Psychological processes underlying Wikipedia representations of natural and manmade disasters

Michela Ferron
Center for Mind/Brain Sciences
University of Trento
Corso Bettini 31, Rovereto, 38068

+39 0462 314050
m.ferron@unitn.it

Paolo Massa
Fondazione Bruno Kessler
Via Sommarive 18
Trento, 38123

+39 0461 314336
massa@fbk.eu

ABSTRACT

Collective memories are precious resources for the society, because they help strengthening emotional bonding between community members, maintaining groups cohesion, and directing future behavior. Studying how people form their collective memories of emotional upheavals is important in order to better understand people's reactions and the consequences on their psychological health. Previous research investigated the effects of single traumatizing events, but few of them tried to compare different types of traumatic events like natural and man-made disasters. In this paper, interpreting Wikipedia as a collective memory place, we compare articles about natural and human-made disasters employing automated natural language techniques, in order to highlight the different psychological processes underlying users' sensemaking activities.

Keywords

Collective memories, traumatic events, man-made disasters, natural disasters, LIWC, automated content analysis techniques.

1.   INTRODUCTION

On 11 September 2001, the US were shocked by one of the worst terrorist attacks in world's history: two hijacked airplanes hit the Pentagon and were crashed into the World Trade Center causing both towers to collapse and killing nearly 3,000 people. On 26 December 2004, a 9.1-magnitude earthquake struck off the west coast of Sumatra, Indonesia. The resulting tsunami devastated the coasts of fourteen countries with waves up to 30 meters high, killing 230,000 people.

Both were highly traumatizing events which left indelible marks on people's consciousness all around the world, but they are inherently different in their origin. In this article we study the representation of man-made and natural disasters in the English Wikipedia, proposing to implement natural language processing techniques to investigate the formation of collective memories of traumatic events in the online encyclopedia.

The concept of "collective memory" was introduced by Halbwachs in the early past century [22], and highlights the collective dimension of memory building. In particular, we intend collective memory as the continuous active process of sense-making and negotiation between past and present [31] and we interpret Wikipedia as a collective memory place [40], where the discursive formation of memories takes place.

Particularly in Wikipedia's article pages about events, it is possible to observe the process of creation and maintenance of a common collective memory of these events. This process happens through direct edits to the article pages themselves and discussions on the associated talk pages. Through this process of coordination and negotiation of different perspectives on the past which sometimes leads to controversial debates and heated discussions, typical of collective remembrance dynamics, Wikipedia users reach a common representation of past events. The particular characteristics of Wikipedia, which allow users to express their ideas and pursue their perspectives in the article and talk pages, provide researchers the possibility to access this complex ecology of discussions unobtrusively and almost in real time, and make it one of the most stimulating Web 2.0 environments for the study of collective memory processes also from a longitudinal point of view.

Investigating the complex dynamics of collective memory formation is particularly important in the case of traumatic events, which shock the inner core of a community's identity [3]. Collective memory plays an important social role, in that through the social sharing of the past, it affects the attitudes toward the present, highlighting collective needs, re-defining cultural identities, leading sometimes to political and institutional changes and persisting for many years and generations [39, 44, 15, 50].

In this paper we focus on the English Wikipedia's articles about natural and man-made disasters, investigating the different representations of these traumatic events and highlighting the diverse psychological processes underlying their shared accounts, exploiting automated content analysis techniques.

Previous studies investigated the effects of single traumatizing events [13, 1, 21, 29, 23, 41, 52, 19, 18] but few of them compared natural and human-made disasters [20], partly because each one is unique in its outcomes and consequences on the psychological and physical health of the involved populations. Undeniably, outlining common effects of different kinds of traumatic events can be difficult, because of the peculiarity of their particular circumstances, the number of dead and injured people, the damage to properties, the exposure to toxic chemicals or to the death of others, the degree of life threatening experiences, etc. Moreover, each individual can react in a different way, and the recovery environments can be very different. However, past research suggested theoretical reasons for hypothesizing some differences in the type, severity and

longevity of consequences of natural and human-made disasters [18]. In this article we implement natural language processing techniques to investigate quantitatively the differences in the language used in the English Wikipedia's articles to describe natural and man-made disasters, reporting how the language used can reflect different underlying psychological processes [51], highlighting the differential nature of traumas. In the next sections we describe the Linguistic Inquiry Word Count (LIWC), a software for computerized text analysis [36], showing that it can successfully be applied to Wikipedia to detect differences between articles about events with different subjects, like articles about traumatic and non-traumatic events, and about recent and old traumatic events. Then we compare articles about natural and human-made disasters, showing the diverse psychological processes that can be extracted from their textual representations.

2. AUTOMATED NATURAL LANGUAGE PROCESSING OF EVENTS IN WIKIPEDIA

The language we use in our daily life reflects our identities, our inner thoughts, fears, desires, and emotions [51]. At the individual level, it is our primary form of communication, through which we translate our internal thoughts and establish social relationships with others, and at the cultural level, it expresses the collective needs, beliefs and memories of a society. We also translate events into words, and the language we use affects the way those events will be later perceived, understood, and recalled, enriching them with moral values and profound meanings, contributing to define the collective identity of a community. Indeed, talking about an event is a sensemaking activity, which acts also as a form of rehearsal, helping to better organize thoughts and memories about the facts [37].

When people face an emotional upheaval they naturally tend to talk about it, and the social sharing serves different psychological purposes, such as making sense of the trauma, seeking for affective and social support, or strengthening the emotional bonding with the other members of the community [13]. Pennebaker and Harber [37] studied the degree of social sharing of two traumatic events: the San Francisco Bay Area earthquake of 1989 and the Persian Gulf War. During a period of three months, they interviewed residents of San Francisco and of Dallas, Texas, asking, among other questions, how frequently in the previous 24 hours they had talked with someone about the earthquake or the war. They found that the frequency of social sharing of these traumatic events was considerably higher during the two weeks immediately following the earthquake or the beginning of the war. According to the authors, and considering also that during the first days the media coverage had been intense, this initial period of social sharing provided the basis for the construction of common experiences and collective memories. Talking about an emotional upheaval helps people to better organize and assimilate facts. When an event is particularly meaningful from an emotional point of view, it determines talking and social sharing in the attempt to understand and make sense of it. In this context, social sharing can also have a therapeutic effect, fostering healing and allowing people to move past the traumatizing experience. Once an emotional upheaval has been cognitively processed and assimilated, people may also forget the trauma [39].

