Selecting significant respondents from large audience datasets: The case of the World Hobbit Project

Alberto Trobia,
University of Palermo, Italy

Abstract:
International projects, online questionnaires, or data mining techniques now allow audience researchers to gather very large and complex datasets. But whilst data collection capacity is hugely growing, qualitative analysis, conversely, becomes increasingly difficult to conduct. In this paper, I suggest a strategy that might allow the researcher to manage this complexity. The World Hobbit Project dataset (36,109 cases), including answers to both closed and open-ended questions, was used for this purpose. The strategy proposed here is based on between-methods sequential triangulation, and tries to combine statistical techniques (k-means clustering) with textual analysis. K-means clustering permitted to reduce data to a small number of ideal-typical respondents: the 'average spectator', the 'die-hard fan', the 'cultured spectator', the 'alternative spectator'. These clusters are the outcome of a cross-validation process. Textual responses corresponding to each cluster, in fact, were also analyzed using a quali-qualitative approach, in order to both refine the clusters and identify meaningful discourse patterns. The methodological mix proposed can be used with confidence, since it proved to yield reliable results.

Keywords: international audiences, audience research methods, mixed methods, large datasets, cluster analysis, content analysis, specificity analysis

1. Introduction. Changing audiences and new research methods

Media and communication technologies, today, are characterized by huge developments, emerging forms, and intricate topologies. This has determined deep changes in consuming practices, and has raised new challenges, both theoretical and methodological, in the field of audience research (Patriarche et al., 2014). Three adjectives seem to summarize the main

directions in this wide field of studies: 'holistic', 'relational', and 'participatory' (ibid.: 11). Practices have changed as a consequence of the growing interactivity new technologies allow and the convergence (Jenkins, 2006) of different media into single devices or platforms. The methodological challenge, on the other hand, has been managed through the adoption of multimethod research designs, in order to try to recognize, classify and interpret people's practices and meanings (Jensen, 2012; Schrøder et al., 2003), and through the introduction of many types of Web applications, tools and facilities, such as online questionnaires, tracking systems, text mining software, and others (Hastall and Sukalla, 2014; Procter et al., 2015).

Various multimethod research designs have been proposed. Triangulation, originally conceptualized by Denzin (1970), that is, 'the use of more than one method or source of data in the study of a social phenomenon so that findings may be cross-checked' (Bryman, 2016: 697) is one of the most popular. According to Cohen and Manion (2000: 254), triangulation is an 'attempt to map out, or explain more fully, the richness and complexity of human behavior by studying it from more than one standpoint'. Denzin's original proposal was to combine different data sources, investigators, theories and methodologies.

Triangulation can be achieved using the same dataset, when – for instance – the latter contains both quantitative and qualitative data (Barker et al., 2008; Trobia 2008a). Analyzing the same data using different methods is what Denzin (1970) calls a 'between-methods' or 'across-method' approach. In this case, the researcher aims at integrating different techniques, in order to compensate for their weaknesses and take advantage of their strengths. 'The challenge is to find the right combination of methods and then bring together the findings, weighing their cumulative evidence' (Findahl et al., 2014: 55).

Few studies, however, have really used triangulation in audience research, 'even if that is a general recommendation mentioned in the literature' (ibid.: 56). Triangulation, in fact, can be troublesome, and some critiques have been raised against it (e.g. Blaikie, 1991). According to Fielding and Fielding (1986: 33): 'One may get a fuller picture, but not a more "objective" one. Similarly, (…) combining [different methods] can add range and depth, but not accuracy'. Nonetheless, pragmatic and less ideological approaches have been proliferating in the last years (see Bergman, 2008; Teddlie and Tashakkori, 2009). Sequential designs, in particular, have gained a good consensus, even among the major critics of triangulation (Blaikie, 1991). Barker's loop, for example, that is 'a form of triangulation *within* the same research design (…) in which the findings from both the qualitative and quantitative approaches are *mutually informing*, and show that triangulation is achievable even at a cheap price' (Barker et al., 2008: 230, emphasis in the original), has been applied in international audience research.

Online questionnaires have proven to be quite useful in contemporary audience research too. The advantages of online questionnaires are manifold, including the possibility of getting larger sample sizes (thus reducing hasty conclusions, caused by a lack of statistical power) and carrying out intercultural and transnational comparisons; they allow an increased flexibility, a higher level of anonymity for respondents and a lower level of

obtrusiveness in the interviewing process (thus reducing social desirability bias), and a less expensive and more efficient data collection. On the other hand, sample bias, dropouts rates, generalizability (external validity), limited control over respondents, reduced nonverbal clues, or fake answers can be an issue (Hastall and Sukalla, 2014; Gaiser and Schreiner, 2011; Wright, 2005).

Wide international projects, online questionnaires, or data-mining techniques now allow social science researchers to gather very large and complex datasets. There is a significant difference, however, between large datasets coming from online questionnaires and large datasets collected through data-mining techniques, since the former are 'made data' (meaning that an informed data construction process has been carried out before data collection), while the latter are 'found data' (Jensen, 2012). The popular expression 'big data' mainly refers to found data.

Large datasets may combine quantitative and qualitative answers. Since a well-known weakness of the survey is that the outcomes are strongly dependent on the questions that the researcher has chosen to ask, and on the ability, truthfulness, and willingness of the respondent to answer them (Tourangeau et al., 2000; Schaeffer and Presser, 2003), the presence of open-ended questions in the questionnaire, instead, allow the researcher to gather more interesting and reliable materials, with more possibilities for exploration. This is the strategy the researchers employed in *The Lord of the Rings* project (Barker et al., 2008) and in the *World Hobbit Project* (see Martin Barker's 'Introduction', in this Section).

Complex datasets can be rather easily analyzed using statistical methods for quantitative data, but they require a much stronger effort when dealing with qualitative data, coming – for example – from responses to open-ended questions. Although data collection capacity is hugely growing, qualitative analysis, conversely, becomes increasingly difficult to conduct for many reasons. QDA (Qualitative Data Analysis) software is often used in order to analyze large textual material. Some authors, however, have criticized its efficacy in multi-method designs (e.g. White et al., 2012). In any case, this is hardly useful when handling open-ended questions, since texts are too short and fragmented for the software to be effective (Seale, 2000). Besides, when datasets are multilingual, it is practically impossible to translate all of the answers. Automatic text analysis can help, but an in-depth look at the actual answers from respondents would still be important for the researcher (Franzosi, 2012). An often neglected problem, moreover, is that large datasets are expected to contain some statistical noise and/or spam, that has to be filtered out. As a consequence, a convenient way to reduce data complexity is strongly required.

2. Building a typology of spectators through K-means clustering

Cluster analysis is a common statistical technique for categorizing individuals and cases. This latter is actually a 'family' of techniques, such as hierarchical models (agglomerative or divisive), iterative partitioning (e.g. k-means clustering), local density or nearest neighbor analysis, and neural nets (e.g. SOM/Kohonen) (Everitt et al., 2001; Garson, 2014).

Cluster analysis is a data reduction technique. It can assist the researcher, when she or he wants to build a typology of objects. This is implemented by 'seeking to identify a set of groups which both minimize within-group variation and maximize between-group variation' (Garson, 2014: loc. 261). In other words, the purpose of cluster analysis is to detect groups of respondents who show similarity to each other when compared to respondents who belong to other groups.

This is the case when trying to segment the audience. Each group, in fact, may be thought of as corresponding to an *idealtypus*, as defined by Weber (1904). As Bauman (2007: 23-24) argues: 'Ideal types are not descriptions of social reality but the tools of its analysis and – hopefully – its comprehension. Their purpose is to force our picture of the society we inhabit to "make sense"; to achieve that purpose, they deliberately postulate more homogeneity, consistency and logic in the empirical social world than daily experience makes visible and allow to grasp'. The theoretical outcomes and the associated advantages of this kind of research are clear. An ideal type is a model of reality. It is able to capture the general meaning of a contingent social structure, omitting trivial details, but orientating the scientific labor.

Reasoning in terms of ideal types also allows the researcher to deal with an emerging and crucial issue in recent audience research, that of 'imagining audiences' (Jensen, 2014). Researchers imagine both the audiences that they seek to measure and interpret, find or make, and the kinds of audiences they might promote through their studies (ibid.: 235). Filmmakers and many other practitioners also need to anticipate their audiences as part of their professional and organizational routines (Ettema and Whitney, 1994). Imagined audiences can be considered quite real. Researchers, in fact, 'make, measure and interpret empirical audiences, in order to conceive of them theoretically and, in some instances, politically and normatively' (Jensen, 2014: 236). This is definitely one of the questions cluster analysis tries to answer.

In order to identify the clusters, some sort of similarity or dissimilarity criterion must be previously defined. Generally, the first step is the construction of a similarity or distance matrix. In addition, a set of variables has to be chosen by the researcher that are responsible for within-group composition and between-group variability. Such variables can't be measured on a categorical level: dichotomous or ordinal variables are at least required.

