

Harry Potter Analysis

Or Pickholz

During the summer break between the second and third year of my Computer Science and Statistics B.Sc degree, I decided to analyze the scripts of the Harry Potter movie series: 1,2,3,4 and 6.

I downloaded Harry Potter scripts 1-3 from "Kaggle", and extracted Harry Potter scripts 4 and 6 using web scraping.

For my analysis I used Python, SQL, R, and libraries such as pandas, sqlalchemy, networkX, BeautifulSoup, etc.

In this document I share some of my conclusions. My code can be found in my GitHub account: <https://github.com/orpick/Harry-Potter-Analysis.git>

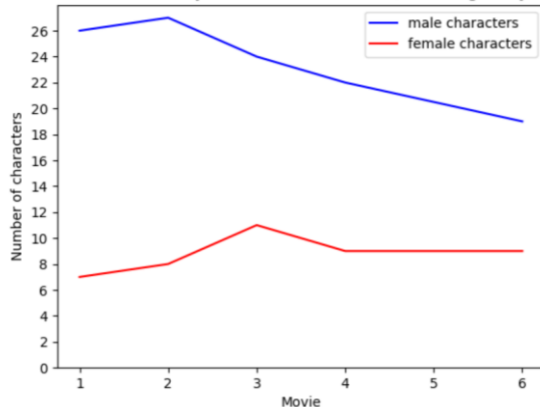
Note: movies 5, 7 part I and 7 part II were not available online in a format different from PDF. Please notice all the analysis and graphs exclude those films.

Gender Equality

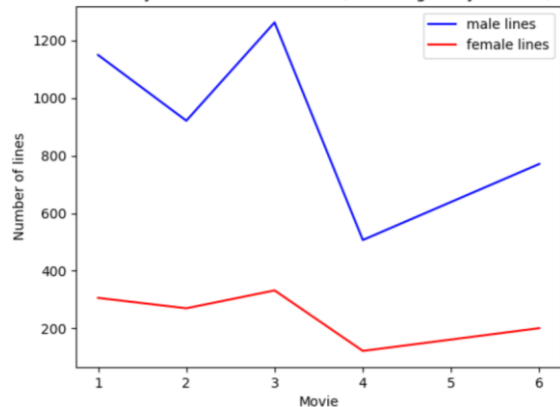
I was interested in exploring the question if there were any gender differences in the movies' scripts:

My first conclusion was that the number of male characters was 2-4 times larger than the number of female characters in all of the movies, and as a result so was the number of lines said by male vs. female characters :

Number of Characters by Gender in Each Movie (Excluding Harry Potter 5)



Lines by Gender in Each Movie (Excluding Harry Potter 5)



However, considering the number of female/male characters, **it is statistically significant that in movies 1-4 female and male characters say an equal amount of lines!** (in Harry Potter 6 the variances differed, so I did not perform a t-test)

*Full R code is on GitHub

Female and Male characters have equal amount of lines in Harry Potter 1

Welch Two Sample t-test

```
data: unlist(HPMale[i]) and unlist(HPFemale[i])
t = 0.018372, df = 12.076, p-value = 0.9856
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -56.17284  57.12888
sample estimates:
mean of x mean of y
44.19231  43.71429
```

Female and Male characters have equal amount of lines in Harry Potter 2

Welch Two Sample t-test

```
data: unlist(HPMale[i]) and unlist(HPFemale[i])
t = 0.02093, df = 16.221, p-value = 0.9836
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -36.17434  36.89656
sample estimates:
mean of x mean of y
34.11111  33.75000
```

Female and Male characters have equal amount of lines in Harry Potter 3

Welch Two Sample t-test

```
data: unlist(HPMale[i]) and unlist(HPFemale[i])
t = 0.99183, df = 25.499, p-value = 0.3306
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -24.06924  68.87227
sample estimates:
mean of x mean of y
52.58333  30.18182
```

Female and Male characters have equal amount of lines in Harry Potter 4

Welch Two Sample t-test

```
data: unlist(HPMale[i]) and unlist(HPFemale[i])
t = 0.93145, df = 26.699, p-value = 0.36
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11.4257  30.4055
sample estimates:
mean of x mean of y
23.04545  13.55556
```

Variances Differ in Harry Potter 6

```
var.test(unlist(HP6Female), unlist(HP6Male), alternative = "two.sided")
```

```
##
## F test to compare two variances
##
## data: unlist(HP6Female) and unlist(HP6Male)
## F = 0.22068, num df = 8, denom df = 18, p-value = 0.03507
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.07343034 0.89016287
## sample estimates:
## ratio of variances
##      0.2206781
```

Spells

I was curious to see if there were any patterns among the spells the three main characters (Harry, Hermione and Ron) cast throughout movies 1,2,3,4 and 6:

Harry:

Spell	Number of Times Casted
Expecto Patronum	6
Lumos Maxima	4
Sectumsempra	3

It is no surprise that Harry uses the "Patronus Charm" the most as he was consistently attacked by Dementors.

Hermione:

Spell	Number of Times Casted
Oculus Reparo	2
Alohomora	2
Petrificus Totalus, Wingardium Leviosa, Bombarda, Finite, Immobulus, Oppugno	1

Ron:

Spell	Number of Times Casted
Alohomora	2
Vera Verto, Wingardium Leviosa, Riddikulus	1

Although it seems as if Hermione casted the most distinct spells among the three, it was Harry that casted the most spells.

It is still not surprising though that Hermione casted this many different spells, given the fact that she is the most educated among the three.