Typically, research on traumatic events has implemented interviews and retrospective self-reports to study the outcomes and consequences of disasters [13], but research has shown that self-reports can lead to biases and memory distortions [48].

Now, with the development of computers, Internet and Web 2.0 environments, we are provided with the unprecedented opportunity to study people's reactions to upheavals as they naturally unfold. The widespread accessibility of people's thoughts and emotions in the digital environments provides researchers with new opportunities toward empirical and quantitative work at large scale. The Internet provides a large amount of data which researchers can collect unobtrusively and almost in real time, and the massive backup into digital archives allows scholars to conduct longitudinal studies on these data, without giving up the spontaneity of interactions [17, 13]. In this context, Wikipedia, where users share their thoughts and perspectives to reach a common view on events, seems particularly appropriated for the study of traumatic events [16, 26, 40].

With this regard, in this paper we investigate if and how automated content analysis tools can be employed to analyze, empirically and quantitatively, the active process of sensemaking and negotiation of meaning taking place in Wikipedia, by examining the patterns of language used in the pages about traumatic events. The Linguistic Inquiry and Word Count (LIWC) was designed by Pennebaker, Francis and Boot [51] to calculate the degree to which people use different categories of words in a text, assigning words to various linguistic and psychological dimensions of language. LIWC has two main features: a processing component, which analyzes each text file on a word-by-word basis, and a dictionary, which is a collection of words defining a particular category. Specifically, LIWC reads a text file analyzing one word at a time, looking for a dictionary match with each target word in the text. If the current word matches a dictionary word, the corresponding word category scale is incremented. After processing all the words in a text file, LIWC calculates the percentage of words for each category. Categories in the dictionary are generally organized hierarchically (see Table 1): for example, all words included in the category "Positive emotions" are also included in the overarching category "Affective processes", which is the sum of "Positive emotions" and "Negative emotions". One exception is "Social processes", which includes verbs suggesting human interactions as well as non-first-person-singular personal pronouns [36]. The simplicity of LIWC's processing strategy was found to perform well in general, but the fact that the program allows for minimal contextual analysis remains a limitation, making it difficult to distinguish between different meanings of the same word [12]. Indeed, while language dimensions are often straightforward, other psychological categories are more subjective and context dependent. Bantum and Owen [6] analyzed emotional expressions in a web-based intervention for women with breast cancer using rater coding, LIWC and another computerized content analysis program called PCAD (Psychiatric Content Analysis and Diagnosis). Although both programs were found to over-identify emotional expression, authors suggested that LIWC performed better that PCAD for identification of emotions in a text. Specifically, LIWC showed a stronger performance in general emotion categories (i.e., affective processes) than in specific types of emotions (i.e., anxiety), and was found to be 12 times more likely to over-identify than to under-identify emotional words.

For example, some words that were frequently coded as emotion by LIWC but not by raters were "good", "hope", "beautiful", "best", and "like". Nevertheless, LIWC allows for limited contextual analysis for particular words, such as "like" and "kind". For example, the word "like" is assigned to the category "affective processes" and "positive emotions" if it is preceded by a pronoun or by a word indicating "discrepancy" (i.e., "would"), otherwise it is categorized as a "filler" (i.e., "youknow"). Although LIWC's accuracy could be certainly improved employing more sophisticated computational strategies for word disambiguation [2, 54], it has proven to perform well for emotion detection, and its content and construct validity was successfully assessed by Pennebaker and colleagues [35; 36]. Moreover, Bantum and Owen found LIWC to have a good convergent and discriminant validity for the analysis of emotional content in a text [6].

Cohn and colleagues [13] used LIWC to analyze blog posts during two months prior to and after the September 11, 2001 attacks, and found signs of psychological changes in the language used by bloggers. Specifically, these changes consisted mainly in an increase of words associated to negative emotions, cognitive processing, social engagement, and psychological distancing from the event during the first days following the attacks. Back and colleagues [5, 4] employed LIWC to analyze the use of emotional words in messages sent to text pagers in US after the September 11, 2001 attacks, and found an increase in language expressing anger in the first hours after the crashes. Keegan [26] selected 127 Wikipedia articles listed under "List of accidents and incidents involving commercial aircraft", classifying them as breaking articles, started within 48 hours of the crash, and non-breaking articles, written more that 2 days after the accident. In order to understand how editors may embody their psychological states in the articles, he processed every revision with LIWC and analyzed the variations in the use of words with emotional valence, of causal, insightful, tentative and discrepant language, and in the use of narrative immediacy (present vs past tense) as articles stabilized over time. The author found an increase in the emotional valence of articles, a decrease in the use of causal language, and a non significant decrease in the narrative immediacy, showing that while breaking articles tend to become more positive over time, non-breaking articles become progressively more negative, and have an accelerating tendency to contain more past-tense that present-tense language over time.

In the following sections we apply LIWC to Wikipedia's articles, to assess whether it can be employed to successfully detect different psychological processes underlying different patterns of language, providing the basis for the implementation of automated content analysis techniques for the study of collective memory processes on Wikipedia.

2.1 Characteristics of traumatic and non traumatic events

As a first step toward the study of the representation of traumatic events in Wikipedia, we tested if LIWC can be applied to the English Wikipedia to detect meaningful differences in the use of words related to the main psychological categories between different kinds of pages. For this reason, we first compared the content of a sample of articles related to traumatic events (bombings, assassinations, earthquakes, etc.) to a sample of articles related to non traumatic events (such as music events, royal weddings, sport competitions).