Whatever the technique chosen to form the clusters, the assessment of the emerging classification must meet five basic criteria (Garson, 2014: loc. 334): each cluster should be easy to understand (meaningfulness); clusters should be clearly separated from each other (separation); all clusters should have a sufficient number of cases (size); the cluster membership should correlate with other variables known from theory or prior research (criterion validity); finally, 'using one set of data to develop the clustering model and then using another set to validate it is recommended' (ibid: loc. 354).

Cluster analysis yields very robust results, and does not require a probability sample, because it classifies objects (individuals, in the study of audiences) in a typology, regardless

of their number. Of course, a certain typological variability must be guaranteed within the sample. This is generally the case when analyzing large datasets (Kraut et al., 2004). Classifications emerging form large datasets using cluster analysis may thus be considered quite reliable, and could effectively counterbalance the lack of a probability sample. For example, analyzing two different subsets of *The Lord of the Rings* dataset, that counted 24,739 cases, the German and Italian teams obtained overlapping results: four clusters of similar spectators (Barker et al., 2008: 229). This kind of information may be used to generate better-designed purposive or probability samples in subsequent research (see Jerslev et al., in this Section, for other sampling issues).

Dealing with variables and not with words, cluster analysis can also be used in those researches where individuals speaking different languages are interviewed, as in the case of transnational audiences, providing that the codebook is consistent across the countries involved. Of course, several difficulties have to be faced when translating all the questions in different languages (Egan and Barker, 2006). Incidentally, the issue of ethnic minorities is gaining an increasing interest and suggesting new methodological strategies in the study of contemporary audiences (e.g. Cola and Brusa, 2014).

There have been several attempts to segment the audience in the cultural field, including cinema, based on different criteria: demographic and behavioral variables; frequency of attendance; attitude or interest towards the arts; values and lifestyles; activities, consumptions and benefits sought. Studying young moviegoers, for instance, Cuadrado and Frasquet (1999) found three differentiated and consistent groups of attendees: the *social*, the *apathetic* and the *cinema-buff*; while, relying on a list of nine basic values, Marchand and Khallaayoune (2010) discussed three interesting segments of audience: the *achievers*, the *hedonists*, and the *emphatics*. Moreover, analyzing *The Lord of the Rings* project dataset, Trobia (2008a) recognized four basic types of spectator: the *enthusiastic fan*, the *critic reader*, the *disappointed fan*, and the *mass spectator*.

Not all clustering techniques are suitable for large datasets, however. K-means clustering is much less computer-intensive and is therefore often preferred when datasets are large (e.g. when the sample size is greater than 1,000 cases) (Garson, 2014: loc. 369). This clustering technique requires the researcher to specify the number of clusters in advance. This number may be suggested by previous knowledge on the subject being studied, by trial and error, or by specific statistical criteria (the gain in between-group variance, the Pseudo F statistic, the Overall R-Squared, the Cubic Clustering Criterion). Nevertheless, these latter are still under development, and their efficacy is controversial (ibid: loc. 1098).

K-means clustering is an iterative technique; that is, cases are shifted from one group to another, confronting Euclidean distances to the mean of the cluster, in order to minimize within-cluster variance and maximize between-cluster variance.

Six ordinal or dichotomous variables from *The Hobbit* dataset (36,109 respondents) were considered for the construction of the clusters:

1) What did you think of the Hobbit films overall?

2) How important was it for you to follow stories and debates around the films?

3) What did you think of *The Lord of the Rings* films overall?

4) Have you read *The Hobbit*?

5) Do you prefer to see films like *The Hobbit* in a cinema?

6) What level of education have you reached?

Other variables were also considered for inclusion, such as age, gender, or other preferred viewing options; these variables, however, didn't prove to yield superior or more interesting results (see Martin Barker and Ernest Mathijs' 'Introduction', in this Section, on the discriminating role of education).

Since all criterion variables were not measured on the same scale, they were normalized, in order to avoid misleading distance measures and to guarantee comparable outputs. Normalization (or standardization) is recommended, because K-means clustering is highly sensitive to unequal variances (ibid: 2806 and 2821) as in our dataset.

Figure 1 shows the plot of means for each cluster. This plot is needed, in order to infer the meaning of each cluster, thus assigning it a label.

The first cluster is marked by a circle in the Figure. The corresponding profile is characterized by a high rating given to the *Hobbit* films overall, a medium-low interest in following stories and debates about the films, a rather high rating given to *The Lord of the Rings* films. Members of this cluster did not read *The Hobbit*, they prefer to see fantasy films in a cinema, they have a medium level of education. This cluster seems to gather the *average spectators*.

**Figure 1:** K-means clustering: Plot of means for each cluster
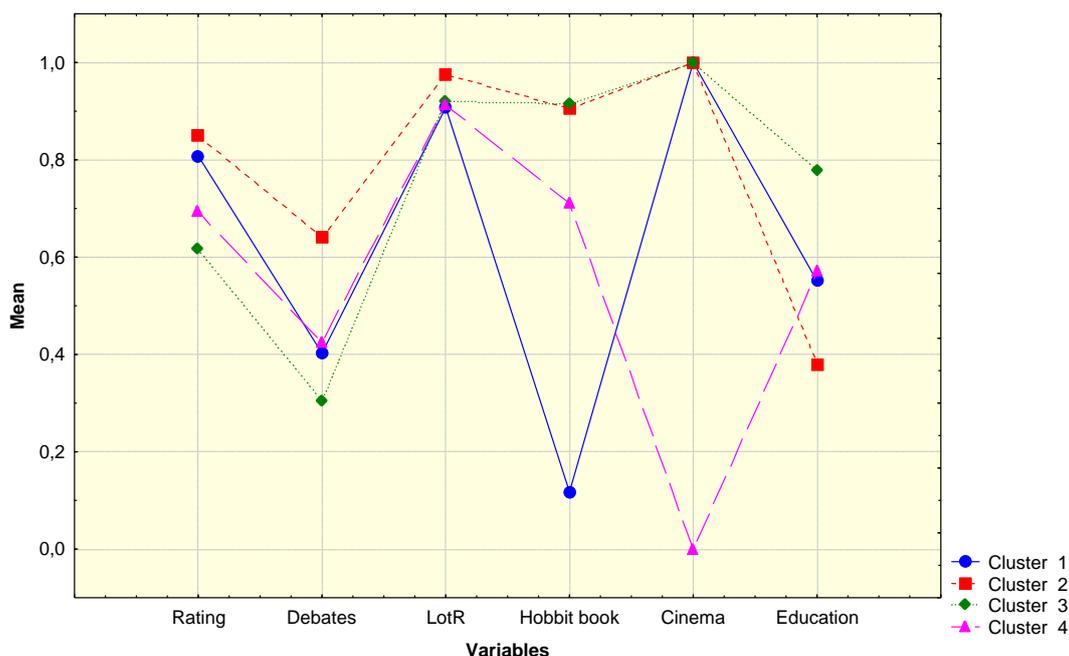
The second cluster is represented by a square. Its profile shows the highest means for each criterion variable, except for education. This is the cluster reasonably containing the *enthusiastic fans*. Interestingly, the members of this cluster have the lowest level of education.

The third cluster is marked by a diamond. This group of spectators is characterized by the lowest rating given to the Hobbit films and the lowest interest in debates about them. Respondents, however, have the highest levels of education, they have read *The Hobbit*, prefer to see films in a cinema, and have seen *The Lord of the Rings* films a bit less than the enthusiastic fans. These respondents could be defined as *cultured spectators*.

Finally, the fourth cluster is represented by a triangle in **Figure 1**. The rating given to the Hobbit films, here, is not high, and the debates are not considered interesting. The level of education is medium. The members of this group have seen *The Lord of the Rings* films might have read the book, but are not used to going to the cinema; this latter is the most important feature of the cluster. Hence, we might call this the group of *non-spectators*.

Figure 2 shows the size of each cluster with respect to the whole sample (36,109 respondents). As we can see, cultured spectators prevail (29.7%), followed by the enthusiastic fan ideal type (28%), while the smallest cluster is represented by the average spectators (19.7%). This picture is quite plausible, since the respondents to the questionnaire were mainly self-selected and interested, for some reason, in the *Hobbit* phenomenon (see Jerslev et al., in this Section). Cluster sizes, in any case, are rather balanced, possibly as a consequence of the huge number of respondents; that guaranteed a good *typological representativeness* of the sample. This latter is achieved, 'when the researcher has sufficient members of all the main categories of interest to be able to describe with confidence their patterned similarities and differences' (Trobia, 2008b: 783). In other words, typological representativeness alludes to having as many types of respondent as possible and as many forms of argumentation as possible (Barker at al., 2008: 222).

