Nouns and Verbs in Professional Reporting of Extreme Events

Anna Karampela
Trinity Centre for Computing and Language Studies
Trinity College Dublin, the University of Dublin
Dublin, 2, Ireland
Email: karampea@tcd.ie

Carl Vogel
Trinity Centre for Computing and Language Studies
Trinity College Dublin, the University of Dublin
Dublin, 2, Ireland
Email: vogel@tcd.ie

Abstract—This corpus-based study tested whether linguistic devices that can evoke negative emotions are featured in Terrorist Attack news articles to a greater extent, in comparison to articles on Natural Disasters and Human Accidents. A total of 210 news reports were analysed in terms of linguistic categories. The results showed that news reports on Terrorist Attacks display the distinctive patterns of use in the distribution of nouns and verbs than the other two categories.

I. INTRODUCTION

The public can be significantly affected emotionally by events that feature numerous casualties and extend nationwide and worldwide. It is sometimes claimed that when a community is stricken by a tragic event, collective shock and grief are experienced [1, p.581]. In such unfortunate incidents, the press generally provides extensive and somewhat diverse media coverage. News articles describing events can impact the reader depending on the way the facts are presented and the linguistic devices that the reporter chooses. As has been argued [2], the syntactic structures of news texts, along with the lexical choices, appear to be of great importance in directing thoughts and emotions in relation to the underlying semantics.

The experience of shock and grief elicited by a tragic event can be perceived more intensely upon reading media coverage of it. It has been noted that “big breaking stories are literally exciting. They’re often about death or the threat of death, or they otherwise create anxiety. Neurological research shows that the more of your personal bandwidth anxiety takes up, the less clearly you think” [3, p.27]. Shock, for instance, might be turned into a state of panic in cases when an article covering the incident implicates further threat via certain syntax, memory triggering references and/or the use of purely negative-connotation vocabulary. The present work aims at examining aspects of linguistic choices (relative preponderance of nominal or verbal descriptions, which in turn, affects syntax) found in online news media platforms when covering incidents on terrorism, natural disasters and human accidents. Articles about terrorist attacks are found to have more verbs than one would expect if there were no interaction between topical event type and syntactic category choices.

A. Previous Analyses

Research on news perception has been dominated by reception processes that are cognitively oriented, while emotional effects have been neglected [4, p.1913]. Thus, specific features of a news story that might produce affective responses have remained open, while the focus on an emotional perspective has been shifted towards entertainment research [5], [6], [7]. In their attempt to direct the attention towards the emotional processes in news media research, [4] focused on suspense, which is defined [8, p.208] as “a noxious affective reaction that characteristically derives from the respondents’ acute, fearful apprehension about deplorable events that threaten liked protagonists, this apprehension being mediated by high but not complete subjective certainty about the occurrence of the anticipated deplorable events”.

Other studies that adopted an emotion-oriented viewpoint, focused on the impacts of emotionally intense images and videos along with the news articles. Some of those explored the prior-receptive phase of emotional news content [9], [10], while others examined the effect of emotionally strong news images during browsing [11]. The post-receptive phase has also been the focus of research [12], [13]. A similar approach was adopted in an endeavour to investigate the impact that the series of emotion-enhanced television news broadcasts has on the viewers’ evaluation of news media content [14]. Further research has shifted the focus towards the attributes of news articles and their effect on the recipients’ perception and emotional responses [15], [16]. Such studies concluded that the recipients’ emotional responses were modulated by news stories about traumatic events [17], such as terrorist attacks. The present work attempts to analyze how linguistic devices are deployed differently across content categories.

B. Aims of the present work

The main hypothesis of the present work is that articles about terrorist attacks employ more linguistic devices that could elicit and intensify emotions such as fear and anxiety. In order to examine this argument, syntactic choices devices will be analysed within each of the three small corpora.

The comparisons within and across the corpora are based on narrative emphasis on nominal and verbal categories. News media coverage is broadcast in order to inform or update the public regarding the progress of an event. Popular culture encourages the idea of a neutral press which approaches any incident with objectivity and distance on the reporter’s part, while a different approach might result in narration reminiscent of genres other than news reports, which could result in

1At times, however, homogeneity of coverage seems evident.
emotional triggers. As noted elsewhere [18], the representation of events in news articles is unnatural. For instance, some note a tendency in newspapers to favour compact nominals, for example in headlines, apparently compressing information into fewer words than otherwise. When such a tendency towards nominalization is not only avoided, but on the contrary, the use of verbs is significantly increased, a sense of narration is attributed to the event description. As [4, p.1913] argue, the enjoyment of a narrative and its immersive potential are important for entertainment, but we may question whether one can similarly enjoy a news story about negative events. In their study, they state that the attributes of a news story that may elicit certain emotional responses still remain open, as well as the emotional potential of the narrative. Thus, they focus on the emotional evocations of more narrative news stories. Our work adopts the notion that the emotional impacts of narrative could be perceived more intensely when the use of significantly more verbal in relation to nominal sequences is observed. We suggest that a corpus where the verb sequence (VS) frequency is significantly larger than that of noun sequences (NS), could trigger the recipients’ emotions in a more robust manner.