Past research on the psychological consequences of traumatic events found a temporary growth in the negative mood after an emotional upheaval [27, 49], an increase in cognitive processing as sensemaking activities take place and people try to comprehend and eventually find a meaning in what happened [14, 38], and an increase in social sharing and social interactions [30, 43, 45]. In this paper we focus particularly on affective, cognitive and social processes, which are represented in LIWC by the corresponding categories and subcategories reported in Table 1. Specifically, in the comparison between articles about traumatic and non traumatic events, we expect to find higher amounts of words related to negative emotions, cognitive and social processes in the former, while we expect to find a higher presence of language related to positive emotions in the latter.

Using Wikipedia's internal categories such as "Events by topic", Britannica Online Encyclopedia [11], History Central [24], and Information Britain [25] we identified 66 articles related to traumatic events, such as "September 11 attacks" or "2004 Indian Ocean earthquake and tsunami". Through Wikipedia's internal categories such as "Sports events", "List of most watched sporting events in 2004", "Largest concerts ever", "Music events" we identified 40 articles about non traumatic events, like "Coronation of Queen Elizabeth II" or "2010 FIFA World Cup". The complete list of articles, along with their LIWC scores is available at http://sonetlab.fbk.eu/data.

We applied LIWC to the text of each article to get a score for and the psychological categories of theoretical interest (Table 1), and computed all measures as percentages of words in relation to the total number of words contained in each text.
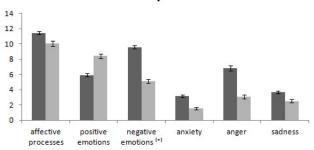
Table 1. LIWC psychological categories considered for analysis and their hierarchical categorization.

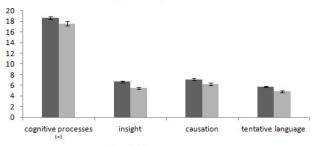| Psychological processes	Examples
Social processes	Mate, guy, boy
Family	Daughter, brother, dad
Friends	Buddy, friend, mate
Humans	Adult, children, girl
Affective processes	Happy, hate, kiss
Positive emotions	Love, party, pleasant
Negative emotions	Hurt, abuse, scary
Anxiety	Worried, afraid, apprehensive
Anger	Kill, aggression, destroy
Sadness	Sad, cry, depression
Cognitive processes	Cause, acknowledge, admit
Insight	Think, assume, interpret
Causation	Because, depend, elicit
Discrepancy	Should, could, if
Tentative	Maybe, apparently, suppose
Certainty	Always, absolutely, clear
Inhibition	Block, abstain, avoid
Inclusive	And, add, along
Exclusive	But, either, without

We applied the arcsine transformation, a linear combination traditionally employed in general linear modeling to analyze a dependent variable when the raw values are proportions or percentages, which consists in taking the arcsine of the square root of a number and transforming the original data in radians [47, 32]. Being a linear combination of variables, the test statistics and their probabilities remain unchanged.

After assessing for the assumption of normality with the Kolmogorov-Smirnov test, we applied a series of independent samples t-tests to compare the scores of the different psychological variables between articles related to traumatic and non traumatic events. Where the assumption of normality was violated, we applied the Mann-Whitney U non parametric test (Figure 1).
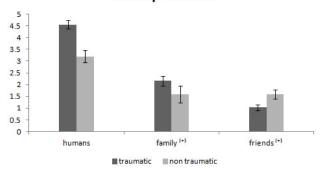

Figure 1. Arcsine percentage of words related to different psychological categories for articles about traumatic and non traumatic events. The symbol $^{(+)}$ indicates variables for which the assumption of normality had been violated, and whose scores were compared through Mann-Whitney U non parametrical test. All differences are statistically significant.

In general, the results showed a significantly higher presence of language related to affective processes ($t_{(104)}=3.90$, $p<.001$),

negative emotions ($U=64$, $p<.001$), cognitive ($U=983.5$, $p=.028$) and social processes for articles about traumatic events. Figure 1 shows the results for the overarching categories and the subcategories of interest. Specifically, the relative number of words expressing anxiety, anger and sadness (e.g. "worried", "hate", "cry") was significantly higher in articles about traumatic events (respectively, $t_{(104)}=6.47$, $p<.001$; $t_{(102.96)}=9.44$, $p<.001$; $t_{(104)}=3.94$, $p<.001$), while language associated with positive emotions was significantly higher in articles about non traumatic events ($t_{(68.56)}=-6.76$, $p<.001$). The amount of words related to cognitive activity, such as those expressing insight, causation and tentative language (e.g. "think", "because", "maybe") was found to be significantly higher in articles about traumatic events (respectively, $t_{(104)}=4.73$, $p<.001$; $t_{(104)}=2.79$, $p=.006$; $t_{(104)}=3.58$, $p=.001$), confirming the presence of sensemaking activities when creating the collective memory of emotional upheavals [38]. With regard to social processes, the t-tests showed a significantly higher amount of words expressing references to other people in general, such as humans (e.g. "adult") in articles about traumatic events ($t_{(104)}=4.38$, $p<.001$). In the same direction, the Mann-Whitney U test showed a significantly higher presence of references to family members (e.g. "daughter), with $U=971.5$, $p=.021$. On the contrary, probably reflecting a more relaxed content, references to friends (e.g. "buddy") were found to be significantly higher in articles about non traumatic events ($U=935.5$, $p=.011$). All other differences for the remaining psychological categories reported in Table 1 were not statistically significant.

2.2 Temporal focus of recent and old traumatic events

Given their encyclopedic nature, Wikipedia's articles about historical events should be generally written in the past tense, and breaking news should not be treated differently from other information or written in news style [55, 56]. Notwithstanding, in Wikipedia, pages about recent traumatic events tend to get created just few hours or days after their happening [16, 26]. However, in articles related to breaking news, a higher use of present tense rather than past tense verbs might reflect that sensemaking processes are ongoing and editors are still trying to understand the facts, organize the sourcing and interpret the events [26]. Indeed, studying the tense of common verbs employed in Wikipedia's articles can tell us more about the temporal focus of their editors [51].