Figure 2: Percentage of members for each cluster (N=36,109)
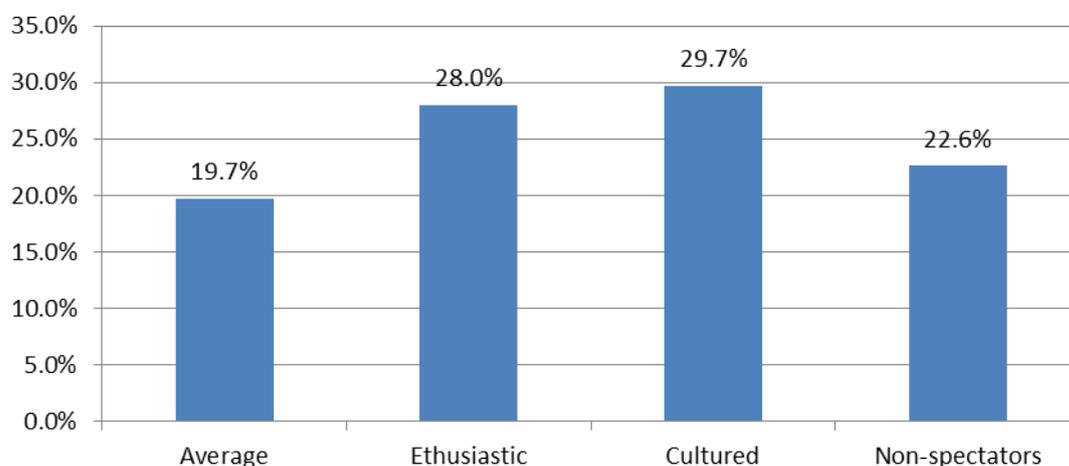
The clusters emerged from the analysis do not appear correlated with sex, with the exception of non-spectators; in this case, female sex prevails (see **Figure 3**). A slight prevalence of female sex is also present in the third cluster (cultured spectators).

Figure 3: Clusters distribution by sex
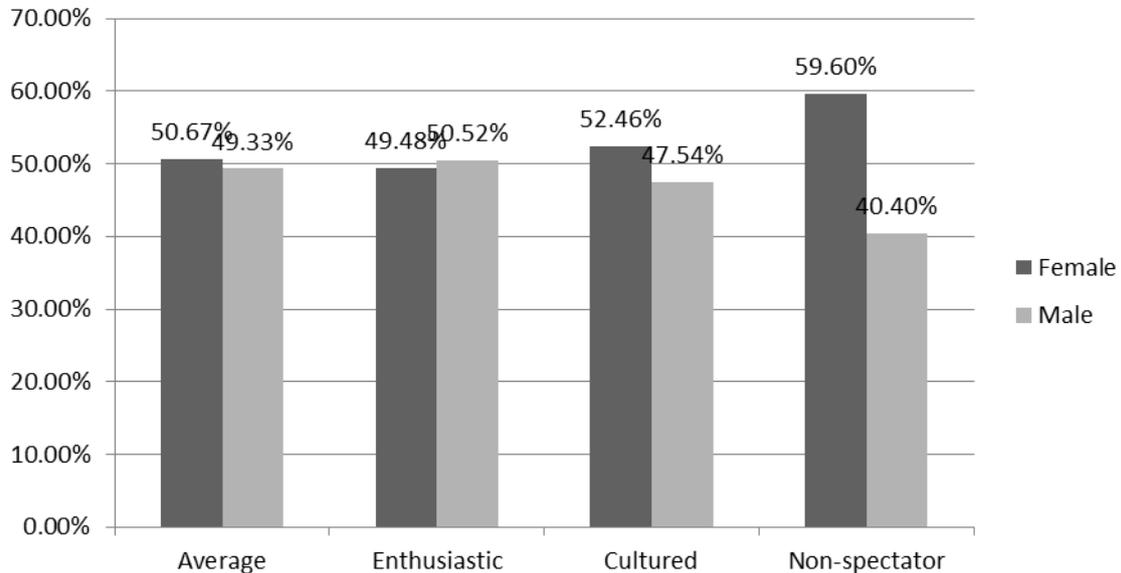
The *Hobbit* research project involved many countries and nationalities. **Table 1** shows the cluster distribution by nationality. Only the first ten nationalities in the datasets are reported. The highest percentages in each row are marked in bold. The cultured spectator is predominant among Russian, British, American, Finnish and Danish respondents; while the enthusiastic fan obtains the highest percentages among Germans, Swedish and Brazilians.

Table 1: Cluster distribution by nationality (first ten nationalities)

Nationality	Average	Enthusiastic	Cultured	Non-spectator	% of sample
Czech	13.99%	31.40%	21.13%	**33.47%**	10.7%
German	14.18%	**43.62%**	26.44%	15.76%	10.2%
Turkish	**37.36%**	22.61%	16.95%	23.08%	8.2%
Swedish	20.25%	**43.25%**	23.38%	13.11%	5.7%
Russian	16.67%	26.12%	**32.60%**	24.61%	5.3%
British	10.43%	18.72%	**49.89%**	20.96%	4.9%
American	5.81%	28.76%	**38.09%**	27.35%	4.7%
Finnish	15.37%	24.78%	**43.68%**	16.17%	4.5%
Brazilian	21.42%	**28.45%**	27.15%	22.98%	3.4%
Danish	27.46%	17.80%	**40.05%**	14.69%	3.3%

3. The *Hobbit* audience: from ideal types to forms of discourse

Having reduced data complexity quantitatively, through cluster analysis, an in-depth qualitative exploration of the groups obtained can now be carried out, given the availability of textual data coming from the answers to the open-ended questions.

A criterion, however, must be chosen, in order to select meaningful cases in large datasets, since the latter usually contain a lot of 'noise' or 'spam'. I suggest that we take into account the distance of each case to the centroid of the cluster to which it belongs. This distance is computed during k-means clustering and can be added to the dataset as a new variable in most statistical packages. The cluster centroid is simply the middle of a cluster, its conceptual core. The term 'centroid', used instead of 'center', alludes to the multiple dimensions implied in its calculation, since it takes into account several criterion variables (Biorcio, 1993: 57). The cluster centroids correspond with the lines drawn in **Figure 1**. The distance to the cluster centroid shows how typical is the case within the cluster to which it is assigned: the more a case resembles the ideal type, the less is such distance; the more a case is atypical, the more is the distance.

In **Table 2**, some descriptive statistics regarding cluster distances are shown. We can see, for instance, that cultured spectators are generally less distant to the ideal type than other types of spectator (see mean values in **Table 2**); while non-spectators are on average more distant. While, if we pass to consider heterogeneity, the enthusiastic fan group show the lowest value (see standard deviation values in **Table 2**).

Table 2: Distances to cluster centroids: descriptive statistics

	Average	Enthusiastic	Cultured	Non-Spectator
Mean	0.2	0.17	0.16	0.25
SD	0.073	0.058	0.065	0.08
Min	0.07	0.09	0.07	0.07
Max	0.53	0.50	0.53	0.60

In order to select the most meaningful cases within each cluster, the minimum distance to the centroid must then be considered. Following this criterion, the first five cases belonging to each cluster, having the shorter distance to the centroid, were extracted from the dataset, and their qualitative responses reported below together with some personal information. Only English speaking respondents were considered, regardless of nationality.

The analysis of such texts has at least two important purposes: first, it provides a basic form of 'between-methods' triangulation (quantitative results, in fact, can be interpreted using qualitative information from open-ended questions); second, the analysis of texts can pave the way to the crucial analysis of discourse patterns and meanings. In addition, the method discussed hereunder can help optimize the choice of the responses in

other languages that might be translated for further analysis, thus avoiding a long and dispersive process of trial and error.

3.1. *The average spectator*

The average spectator's qualitative answers are characterized by a high number of blank cells. As a consequence, some cases were replaced by others containing textual data, until the total of five respondents per cluster was attained. Actually, two more cases were added, in this case, in order to get a more detailed representation of the opinions expressed. The average spectator, however, did not generally talk at length about his/her experience of the film. The following quotations summarize their comments about *The Hobbit* films.

> I had an amazing time watching the movies, but I bet I could've enjoyed better if I had read the books as well. I was expecting more of the last movie, but I can't really say why. Maybe it's because there wasn't enough plot as there were battles scenes. I think it's a great story for the children. It teaches them to overcome fears, to deal with differences, to respect the other, to know when to listen or to speak, to be part of a group. Also, I love fantasy worlds, where our imaginations take us to different realities (#4461, distance: 0.08, Brazilian, 16-25 years, female, vocational qualification).

> Fantastic scenery and an epic storyline. But I do think it should have been two movie instead of three. It was a little too slow at times. The elf and dwarf love story was pointless in my opinion. Smaug had a surprisingly short amount of screen time (…). I think my opinions about the *Hobbit* are pretty general. I think most people who are not hard-core Tolkien fans were a little bit disappointed in paying for 3 movies. (…) I like stories that use a lot of imagination. I love when a story can create a vivid and wonderful world around it (#12606, distance: 0.08, Finnish, 16-25 years, male, vocational qualification).