I also looked at the most used spell in general throughout Harry Potter 1,2,3 and 6:

Movie	Most used spell(s)	Number of times used	Common caster
Harry Potter 1	Alohomora	4	Hermione
Harry Potter 2	Arania Exumai, Vera Verto	3	Harry, McGonagall (respectively)
Harry Potter 3	Riddikulus	8	Lupin
Harry Potter 6	Sectumsempra	3	Harry

In Harry Potter 1, "**Alohomora**" is the most used and a relatively simple spell, so no wonder it was used the most in the first movie by a first-year student, Hermione.

In Harry Potter 2, "**Arania Exumai**" and "**Vera Verto**" were used the same number of times. "**Arania Exumai**" was used the most by Harry, when he and Ron faced Aragog's family. "**Vera Verto**" was used the most by professor McGonagall while she taught this spell in class.

In Harry Potter 3 there's a whole scene where Remus Lupin teaches the students the "**Riddikulus**" spell, so it is understandable why it's the most used spell in that movie.

In Harry Potter 6, Harry is fascinated by the Half Blood Prince's accuracy in potion making, so when he discovers his special spell "**Sectumsempra**" he infers that this spell is powerful and uses it.

*In Harry Potter 4 script there were no spells! I assume it's because the script I used was an initial version that was later changed.

Common Connections

I examined who the three main characters referred to the most, excluding references to one another. It was not surprising that all three referred to Hagrid a lot, given the fact that many scenes took place at Hagrid's Hut and that Hagrid was their friend:

Harry:

Character's Name	Number of References
Hagrid	34
Dumbledore	21
Dobby	16

Dumbledore is a huge part of Harry's life, so the fact that he's referred to him a lot is not a surprise as well. In addition, Dobby is also in the list due to his many attempts to make Harry leave Hogwarts in Harry Potter 2.

Hermione:

Character's Name	Number of References
Dumbledore	7
Hagrid	7
Mcgonagall	4

Hermione referred to schoolteachers more than anyone other than Harry or Ron, which suits her studious personality.

Ron:

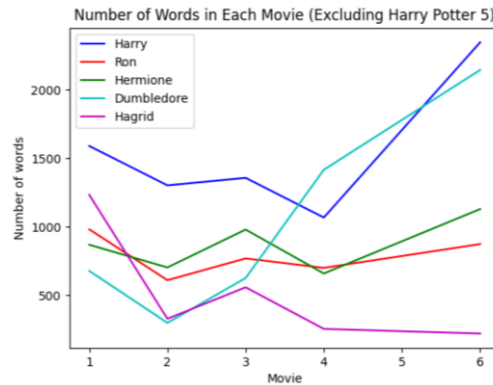
Character's Name	Number of References
Malfoy*	7
Dad	7
Hagrid	5

*Draco Malfoy

Ron's references to his dad makes sense since Ron's family has a big part throughout the series, and his father also works at the ministry of magic.

Comparing Movies

I tried to compare between the number of words said by 5 main characters (Harry, Hermione, Ron, Dumbledore and Hagrid) in order to understand their involvement in the plot throughout the series.



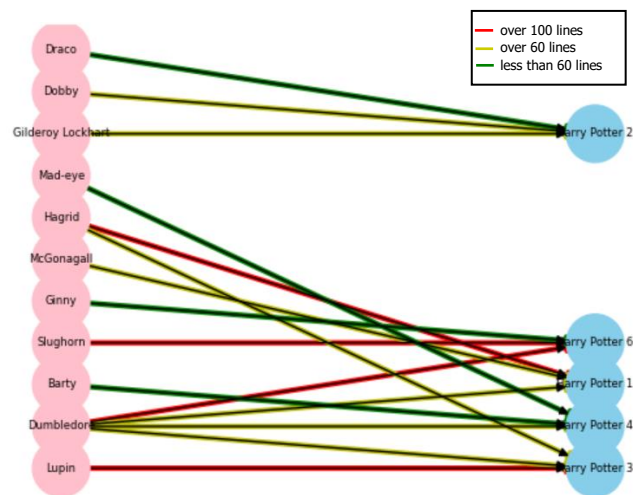
Harry speaks the most in all the movies except for Harry Potter 4 – in that movie Harry mainly participates in the Triwizard Tournament by himself and less connects with other characters.

As Harry grows up Dumbledore gets more and more involved in his life, therefore the consistent incline in Dumbledore words in the movies.

Ron and Hermione speak pretty much an equal amount of words, although from watching the movies we get the feeling that Hermione talks a lot more.

It seems that in later movies Hagrid gets a smaller part. It makes sense since as Harry, Hermione and Ron grow up, they are more focused on complicated issues and spend more time solving them instead of hanging around with Hagrid.

Another thing I was interested in examining was who were the top three characters in each movie which had the most lines, excluding Harry, Hermione and Ron:



We can see that only Dumbledore and Hagrid belonged to this category in more than one movie. This is reasonable since the plot gets more complex throughout the films, and as a result we are exposed to a variety of important characters as the story evolves.

Another interesting thing is that in Harry Potter 2, all top three characters are unique to this movie. Even Dumbledore is not a top character!

Voldemort References

In order to examine how the conflict with Voldemort evolved throughout the series, I looked at the number of 'Voldemort', 'Tom Riddle', 'the Dark Lord' references in each movie:



It is not surprising that as the plot evolves, there are more references to Voldemort, especially after Harry Potter 4, where he comes back to life.

In Harry Potter 3 the main plot is about the prisoner of Azkaban, so the very few references are reasonable.