In this second phase, we focused specifically on traumatic events by analyzing the differences between those happened before the founding of Wikipedia in 2001 (old traumatic events) and those occurred after this threshold (recent traumatic events). Since we restricted our analysis to traumatic events, the differences discernible from the text alone were smaller and more nuanced, and therefore we decided to analyze the articles as they were at their 500th edit. This threshold allowed us to restrict our dataset to a group of articles with a sufficiently large amount of text, without removing too many pages. At the same time, by focusing on early versions, the process of collective memory creation was still in its beginning and so the presence of psychological traits in the texts could be larger. Later on, it is possible that with time, after hundreds of additional edits, the encyclopedic nature of Wikipedia would reduce the presence of psychological traits in texts. Requiring to have at least 500 edits reduced our dataset of traumatic events to 55 Wikipedia articles, 26 of which were old

traumatic events and 29 were recent accidents and disasters (the complete list is released as well at http://sonetlab.fbk.eu/data).

As in the previous analysis, we applied LIWC to each text, but in this case we focused on the temporal dimension of articles and hence on the LIWC scores for linguistic categories about the tenses of common verbs, computing these measures as percentages of words in relation to the total number of words contained in each text. We applied the arcsine transformation, and after testing for normality with the Kolmogorov-Smirnov test, we applied independent samples t-tests to compare the presence of past, present and future tenses in the articles about recent and old traumatic events (Figure 2).
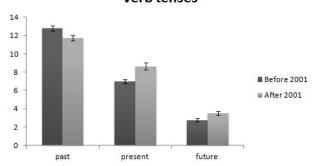
Verb tenses

Figure 2. Arcsine percentage of past, present and future tense verbs in articles about recent and old traumatic events. All differences are statistically significant.

The graphs in Figure 2 show a significantly higher presence of past tense verbs in articles about traumatic events happened before 2001 ($t_{(53)}$=2.59, p=0.12), and a significantly higher amount of present and future tense verbs in articles related to more recent events (respectively, $t_{(43.67)}$=-3.64, p=.001; and $t_{(53)}$=-2.56, p=.013).

In summary, our results actually confirm that LIWC has the ability to successfully detect different types of psychological processes in articles about traumatic and non traumatic events, and a different use of common verb tenses in articles about recent and old traumatic events. This confirms LIWC as a suitable and effective tool for the automated analysis of Wikipedia's articles, allowing us to go one step further in the study of the collective representation of man-made and natural disasters in Wikipedia, as explained in the next section.

3. REPRESENTATION OF NATURAL AND HUMAN-MADE TRAUMATIC EVENTS

Clearly, understanding how people react to collective emotional upheavals is crucial to a better comprehension of the consequences on the physical and psychological health of many involved communities. In fact, for a long time a great amount of past research has been focusing on the outcomes of particular disasters. For instance, Adler's study of the effects of the 1942 Cocoanut Grove fire [1] showed a prolonged persistence of nervousness and anxiety up to 11 months after the fire, while Green [21] found evidence of emotional problems up to 15 months after another nightclub fire at Beverly Hills. Other studies investigated longer-term effects of traumatizing events and found an increase of psychiatric, psychological, and work-related

problems up to several years after a marine explosion, and other psychological problems up to 10 years after a mine cave-in [29, 23, 41]. In their investigation on the consequences of the dam collapse and flood at Buffalo Creek in West Virginia, Titchener and Kapp [52] reported high rates of emotional problems like anxiety, depression and personality changes, while other studies showed also evidence of hostility, in addition to sleep disturbances and psychiatric problems up to more than 2 years after the accident [19, 18]. Pennebaker and Harber [37] studied the aftermath of the Loma Prieta Erthquake, which shocked the San Francisco Bay area in 1989, and found an increase in self-reported illness episodes, quake-related dreams, arguments with family members and co-workers, and also aggravated assaults, in the first weeks following the earthquake.

As stated before, earlier research showed that immediately after a traumatic event nearby residents tend to talk more about the accident, and this increased frequency of social sharing can provide a basis for the future construction of collective memories [37, 33]. With time these collective memories, conceived starting from the immediate responses to disasters, will influence the cultural attitudes toward the current society, directing present and future behavior to ensure the social cohesion and the continuity of the community, for instance engendering protective responses after a threatening traumatic event [39; 44, 43]. But what kind of memories are these? Is it possible to detect meaningful characteristics of traumatizing events associated to different psychological responses in the aftermath of these upheavals and different types of collective narratives in Wikipedia?

Clearly, traumatic events can be studied from several perspectives and levels of analysis. Disasters can highly differ in nature: they can be meteorological or climatic events, technological mishaps, biological hazards, political or economic crises, all posing potentially serious consequences on both the social and psychological levels, like collective migration, social disruption, loss of property or life, trauma, distress and shock. Of course, this variability makes disasters hard to define. In fact, it is the precise combination of different characteristics – like the nature of the event, its impact and the victims' responses – that makes each disaster unique.

In this diversified framework, identifying the conditions leading to particular effects and social responses becomes crucial to understand the various psychological implications of traumatic events, and eventually to provide a basis for the prediction of short and long term consequences [7]. The analysis of collective emotional upheavals in terms of their characteristics, such as natural or man-made traumatic events, can provide researchers with additional tools to predict the extent of trauma, the consequences on individuals' psychological health and on communities' social cohesion.

Of course, there are numerous variables mediating disasters' responses, and this makes it is difficult to compare natural and human-made disasters, partly because each emotional upheaval is unique, and so are the subsequent consequences [20]. It is challenging to assess common outcomes of disasters of a certain kind because there is a considerable variability in the nature of the physical agent, its impact and the emotional responses it causes. Each individual can react in a different way, the exposure to life-threatening situations may vary among the population, and the recovery environments can be very dissimilar. Notwithstanding,

there are theoretical reasons for hypothesizing different outcomes of natural and man-made catastrophes [20].

