배우/제작진의 스캔들과 영화 평점별 리뷰의 상관관계

중국언어문화전공 201701981 신희진

중국언어문화전공 201703472 진예솔

**Ⅰ. 조사의 배경**

최근 영화계에서는 영화의 외적인 측면에서 나타난 해당 배우나 제작진의 스캔들이 영화 불매운동 또는 영화 리뷰의 평점 테러 등으로 이어지는 현상이 나타나고 있다. 변성현 감독의 〈불한당: 나쁜 놈들의 세상(2016)〉은 개봉 전 감독의 SNS에 올라온 정치적 발언과 특정 지역 비하 발언 등의 논란으로 인해 손익분기점의 절반도 넘기지 못했을 정도로 흥행에 실패하였다. 그러나 독특한 촬영기법의 활용 및 타 누아르 장르의 영화와는 다른 스토리 전개 등 영화 내적인 요소로 인해 극찬을 보내는 관람객도 있어, 이는 곧 한국 영화계에서 보기 드물게 영화에 대한 평이 극명하게 갈리게 되었다. 따라서 이 영화 〈불한당: 나쁜 놈들의 세상(2016)의 영화 평점과 리뷰 내용 분석을 통해 관람객의 평가 양상을 파악하고, 영화의 외적인 스캔들이 해당 영화의 평점과 리뷰에 어떠한 영향을 끼치는지 알아보고자 하였다.

다만, 하나의 영화만을 그 대상으로 하는 것은 다소 조사 결과의 신뢰도가 낮을 수 있다고 판단해, 영화의 외적인 요소인 배우 및 제작진의 스캔들이 해당 영화의 평점과 리뷰에 영향을 주었을 것으로 판단되는 다음의 세 영화를 그 대상에 함께 포함하였다. 홍상수 감독의 〈지금은맞고그때는틀리다(2015)〉와 〈밤의 해변에서 혼자(2016)〉는 감독과 주연 배우인 김민희의 불륜 관계로 인해 논란이 거셌던 영화였다. 이로 인해 관람객 리뷰에는 ‘불륜을 영화로 합리화해서는 안 된다’는 의견과 ‘불편하면 보지 않으면 된다’는 의견이 공존하게 되었다.

또한, 니키 카로 감독의 〈뮬란(2020)〉은 많은 이들의 호평을 얻은 디즈니의 애니메이션 실사화 영화인 〈알라딘(2019)〉과 달리, 미투 운동과 정치적 올바름 등을 반영해 원작과 큰 내용적 차이가 생겨났다는 점이나 주연 배우였던 유역비의 홍콩 경찰 옹호 발언과 중국 자본 투자의 영향으로 중국 내 소수민족인 위구르족 탄압 단체가 엔딩 크레딧에 언급되었다는 점 등으로 인하여 혹평을 받게 된 영화이다. 따라서 지금까지 제시한 영화의 평점 및 리뷰를 수집하여 해당 영화 관람객의 평점별 언어 사용 등을 분석하고, 리뷰 내용 간 공통점과 차이점을 알아보고자 한다.

**Ⅱ. 데이터 수집 및 분석 방법**

먼저 해당 프로젝트의 조사 대상이 된 영화는 위에서 언급한 4편이며, 조사에 사용할 데이터는 네이버 영화의 평점 항목 중에서도 관람객 평점과 리뷰 내용을 대상으로 하여 웹 스크래핑 방식을 활용해 해당 영화의 리뷰를 수집하였다. 이를 통해 변성현 감독의 〈불한당: 나쁜 놈들의 세상(2016)〉의 평점과 리뷰 12,580개(공백 제외 시 12,468개), 홍상수 감독의 〈지금은맞고그때는틀리다(2015)〉와 〈밤의 해변에서 혼자(2016)〉의 평점과 리뷰 4,470개(공백 제외 시 4,466개), 니키 카로 감독의 〈뮬란(2020)〉의 평점과 리뷰 4,520개(공백 제외 시 4,313개)로 총 21,570개(공백 제외 시 21,247개)의 리뷰를 수집할 수 있었다. → GitHub 내 각각 1\_1, 1\_2, 1\_2\_1, 1\_2\_2, 1\_3으로 시작하는 웹스크래핑 관련 ipynb, csv, excel 파일 참조