> I had a double feeling for the *Hobbit* movie. At one side because I would already know the end, at the other side because it was a honor to enter Middle-earth again. *The Fellowship* and *Return of the King* are timeless and classical movies that has changed my own life and my family's lives. Every time we put these movies to see we are drawn in to a magical wonderland where the music, landscapes appeal to the eye and ear. In regards to the *Hobbit* movie I would say I enjoyed it but I feared it would never get at the same level of *Lord of the Rings*. So, I had to adjust myself and lower my expectations. I loved part 1 yet it felt to lengthy, I love part 2 but felt it was also stretched and for 3 I did have great expectancy but tragically it didn't meet my expectation. (…) The third installment suffered due to lack of GCI quality (…) Also, the cutting was at times very rushed, I couldn't enjoy certain scenes it went too

quick or too chaotic in sequence. (…) It had some great scenes but some cancelled their greatness each time. (…) If I could give Peter Jackson an advice I would say to him to balance his close up shots with the shots down on the larger screen. The ending of *The Hobbit* was also not at all adding to anything, most people left cinema with oh well that was it! (…) I believe they could have better have made it into 2 parts. (…) If they shot a new movie of Middle-earth even with a crap story I would still buy it and go see it. (…) Evil can then only be won by sacrifice, love, determination. So, Love, Friendship, Community, Being One, Bravery, Heroic, etc. (#31557, distance: 0.08, Dutch, 26-35 years, male, vocational qualification).

Found it a bit more humorous than *LOTR* and easier to watch also, more going on to keep your attention, where as in *LOTR* I generally miss out two towers as I just get bored with it. Not sure it took 3 movies to tell the story (#29196, distance: 0.08, British, 26-35 years, female, vocational qualification).

The film is awesome. Helps to broaden one's imagination. I didn't notice [any reason of disappointment]. For me, it is all about being inspired to achieve whatever set goals you have. I love to use my imaginations a lot so I think the film suits people like me (#1406, distance: 0.09, Nigerian, 26-35 years, male, vocational qualification).

Mostly visually pleasing with a few moments that felt a like they haven't been polished less than the rest of the movie. Some moments in the story felt like they were here just to add length to the final product. The rhythm of the movie is a bit weird since there are moments where you are immersed a lot and others during which I felt completely apart. (…) The overkill/overuse of CGI and the lengthening of a story to create more shots with less interesting coverage. I know for a fact that there's a lot of people that share my opinion (#31540, distance: 0.09, Russian, 16-25 years, male, vocational qualification).

It was a frivolous adventure, lacking the real grit and seriousness that *LOTR* had in abundance. I always enjoy movies that are about a journey, where the start is much different than the end. It was good, but not great. CGI was seamless. And some of the action scenes were a little soulless. (…). Another thing that bothered me was towards the end of the film, Thranduil had an unreasonable change of heart. He went from having utter disdain for dwarves and interracial relationships. Then suddenly, he had changed his mind. People do not change their minds about their racism easily. What sparked that change? (…) I suppose the depictions of orcs are heavily one sided. We only ever see it from the elves/humans point of view. They're always the bad guys. But, there is no good

or bad - those are only perceptions. So yes, the entire movie itself is skewed against the orcs. It raises the issue of incompetent governance, with the Master of Laketown being more interesting in money, power and influence than the actual welfare of his people. It also raises the issue of refugees. The dwarves of Erebor became refugees after the dragon Smaug took the mountain, the people of lake town practically became refugees (…). I have no participated in recent debates and I cannot remember any (#13028, distance: 0.09, Australian, 26-35 years, male, vocational qualification).

Double feelings and a general sense of disappointment seem to emerge, for various reasons, in the group of average spectators ('It was good, but not great', 'I can't really say why'; 'It didn't meet my expectation'). Both content issues and technical criticisms were addressed to the Hobbit movies.

The most frequent reason for discontent is perhaps the choice of dividing the story into three chapters instead of one or two ('I think most people who are not hard-core Tolkien fans were a little bit disappointed in paying for three movies'). These spectators have perceived some problems in the development of the plot or in the space given to certain characters ('There wasn't enough plot as there were battles scenes'; 'Some moments in the story felt like they were here just to add length to the final product'). Editing, rhythm and CGI (Computer-Generated Imagery) issues were also raised ('The cutting was at times very rushed', 'The rhythm of the movie is a bit weird'; 'The third installment suffered due to lack of GCI quality'). The successful films of *The Lord of the Rings* were sometimes used as a benchmark for the assessment of the spectators' personal experience ('It would never get at the same level of *Lord of the Rings*'; 'Found it a bit more humorous than *LOTR* and easier to watch also').

In many responses an emphasis on values ('Evil can then only be won by sacrifice, love, determination'; 'It's a great story for the children. It teaches them to overcome fears, to deal with differences, to respect the other, to know when to listen or to speak, to be part of a group') or on wider political issues ('It raises the issue of incompetent governance'; 'It also raises the issue of refugees') is added. Other spectators have deemed the idea of 'escape' as crucial to enjoy the films ('Our imaginations take us to different realities'; 'Helps to broaden one's imagination'; 'I always enjoy movies that are about a journey'). This seems to recall Herbert Blumer's pioneering research on spectatorship, which dates back to the early 1930s (Blumer, 1933), later reprised by Stacey (1994). In fact, 'his observations of strong emotional experiences, and identification as "losing oneself" have links to both previous and later scholarship on film (and television)' (Gripsrud, 1998: 206).

Interestingly, some respondents feel to belong to a wider and general community: they say they share their own opinions with many other spectators ('There's a lot of people that share my opinion'; 'Most people left cinema with oh well that was it!'). They clearly recognize, however, that they are not die-hard fans ('I think my opinions about the Hobbit

are pretty general'). The picture emerged from quantitative analysis seems to have here a first confirmation.

The analysis of word frequencies can reveal other information about the cluster. It allows the researcher to cover an entire corpus of comments, in order to search for relevant keywords. This is a typical content analysis application (Krippendorff, 2004). An interesting technique, within content analysis, is linguistic specificity analysis, that is used in order to check 'which lexical units (words, lemmas or categories) are typical or exclusive in a text or a corpus subset defined by a categorical variable' (Lancia, 2014: 129). Subsets can be compared either with the rest of the corpus or with another subset. This simple but powerful text mining technique recognizes and grades both overused and underused words for each subset (ibid.), thus providing interesting insights.

For our purposes, all the answers to the open-ended questions within each cluster were assembled in a single text. A long process of cleaning, correction, lemmatization and disambiguation of the corpus was necessary, in order to obtain reliable results. For example, the word 'movie' was converted into 'film'; the expression '*Lord of the Rings*' became *LOTR*; some equivalent expressions such as 'Battle of the Five armies' and 'Battle of the 5 armies' were converted into a single word: 'Battle_of_the_Five_Armies', and so forth. Stop-words ('and', 'with', 'the' etc.) were not considered in the analysis. The categorical variable used in specificity analysis was cluster membership. Part-whole analysis of typical words was carried out.[1]

Table 3: Average spectator: specificity analysis and word cloud

Ten most typical words	Freq.	Ten most atypical words	Chi-squared (p < 0.05)
Film	1549	LOTR	323.38
Think	338	book	100.76
Hobbit	333	Tolkien	64.60
People	269	Tauriel	25.32
Good	251	change	25.02
Trilogy	205	adaptation	23.99
Felt	169	triangle	20.55
Long	168	Beorn	19.09
Enjoy	144	Peter_Jackson	17.98
Know	136	addition	17.36

For each cluster, a 'word cloud' was also generated (see **Tables 3, 4, 5, & 6**)[2]. 'Word clouds provide a visual display of word counts from one or more texts. The more frequently a word appears, the larger the word is displayed' (Ignatow and Mihalcea, 2016: loc. 1650). Word clouds have been criticized because they do not provide context for readers to understand how words are used within a text. However, 'two or more clouds can be shown together to contrast word usage across corpora or documents' (ibid.: 1661). Besides, word clouds could also be a valuable tool in a quali-quantitative research design.

The results of specificity analysis seem to confirm the picture coming from the qualitative approach. The average spectator enjoyed the *Hobbit* films, he/she found them 'good', but thought they could have been better. The choice of shooting a long trilogy of films was not welcomed. Words like 'people' and 'friend' (see the word cloud) suggest that these films were experienced socially. Significantly, the less used words[3] are those regarding the book, its adaptation, and the changes in the original story (for example, the addition of Tauriel and her love triangle with Kíli and Legolas).