Andrew Baum [7], in reviewing previous research on natural and man-made disasters, proposed that accidents caused by man may have more important effects in terms of longevity and severity. Starting from the analysis of the accident at Three Mile Island, which despite causing little physical damage to the population had a long aftermath of more than three years, Baum and colleagues [8] proposed a distinction between natural and technological disasters in terms of their characteristics:

- Suddenness: although there is still variability in the suddenness of natural and man-made disasters, in general both can have a rapid onset. While a tornado or a storm can take days to form, an earthquake can give little warning to the population. Industrial explosions or air crashes are usually sudden and unexpected. There are also human-made disasters with a slow onset, like heat waves, the poisoning of the oceans or economic crisis, but usually they are not considered as traumatic events.

- Power: both natural and technological disasters can be highly powerful and threatening, causing death and destruction.

- Visible damage: while generally natural disasters cause damage to the environment and destruction of properties, this is not always true for technological disasters, as the Three Mile Island or the Chernobyl accidents, where invisible damage is related to illness in the long-term.

- Predictability: thanks to the improvement of our forecasting abilities, some natural disasters can now be prevented to a certain extent (e.g., storms or tornadoes). On the contrary, technological disasters are not predictable. They are also usually sudden, leaving little or no time for evacuation.

- Low point: natural disasters usually have a clear and identifiable low point, when the worst has already happened and after which the focus can be moved to recovery efforts; for some technological catastrophes, like those involving radiation or toxic leaks, it is more difficult to identify a specific low point.

- Perception of control: natural disasters are usually perceived as uncontrollable. There may be political controversies related to the management of the disaster, but they mainly depend on our inability to control the natural elements. On the contrary, technology itself is the manifestation of human control over the environment, so when it fails, the disaster is perceived as a loss of control. Losing control, meaning not having control when you have expectations for it, seems to cause different psycho-physiological consequences than does not having control when you do not expect to have it [10, 58]. Baum and colleagues [9] showed that in the former case the loss of control can be related to stress arousal, while in the latter, when control is not expected, there is more helplessness and passive behavior.

- Extent of effects: for natural disasters it is usually bounded to the people directly involved and to a limited area around the accident; for technological disasters, the effects may be broader and involve also people not directly affected by the accident, due to a general loss of confidence and credibility.

- Persistence of effects: according to Baum and colleagues [8], it seems that for natural disasters the effects may be limited to the short-term, even if their extent depends upon the individual experience of the trauma (i.e. loss of property or of loved ones); with regard to technological disasters, the effects seem to be more prolonged in time, particularly for toxic accidents.

In this article we compare the collective representations of natural and human-made traumatic events analyzing the content of the related English Wikipedia's articles, and particularly the language associated with affective, cognitive and social processes. We have already shown that, in general, articles related to traumatic events contain higher amounts of words associated to negative emotions, cognitive and social processes. In the comparison between articles about natural and man-made disasters, drawing from Baum's theoretical distinctions [8], we expect a distinctive presence of words related to anger, anxiety and sadness. Since natural disasters are generally perceived as uncontrollable and may engender more passive responses, we expect a higher presence of language related to sadness in the associated articles. On the contrary, in case of human-made disasters, the sudden loss of control may provoke more focused anger and anxiety, because technology is not supposed to fail and these kind of accidents can, in theory, be prevented. Moreover, especially in case of terrorist accidents, subsequent reactions may find expression in anger and blame because the causes can be directly found in human agencies.
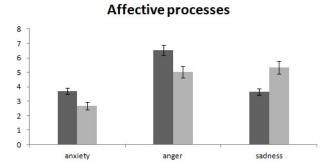
With regard to cognitive processes, although the search for causes and accountability is a typical consequence also of natural disasters [44], we expect a higher presence of language related to cognitive processes in articles about human-made disasters, because in this case the loss of control is unexpected and may require more exhaustive and comprehensive explanations.

We also expect a higher amount of words related to social processes in articles about man-made traumatic events, partly because this kind of collective upheaval may affect more deeply people's social lives, triggering an increased orientation toward others, and partly because being their origin inherently human, there may be more references to other people.
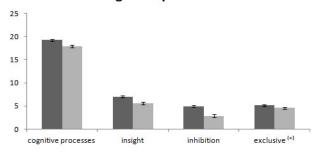
We classified the 55 articles in our dataset as natural or human-made traumatic events, exploiting the categorizations made by the users through Wikipedia's internal lists and categories, like "List of natural disasters by death toll", "2004 natural disasters in the United States", "Man-made disasters", or "Terrorist incidents". Articles belonging to "Man-made disasters", "Mass murder", "Assassinations", and other similar categories were classified as human-made disasters, while articles assigned by Wikipedia users to categories like "1993 natural disasters" or "2008 Atlantic hurricane season" were classified as natural disasters (the complete list of articles with their related categories is released at http://sonetlab.fbk.eu/data). Again, we considered the versions of the texts after 500 revisions, in order to analyze the articles in their early stages. Out of 55 articles, 19 were about natural disasters and 36 were about human-made traumatic events.

We applied LIWC to each article to get a score for the different psychological categories reported in Table 1 and computed all measures as percentages of words in relation to the total number of words contained in each text. We applied the arcsine transformation, and tested for the assumption of normality with the Kolmogorov-Smirnov test: where it was not violated we applied independent samples t-tests to compare the presence of words related to different categories in the articles about natural

or man-made disasters, otherwise we applied the Mann-Whitney U non parametric test (Figure 3).
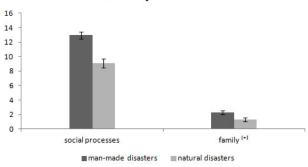

Figure 3. Arcsine percentage of words related to different psychological categories for articles about man-made and natural disasters. The symbol (+) indicates variables for which the assumption of normality had been violated, and which scores were compared through Mann-Whitney U non parametrical test. All differences are statistically significant.

