

# Overcoming the Language Barrier: American Sentiment Towards South Korean Comedy-Horror Movies

Using Python for sentiment analysis on movie reviews, we can investigate the public audience sentiment towards *The Host* and *Parasite*, two Korean films directed by Bong Joon-Ho.



Image by Mary Shin

It's an understatement to say that *Parasite* (2019) was a success. Director Bong Joon-Ho's latest masterpiece earned more than \$258 million worldwide ([Box Office Mojo](#)) and won 4 Oscars ([IMDb](#)) on its way to become one of the most acclaimed films of the 21st century.

Does this mean *Parasite* is Director Bong's best work so far? Although I'll leave that up to your own opinion, what I want to emphasize is that more often than not, our previous work always sets us up for future success. Before his 2019 multi-Oscar-winning film, Director Bong already had an impressive arsenal of movies, including *Snowpiercer*, *Okja*, *Mother*, and *The Host*.

So my real question is, when it comes to American perception of these Korean movies: **How did people rate *Parasite* compared to his previous work?**

## Enter: *The Host* (2007)

What came a decade before *Parasite* was the thrilling, hilarious, heart-wrenching rollercoaster of emotions that is *The Host*. I chose to focus on the differences between movie reviews of *Parasite* and *The Host* because although they have completely different plots, both movies fall under the open-ended film genre of Comedy-Horror (and of course, they both star the amazing Actor Song Kang-Ho).

When looking at movie reviews, there are two categories to consider: 1) film critic and 2) public audience. Most of my research for this topic relies on the American movie review site, [Rotten Tomatoes](#).

First, I'll briefly mention the differences between how film critics have reviewed each movie, before diving into the bulk of my research on public audience reviews.

## Critic Reviews of *The Host* and *Parasite*

Although reviews sites like Rotten Tomatoes and also [Metacritic](#) both showcase film and television critic reviews, my access to many of these reviews was limited by paywalls and archived web pages. Therefore, I resorted to my dear friend, [Nexis Uni](#), which is an academic search engine featuring more than 17,000 news, business, and legal sources.

From this database, I was able to run search queries for newspaper, magazine, and blog articles that were movie reviews of each film, download the document results, and clean them (extract the review text). I ended up with 27 critic reviews of *The Host* and 261 of *Parasite*.

Using these two lists of review texts, I ran sentiment analysis (SA)—a broad term used to describe Natural Language Processing (NLP) approaches to quantifying the types and degree of emotions expressed in text—and specifically used [VADER](#) (Valence Aware Dictionary and sEntiment Reasoner), which is a lexicon and rule-based sentiment analysis tool.

Running SA using VADER allowed me to get polarity (negative and positive) scores, and I computed the mean compound (normalized,

weighted composite score that runs from -1 very negative to 1 very positive) sentiment scores for each list. I ended up with mean compound sentiment scores of 0.108 for critic reviews of *The Host*, and 0.878 for that of *Parasite*. Similarly, retrieving polarity scores using another Python NLP processor called TextBlob, I calculated mean polarity sentiment scores of 0.058 for *The Host* and 0.190 for *Parasite*.

Results from using both SA tools show that there was, on average, **higher positive sentiment across critic reviews of *Parasite*.**

However, VADER does have its limitations because it only looks at each specific word and fails to consider certain nuances or context clues. In addition, due to the large discrepancy in the size of the data sets (27 vs. 261), this is not the most valid comparison. Nevertheless, that in and of itself is a noteworthy observation: there's generally much more "buzz" about *Parasite* than about *The Host*, and we can overall conclude that there has been less attention devoted to the latter movie in mainstream American media.

If this is the case for critic reviews, what can we understand about public audience reviews?

## Public Audience Opinion Via Rotten Tomatoes

I chose Rotten Tomatoes, an American website that aggregates reviews, because of its immense popularity and also my personal familiarity with it. After getting the urls of both movies on the site, I web-scraped the public audience reviews and collected the date, score, and text of each review.