3.2. *The enthusiastic fan*

Looking at the most representative comments in the second cluster, the fan's status clearly emerges. However, in light of the actual content of the texts selected, that do not omit advice and criticism, we'd better use the term 'die-hard' fan, instead of 'enthusiastic'. The passages below were cut out from the answers of the most ideal-typical respondents (those having the shortest distance to the centroid) belonging to the second cluster.

> I loved the feel of them. Being taken back to Middle-earth. There were moments that annoyed me, variations from the original story that I felt were handled better on *LOTR* than in this. However there were moments where it captured the essence of the book so well I found myself welling up. (...) Visually I found they were very pleasing. (…) One of the main reasons I love the books is the cultural and historical relevance. Finally while many didn't like the *LOTR* tie

in moments (Gimli being shown as a boy, and the reference to Aragorn in the last film) I greatly enjoyed it. I nearly lost it in the best way at the end of the third film, when Gandalf knocks. (…) I don't enjoy how over the top his fight scenes were. They made he escape in the barrels from Mirkwood something it didn't need to be. (…) I play an online game, *Lord of the Rings* online. It was there I met my girlfriend who lived across the ocean in America. Through our shared love of Tolkien as a writer and a scholar and the worlds that not only he created but the ones he drew inspiration from we found one another and eventually she moved here to study creative writing. The first time I visited her we saw the release of an unexpected journey. (…) Tolkien's works have shaped my life this way and while I as many do have mixes feelings on the hobbit films, first and foremost the memories attached to them will be what remains with me (#4421, distance: 0.09, British, 26-35 years, male, vocational qualification).

I loved them, overall, there was great attention to details and scenes that illuminated different aspects of Middle-earth and the characters than we have seen before, in *LOTR*, *The Battle of Five Armies*. (…) There are too many liberties taken in that movie especially, I do not mind the subplot with the white council, seeing as that is taken from actual source material, but the never ending chase of the dwarves by Azog is too much. Additionally, the Tauriel-Kili romantic subplot only serves to belittle the eventual friendship between Legolas and Gimli, as well as the relationship between Aragorn and Arwen, since they are so special because they are unique and, for Legolas and Gimli, without precedents. It definitely puts into sharp relief the reason why Christopher Tolkien and the Tolkien estate refuse to sell the rights to Silmarillion, and why they are justified in doing so. (…) Definitely the active debate about the Kili-Tauriel romance (…) has many interesting viewpoints from different angles. (…) I am able to appreciate all the work Tolkien has done, and I do get a bit annoyed when it is disregarded (#4587, distance: 0.09, Danish, 16-25 years, female, vocational qualification).

Amazing, mind and eye candy. Followed the book OK, liked the new additions from his other books. Not as good as *LOTR*.   No Tom Bombadil! Didn't follow hobbit book enough, did like some of the additions but we are fans, knowing the stories. That's the way the movie should be. Not a purist, but it does take away from the movie. Pulls my attention when storyline and distracts from the movie. Don't think *LOTR* ever did that. *Hobbit* did. *Hobbit* was good but no *LOTR*. I'm part of the ALEP group. We are a Tolkien enthusiast group that gets together every 3 yrs in a non convention style Shire/Middle-earth gathering (#9833, distance: 0.09, American, 46-55 years, male, vocational qualification).

As films, they were exciting, fun, emotional, visually fascinating, aurally thrilling: an immersive experience. Very enjoyable as long as I shut out that part of my brain that read ALL of Tolkien's works. Super Dwarves! Super Elves! Silly Goblin King. A Shire hobbit learns to use a sword by watching others! Thranduil, a Woodland elf; Legolas, a Woodland elf; the equals or betters than the Eldar, and even lay claim to be among them. Super Bard! Galadriel more powerful than Gandalf? Physics negated by, maybe, magic and luck. Gold. Yeah, gold; too much of it, and too little comprehension of the fluid properties of molten metal. The 'themes' Peter Jackson inserted are standard: love, loyalty, greed, redemption. No treatment of them was particularly illuminating. But this is an action movie, and as such I'm glad he kept some of the themes of the true *Hobbit*. (…) Most prefer to just enjoy rather than be critical. I did enjoy the films when I put away judgements. People who did not read the books enjoy the movies most, and, most likely, would not like the books as much as people who read them before the Age of Computer Games (which I also enjoy). (…) First read *The Lord of the Rings* in 1973, *The Hobbit* soon afterward. Each many times since. Read *The Silmarilion* around 1980. Read several times since. Read Letters, histories, etc. Listened to Nicol Williamson's record of *The Hobbit* several times; then to the Mind's Eye cassette tapes of *The Lord of the Rings*, literally, every night before bed until I found the BBC recordings, for around 15 years. Never learned the any of the Middle-earth languages. I was extremely excited to hear that there was going to be a LOTR movie. I have followed TheOneRing.net since 1999. I was glad to see *LOTR* and *Hobbit* on the big screen but was disappointed ultimately. Would go seen a new version of each if they followed the books more closely (#10763, distance: 0.09, American, 56-65 years, male, vocational qualification).

Not as much Beorn is the BOTFA, but I'm sure the extended editions will amend that. The whole 'It's the small things…' theme is a great one that sometimes go unnoticed. I know people in my circle of friends who agree with me and I can imagine there are others who do too, (Tolkien fans or not). I imagine they would come from many different walks of life. (…) The Ian McKellen crying about the use of CGI and not being able to act properly is always a good debate/argument. It has changed my life. I now can view things differently and use an imagination I never knew I had. I was given *The Hobbit* to read whilst I was in a mental health ward and it completely got me through (#15137, distance: 0.09, Australian, 16-25 years, female, vocational qualification).

The first thing to notice in these comments is the opening words. These are always positive and often enthusiastic ('I loved them, overall'; 'I loved the feel of them'; 'Amazing, mind and

eye candy'; 'They were exciting, fun, emotional, visually fascinating, aurally thrilling: an immersive experience'). A passage, however, is particularly enlightening: 'Most prefer to just enjoy rather than be critical. I did enjoy the films when I put away judgments. People who did not read the books enjoy the movies most'. This group of spectators really loved the films of *The Hobbit*, but not as much as those of *The Lord of the Rings* ('Not as good as LOTR'). Tolkien's book and the *Lord of the Rings* films, in fact, are constantly evoked as a point of reference and assessment.

They define themselves as fans ('[We] did like some of the additions but we are fans, knowing the stories'). They're more than fans, actually, since many say the books and the films even changed their own lives ('First read *The Lord of the Rings* in 1973, *The Hobbit* soon afterward'; 'Tolkien's works have shaped my life'; 'It has changed my life. I now can view things differently'). The status of fan is confirmed by the activities these respondents often carry out. These are typically social activities ranging widely across media ('The active debate about the Kili-Tauriel romance (…) has many interesting viewpoints from different angles'; 'I'm part of the ALEP group. We are a Tolkien enthusiast group that gets together every three years'; 'I play an online game, *Lord of the Rings* online. It was there I met my girlfriend who lived across the ocean in America'; 'I have followed TheOneRing.net since 1999') (Baym, 2000; Miller, 2012; Procter et al., 2015).

Table 4: Die-hard fan: specificity analysis and word cloud

Ten most typical words	Freq.	Ten most atypical words	Chi-squared (p < 0.05)
love	1199	long	79,42
think	1096	poor	33,24
people	947	action	28,87
Tauriel	739	bore	28,76
Kíli	501	narrative	28,41
feel	493	sequence	25,97
time	484	drag	25,50
know	446	stretch	24,55
world	439	bloat	21,02
enjoy	439	violence	20,01

'Love', with 1,199 occurrences, is the most typical word in the 'die-hard' fans cluster (see **Table 4**). The role of emotions ('feel', 'enjoy', 'amaze') and the social dimension of film experience ('people', 'share', 'debate', 'world') also seem to come out here, thus confirming the profile characteristics derived from qualitative analysis. Moreover, as we can see in the word cloud too, we can find characters and places related to Tolkien's world ('Kíli', 'Legolas', 'Thorin', 'dwarves', 'Middle-earth', 'The Battle of the Five Armies'). Interestingly, Tauriel – who does not appear in the books – is the most quoted character. Being a controversial character, she was presumably at the center of many debates among fans.

As a further confirmation of the cluster identity, the underused word list is essentially populated by negative terms: 'long', 'poor', 'bore', 'drag', 'stretch', 'bloat', 'violence'. These are the less used words in the 'die-hard' fan cluster (p < 0.05).

3.3. *The cultured spectator*

The cultured spectator has a high level of education, he/she prefers to view films in a cinema, has read Tolkien's books and has seen the *Lord of the Rings* trilogy, but he/she did not actively participate in the debates about the *Hobbit* films, and has given them the lowest ratings.