The results reported in Figure 3 confirm our expectations, showing in general significantly higher amounts of emotional language related to anger and anxiety, cognitive and social processes for articles about man-made traumatic events. In particular, t-tests results showed that words expressing anxiety and anger (e.g., "worried", "hate") were significantly higher in articles related to human-made disasters (respectively, $t_{(53)}=2.79$, p=.007; $t_{(53)}=2.63$, p=.011), while on the other hand the amount of words associated to sadness (e.g., "cry") was significantly higher in articles about natural disasters ($t_{(53)}=-3.812$, p<.001). The language expressing cognitive activity, insight (e.g., "because", "think"), inhibition (e.g., "prevent", "forget") and exclusion (e.g., "but", "either") was, as expected, more present in articles about

man-made traumatic events (respectively, $t_{(53)}=3.47$, p=.001; $t_{(53)}=4.51$, p<.001; $t_{(53)}=5.57$, p<.001; U=226.5, p=.041). Moreover, the amount of words referring to social processes and to family (e.g., "they", "son") was significantly higher in articles related to human-made disasters (respectively, $t_{(53)}=4.79$, p<.001; U=211, p=.020).

In summary, our analyses confirmed that natural and man-made disasters are represented with different language patterns in Wikipedia, suggesting that distinctive psychological and sensemaking processes may underlie users' collaboration to the editing of these articles. In particular, articles about human-made and natural traumatic events seem to be characterized by specific patterns of emotional language, which in the frame of a dimensional approach to emotions [46, 28] have a similar negative valence, but opposite levels of arousal. The language patterns associated to different affective processes in these articles seem to confirm that traumatic events caused by human agencies may be more stressful, trigger more anxiety and focused anger than natural disasters, which instead may be characterized by more passive responses. Although the urge to understand the causes and find a convincing explanation is crucial for both these types of disasters, it may be even more critical for man-made traumatic events, where the loss of control is perceived as a serious damage to the community and the need to blame somebody for failure is more pressing. Wikipedia's articles about human-made disasters are also characterized by a higher presence of words reflecting social processes. This, along with previous research showing an increase in social sharing and social interactions after traumatizing events [37, 30, 45], may suggest more insidious and profound effects of disasters caused by human agencies, which may lead to a language expressing more references to other people and family members. However, to this end it is difficult to draw decisive conclusions, given that in case of man-made disasters the causes are inherently human, and therefore the related articles may contain more references to a killer's social relations or a bomber's family.

Nevertheless, it is important to consider that Wikipedia is intended to be an online encyclopedia, and to express a neutral perspective on events, without biases or emotional content (NPOV, neutral point of view; 51). Still, the fact that such peculiarities in the use of language emerged from articles about natural and man-made disasters are symptomatic of clearly different psychological processes underlying users' collaboration.

4. CONCLUSIONS

Collective memories are precious resources for the society and serve many purposes, like maintaining groups cohesion, strengthening emotional bonding between members, directing future behavior or even mobilizing collective action and leading to social and political changes [53]. Understanding how they are formed through the public discourse is crucial to a better comprehension of the current society. To this regard, it seems even more important to understand how people react to traumatic events, because the way they engage in sensemaking activities, interpret and organize the facts in their memory will affect future beliefs, values and possibly lead to cultural and political changes. However, few empirical studies have attempted to investigate how communities form their collective memories on a large scale [34].

In this paper we employed automated content analysis techniques to compare articles related to natural and human-made disasters,

showing that different patterns of language characterize these articles. The words employed to describe particular events in Wikipedia reflect users' internal thoughts, emotions and sensemaking activities, and can tell us about the psychological processes at the basis of users' collaboration to articles' editing. In particular, articles about traumatic events caused by human agencies are characterized by an emotional language expressing anxiety and anger, and by a greater presence of words referring to cognitive activity and social processes, with respect to articles about natural disasters. This suggests that even in Wikipedia, where articles are supposed to be written without biases and from a neutral point of view, the collective representation of different types of traumatic events shows diverse psychological processes. Understanding how different characteristics of traumatic events, such as natural or human-made disasters, are related to particular psychological processes can provide a basis for the prediction of typical social responses, and short and long term effects on individuals and communities.

Confirming the theoretical characterization of natural and technological disasters proposed by Baum [8], our analysis showed that man-made traumatic events may cause more insidious effects. Although it seems reasonable to find expressions of anger in association to any loss, human-made disasters may engender more anxiety and focused anger because in this case there is someone to blame, the accident could have been prevented and the victims could have been saved. For similar reasons, although the need to comprehend the causes and to find a convincing explanation is understandably present in the aftermath of every traumatizing experience, it may be even stronger where human agencies are responsible. A slightly different explanation could be advanced with regard to the higher presence of words expressing social processes in articles related to man-made disasters. On the one hand, along with previous research findings showing an increase in social sharing and interactions after emotional upheavals, this result suggests that in case of man-made disasters the traumatizing experience may be more insidious, resulting in a language expressing more references to other people and family members. On the other hand, however, such conclusion may be deceiving, given that articles about these kinds of events may simply contain more references to the human agencies at the origin of the disasters.

Indeed, as Pury intelligibly showed [see 5; 42; 4], automated text analysis tools should be carefully employed when applied to large datasets, especially if they contain automatically generated messages, in order to prevent confounds and hazardous conclusions. Notwithstanding, the massive availability of digital data offers an unprecedented opportunity, if exploited with caution, to access people's thoughts and feelings, allowing scholars to empirically investigate social theories on trauma in ways that would have been difficult just few years ago.