II. MATERIALS AND METHODS

A. Materials and sources

A total of 210 news articles were extracted from seven online news platforms. Ten distinct events were chosen for each of the three corpora on Terrorist Attacks, Natural Disasters and Human Accidents. To focus on news articles about events of similar significance, the event selection was based on the number of people affected. Terrorist and natural disasters articles involved dozens of people in all cases. The events were selected chronologically starting from 2016. The third category of events, accidents caused by humans, had as main criterion the involvement of at least five people, beginning from 2017.

The news sources selected are online news web-pages of popular institutions in English speaking countries and were constant among all three event categories. These were CNN, BBC, Fox News, The Guardian, Huffington Post, Washington Post and The Daily Mail. The selection intends to incorporate “broadsheet” and “tabloid” news treatments. The article selection was not based on an average text length, as an event presented by CNN might consist of 600 words, whereas the same incident presented by Washington Post might consist of 40. Therefore, each corpus features news articles of various word lengths. The event and word count for each of the three corpora is presented in Table I.2

| Table I. Event and word count of Terrorist, Natural Disaster and Human Accident corpora |
|---|---|---|
| Corpus                | Event Count | Word Count |
| Terrorist Attacks     | 10          | 86533      |
| Natural Disasters     | 10          | 56613      |
| Human Accidents       | 10          | 41163      |

B. Analysis procedures

This section explains further our focus on nouns and verbs and reasoning regarding potential emotional impact for readers.

As mentioned in the previous section, we understand the carving up of predications into nominal and verbal sequences has an impact on narrative. A greater sense of narrative is attributed to news texts that feature a greater number of verb sequences in relation to noun sequences, in that nouns highlight “things” while verbs emphasize “happenings”.3

The element of narrative, “story telling”, observed in news articles might result to an intensified emotional impact for the reader when seen in news articles, and inasmuch as such news is negative, it is negative emotions that are aroused.4 Narrative can be observed both through the reporters’ choice of words and through the words of witnesses incorporated into an article. We assume that an extensive use of verb sequences (VS) compared to a lesser use of noun sequences (NS) results to a form of narrative dynamism and could potentially affect the reader emotionally. We take dominance of verbal sequences to be a proxy measure of narrative prominence.

To examine the extent to which the element of narrative is present in each of the three corpora, a count of nouns, verbs and other items was conducted. For identifying the NS and VS to count in each corpus, TreeTagger [19], [20] was used; this is a tool for annotating text with part-of-speech (POS) and lemma information. POS sequences were manually inspected to identify those constituting noun phrases (e.g. “Canadian junior hockey team bus”) and given the label, “NS” and sequences of verbal elements (e.g. “has destroyed”) given the label “VS”. The additional category of “Other words” marks the complement category of remaining lexical tokens.

III. RESULTS

The analysis of the present work begins by investigating the occurrences of noun sequences (NS) in relation to verb sequences (VS) and other sequences across the three corpora. This is useful in determining whether the element of narrative is attributed to the corpora through the use of more verb sequences in relation to noun sequences.

Firstly, we test whether there is an interaction between the NS/VS occurrences and the three event types. A $\chi^2$ test is performed with the null hypothesis being that there is no relationship among the article types and the sequences. Table II presents the total numbers of NS and VS occurrences across the Terrorist, Natural Disaster and Human Accident corpora respectively. The penultimate column ‘Other words’ represents the words that remain in each corpus.

| Table II. Observed nominal vs verbal sequences in the three corpora in terms of event type. |
|---|---|---|---|
| Type                | Nouns | Verbs | Other words | Total words |
| Human Accident      | 2383  | 2595  | 36185       | 41163       |
| Natural Disaster    | 3103  | 4929  | 49591       | 56613       |
| Terrorism           | 4090  | 8693  | 73750       | 86533       |

This word choice, “things” and “happenings”, both nominals, makes clear that nominalizations can encode events. Nonetheless, verbs, including deverbal are more dynamic, hence more narrative.

4That news is, by default, bad, follows from the contrapositive of the conditional implicit in the popular expression that “No news is good news”.