> Visually they are beautiful films, but I love the book, so was very disappointed by the things that were added (and some things that were cut). If I'd never read the book maybe I would like the films more. Too much padding and characters which aren't in the book. Events happening/characters behaving differently in the film to the book. Making one little book into 3 films was not necessary. Why does every fantasy film now have to be part of a trilogy... I don't know. I sound like a grumpy old Tolkien purist, but I'm not really! People who grew up with the book may agree with me (#950, distance: 0.07, New Zealander, 36-45 years, female, university degree).

Beautiful, great acting, but a poor story line. They tried too hard to tie it back in with the *LOTR* trilogy. It severely disrupted the story-telling process. (…) There were so few female characters. In the last film, we had a grand total of 4 named characters, two of whom were children whose only lines of dialogue were 'AAA!' and 'Daa!' I like how a blockbuster made-for-the-money movie went with the 'gold is evil' theme. Ironic. Sure. I know my dad agrees with parts. I know my mother enjoys them with me. I'm going to talk about it with my colleagues (#989, distance: 0.07, Dutch, 16-25 years, female, university degree).

Mostly underwhelmed. Some truly great acting moments and wonderful fantastic images, but overall too much emphasis on the spectacle and too little on the story. The overabundance of CGI didn't help. Some of the embellishment doesn't work for me (especially the climax of the second movie) as they distract from the core of the story. Also, the visual effects feel rushed and look cheap. Overall, the story doesn't convey danger very well, because the characters survive ludicrous stunts (very different from the *Lord of the Rings*). I felt Peter Jackson indulged himself a bit too much in some of the action scenes (rollercoaster camera movements, etc.). Spectacle over substance. Technology over story (3D, HFR). [People with whom I'd share my ideas:] Friends and colleagues, people with similar interests. What people thought of the adaptation of the book (#2979, distance: 0.07, German, 36-45 years, male, university degree).

The movies did not capture the magic of *The Hobbit*, mainly because they crammed to much meaningless action in the movie. Less would have been more. Radagast the Brown: a second coming of Jar Jar Binks, it ruined the whole first movie. If there is a director's cut of the first movie please cut that part. The encounter of Bilbo and Smaug; the best example where story has been replaced by action, it was kind of *Transformer: Age of Extinction* where one was just hoping the film would end, it dragged and dragged on. Silly and useless. The hobbit falls in the same trap as so many blockbusters these days. Good story telling is replaced by CGI effects and action. It is boring and unimaginative and once the movie has been seen its forgotten. A good story and good character will be remembered CGI not (case: 3288, distance: 0.07, Swiss, 36-45 years, male, university degree).

As a Tolkien fan, the films are a must see for me. However, I was (and still am) dissatisfied with the decision to make 3 movies out of the 200-pages book. Artificially added battles (battle with orcs in the mines in *Unexpected Journey*), extremely long action scenes (barrels scene in *Desolation of Smaug*), useless

and even awkward characters (Tauriel, Legolas) really harm the overall experience. Action scenes, although technically advanced, are essentially boring, as the informed viewer knows they just serve as filler to extend the footage. As nobody dies, the heroes are too perfect and immortal and the villains totally useless. Recurring characters from *LOTR*, mainly Legolas, just do not work and do not fit in the movie. Overall, the *Hobbit* trilogy lacked the atmosphere of *LOTR* trilogy and failed to get me emotionally invested in the story and the characters. However, still, if Jackson makes more Tolkien movies, I will definitely watch them. Every trip to Middle-earth is a great experience for me. (…) I liked the character [of Radagast] in the book. I imagined all wizards as wise and dignified. Radagast, focusing his effort on animals and plants rather than humans, might have seen the truly important things in the world - saving of the nature, which is more important than power struggles between the warlords. However, the film depiction of a near-crazy, irresponsible, weak-minded and half-witted homeless guy with the shit in his hair (seriously, why humiliate him so much?), used as comic relief, really offended me. Only the Battle of the Five Armies raise any broader issues, albeit, sufficiently. The corrupting power of wealth and power was addressed beautifully and satisfactorily. Also the final realization of Thorin and Thranduil, each in his own way, of the truly important thing above their immediate personal and political interests is able to convey a message. Tolkien fans who enjoy the books (and expanded universe) may share the ideas, perhaps. (…) I read the *Hobbit* book (as well as *LOTR* and all other Middle-earth books) several times and consider myself a Tolkien fan, although I was never a general fantasy genre fan. Interested in literature and history, reads a lot (usually more serious literature than fantasy) (#4602, distance: 0.07, Slovakian, 26-35 years, male, university degree).

This cluster is basically made up of (heavy) readers. One of the respondents describes himself as follows: 'Interested in literature and history, reads a lot (usually more serious literature than fantasy)'. The cultured spectator has read both *The Hobbit* and *The Lord of the Rings*, in some cases more than one time, but he/she is not necessarily a fantasy-addicted: 'I read the *Hobbit* book (as well as *LOTR* and all other Middle-earth books) several times and consider myself a Tolkien fan, although I was never a general fantasy genre fan'. These readers did not like the films, and they generally mention the weakness of the story as the main reason for disappointment ('Beautiful, great acting, but a poor story line'; 'Too much emphasis on the spectacle and too little on the story'; 'Good storytelling is replaced by CGI effects and action'). The other issue this kind of spectator raises is the bad adaptation of the book ('I love the book, so was very disappointed by the things that were added (and some things that were cut). If I'd never read the book maybe I would like the films more. Too much padding and characters which aren't in the book. Events happening/characters

behaving differently in the film to the book (…). I sound like a grumpy old Tolkien purist, but I'm not really!'; 'They crammed to much meaningless action in the movie. Less would have been more'). Some comments let us think these are competent spectators indeed; for example, when some paradoxes are noticed ('I like how a blockbuster made-for-the-money movie went with the 'gold is evil' theme. Ironic') or the lack of well-developed female characters are highlighted ('There were so few female characters'). As we'll see, this topic returns in the last cluster.

Although the cultured spectator did not participate to the debate on the Hobbit films in the media, an interesting characteristic of this group is sociability. The cultured spectator tends to talk about the films with other people ('I'm going to talk about it with my colleagues'; '[People with whom I'd share my ideas:] Friends and colleagues, people with similar interests. What people thought of the adaptation of the book'). He/she might act as an opinion leader or might ask other people for advice. An upgraded version of the classic 'two step flow of communication' model (Katz and Lazarsfeld, 1955) seems to operate in this case. Considering the role currently played by social networking sites such as Facebook, Twitter and others, some scholars have used the expression: 'multistep flow' (e.g. Davis, 2009). 'In the multistep flow of communication, a social network receives information from the media that is, in turn, disseminated throughout a community by leaders within it. Likewise, the public may receive information from media and carry it back to the opinion leaders for evaluation or confirmation' (Jowett and O'Donnell, 2012: 395). This would be worth studying through social network analysis techniques (Smith et al., 2014).

Table 5: Cultured spectator: specificity analysis and word cloud

Ten most typical words	Freq.	Ten most atypical words	Chi-squared (p < 0.05)
book	3131	Kíli	146,32
LOTR	2272	amaze	125,74
scene	1153	Tauriel	103,02
felt	982	Fili	88,55
long	934	love	76,96
action	658	Thorin	67,17
read	616	die	64,80
original	507	Battle_of_the_Five_A	43,55
battle	504	perfect	40,39
plot	471	death	39,30

Specificity analysis (see **Table 5**) confirms that the cultured spectator uses the book as a point of reference, when talking about the Hobbit films ('book', 'original', 'read'). *The Lord of the Rings* is also often mentioned, presumably referring to both the books and the films. Criticisms are raised about the adaptation of the book: the films are too 'long', 'stretched' and 'boring'; they contain 'unnecessary' scenes, but at the same time they 'lack' something. 'Ridiculous' is another negative term we glimpse in the word cloud. Specificity analysis makes us notice that the story ('scene', 'sequence', 'plot') and the *mise-en-scène* ('action', 'battle', 'effects') mostly drawn the attention of this type of spectator.

If we look at the most underused words, two groups of words seem to come out at least. The first one, significantly, that has to do with the film enjoyment (e.g. 'amaze', 'love'). The second one might be connected to the theme of death (e.g. 'die', 'death'): Kíli and Fili, in fact, are killed, while defending Thorin, who is mortally wounded, in *The Battle of the Five Armies*. They are the only characters in Bilbo's team that die in *The Hobbit*. Perhaps because she is an added character, not present in the original book, Tauriel also was not mentioned (and maybe appreciated) by the cultured spectator.

3.4. *The non-spectator*

The non-spectator's answers are characterized by a high number of blank cells in the dataset. Missing data are, in fact, quite frequent. This is a sort of residual group, characterized by the fact that its member do not prefer to see films in a cinema. Besides, as we have learned from cross-tabulations, there is a significant prevalence of women over men in this group. As for the rest, the quantitative analysis does not bring to light any other peculiar traits. This is one of those cases where qualitative answers might give us much more information.