위에 제시한 방법으로 수집된 감독별 영화 리뷰는 konply 한국어 형태소 분석기를 사용하여 해당 리뷰 언어 분석의 기초를 마련하였다. 이렇게 정리된 감독별 영화 리뷰의 형태소 자료는 다시 Word Cloud Generator와 Excel의 피벗 테이블을 활용하여 시각화를 진행하였다. 다만 본격적인 시각화를 진행하기에 앞서, 조사와 접사 등 한국어의 언어적 특성상 어떤 문장에서든 빈번히 나타나는 형태소를 제거하고 시각화를 진행하는 것이 더 효과적이라고 판단하였다. 따라서 품사별로 정리된 형태소를 다시 실제적 의미를 담고 있는 실질 형태소와 문법적 역할을 담당하는 형식 형태소로 구분하고, 이 중에서 실질 형태소인 Adjective, Adverb, Alpha, Exclamation, Foreign, Hashtag, Korean Particle, Modifier, Noun, Number, Verb의 총 11가지 품사만을 시각화에 활용하였다. → GitHub 내 각각 2\_1, 2\_2, 2\_3으로 시작하는 형태소분석 관련 ipynb, csv, excel 파일 참조

Excel의 필터 기능을 사용하여 위에서 제시한 11가지의 품사만을 선별한 뒤로는, 감독별 영화 리뷰를 다시 평점에 따라 1점부터 3점까지, 4점부터 7점까지, 8점부터 10점까지의 세 단계로 나누어 Word Cloud Generator(<https://www.jasondavies.com/wordcloud/>)를 통해 시각화 이미지를 도출하였다. 해당 이미지 도출 전 재조정한 설정값은 Spiral: Archimedean, Scale: √n, 5 orientations from 0° to 0°, Numbers of words: 250이다. → GitHub 내 각각 3\_1, 3\_2, 3\_3으로 시작하는 Word Cloud 관련 zip 파일 참조

분석은 첫째로 Word Cloud 시각화 이미지를 통해 평점별 주요 키워드가 나타나는 양상을 비교하였으며, 다소 특이한 키워드의 경우에는 일부 원문 분석도 진행하였다. 둘째로 Excel의 피벗 테이블에서 위에서 제시한 평점별 구간에 따라 Word Cloud에서 나타난 주요 키워드의 구체적인 빈도를 파악하고, 시각화에서 나타나지 않는 차이를 추가로 분석하였다. → GitHub 내 각각 4\_1, 4\_2, 4\_3으로 시작하는 피벗테이블 관련 excel 파일 참조

**Ⅲ. 데이터 분석 결과**

**1. Word Cloud 분석**

가. 불한당: 나쁜 놈들의 세상(2016)

|  |  |
| --- | --- |
|  | 불한당: 나쁜 놈들의 세상(2016)  평점 1-3 |
|  | 불한당: 나쁜 놈들의 세상(2016)  평점 4-7 |
|  | 불한당: 나쁜 놈들의 세상(2016)  평점 8-10 |

평점 1-3점에서는 ‘일베’, ‘홍어’, ‘전라도’, ‘SNS', ‘트위터’, ‘인성’, ‘새끼’, ‘망한’, ‘쓰레기’, ‘재미없어요’, ‘재미없음’ 등의 키워드가 나타났다. 영화 자체를 폄하하는 키워드도 나타나지만, 감독의 언행 및 정치 성향 등 논란이 되었던 주제와 관련한 키워드가 다수를 이루었다. 평점 4-7점에서는 ‘논란’, ‘그냥’, ‘킬링타임’, ‘그럭저럭’, ‘짬뽕’, ‘아쉽다’, ‘수준’, ‘솔직히’, ‘비슷한’, ‘괜찮은’, ‘그저’, ‘캐스팅’, ‘브로맨스’, ‘조폭’, ‘깡패’ 등의 키워드가 나타났다. 평점 1-3점이나 8-10점보다는 다소 영화 내적인 요소인 캐스팅이나 스토리와 관련한 키워드가 두드러지는 양상을 보였다. 평점 8-10점에서는 ‘느와르’, ‘연기’, ‘스토리’, ‘연기력’, ‘좋았습니다’, ‘였습니다’, ‘봤습니다’, ‘재밌었어요’, ‘좋았어요’, ‘보세요’ 등의 키워드가 나타났다. 평점 4-7점과 비슷하게 영화의 내부적 요소인 스토리나 배우의 연기력에 대해 언급하고 있지만, 주로 ‘좋다’, ‘재밌다’ 등의 긍정적인 감정 표현이 두드러진다.