Because I wanted a similar timespan, and *Parasite* came out 2 years ago in 2019, I also limited reviews of *The Host* to within 2 years. Since it was released to the American public in 2007, I filtered the reviews to only include the ones that were posted between 2007 and 2009. This also "protects" *The Host* reviews from becoming biased due to a reviewer's familiarity with Director Bong's recent film(s) that could potentially affect how they view his earlier work. After scraping, cleaning, and organizing the reviews, I ended up with 4,855 of *The Host* and 4,409 of *Parasite*.

With this data, I ran several analyses using Python to find out **whether or not *Parasite* was more well-received (higher sentiment and**

more positive reviews) than *The Host*:

## 1) Dividing into Positive and Negative Reviews

The first analysis I conducted was to split up the reviews of each movie into positive and negative ones. Since the reviews were each scored on a 5-star rating scale (0 to 5), I wrote code to go through each review and divide them into positive if it's greater than or equal to 4 stars, negative if it's less than 3 stars, and neutral for all else.

My results are as follows: for audience reviews of *The Host*, about 49% of them were positive, and about 51% negative and neutral. For *Parasite*, about 79% of reviews were positive, and 21% negative and neutral.

**That is, about half of *The Host* reviews were positive, compared to more than three-quarters of *Parasite* reviews being positive.**

That's a huge, significant difference! The audience scores for *Parasite* are overwhelmingly more positive than negative or neutral, while it's pretty much divided along the middle for *The Host*. Let's take a quick look at a few samples of positive and negative reviews.

For *The Host*:

*"I liked this a lot more than Cloverfield. Great cinematography on top of a fun action packed monster drama. South Korea knocks it out of the park again."*

*"A Korean comedy-horror movie that fails miserably on both counts. But then maybe it's me; this was the highest grossing film in Korean movie history as [of] January of this year."*

For *Parasite*:

*"Parasite completely lives up to being the Oscar winner that it is...It's a masterpiece that deserves the popularity it got and Bong Joon-Ho shall be remembered for it for days to come."*

*“Parasite is fundamentally no different from a B-list horror movie where the situation only happens because of continued awful decisions on the part of the main cast...it’s not something you are ever likely to watch twice and I seriously doubt it’s something that will be remembered as anything other than ‘that foreign film that won best picture that one time.’”*

At this point, it’s worth asking: **What exactly is the public audience saying about these two movies? Are there specific words in the reviews that stand out for positive versus negative ratings?**

## 2) Keyness Analysis

For the second part of my data analysis, I focused on what’s called keyness analysis, which represents the comparison of normalized frequencies of words or phrases in two corpora (collections of texts). This is useful because it allows us to compare the most frequently used words or phrases in two differently sized data sets, which is perfect for my case! Therefore, using keyness analysis here can effectively help us understand what kind of lingo is used in positive and negative reviews.

My results showed that for reviews of *The Host*, the word “monster” occurred about 79 times more frequently among positive reviews than negative, and “korea” about 15x more. On the other hand, some of the words that occurred more frequently among negative reviews were “bad” at about 135x more, and “translation” at around 26x more.

### *The Host*



Image by Mary Shin; Source: Rotten Tomatoes

For *Parasite* positive reviews, the word “bong” (Director Bong) occurred about 110 times more frequently than in the negative reviews,

and “best” appeared around 49x more frequently. Among negative reviews, “boring” occurred 94x more frequently than in positive reviews, and “overrated” at about 74x more.

### *Parasite*



Image by Mary Shin; Source: Rotten Tomatoes

In addition, “bong” came up at the top of the most frequently used words for positive *Parasite* reviews, but not towards the top for positive reviews of *The Host* (i.e., he’s perhaps receiving more credit for the recent film)! It’s also worth noting that among positive reviews for both movies, there’s a lot more attention to them being a **Korean** film and more mentions of the South Korean film and entertainment industry in general.

