Speech
Sentence	about - preposition
tended to have	about - preposition
more fantasies	about - preposition
aboutviolencean	

d to think

violence in real

the debates

over video

game-playing and -

andviolenceapp conjunction

ear to be

coming back.

of a violent

outcome. Not

allviolenceis the

result of violent

game

all -

determiner

Words appearing to 1R of ' violence' are mostly verbs and nouns.

Grammati

Sentence

cal Part of

Speech

does the ubiquity

of video

gameviolencebe

get real-life

violence?

beget -

verb

frequently

embed cues that

effectively

frameviolenceen

enacted -

acted against

verb

seemingly social

beings as

ESSENTIAL

FACTS ABOUT

Games

andViolenceFacts

Facts -

, common sense,

noun

and numerous

studies

When fixing the sort levels as mentioned before while searching for ' violence' to look for concordances, we see that 1R to violence were verbs and nouns, the sentence did sound in perfect context and were meaningful.

When the sort levels are changed further, we see that words appear to be out of context thus deeming the entire sentence meaningless.

Thus, ' violence' along with the changing sort levels makes and breaks sentences in and out of context.

When searching for ' study\*', we see that the number of concordance hits are 134. The results show different forms of the word ' study' such as "

<https://assignbuster.com/corpus-analysis-on-violence-in-video-games/>

studying”, “ study’s” wherein the sentences containing these words make perfect sense.

For eg, particularly problematic in the context of studying antisocial behaviours, such as aggression

subsequent increases in physical aggression,” the study’s lead author Jay Hull said in

Collocations:-

Collocations are an arrangement of words or terms that follow more generally than would be predicted by chance ([https://www.jstor.org/stable/27919940?seq=1#page\\_scan\\_tab\\_contents](https://www.jstor.org/stable/27919940?seq=1#page_scan_tab_contents)).

The understanding to use a particular language in this case, English, comprises of a recognition of a peculiar feature of the language known as collocation.

Collocation is that practice of the language whereby two or more words work together in a certain speech or writing.

On changing the window span from 1L to 1R and minimum collocate frequency to 3, keeping the collocate measure set to MI and searching for the term ‘ violence’, the total number of collocate types obtained is 32 and total number of collocate tokens is 331.

We see various combinations of words and types of collocations. Namely, there would a noun and noun combination, preposition and noun combinations.

<https://assignbuster.com/corpus-analysis-on-violence-in-video-games/>



Listing some of the examples below:

Noun and noun combination when ' violence' is searched:

	Grammatical
Sentence	Part of
	Speech

might

frequently

embed cues

that effectively frame

frame violence violence -

enacted noun + noun

against

seemingly

social beings

for the same

reasoning as

between

violence and

real-life crime

between

violence -

preposition

+ noun

Clusters/N-Grams:-

The Clusters tool is a much better enhancement to the Collocation tool in Antconc. It is often a much better of showing sorted groups of co-occurring words.

When we search for ' violent' in the Clusters tool in Antconc with cluster size set minimum to 2 and maximum to 3, we see combinations of words such as ' violent video', ' violent crime',

' violent media', ' violent video games', ' violent behaviour' , etc.

Attaching a small output of the result showing the frequency and range of the above mentioned clusters of words.

Cluster	Rang e	Frequenc y
violent video	23	341
violent crime	6	14
violent media	9	19
violent video games	23	200

violent                    3        4  
behaviour

On checking for N-Grams through Antconc, we see the total number of n-gram types to be 95085 and tokens to be 127507.

Most noticeable N-Grams are ' violent video games'.

Keywords:-

Keywords are words that occur in a particular speech or text more often than you would normally expect (<https://ota.ox.ac.uk/documents/searching/handbook.html>).

Keywords are calculated based on statistical tests which generally compare word frequencies in a speech or text against their predicted frequencies that are derived from a much larger corpus which we would use as a reference for normal language usage.

Keyness would then be the trait a word or a given phrase has of being " key" in context.

After having attached the BNC\_Wordlist file in Antconc and running the Keyword List tool, we see words sorted out based on their keyness and their frequency.

Words that show up in the top 20 results namely are ' violent', ' game', ' aggression', ' video'.

Attaching an output of the result showing the keyness and their frequency of the above mentioned words.

Keyword	Keyness	Frequency
violent	+ 5646.03	605
game	+ 1784.01	344
aggression	+ 1196.19	148
video	+ 6985.15	831

With a high keyness score, it could be assumed in a way that playing violent games can often lead to violent aggression in our day-to-day lives.

Having chosen ' violent' from the keyword results list and checking them over in the concordance tool, we see the number of hits to be 605. Also, we have set the sort level to 1L, 2L and 3L, we see collection of words wherein ' association' is being connected to violence. Also, words such as ' disengagement', ' embedded', ' link' are in a way show strong associations with the word violence.

Conclusion:-

By concluding this essay, we can see that this corpus that I have chosen to study can offer heaps in revealing definite and distinct information about certain categories and recurrence of collocations of a given word/words in my chosen corpus which is violence in video games.

What could be affirmed in a broader and greater-scale corpus study of the closeness and distinctions in checking for violence in video games is certainly more than what was able to be stated here.

While my understanding of the closeness and distinctions has expanded considerably, I remain uncertain of the actual usage of violence in video games and how drastic the results would be when a larger corpus would be considered.

#### References:-

- Anthony, L. ( 2014). AntConc 3. 4. 1. Retrieved from <http://www.antlr.lab.sci.waseda.ac.jp/index.html>
- British National Corpus (BNC). ( 1999). The BNC sampler corpus. Oxford, UK: Oxford University Computing Services [= BNC Sampler].
- Hunston, S. ( 2002). Corpora in applied linguistics. Cambridge, UK Cambridge University Press.
- <https://ota.ox.ac.uk/documents/searching/handbook.html>
- <https://pdfs.semanticscholar.org/fa9a/84ae6215251b879569c4f902e14d71fb49f9.pdf>
- [https://www.jstor.org/stable/27919940?seq=1#page\\_scan\\_tab\\_contents](https://www.jstor.org/stable/27919940?seq=1#page_scan_tab_contents)