‘감독’이란 키워드는 모든 평점에서 주요 키워드로 등장하는데 평점이 낮아질수록 더 크게 나타나고, 주연배우의 본명인 ‘임시완’, ‘설경구’의 키워드는 평점이 높아질수록 더 크게 나타나는 경향을 보였다. ‘신세계’, ‘프리즌’, ‘무간도’ 등 느와르 장르의 다른 영화제목 역시 모든 평점 구간에서 나타나고 있으나, 실제 원문을 분석하면 완전히 반대되는 문맥을 나타내고 있다. 높은 평점에서는 ‘불한당이 ㅇㅇ 영화보다 낫다’ 등의 의미로 언급되지만, 낮은 평점에서는 ‘감히 ㅇㅇ 영화와 비교를 하느냐’ 등의 의미로 언급되고 있다.

나. 지금은맞고그때는틀리다(2015), 밤의 해변에서 혼자(2016) (이하 ‘홍상수 영화’로 지칭)

|  |  |
| --- | --- |
|  | 〈지금은맞고그때는틀리다(2015)〉, 〈밤의 해변에서 혼자(2016)〉  평점 1-3 |
|  | 〈지금은맞고그때는틀리다(2015)〉, 〈밤의 해변에서 혼자(2016)〉  평점 4-7 |
|  | 〈지금은맞고그때는틀리다(2015)〉, 〈밤의 해변에서 혼자(2016)〉  평점 8-10 |

평점 1-3점에서는 ‘불륜’, ‘쓰레기’, ‘아깝다’, ‘미화’, ‘합리화’, ‘최악’, ‘더럽다’, ‘역겹다’, ‘싫다’, ‘간통죄’, ‘자식’, ‘부인’, ‘열자’ 등의 키워드가 나타났다. 홍상수 감독의 불륜 스캔들과 관련한 키워드가 주를 이루고 있다. 이중 ‘열자’의 키워드는 리뷰 작성 시 최소 기준인 열 글자를 쓰기도 귀찮다거나 시간이 아깝다는 의미에서 나타난 것이다. 평점 4-7점에서는 ‘그래도’, ‘좋았다’, ‘조금’, ‘지루한’, ‘변명’, ‘매력’, ‘틀리다’, ’사실’, ‘찌질’, ‘해변’ 등의 키워드가 나타났으며, 평점 8-10점에서는 ‘연기’, ‘정재영’, ‘배우’, ‘좋았다’, ‘대사’, ‘장면’, ‘일상’, ‘감정’, ‘봤습니다’, ‘좋았어요’, ‘합니다’ 등의 키워드가 나타났다.

전반적으로 ‘영화’, ‘감독’, ‘사랑’ 등의 키워드는 모든 평점 구간에서 많이 등장하고 있다. 스캔들과 관련된 인물인 ‘홍상수’와 ‘김민희’ 역시 모든 평점 구간에서 나타나지만 주로 높은 평점에서 더 크게 나타나고 있다. 낮은 평점에서 영화 외적인 요소인 스캔들에 대한 키워드가 많이 나타남에도 당사자의 이름이 적게 나타나는 것은 인물의 이름을 언급하기보다 ‘불륜’이나 ‘바람’ 등 다른 단어를 통해 언급하고 있기 때문으로 보인다. 이 때문에 ‘불륜’이라는 키워드는 낮은 평점에서 유독 크게 나타나고 있다. 4-7점과 8-10점에서 비교적 크게 나타나는 단어로는 ‘연기’와 ‘정재영’이 있는데, 배우 정재영의 경우 스캔들과는 관련 없이 해당 영화의 주연배우로서만 관련되어 있기 때문으로 보인다.