In particular, Wikipedia seems particularly appropriated for the study of the psychological processes underlying the formation of shared narratives about different kinds of traumatic events from two perspectives. On the one hand, the immediate availability of large amounts of data and the fact that the Mediawiki open source web platform powering Wikipedia records any change made by any user to any article or talk page, allow researchers to study the immediate aftermath of single traumatic events. In this way, scholars can assess unobtrusively the presence of different coping strategies, the extent to which various psychological processes are represented in articles and talk pages in the immediate aftermath

of an upheaval, and how they may vary in time. On the other hand, the study of how collective memories of traumatic events are formed in Wikipedia through debates and discussions, and eventually represented in shared narratives can provide a further high-level perspective on the collective remembrance of emotional upheavals. Collective memories remain alive for years and generations and can influence not only the remembrance of the past, but also attitudes toward current society [39]. Especially in events of trauma, the study of collective memories can provide insights on current desires, needs and identities of the society, and help to understand how the past can affect the present [50].

In this paper, employing automated content analysis techniques to analyze the language used in the English Wikipedia's articles about traumatic events, we showed that the shared representation of man-made and natural disasters reveal different kind of underlying psychological processes. Our results suggest that these traumatic events may engender diverse psychological reactions in individuals and communities, possibly requiring different recovery strategy.

The use of automated content analysis techniques, especially if extended to different online environments, could be useful to determine unobtrusively the extent of distress engendered by collective emotional upheavals and the psychological processes involved in the immediate aftermath and in the construction of the collective representations of these events.

5. REFERENCES

[1] Adler, A. 1943. Neuropsychiatric complications in victims of Boston's Cocoanut Grove disaster. *Journal of American Medical Association*, 123, 1098-1011.

[2] Agirre, E., and Edmonds, P. 2006. *Word sense disambiguation*. Dordrecht, The Netherlands: Springer.

[3] Alexander, J. C. 2004. *Toward a Theory of Cultural Trauma*. In Cultural Trauma and Collective Identity, J. C. Alexander, R. Eyerman, B. Giesen, N. J. Smelser, and P. Sztompka, Eds. University of California Press, Berkeley, 1-29.

[4] Back, M. D., Küfner, A. C. P., and Egloff, B. 2011. "Automatic or the people?": Anger on September 11, 2001, and lessons learned for the Analysis of large digital data sets. *Psychological Science Online*, 22, 837-838.

[5] Back, M. D., Küfner, A. C. P., and Egloff, B. 2010. The emotional timeline of September 11, 2001. *Psychological Science*, 21, 1417–1419.

[6] Bantum, E., and Owen, J. 2009. Evaluating the validity of computerized content analysis programs for identification of emotional expression in cancer narratives. *Psychological Assessment*, 21, 1, 79-88.

[7] Baum, A. 1987. Toxins, technology, and natural disasters. In. *Cataclysms, crises, and catastrophes*, G. R. Van den Bos, B. K. Bryant, Eds., American Psychological Association, Washington, DC, 5-54.

[8] Baum, A., Fleming, R., and Davidson, L. M. 1983. Natural disaster and technological catastrophe. *Environment and Behavior*, 15, 3, 333-354.

[9] Baum, A., Fleming, R., Reddy, D. M. 1986. Unemployment stress: Loss of control, reactance and learned helplessness. *Social Science & Medicine*, 22, 5, 509-516.

[10] Baum, A., and Gatchel, R. J. 1981. Cognitive determinants of reaction to uncontrollable events: Development of reactance and learned helplessness. *Journal of Personality and Social Psychology*, 40, 1078-1089.

[11] *Britannica Online Encyclopedia*. 2011. Britannica Online Encyclopedia. Availble at http://www.britannica.com/

[12] Chung, C. K., and Pennebaker, J. W. 2007. The psychological function of function words. In *Social communication: Frontiers of social psychology*, K. Fiedler, Ed. Psychology Press, New York, 343-359.

[13] Cohn, M. A., Mehl, M. R., and Pennebaker, J. W. 2004. Linguistic Markers of Psychological Change Surrounding September 11, 2001. *Psychological Science*, 15, 10, 687-693.

[14] Davis, C. G., and Nolen-Hoeksema, S. 2001. Loss and meaning: How do people make sense of loss? *American Behavioral Scientist*, 44, 726–741.

[15] Ferron, M., and Massa, P. 2011. Collective memory building in Wikipedia: the case of North African uprisings. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (Mountain View, California, 3rd October 03-05, 2011). Wikisym 2011. ACM, New York, NY.

[16] Ferron, M., and Massa, P. 2011. Studying collective memories in Wikipedia. *3rd Digital Memories Conference*, Prague, March 14-16, 2011. Available at http://www.inter-disciplinary.net/wp-content/uploads/2011/02/ferrondmpaper.pdf

[17] Garde-Hanse, J., Hoskins, A., and Reading, A. 2009. *Save As...Digital Memories*. Palgrave Macmillan, Basingstoke.

[18] Gleser, G. C., Green, B. L., and Winget, C. 1981. *Prolonged Psychosocial Effects of Disaster: A Study of Buffalo Creek*. Academic Press, New York.

[19] Gleser, G. C, Green, B. L., and Winget, C. N. 1978. Quantifying interview data on psychic impairment of disaster survivors. *Journal of Nervous and Mental Disease*, 166, 209-216.

[20] Green, B. L. 1996. Traumatic stress and disaster: mental health effects and factors influencing adaptation. In *International review of psychiatry*, vol. 2, E. Liehman, and C. Nadelson, Eds., American Psychiatric Press, Washington, DC, 177-210.

[21] Green, B. L. 1980. *Prediction of long-term psycho-social functioning following the Beverly Hills fire*. Doctoral thesis, University of Cincinnati.

[22] Halbwachs, M. 1992. *On Collective Memory*. University of Chicago Press, Chicago.

[23] Henderson, S., and Bostock, T. 1977. Coping behavior after shipwreck. *British Journal of Psychiatry*, 131, 15-20.

[24] *History Central*. 2011. History Central. Available at http://www.historycentral.com/

[25] *Information Britain*. 2011. Information Britain. Available at http://www.information-britain.co.uk/

[26] Keegan, B. 2011. Breaking News, Breaking Planes, and Breaking Hearts: Psycholinguistics and Sensemaking in Collaborative Accounts of Catastrophe. *International Communication Association*, Boston, MA.