Let’s look at a small random sample of the context around “korean”, for *The Host*:

“it wouldn’t be korean without over-the-top melodrama”

“recommended korean monster-movie mayhem”

“surprisingly heartfelt korean monster film”

And for *Parasite*:

“like a korean quentin tarantino”

*“high bar set by korean cinema”*

*“kfilm kpop kdrama korean entertainment is powerful”*

These results show that the American public is regarding “Korean cinema” as its own, distinct category, as opposed to generalizing Korean cinema as Asian entertainment, for example.

Furthermore, “translation” only showed up at the top most common words for negative reviews of *The Host*, including statements like:

*“terrible translation”*

*“the translation doesn’t really work”*

*“the English translation was absolutely horrific”*

I find this fascinating because it means that part of the reason why these ratings were negative was due to perceived problems with the translation, and this aligns with what Director Bong has famously said: **“Once you overcome the 1-inch tall barrier of subtitles, you will be introduced to so many more amazing films.”** Perhaps aversion to subtitles has decreased since 2007? That’s a research question for another time.

Now, it’s time to go even deeper. Let’s try to further quantify these reviews with sentiment analysis!

### 3) Sentiment analysis using NRC Emotion Lexicon

Because of VADER’s limitations, it’s worth looking at other lexicons for sentiment analysis. The NRC Emotion Lexicon (EmoLex) is another tool that can be much more helpful if we want to more accurately capture the valence of emotions used throughout texts. So for the third and last part of my analysis, I used the NRC EmoLex—a list of English words and their associations with 8 basic emotions (anger, fear, disgust, sadness, anticipation, trust, surprise and joy) and 2 sentiments (negative and positive).

Using crowdsourcing, over 14,000 words have been classified as to which of these 10 categories they are judged to belong to. For instance, the word “powerful” is found in: positive, trust, joy, anticipation, fear, anger and disgust. Whereas the word “foreign” only occurs in the negative category. This reflects the fact that words can have multiple senses or emotions based on the context.

The EmoLex is useful for us because it can try to capture more specifically the emotions used across the audience reviews and help us understand which emotions are prevalent for reviews of each movie. Hence, we can ask: **for each of the 8 emotions and 2 sentiments, do the reviews of *Parasite* or *The Host* have a higher average count of words that fall under that category?**

To answer this question, I coded a function to score every single audience review with the 10 different categories in the EmoLex so that for each review, the words are organized under the 8 emotions and 2 sentiments. Then, I counted up the number of words that fall under each category and found the average counts of words throughout reviews of both movies.

In terms of positive or negative sentiment, here are the average count of words, by movie:

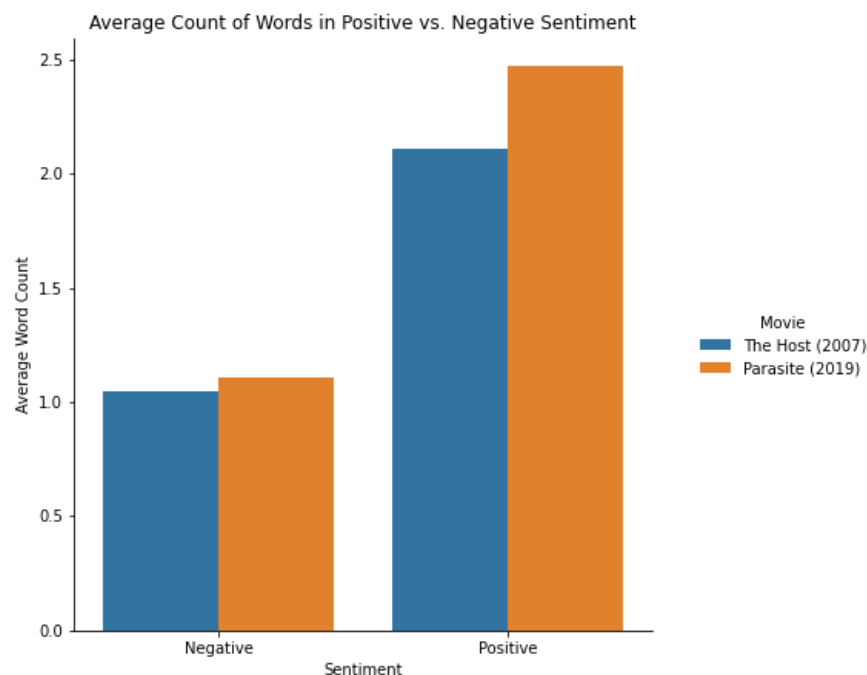


Chart by Mary Shin; Source: Rotten Tomatoes



You can see that the average count of words under **the positive sentiment category increased for *Parasite* than *The Host***, and there was also a slight increase (but way smaller) for negative sentiment words.

Here are some of the words that were driving the increase in positive sentiment throughout reviews of *Parasite* when compared to *The Host*:

*masterpiece unexpected clever absolute art script worthy  
perfect captivating beautiful director knowing powerful friend praise  
working reading creative fascinating*

Image by Mary Shin

For the negative sentiment category, these are the words that occurred more frequently in reviews of *Parasite*:

*unexpected unpredictable struggle shock leave outstanding  
conflict unbelievable feeling doubt immediately*

Image by Mary Shin

And negative sentiment words that occurred more frequently in reviews of *The Host*:

*bad shit crap silly terrible stupid hell annoying dumb  
horrible lost weird case disappointed gore hate fear*

Image by Mary Shin

Looking at the *Parasite* negative sentiment words, we can see how EmoLex has its own limitations—words like “outstanding” and “feeling” are included under this sentiment category, even though they may not directly express negativity. This reflects the difficulty of sentiment analysis: words can have multiple sentiments and meanings based on the context, and language is very idiosyncratic!

What’s also noteworthy here is that the negative sentiment words for *Parasite* reviews are more about the “shock factor”, i.e. the film being unexpected or surprising. And I believe words like “conflict” or “struggle” might actually be about the plot of *Parasite*. On the other hand, for negative sentiment words that occurred more often in

reviews of *The Host*, they're more degrading and derogatory, like "shit", "crap", or "stupid".

After running the same type of analysis to investigate words under the other 8 emotion categories (anger, disgust, sadness, fear, joy, trust, surprise, and anticipation), here's what I found:

**On average, the count of words under Trust, Surprise, Joy, Anticipation, as well as Anger and Sadness, were all higher for reviews of *Parasite*. By contrast, words that fall under Fear and Disgust occurred much more frequently in reviews of *The Host*.**

Hand-picking a few of the emotion categories now, let's take a look at some of the key words in Trust that were more frequently used in reviews of *Parasite*:

*worthy director perfect powerful friend praise system brilliant  
humanity excellent true winning engaging important deserve favorite  
feeling title personal constantly doubt moral*

Image by Mary Shin

In contrast, for the Fear and Disgust categories, these are some of the words used more often in reviews of *The Host*:

*bad terrible shit hell crap horrible nasty gore hate hanging weird*

Image by Mary Shin

The numeric differences are small, but nonetheless, I discovered that **audience reviews of *Parasite* generally contain more positive sentiment words, as well as less derogatory words under the negative sentiment category, than that of *The Host*.**

## Major Takeaways

Through my analysis, we can start to understand how *Parasite* and *The Host* were received in American culture by using data scraped from public movie review sites. I looked through film critic and public audience reviews of each movie, and quantified the comparisons between the reviews by running sentiment analysis.

Overall, my main finding is that there's a higher percentage of positively rated reviews and higher positive sentiment scores for *Parasite* than *The Host*.

Although Director Bong had phenomenal earlier films, *Parasite* was the one that managed to break the language barrier by becoming a popular, overall highly rated, and well-known South Korean movie in mainstream American culture. In 2007, issues with English translations, dubbing, and subtitles were prevalent and perhaps posed limitations for monolingual American audiences to overcome those barriers. But the times are different now, and I hope that *Parasite* is only the beginning of “foreign” films and directors basking in the spotlight they deserve.