다. 뮬란(2020)

|  |  |
| --- | --- |
|  | 〈뮬란(2020)〉  평점 1-3 |
|  | 〈뮬란(2020)〉  평점 4-7 |
|  | 〈뮬란(2020)〉  평점 8-10 |

평점 1-3점에서는 ‘중국’, ‘홍콩’, ‘평점’, ‘free’, ‘hongkong’, ‘위구르’, ‘공산당’, ‘인권’, ‘자본’, ‘쓰레기’, ‘보이콧’, ‘최악’, ‘아깝다’, ‘2’, ‘3’ 등의 키워드가 주로 나타났다. 영화 개봉 당시 문제가 되었던 주연배우 유역비의 홍콩 경찰 지지 논란과 관련하여 홍콩 및 중국 내 소수민족과 중국 공산당에 관한 키워드가 주로 언급되었다. 여기서 2와 3이라는 키워드는 2D, 3D에서 나타난 것이며 원작인 2D 영화와 3D 영화인 해당 작품을 비교하고자 사용된 것으로 나타난다. 평점 4-7점에서는 ‘아쉬움’ 솔직히’ ‘그저’, ‘논란’, ‘나름’, ‘자체’, ‘수준’, ‘중국무협, ‘실사영화’, ‘입니다’ 등의 키워드가 나타났다. ‘논란’이라는 키워드가 나타나긴 하지만 평점 1-3점과 달리 논란과 관련한 키워드들이 직접적으로 나타나지는 않는다. 대신 영화의 특성 및 작품성과 관련한 키워드가 주를 이룬다. 평점 8-10점에서는 ‘액션’, ‘감동’, ‘감상’, ‘입니다’, 재미있게’, ‘좋은’, ‘좋았어요’, ‘봤습니다’, ‘였습니다’ 등의 키워드가 나타났다. 주로 ‘감동’, ‘좋다’ 등의 긍정적인 키워드가 나타나며 공손한 존대어 등의 어미 역시 주요 키워드에 포함되었다. 4-7점과 8-10점에서 ‘입니다’라는 키워드가 크게 나타나는 것을 통해 평점에 따라 리뷰 작성자들의 말투 역시 달라진다는 사실을 유추할 수 있었다.

라. 전반적인 분석

평점이 낮을수록 스토리나 연기력, 작품성 등 내적인 요소보다는 제작진의 스캔들에 관련한 키워드가 다수 나타났다. 반면 평점이 높아질수록 배우의 연기력이나 감독의 연출 등 영화와 관련한 키워드가 주를 보이며, 긍정적인 키워드를 통해 리뷰 작성자의 감정이나 감상을 드러내는 양상도 나타났다.

평점 4-7점에서는 낮은 평점과 높은 평점에서 나타나는 키워드를 모두 볼 수 있다. 주로 영화 내적인 요소나 감정을 나타내는 키워드가 이러한 양상으로 나타났다. 반면 평점 4-7점에서 적게 나타나는 것은 영화와 관련된 스캔들에 관한 키워드이다. 평점이 낮은 경우에는 논란을 비판하고자 다수 언급하고, 평점이 높은 경우에는 ‘ㅇㅇ 논란은 아쉽지만’이라는 말을 덧붙이며 키워드가 언급되는 것을 확인할 수 있다. 높은 평점에서 가장 스캔들과 관련한 키워드가 적게 나타날 것이라는 기존 예측과는 달리, 평점 4-7점 구간에서 영화 내적 요소에 집중된 키워드가 가장 많이 나타남을 확인하였다.

또한 평점에 따라 리뷰를 작성하는 말투 역시 달라지는 것을 알 수 있었는데, 평점이 낮은 리뷰에 비해 평점이 높은 리뷰에서 ‘~니다’ 나 ‘~요’의 어미가 붙은 단어가 주로 나타난다. 이를 통해 비교적 공손하고 정제된 말투를 이용하여 영화 및 제작진에 대한 긍정적인 태도를 보여주고 있음을 알 수 있다. 일부 낮은 평점에서도 ‘~요’, ‘~니다’ 등의 어미가 붙은 키워드가 나타나기는 했지지만, 이와 함께 사용되는 어휘를 보면 ‘보지+마세요’, ‘노잼+입니다’ 등 부정적인 감정을 담고 있는 것을 알 수 있었다. 또한 이러한 말은 영화나 제작진을 비판하고 리뷰를 보는 다른 네티즌들에게 의견을 전달하는 두 가지 목적이 있음을 보여준다.