> Very entertaining films, although I didn't agree with all the changes from the book (e.g. the Dwarf/Elf love triangle). The CGI - I thought *LotR* looked better with Orcs etc. in masks. Parts of the hobbit films looked like a computer game.

Adding female characters (good) but making them a romantic interest (bad) (#7070, distance: 0.08, British, 26-35 years, female, vocational qualification).

I was surprised at all the added content and storyline. It was very well done but did and didn't like all the deaths at the end. In American films they all tend to have happy ending. I was expecting that. I do love to see films that give you unexpected. (…) I am a disabled shut in. My world is 4 walls. This has allowed me to escape of those walls (#9750, distance: 0.08, American, 46-55 years, female, vocational qualification).

Good, but not nearly as good as *LotR*. The death of Smaug. It didn't seem fitting after all that buildup. Behind the scenes stories. It is a little disappointing that the world that Tolkien created, is so black-and-white, which I think is the reason that the *Song of Ice and Fire* has become such a success. We want to identify with characters (#19915, distance: 0.10, Austrian, 16-25 years, male, vocational qualification).

Unnecessary to have three films I feel but I could be wrong depending on the third film. Little bit too much CGI, preferred men in orc costume like *LOTR* trilogy. Just life lessons, stand your ground. Just because you are small does not make you brave, be compassionate fight for good etc. (…) Adding of characters not in the book, and CGI goblins. I enjoy the hero's journey story lines (#929, distance: 0.10, Irish, 26-35 years, male, university degree).

They had their moments, but there were too many battle scenes for me. Some of the crucial scenes (in my opinion) were only in the extended versions (case: 4602, distance: 0.07, Slovakian, 26-35 years, male, university degree) (#17327, distance: 0.10, Czech, 26-35 years, female, university degree).

Very enjoyable, not enough strong female characters, gender equality, brotherhood, politics. Many of my friends agree with my views the difference between the book and film, lack of female equality (#18881, distance: 0.10, British, 16-25 years, female, university degree).

Breathtaking fight scenes (especially in the *Battle of the Five Armies*), and an wonderful experience of *The Hobbit* as a film trilogy. The depiction of the character Tauriel. I understand the need for strong female characters in a film dominated by males, but the context in which she is presented diminishes her impact. Tauriel is a captain of the Mirkwood elves, yet her character arc revolves around being the love interest of both Legolas and Kili. (…) [Broader themes:] The depiction of female characters in entertainment (…), the

implications of different forms of government (see societies of Mirkwood, The Shire, Rivendell, Erebor pre-Smaug). (…) The *LOTR*/*Hobbit* fandom on tumblr can be very vocal about many issues, and that's where I learned about most these issues (#23264, distance: 0.10, Canadian, 16-25 years, female, university degree).

The most interesting aspect emerging from the answers above is the significant presence of passages discussing the role of female characters. Some respondents consider this presence as unfair, as instrumental in capturing more audience ('Adding female characters (good) but making them a romantic interest (bad)'; 'I understand the need for strong female characters in a film dominated by males, but the context in which she [Tauriel] is presented diminishes her impact'), although this has not been confirmed by empirical research (Lindner et al., 2015). Oddly, others denounce the lack of important female roles, and the lack of social and political commitment in general ('Not enough strong female characters, gender equality, brotherhood, politics').

One of the reasons why some people do not (prefer to) see films in a cinema is disability. This is the case of one respondent here ('I am a disabled shut in. My world is four walls. This has allowed me to escape of those walls'). Other answers emphasize the need of seeing the extended versions of the films instead ('Some of the crucial scenes (in my opinion) were only in the extended versions'). These versions, in fact, are rarely screened, and made available only on DVD and Blu-Ray.

In conclusion, it seems that this kind of spectator is characterized by some sort of social commitment, as a consequence of some sexual, social or 'political' (I allude to film distribution policies here) discrimination (Gwynne and Muller, 2013; Aitchison, 2003; Drake, 2008). This, of course, may be an actual discrimination or a perceived one. The non-spectator, in any case, has a different eye when viewing a film. He/she is competent, but unorthodox. In light of these findings, the label 'alternative spectator' seems more appropriate, rather than 'non-spectator'.

Specificity analysis shows us that the alternative spectator mainly addressed his/her comments to the love triangle between Tauriel, Kíli and Legolas, while the overuse of the term 'Tolkien' is presumably connected either to the fact that this relationship is not present in the books, or to wider issues concerning the adaptation of the book (see the words 'bad', 'poor'. The term 'literature' can be glimpsed in the word cloud too). Other words seem to focus on the battle against Azog, in which Thorin's company was dramatically involved.

Much more interesting is the picture that comes out from the analysis of underused words. A group of words ('power', 'nerds', and 'gender') seems to refer to the social and 'political' questions previously discussed, ideally retracing some of the findings emerged in the qualitative analysis of responses. These words are underused because they are presumably felt as 'discriminatory', in many senses, by this kind of spectator. Other untypical words allude to both technical aspects ('frame', 'rate', 'split', 'stretch') and the appreciation of the films ('enjoyable', 'badly').

Table 6: Alternative spectator: specificity analysis and word cloud

Ten most typical words	Freq.	Ten most atypical words	Chi-squared (p < 0.05)
Tolkien	813	power	11,85
Tauriel	532	rate	11,59
Legolas	343	frame	11,30
Kíli	329	split	9,55
Thorin	321	nerds	9,36
dwarf	167	gender	9,32
bad	159	enjoyable	8,19
point	153	protagonist	8,00
Azog	101	badly	7,84
poor	93	stretch	7,19

4. Conclusions

Large and complex datasets could be rather difficult to analyze, when they include qualitative data. In this contribution, I suggested a strategy that might allow the researcher to manage such complexity, identifying a small number of ideal-typical respondents for further analysis. The Hobbit Research Project dataset, containing answers to both closed and open-ended questions, was used for this purpose.

The strategy proposed here is based on between-methods sequential triangulation, and tried to combine statistical techniques (K-means clustering) with both qualitative and quantitative textual analysis.

K-means clustering was used to classify the subjects. Four groups of respondents, differently distributed in the countries that participated in the project, emerged from the analysis. They were first labeled as 'average spectator', 'enthusiastic fan', 'cultured spectator', and 'non-spectator'. A cross-validation process was then settled to assess the reliability of these ideal-types. Textual responses corresponding to each cluster were initially analyzed using a qualitative approach, in order to both refine the clusters and identify meaningful discourse patterns. This led us to reconsider the definitions of two clusters: the enthusiastic fan was then renamed as 'die-hard' fan, while the non-spectator was renamed as 'alternative spectator'.

The analysis produced many theoretical and methodological insights. From a sociological viewpoint, a multifaceted profile of the contemporary spectator emerges. Viewing a film is still a social experience. This latter is now accomplished not only face-to-face but also online, through social networking sites, for instance. These spectators appear experienced and cunning. They are the expression of several 'screen cultures', that is 'cultural formations built around the consumption of texts and processes delivered through visuals on a screen: video games, DVDs, home computer games, and the continual development of new modes of access to visual entertainment through digital and online media' (Turner, 2006: 61). Some spectators also form part of interpretive communities, that often link spectatorship to social activism (e.g. the alternative spectator) (Pozo, 2013).

An additional quantitative check on textual responses was finally made, in order to further validate the results, that is, sensitivity analysis. The outcome of this additional analysis confirms that the methodological mix proposed in this contribution can be used with confidence, since it yields robust results. The analysis of typical and untypical words, indeed, strengthened and deepened the picture emerged from cluster analysis and qualitative content analysis, without contradicting its core meaning. The word clouds and the statistical clusters, moreover, almost overlap. This can be considered as a good example of 'internal validity'. A similar strategy was used in the *Lord of the Rings* Project (Trobia, 2008a) and also proved to generate robust and stable results. In this case, we can speak of 'external validity'.

The analysis reported in this contribution concerned English language responses only. Nonetheless, the potential of the methodological mix illustrated here goes beyond. When many responses in other languages are present in the dataset, the method discussed here can help optimize the choice of those that have to be translated for further analysis, thus avoiding a long and dispersive process of trial and error.

Biographical note:

Alberto Trobia is associate professor of Sociology and Social Research Methods at the University of Palermo (Italy). His main research interests are in the field of mixed methods and computational social sciences (with an emphasis on text analysis and social network analysis). He is also interested in popular culture, mainly popular music and cinema (he

participated in the *Lord of the Rings* Research Project). He has written some entries for the *Encyclopedia of Survey Research Methods* (Sage, 2008). Contact: alberto.trobia@unipa.it.

References:

Aitchison, C. (2003) 'From leisure and disability to disability leisure: developing data, definitions and discourses', in *Disability & Society*, 18(7): 955-969.

Barker, M., Mathijs, E., Trobia, A. (2008) 'Our methodological Challenges and Solutions', in M. Barker and E. Mathijs (eds.), *Watching the Lord of The Rings*, New York: Peter Lang: 213-239.