[27] Koss, M.P., and Kilpatrick, D.G. 2001. Rape and sexual assault. In *The mental health consequences of torture*, E. Gerrity, T. M. Keane, and F. Tuma, Eds., Kluwer Academic/Plenum, New York, 177-193.

[28] Lang, P. J., Greenwald, M. K., Bradley, M. M., and Hamm, A. O. 1993. Looking at pictures: affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30, 3, 261-273.

[29] Leopold, L. R., and Dillon, H. 1963. Psycho-anatomy of a disaster: a long term study of post-traumatic neurosis in survivors of a marine explosion. *American Journal of Psychiatry*, 119, 913-921.

[30] Mehl, M. R., and Pennebaker, J. W. 2003. The social dynamics of a cultural upheaval: Social interactions surrounding September 11, 2001. *Psychological Science*, 14, 579-585.

[31] Olick, J. K., and Levy, D. 1997. Collective memory and cultural constraint: Holocaust myth and rationality in German politics. *American Sociological Review*, 62, 6, 921-936.

[32] Osborne, J. W. 2010. Improving your data transformations: applying the Box-Cox transformation. *Practical Assessment, Research & Evaluation*, 15, 12, 1-9.

[33] Pennebaker, J. W. 1997. *Opening up: The healing power of emotional expression*. Guilford, New York.

[34] Pennebaker, J.W., and Banasik, B. 1997. On the creation and maintenance of collective memories: History as social psychology. In *Collective memory of political events: Social psychological perspectives*, J.W. Pennebaker, D. Paez, and B. Rime, Eds. Erlbaum Publishers, Mahwah, NJ.

[35] Pennebaker, J.W., and Francis, M.E. 1996. Cognitive, emotional, and language processes in disclosure. *Cognition and Emotion*, 10, 601-626.

[36] Pennebaker, J. W., Francis, M. E., and Booth, R. J. 2001. *Linguistic inquiry and word count (LIWC): LIWC 2001*. Lawrence Erlbaum Associates, Mahwah.

[37] Pennebaker, J. W., and Harber, K. 1993. A social stage model of collective coping: The Loma Prieta earthquake and the Persian Gulf War. *Journal of Social lssues*, 49, 125-145.

[38] Pennebaker, J. W., Mehl, M. R., and Niederhoffer, K. G. 2003. Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54, 547-577.

[39] Pennebaker, J. W., Paez, D., and Rimé, B. 1997. *Collective Memory of Political Events: Social Psychological Perspectives*. Lawrence Erlbaum Associates, Mahwah.

[40] Pentzold, C. 2009. Fixing the floating gap: The online encyclopaedia Wikipedia as a global memory place. *Memory Studies*, 2, 2, 255-272.

[41] Ploeger, A. 1972. A 10-year follow up of miners trapped for two weeks under threatening circumstances. In *Stress and Anxiety*, C. D. Spielberger, and I. G. Sarason, Eds., Hemisphere, Washington, DC.

[42] Pury, C. L. S. 2011. Automation can lead to confounds in text analysis: Back, Küfner, and Egloff (2010) and the not-so-angry Americans. *Psychological Science*, 22.

[43] Pyszczynski, T. A., Solomon, S., and Greenberg, J. 2002. *In the wake of 9/11: The psychology of terror*. American Psychological Association, Washington, DC.

[44] Reser, J. P. 2004. The experience of natural disasters: psychological perspectives and understanding. In *International Perspectives on Natural Disasters: Occurrence, Mitigation, and Consequences*, J. P. Stoltman, J. Lidstone, and L. M. DeChano, Eds. Kluwer Academic Publishers, Dordrecht, 369-384.

[45] Rimé , B., Finkenauer, C., Luminet, O., Zech, E., and Philippot, P. 1998. Social sharing of emotion: New evidence and new questions. In *European review of social psychology*, vol. 9, W. Stroebe, and M. Hewstone, Eds.Whiley, Chichester, England, 145–189.

[46] Russell, J., 1980. A circumplex model of affect. *Journal of Personal and Social Psychology*, 39, 1161-1178.

[47] Sokal, R. R., and Rohlf, F. J. 1981. *Biometry*, 2nd Ed. W.H. Freeman, New York.

[48] Stone, A. A., Turkkan, J. S., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., and Cain, V.S. 2000. *The science of self-report: Implications for research and practice*. Erlbaum, Mahwah, NJ.

[49] Stroebe, M. S., Hansson, R. O., Stroebe, W., and Schut, H. 2001. *Handbook of bereavement research: Consequences, coping, and care*. American Psychological Association, Washington, DC.

[50] Sturken, M. 1997. *Tangled Memories: The Vietnam War, the AIDS Epidemic, and the Politics of Remembering*. University of California Press, London.

[51] Tausczik, Y., and Pennebaker, J. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29, 1, 24.

[52] Titchener, J. L., and Kapp, F. T. 1976. Family and character change at Buffalo Creek. *American Journal of Psychiatry*, 133, 295-299.

[53] Wang, Q. 2008. On the cultural constitution of collective memory. *Memory*, 16, 3, 305-317.

[54] Weik, M. 1996. *Communications standard dictionary*. Chapman & Hall, New York.

[55] Wikipedia: The free encyclopedia. 2012. *Wikipedia: Writing better articles*. Available at http://en.wikipedia.org/wiki/Wikipedia:1SP

[56] Wikipedia: The free encyclopedia. 2012. *Wikipedia: What Wikipedia is not*. Available at http://en.wikipedia.org/wiki/Wikipedia:What_Wikipedia_is_not

[57] Wikipedia: The free encyclopedia. 2012. *Wikipedia: Neutral point of view*. Available at http://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view

[58] Wortman, C. B., and & Brehm, J. W. 1975. Responses to uncontrollable outcomes: An integration of reactance theory and the learned helplessness model. In *Advances in Experimental Social Psychology*, L. Berkowitz, Ed., Academic, New York, 278-336.