**2. Excel 피벗 테이블 분석**

가. 불한당: 나쁜 놈들의 세상(2016)

필터 기능을 이용하여 Noun을 체크한 뒤, 명사를 통해 주요 키워드를 분석한 결과는 다음과 같다. 평점 1-3점에서는 스토리(1.576%), 생각(1.346%), 수준(0.987%), 별로(0.926%), 쓰레기(0.777%), 재미(0.654%), 돈(0.611%), 무간도(0.603%), 짬뽕(0.524%), 인성(0.521%), 여자(0.517%), 일베(0.460%) 등의 키워드가 타 구간에 비해 높은 빈도를 보였다. 이중 ‘여자’와 ‘일베’의 경우에는 영화와는 전혀 관련이 없는 감독의 발언과 관련한 키워드이다. ‘스토리’, ‘생각’, ‘재미’ 등 스캔들과 관련성이 적은 키워드는 4-7점 평점 구간과 비슷한 수치를 보인 반면, ‘수준’, ‘쓰레기’, ‘돈’, ‘인성’, ‘여자’ 등의 키워드는 다소 큰 차이를 보였다. 특히 ‘일베’의 경우엔 평점 4-7점에서의 수치가 매우 낮아 같은 표시 형식에서 수치를 확인할 수 없었다.

평점 4-7점에서는 연기(3.911%), 배우(9.161%), 설경구(2.901%), 신세계(2.283%), 그냥(1.503%), 느낌(1.373%), 내용(1.356%), 느와르(1.156%), 연기력(1.096%), 평점(1.032%), 연출(0.892%), 반전(0.646%), 결말(0.621%), 장면(0.611%), 작품(0.607%) 등의 키워드가 타 평점에 비해 높게 나타났다. Word Cloud에서 분석한 결과와 같이 주로 영화 내적인 요소인 연기, 스토리, 장르 등과 관련한 키워드가 다수를 이룬다. 다만 여기서 ‘그냥’이라는 키워드는 1-3점과 4-7점 평점에서 모두 높게 나타났는데 조합되는 어휘의 차이를 볼 수 있다. 1-3점에서는 ‘그냥’이라는 키워드가 ‘쓰레기’, ‘노잼’, ‘다른거 봐라’, ‘영화가 별로’ 등의 말과 붙어서 사용되는 반면, 4-7점에서는 ‘그냥+그래요’, ‘그냥+저냥’과 같이 사용되고 있었다.

평점 8-10점에서는 정말(0.766%), 최고(0.650%), 인생(0.327%), 홍어(0.309%) 등이 타 평점에 비해 높게 체크되었으며 느와르(1.055%), 평점(1.053%), 연출(0.851%) 등의 키워드는 4-7점과 비슷한 수치를 보였다. 여기서 ‘홍어’라는 키워드는 변성현 감독이 ‘홍어’를 비하 목적이 아닌 순수한 의도에서 사용하였다는 의미로 스캔들이 루머임을 주장하며 감독을 옹호하고자 사용되었다. 여기서 ‘자체’라는 단어는 8-10점에서 36번째로 자주 나타난 명사이지만, 수치는 1-3점에서 0.513%, 4-7점에서 0.835%, 8-10점에서 0.418%로 가장 낮게 나왔다. 주로 ‘영화 자체만 놓고 평가합시다’라는 의미에서 자주 사용되었다.

필터 기능을 이용하여 Adjective, Adverb, Alpha, Exclamation, Foreign, Hashtag, Korean Particle, Modifier, Noun, Number, Verb를 체크하고, 실질형태소 중심으로 키워드를 분석했을 때는 영화 ‘불한당’과 비교되는 누아르 장르의 타 영화에 대한 언급 빈도 역시 파악할 수 있었다. 앞서 평점이 낮은 구간과 높은 구간에서 이러한 비교하는 문장이 자주 언급되며 그 내용이 반대를 이룬다고 언급하였는데, 상대빈도 수치에서는 대부분 4-7점에서 언급이 많은 것을 볼 수 있었다. (키워드 ‘신세계’ : 1-3점에서 1.802% / 4-7점에서 2.283% / 8-10점에서 0.563%, 키워드 ‘프리즌’ : 1-3점에서 0.292% / 4-7점에서 0.796% / 8-10점에서 0.150%, 키워드 ‘무간도’ : 1-3점에서 0.603% / 4-7점에서 0.403% / 8-10점에서 0.097%) 또한 ‘수준’ 같은 다소 중립적인 단어가 혹평 혹은 비판을 하는 데에 잘 사용된다는 점을 수치를 통해서도 알 수 있었다. (1-3점: 0.987% / 4-7점: 0.246% / 8-10점: 0.089%)

나. 지금은맞고그때는틀리다(2015), 밤의 해변에서 혼자(2016)