Bauman, Z. (2007) *Consuming Life*, Cambridge, Polity.

Baym, N.K. (2000) *Tune In, Log On: Soaps, Fandom, and Online Community*, Thousand Oaks, CA: Sage.

Bergman, M.M. (ed.) (2008) *Advances in Mixed Methods Research*, London: Sage.

Biorcio, R. (1993) *L'analisi dei gruppi*, Milano: Franco Angeli.

Blaikie, N. (1991) 'A Critique of the Use of Triangulation in Social Research', in *Quality & Quantity*, 25: 115-136.

Blumer, H. (1933) *Movies and Conduct*, Macmillan: New York.

Bryman, A. (2016) *Social Research Methods*, 5th ed., Oxford: Oxford University Press.

Cohen, L., Manion, L. (2000) *Research Methods in Education*, 4th ed., New York: Routledge.

Cola, M., Brusa, M.M. (2014) 'Researching Ethnic Minority Groups as Audiences: Implementing Culturally Appropriate Research Strategies', in Patriarche et al. (2014): 107-123.

Cuadrado, M., Frasquet, M. (1999) 'Segmentation of Cinema Audiences: An Exploratory Study Applied to Young Consumers', in *Journal of Cultural Economics*, 23: 257–267.

Davis, D.K. (2009) 'Two-step and multistep flow', in S.W. Littlejohn (ed.), *Encyclopedia of communication theory*, Thousand Oaks, CA: Sage: 969-971.

Denzin, N.K. (1970) *The Research Act in Sociology: A Theoretical Introduction to Sociological Methods*, London: Butterworths.

Drake, P. (2008) 'Distribution and marketing in contemporary Hollywood', in J. Wasko and P. Mcdonald (eds.) *The Contemporary Hollywood Industry*, Oxford: Wiley-Blackwell: 63-82.

Egan, K., Barker, M. (2006) 'Rings Around the World: Notes on the Challenges, Problems and Possibilities of International Audience Projects', in *Participations*, 3(2), http://www.participations.org/volume%203/issue%202%20-%20special/3_02_eganbarker.htm

Ettema, J.S., Whitney, D.C. (eds.) (1994) *Audiencemaking: How the Media Create the Audience*, Thousand Oaks, CA: Sage.

Everitt, B.S., Landau, S., Leese, M. (2001) *Cluster analysis*, 4th ed., London: Arnold.

Findahl, O. (2011) *Swedes and the Internet 2011*, Stockholm, The Internet Infrastructure Foundation, https://www.iis.se/docs/SOI2011-en.pdf.

Findahl, O., Lagerstedt, C., Aurelius, A. (2014) 'Triangulation as a Way to Validate and Deepen the Knowledge about User Behavior. A comparison between Questionnaires, Diaries and Traffic Measurements', in Patriarche et al. (2014): 54-69.

Franzosi, R. (2012) 'The Difficulty of Mixed-method Approaches', in *Sociological Methodology*, 42: 79-81.

Gaiser, T.J., Schreiner, A.E. (2011) *A Guide to Conducting Online Research*, Thousand Oaks, CA: Sage.

Garson, G.D. (2014) *Cluster Analysis*, Asheboro, NC, Statistical Publishing Associates, e-book edition.

Gripsrud, J. (1998) 'Film Audiences', in J. Hill and P. Church Gibson (eds.), *The Oxford Guide to Film Studies*, Oxford: Oxford University Press: 202-211.

Gwynne, J., Muller, N. (2013) *Postfeminism and contemporary Hollywood cinema*, New York: Palgrave Macmillan.

Hastall, M.R., Sukalla, F. (2014) 'Digging the Web. Promises and Challenges of Using Web 2.0 Tools for Audience Research', in Patriarche et al. (2014): 176-195.

Ignatow, G., Mihalcea, R. (2016) *Text mining. A Guidebook for the Social Sciences*, Thousand Oaks, CA: Sage, e-book edition.

Jenkins, H. (2006) *Convergence Culture: Where Old and New Media Collide*, New York: New York University Press.

Jensen, K.B. (2012) *A Handbook of Media and Communication Research: Qualitative and Quantitative Methodologies*, 2nd ed., London: Routledge.

Jensen, K.B. (2014) 'Audiences, Audiences Everywhere–Measured, Interpreted and Imagined', in Patriarche et al. (2014): 226-240.

Jowett, G.S., O'Donnell, V. (2012) *Propaganda & Persuasion*, 5th ed., Thousand Oaks, CA: Sage.

Katz, E., Lazarsfeld, P.F. (1955) *Personal Influence: The part played by people in the flow of mass communications*, New York: The Free Press.

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., Couper, M. (2004) 'Psychological Research Online: Report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet', in *American Psychologist*, 59(2): 105-117, http://kraut.hciresearch.org/sites/kraut.hciresearch.org/files/articles/kraut04-PsychologicalResearchOnline.pdf

Krippendorff, K. (2004) *Content Analysis. An Introduction to Its Methodology*, 2nd ed., Thousand Oaks, CA: Sage.

Lancia, F. (2014) *T-Lab 9.1. Tools for Text Analysis. User's Manual*, http://www.mytlab.com/Manual_en.zip

Lindner, A.M., Lindquist, M., Arnold, J. (2015) 'Million Dollar Maybe? The Effect of Female Presence in Movies on Box Office Returns', in *Sociological Inquiry*, 85(3): 407-428.

Marchand, J., Khallaayoune, Z. (2010) 'LOV and the big screen: A value-system segmentation of movie goers', in *Journal of Targeting, Measurement and Analysis for Marketing*, 18, 3-4: 177–188.

Miller, K. (2012) *Playing Along: Digital Games, YouTube, and Virtual Performance*, New York: Oxford University Press.

Patriarche, G., Bilandzic, H., Jensen, J.L., Jurišić, J. (eds.) (2014) *Audience Research Methodologies. Between Innovation and Consolidation*, New York: Routledge.

Pozo, D. (2014) 'Feminist Film Theory, core concepts', in E. Branigan and W. Buckland (eds.), *The Routledge Encyclopedia of Film Theory*, New York: Routledge: 187-194.

Procter, R., Voss, A., Lvov, I. (2015) 'Audience research and social media data: Opportunities and challenges', in *Participations*, 12(1): 470-493,

Schaeffer, N. C., Presser, S. (2003) 'The Science of Asking Questions', in *Annual Review of Sociology*, 29: 65-88.

Schrøder, K., Drotner, K., Kline, S., Murray, C. (2003) *Researching Audiences*, London: Arnold.

Seale, C. (2000) 'Using computers to analyse qualitative data', in D. Silverman, *Doing Qualitative Research: A Practical Guide*, London: Sage: 155-174.

Smith, M.A., Rainie, L., Shneiderman, B., Himelboim, I. (2014) 'Mapping Twitter Topic Networks: From Polarized Crowds to Community Clusters', in *Pew Research Internet Project*, February 20, http://www.pewinternet.org/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters/

Stacey, J. (1994) *Star Gazing: Hollywood Cinema and Female Spectatorship*, London: Routledge.

Teddlie, C., Tashakkori, A. (2009) *Foundations of Mixed Methods Research. Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*, Thousand Oaks, CA: Sage.

Tourangeau, R., Rips, L. J., Rasinski, K. (2000) *The Psychology of Survey Response*, Cambridge: Cambridge University Press.

Trobia, A. (ed.) (2008a) *Sociologia del Cinema Fantastico*, Torino: Kaplan.

Trobia, A. (2008b) 'Sampling', in P.J. Lavrakas (ed.) *Encyclopedia of Survey Research Methods*, Thousand Oaks, CA: Sage: 783-784.

Turner, G. (2006) *Film as Social Practice*, 4th ed., New York: Routledge.

Weber, M. (1904) *Objectivity of Social Science and Social Policy*, http://anthropos-lab.net/wp/wp-content/uploads/2011/12/Weber-objectivity-in-the-social-sciences.pdf

White, M., Judd, M. D., Poliandri, S. (2012) 'Illumination with a Dim Bulb? What do Social Scientists Learn by Employing Qualitative Data Analysis Software in the Service of Multimethod Designs?', in *Sociological Methodology*, 42: 43-76.

Wright, K.B. (2005) 'Researching Internet-Based Populations: Advantages and Disadvantages of Online Survey Research, Online Questionnaire Authoring Software Packages, and Web Survey Services', in *Journal of Computer-Mediated Communication*, 10(3), http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2005.tb00259.x/full.

Notes:

[1] Specificity analysis was carried out using the software T-Lab (http://tlab.it/en/presentation.php).

[2] The word clouds shown in this contribution have been generated using an online tool available at http://www.wordclouds.com/.

[3] Note that the chi-squared test is reported in the tables, when considering underused words.