2The events and sources for each corpus are presented in Tables 2, 3 and 4, in an online archive: https://www.scss.tcd.ie/clg/SensationalismData/.

3This word choice, “things” and “happenings”, both nominals, makes clear that nominalizations can encode events. Nonetheless, verbs, including deverbal are more dynamic, hence more narrative.
least one noun (subject) and at least one verb (predicate), but English verbs may have more arguments than intransitive verbs, allowing direct and indirect objects, for example. Reflection on basic sentences of English would lead one to expect more nouns than verbs. However, in the case of Terrorist Attack news articles the number of verbal sequences appears to be greater than that of nominal sequences by more than double. Whereas, in both the cases of Natural Disaster and Human Accidents corpus the difference in the use of further verbal in relation to nominal sequences is not as remarkable.

The test of interaction between the row and column categories is significant: \( \chi^2 = 541.7, \text{df} = 4, p < 2.2e - 16 \). The \( \chi^2 \) results suggest that the observed values are significantly different from what would be expected in the case that there was no interaction between the event categories and the parts of speech examined. Thus, the null hypothesis may be rejected.

Insight into the magnitude of the difference between the observed values and those expected if there were no interaction is presented in Table III, with the juxtaposition of standardized residuals. Residuals of magnitude between 2 and 4 are significant at the \( p \leq 0.05 \) level and magnitudes greater than 4 are significant at the \( p \leq 0.01 \) level; the polarity of the residuals indicate the direction of divergence from expectation under the assumption of no interaction between column and row variables – positive values indicate divergence in excess of expectation; negative values, smaller than expected values.

<table>
<thead>
<tr>
<th>Type</th>
<th>NS</th>
<th>VS</th>
<th>OtherWords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Accident</td>
<td>5.2832014</td>
<td>-17.0625286</td>
<td>4.1589472</td>
</tr>
<tr>
<td>Natural Disaster</td>
<td>2.9796851</td>
<td>-0.7406077</td>
<td>-0.4954775</td>
</tr>
<tr>
<td>Terrorism</td>
<td>-6.0539581</td>
<td>12.3671305</td>
<td>-2.4676755</td>
</tr>
</tbody>
</table>

With regards to the Terrorist Attack news articles, it can be argued that nominal sequences and other sequences are used significantly less than one would expect, with their residuals at \(-6.0539581\) and \(-2.4676755\) respectively. The converse conclusion can be drawn from the Terrorist events verbal sequences value (12.3671305), as it is notably above the 2 to 4 magnitude (\( p < 0.01 \)). They appear to be used remarkably more than one would expect if there were no interaction between article content category and part-of-speech distributions. The residuals observed in the Natural Disaster corpus indicate that the use of NSs at 2.9796851, is more than one would expect if there was no interaction between article categories and POS category distributions (\( p < 0.05 \)). In opposition, the VSs and other sequences appear to be used less than expected, at \(-0.7406077\) and \(-0.4954775\) respectively. Finally, it can be concluded that Human Accident news articles use NSs (5.2832014) and other sequences (4.1589472) more than one would expect, while the VSs appear to be used significantly less (-17.0625286) than would be expected if there were no interaction between event type and syntactic category. A general observation regarding the NS and VS occurrences found in the three corpora is that there is a significantly greater use of VS found in the Terrorist Attacks news articles, in opposition to the Natural Disaster and Human Accident news texts.

A second examination of the nominal and verbal sequence uses found within the three corpora was similarly conducted in relation to the seven news sources. Such an analysis might be useful in order to decide whether the greater use verbal sequences is related to certain news sources or is due to chance. Table IV displays the observed nominal and verbal sequence uses in relation to other sequences found in the online news platforms across the three corpora.

<table>
<thead>
<tr>
<th>Source</th>
<th>NS</th>
<th>VS</th>
<th>OtherWords</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBC</td>
<td>1101</td>
<td>1663</td>
<td>15395</td>
</tr>
<tr>
<td>CNN</td>
<td>1280</td>
<td>2116</td>
<td>19630</td>
</tr>
<tr>
<td>Daily Mail</td>
<td>2367</td>
<td>4361</td>
<td>43075</td>
</tr>
<tr>
<td>Fox News</td>
<td>1031</td>
<td>1748</td>
<td>16902</td>
</tr>
<tr>
<td>Huffington Post</td>
<td>1136</td>
<td>1892</td>
<td>18085</td>
</tr>
<tr>
<td>The Guardian</td>
<td>1279</td>
<td>2163</td>
<td>21480</td>
</tr>
<tr>
<td>Washington Post</td>
<td>1382</td>
<td>2274</td>
<td>23951</td>
</tr>
</tbody>
</table>

The observed values shown in Table IV suggest that across the seven news sources, verbal sequences are generally used more than nominal. A \( \chi^2 \) test reveals that the interaction between news sources and nominal versus verbal sequences use is significant: \( \chi^2 = 80.74, \text{df} = 12, p = 2.981e - 12 \). Therefore, the H0 that there is no relationship between the row and column variables can be rejected. The residual values of nominal and verbal sequences occurrences throughout the news sources are presented in Table V.