필터 기능을 이용하여 Noun을 체크한 뒤, 명사를 통해 주요 키워드를 분석한 결과는 다음과 같다. 평점 1-3점에서는 불륜(1.737%), 예술(1.465%), 여자(0.802%), 평점(0.595%), 포장(0.607%) 등의 키워드가 타 구간보다 높은 빈도로 사용되었다. ‘불륜’의 경우, 해당 영화가 직면한 스캔들의 가장 중점적인 키워드라는 점에서 아주 높은 빈도로 사용되었음을 알 수 있었다. 또한, ‘예술’은 ‘이걸 예술이라고 하는 거냐’, ‘평점’은 ‘평점은 왜 이렇게 높냐,’ ‘포장’은 ‘포장 잘한다’ 등의 문맥에서 주로 사용된 단어였다.

4-7점에서는 연기(2.475%), 생각(1.813%), 배우(1.321%), 이야기(1.133%), 현실(0.758%) 등의 키워드의 빈도가 높게 나타났다. 〈뮬란(2020)〉과 마찬가지로 Word Cloud에서 살펴본 것과 같이, 해당 평점 구간에서는 주로 연기나 배우, 스토리 등 영화 내적인 요소에 관한 키워드나 자신의 의견을 말하는 단어가 주를 이루고 있었다.

8-10점에서는 정재영(1.005%), 작품(0.728%), 역시(0.635%), 매력(0.547%), 최고(0.472%) 등의 키워드가 자주 등장했다. 이에 대해서는 긍정적인 어감을 가진 단어가 많이 쓰임을 알 수 있었다. ‘정재영’의 경우엔 주연 배우의 연기력을 칭찬하기 위해 사용한 경우가 대다수였다.

필터 기능을 이용하여 Adjective, Adverb, Alpha, Exclamation, Foreign, Hashtag, Korean Particle, Modifier, Noun, Number, Verb를 체크하고, 다른 실질형태소 중심으로 키워드를 분석했을 때는 감탄사 Exclamation과 한글 자모 Korean Particle에서 유의미한 분석 결과를 볼 수 있었다. 기본적으로 1-3점의 평점에서 어휴(0.052%), 아이고(0.019%), 엑(0.009%), 에잇(0.005%) 등의 부정적 이미지의 감탄사가 사용되었기 때문이다. 또한, 한글 자모 역시 주로 비웃음을 나타내는 ㅋ(0.299%), 아쉬움을 나타내는 ㅠ(0.102%), ㅠㅠ(0.199%), 비속어를 적은 ㅅㅂ(0.110%) 등이 1-3점의 구간에서 나타났다. 반대로 8-10점의 평점에서는 기분 좋음을 나타내는 ㅎㅎ(0.183%), ㅎㅎㅎㅎ(0.021%) 등이 주로 등장하여, 한글 자모를 통해서도 평점에 따른 리뷰 작성자의 태도 차이를 엿볼 수 있었다.

다. 뮬란(2020)

필터 기능을 이용하여 Noun을 체크한 뒤, 명사를 통해 주요 키워드를 분석한 결과는 다음과 같다. 평점 1-3점에서는 타 평점 구간과는 달리 중국(2.635%), 영혼(2.507%), 그냥(1.339%), 디워(1.159%), 무협(0.955%), 돈(0.572%), 별로(0.610%), 실망(0.544%) 등의 키워드의 빈도가 더 높게 나타났다. 이 중에 ‘디워’라는 키워드는 해당 구간에서만 나타났는데, 영화 ‘디워’ 수준으로 망한 영화라는 의미로써 사용된 것이다. ‘영혼’이란 키워드는 자본 및 중국에 영혼을 판 영화란 의미에서 사용되었다.

평점 4-7점에서는 원작(2.991%), 감상(2.138%), 액션(1.743%), 스토리(1.611%), 애니(1.597%), 생각(1.524%), 실사(1.256%), 느낌’(1.165%), 상미(0.923%), 노래(0.920%), 장면(0.869%), 캐릭터(0.859%), 배우(0.859%) 등의 키워드의 빈도가 타 평점보다 높게 나타났다. 앞서 Word Cloud에서 보았던 영화 내적 요소와 관련된 주요 키워드가 모두 평점 4-7점에서 가장 많이 나타남을 알 수 있다. 특이점은 원작이나 애니, 실사 등의 키워드가 자주 나타나면서 원작인 애니메이션과 실사영화인 이 작품을 비교하는 리뷰가 많다는 것이다. 여기서 ‘상미’라는 키워드는 영상미를 의미하는 것이며, 영상미가 괜찮다 혹은 별로다 등 관련한 내용은 다양하게 분포하고 있다.

평점 8-10점에서는 은근(1.224%), 감동(1.214%), 평점(0.732) 등의 키워드가 높게 나타났다. 특이점은 ‘논란’이란 키워드가 유일하게 높은 평점에서 주요 키워드 50위 안에 포함되었지만 언급된 빈도는 가장 적다. (1-3점: 0.426, 4-7점: 0.465 / 8-10점: 0.369)

라. 어미 분석

필터 기능을 이용하여 Adjective, Verb를 체크한 뒤, 각 평점 구간별로 상위 7개의 문장의 끝에 쓰인 것으로 추측되는 서술어를 추출한 결과는 다음과 같다.

|  |  |
| --- | --- |
| 불한당: 나쁜 놈들의 세상(2016) | |
| 평점 구간 | 서술어 (수치 단위 : %) |
| 1-3 | 없다 (0.196) > 하네(0.282) > 준다 (0.237) > 입니다 (0.225) > 해라 (0.222) > 아님(0.197%) > 같네요 (0.194) |
| 4-7 | 입니다(0.498) > 없다(0.485) > 했다 (0.423) > 같다 (0.423) > 아쉽다 (0.389) > 합니다 (0.274) > 봤습니다 (0.251) |
| 8-10 | 입니다(0.554) > 아깝다(0.352) > 봤습니다 (0.320) > 합니다 (0.297) > 봤어요(0.242) > 좋았어요(0.222)> 없다(0.205) |

|  |  |
| --- | --- |
| 지금은맞고그때는틀리다(2015), 밤의 해변에서 혼자(2016) | |
| 평점 구간 | 서술어 (수치 단위 : %) |
| 1-3 | 틀리다(0.402) > 역겹다(0.353) > 모르겠다(0.324) > 없다(0.293) > 된다(0.281) > 아깝다(0.226) > 하다(0.216) |
| 4-7 | 같다(0.889) > 입니다(0.529) > 없다(0.478) > 틀리다(0.462) > 했다 (0.444) > 합니다(0.408) > 모르겠다(0.359) |
| 8-10 | 틀리다(0.441) > 봤습니다(0.436) > 좋았다(0.431) > 좋다(0.405) > 입니다(0.358) > 있다(0.343) > 한다(0.337) |

|  |  |
| --- | --- |
| 뮬란(2020) | |
| 평점 구간 | 서술어 (수치 단위 : %) |
| 1-3 | 입니다(1.188) > 없음 (0.332) > 합니다 (0.320) > 사라질듯 (0.244) > 있다(0.232) > 없다 (0.232) > 아님 (0.228) |
| 4-7 | 입니다(1.188) > 합니다 (0.320) > 봤습니다(0.456) > 했다 (0,445)> 좋았어요 (0.350) > 보세요 (0.339) > 없음 (0.338) |
| 8-10 | 입니다(0.554) > 봤어요(0.515) > 봤습니다(0.422)> 재밌었어요 (0.302) > 재밌어요 (0.285) > 있네요 (0.230)> 좋았어요 (0.219) |

앞선 Word Cloud 분석에서 평점에 따라 리뷰에 나타나는 어미의 차이가 있다고 언급하였는데, 수치를 가지고 비교하면 완전히 양분되는 결과가 나타나지는 않음을 알 수 있다. 그러나 객관적인 수치를 통해 낮은 평점일수록 부정적인 어감의 단어와 함께 반말의 어미를 자주 사용하고, 높은 평점일수록 긍정적인 어감의 단어와 함께 존댓말의 어미를 잘 사용한다는 경향은 도출이 가능하다.

세 가지의 분석대상 중 데이터양이 가장 많았던 불한당의 경우에서 특히 이러한 경향이 잘 나타난다. 높은 평점에서 2번째의 '아깝다(0.352%)'를 제외하면 모두 '-니다' 혹은 '-요'와 같이 경어체의 서술어가 상위권에 속했다. 해당 데이터와 일부 원문 분석을 종합하여 제작진과 영화에 대한 호평을 남기거나 다른 네티즌에게 영화를 추천하여 즐거움을 공유하고자 하는 의도가 있음을 알 수 있었다. 반면 낮은 평점에서는 4번째의 '입니다(0.225%)와 7번째 '같네요(0.194%)'를 제외하면 모두 존대어보다는 평이하고 단정적인 말투의 서술어가 상위권을 차지하였다. 원문 분석과 데이터를 조합해 비교한 결과, 주로 영화와 제작진을 비판하고자 하는 의도가 강하기 때문에 이런 데이터가 나온 것으로 확인하였다.