<table>
<thead>
<tr>
<th>Sources</th>
<th>Nouns</th>
<th>Verbs</th>
<th>Other Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBC</td>
<td>5.1321609</td>
<td>1.6326262</td>
<td>-1.7847583</td>
</tr>
<tr>
<td>CNN</td>
<td>2.4186263</td>
<td>1.9991850</td>
<td>-1.2359050</td>
</tr>
<tr>
<td>Daily Mail</td>
<td>-4.3632222</td>
<td>-0.3183168</td>
<td>1.1675929</td>
</tr>
<tr>
<td>Fox News</td>
<td>0.2642347</td>
<td>0.3918444</td>
<td>-0.1902771</td>
</tr>
<tr>
<td>Huffington Post</td>
<td>1.1789824</td>
<td>0.7959706</td>
<td>-0.5443669</td>
</tr>
<tr>
<td>The Guardian</td>
<td>-0.4405577</td>
<td>-0.6372203</td>
<td>0.3120985</td>
</tr>
<tr>
<td>Washington Post</td>
<td>-1.3824012</td>
<td>-3.1467024</td>
<td>1.3462518</td>
</tr>
</tbody>
</table>

Table V suggests that the only instances of significantly greater use of nouns is found across the CNN and BBC news articles, with Pearson residuals of 2.4186263 (\( p < 0.05 \)) and 5.1321609 (\( p < 0.01 \)) respectively. In opposition, the nouns in the Daily Mail are fewer than expected (\(-4.3632222\)) and verb use across the Washington Post news articles is less (\(-3.1467024\)) than one would expect if H0 were accepted.

The multidimensional interactions of nominal, verbal and other sequences in terms of event types and sources are explored in Figure 1. Starting with the Terrorist Attack corpus, Fig. 1 suggests that VSs are used more than would be expected by the majority of news sources, that is BBC, the Daily Mail, Fox News, Huffington Post and The Guardian. NS and other sequences appear to be used less than expected in all news sources, with the exception of Other Words use by the Daily Mail, which seems to be using other sequences significantly more than would be expected if there were no relationship between the horizontal and vertical axis variables.

The news sources covering Natural Disaster events that display greater use than expected in VS are CNN and Huffington Post, whereas the Daily Mail and Washington Post use verb sequences significantly less than expected. NS are used more than expected by the BBC, CNN and Huffington Post while they seem to be used less than expected by the Daily Mail and Washington Post. Other sequences seem to be used significantly less than expected by the Daily Mail and
Fig. 1. Pearson’s residuals association plot of Nouns, Verbs and Other Sequences (Other Words) in terms of event types and sources. The vertical axis figures event types on the left and the nouns, verbs and other words residuals on the right of the figure. The horizontal axis features the seven news sources. Red colour (below horizontal axis) indicates values from -27 to -4, pink (below horizontal axis) indicates values from -4 to -2, grey indicates values from -2 to 2, light blue (above horizontal axis) represents values from 2 to 4 and dark blue (above horizontal axis) represents values greater than 4.

Washington Post, while they appear to be used remarkably more by CNN, Huffington Post and The Guardian.

In the Human Accidents corpus, verbal sequences are used significantly less than expected by chance across the majority of news sources, that is CNN, Daily Mail, Fox News, Huffington Post and The Guardian, whereas their use is not significantly different for the BBC and Washington Post. NS appear to be used more than expected by CNN and The Guardian ($p < 0.05$), and significantly more than expected by the BBC, Fox News and Washington Post ($p < 0.01$). The Daily Mail uses NS significantly less than expected.

The interaction plots in the next two figures display the means of NS and VS use within each news source across the three corpora. Figure 2 illustrates the means of nominal use, while Figure 3 shows the means of verbal sequence use within each of the seven sources that describe events on Terrorism, Natural Disasters and Human Accidents respectively.