다만 '입니다'의 경우엔 다수의 평점 구간에서 상위권을 차지하였는데, 이는 '-이다'보다는 경어체이지만 다소 단정 짓는 경향을 보이는 어투이다. 여기에는 각 평점에 따라 '수작/명작+입니다' 혹은 '노잼/쓰레기+입니다' 등 완전히 다른 단어와 조합되어 문맥에 따라 긍정/부정/중립적인 의미를 전달함을 알 수 있었다.

**Ⅳ. 조사의 결론 및 시사점**

조사를 진행하기 이전에는 배우나 제작진의 스캔들에 따라 영화 평점별 리뷰에 사용된 단어에 대해 단순히 긍정적이거나 부정적인 어감의 단어 이용 차이 정도만을 예상하였으나, 실제 데이터 분석 이후에는 다음과 같은 조금 더 세밀한 결론을 내릴 수 있었다.

첫째, 같은 단어라 할지라도 평점별로 다른 문맥에서 사용되곤 한다는 것이다. 이것의 예시로는 ‘수준’과 같이 중립적 이미지를 가진 단어가 낮은 평점의 리뷰에서는 ‘말도 안 되는 수준’과 같이 사용되었지만, 높은 평점의 리뷰에서는 ‘수준 높은 연기’ 등으로 사용된 것을 들 수 있다.

둘째, 평점이 낮은 리뷰일수록 스캔들과 관련한 키워드가 해당 논란에 대한 중점적 단어만이 아닌 그와 관련성이 비교적 느슨한 단어도 함께 등장한다는 것이다. 또 낮은 평점의 리뷰에서는 영화 자체보다는 해당 스캔들과 관련한 단어가 많이 쓰인 만큼, 영화 외적인 요소와 관련한 단어를 많이 볼 수 있었다. 이와는 반대로 평점이 높은 리뷰에서는 ‘논란’ 혹은 해당 논란의 중심이 되는 키워드만 주로 사용되었고, 주로 배우의 연기나 영화의 내용과 배경 음악 등 영화 내적인 요소에 집중하였다.

셋째, 평점에 따라 어미 사용에 따른 어투의 공손함에서 차이가 있었다는 점이다. 이는 다른 품사의 분석 결과에 비해 뚜렷하게 드러나지는 않지만, 낮은 평점의 리뷰와 비교했을 때, 높은 평점의 리뷰에서 비교적 ‘-다’보다 ‘-니다’가 상대적으로 많이 등장한 것을 통해 알 수 있는 부분이다.

결과적으로 이 프로젝트를 진행하고 보니 본래의 조사 목적과 부합하는 분석 결과가 나오게 되어 전반적으로는 상당히 만족스러운 결과물이 도출되었다고 말할 수 있을 것이다. 하지만 조사의 과정 중에 아쉬운 점도 존재했는데, 이는 대부분 사람이 직접 해당 영화 평점의 리뷰를 읽고 언어 분석을 수행하는 것이 아니라는 점에서 비롯되었다. 예를 들면, 신조어나 오타가 교정되지 않은 단어, 의도한 띄어쓰기가 들어간 단어 등 일반적으로 사전에 등재된 단어가 아닌 경우 형태소 분석기에서 명확하게 않는다는 것과 같은 것이 바로 그것이다. 물론, 이렇게 많은 양의 데이터를 사람이 직접 다루는 것에 비하면 이처럼 형태소 분석기를 사용해 언어 분석을 하는 것이 비교할 수 없을 만큼 편리하고 빠른 방식이긴 하지만, 그 정확도가 살짝 떨어진다는 점은 분명 아쉬웠다. 그리고 이러한 문제점을 해결하기 위해서, 형태소 분석기 역시도 더 많은 데이터를 수집하여 끊임없이 신조어나 자주 발생하는 오타 등을 자체적으로 판단하고 수정할 수 있도록 하여야 할 것이다.