The two plots indicate that the mean values of NS use within each news source varies in comparison to the means of VS use, which seems to be consistent across the corpora. Starting with the verbal sequences, Fig. 5 suggests that the means of verbal sequence use do not differ significantly within each of the three corpora. That is, all seven news sources seem to have low means of VS use at 0.06 to a little over 0.07 for news texts that describe Human Accidents. All sources that describe events on Natural Disasters appear to have slightly higher means of VS use between 0.08 to 0.09 and lastly, the means of VS use for Terrorist events are between 0.09 and just above 0.10. The plots show that verbal sequences are mostly used by all news sources when Terrorist incidents are described, while the mean values are less when events on Natural Disasters and Human Accidents are the main topics.

The mean values regarding nominal sequences presented in Figure 2, suggest that the distribution of nominal sequences within the news sources is not consistent across the three corpora. Starting with the Human Accidents corpus, the mean of NS use within each news source stands between 0.055 to slightly above 0.065, while the mean values for the Natural Disasters corpus lie between just over 0.050 and 0.065. The means of NS use across the seven news sources for events on Terrorism appear to have a more significant deviation with the values between 0.045 to 0.065. It is, therefore, suggested that the nominal sequences distribution is significantly different among the news sources regarding each corpus.

IV. Discussion

This study has explored whether linguistic devices that elicit emotional distress to the reader are featured more in articles on Terrorism compared to articles on Natural Disasters and Human Accidents. The results support such a hypothesis.

The analysis of nominal and verbal sequences across the three corpora was conducted in order to examine the potential narrative featured in the news texts. The element of narrative is arguably important for entertainment, it is however, questionable whether a news story about negative events can be similarly enjoyed [4, p.1913]. The latter argument indicates that news reports on unfortunate incidents do not constitute a form of entertainment, nor are they constructed.
The findings of this study suggest that preference for verbs versus nouns in narrative description across Terrorist news articles of each news source is significant. Verbs’ potentially emotional effect is achieved through narrative (nominal versus verbal sequences). As indicated above, the purpose of such devices’ use is not the negative emotional impact that is created; their use is the reporters’ attempt to construct news texts that are appealing. However, this argument appears to lead to a paradox. Conventionally, anything that might be considered appealing does not entail indications of disturbing images or distressing vocabulary. Therefore, the question might arise as to how does one not look away when exposed to such texts and, most importantly, why? [21, p.27]. A possible explanation for such phenomenon could be the readers’ need for reassurance. Upon reading a news text one cannot guess the potential negative emotions that might be evoked via the reporter’s choice of words; in contrast, the negative emotional affect needs to be experienced by the reader in order to be realized. Once anxiety, fear or panic are generated within the reader, the latter might feel the need to continue reading in the hope of encountering a positive resolution. In the case that one chooses to terminate the reading, the feeling of anxiety might be experienced more intensely due to the uncertainty that is generated by not knowing what exactly happened and whether it is now worse in the reader’s imagination than in reality.

V. Conclusion

The findings of this study are significant in light of the fact that a great deal of emotional impact of texts is due to the linguistic devices within them – the way things are expressed has impact additional to the impact of the underlying content. A greater relative frequency of verbs were observed more within Terrorist Attack news articles than in Natural Disaster and Human Accident news reports and thus, it seems plausible that the readers might experience a greater sense of anxiety, fear and panic upon reading a news story about terrorist acts. News about terrorist acts may require more verbs, or it may be that using more verbs is a linguistic choice underlying a functional strategy for creating an appealing and emotionally exciting textual environment. However, we noted different degrees of difference in the relative frequency of verbs across news categories as a function of news source, the tabloid news source, The Daily Mail most striking in the greater than expected use of verbs in Terrorist Attack articles and less than expected use of verbs in Human Accident articles (see Fig. 1). This difference in source which aligns with pre-theoretic notions of sensationalism associated with the sources suggests that there is at last some element of functional strategy (the fact that all sources used more verbs and less nouns than one would expect with no interaction suggests that there is also an element explained by terrorist acts requiring more verbs than the other categories). Our work contributes to research in corpus-based data mining and analysis, and in particular, text analytic contributions to emotion analysis.
We have analysed news articles describing incidents that belonged to three distinct event categories, however the examination of a greater number of corpora and text types is necessary, as is study of a wider range of linguistic devices: metaphors, capitalization, exclamation points, formatting devices like font, point size, and so on. Multi-modal data is also important. Audio-visual material is widely implemented in reports posted on online news platforms and has been examined in previous research throughout the past years [21], [22], [23]. Even within the more restricted focus on nominals and verbal expressions taken in this work, it would be illuminating to explore subcategories, such as within verbal aspect, stative versus event based descriptions [24]. Given our sequence-oriented focus, it would be useful to explore links with the event-sequence-as-strings account of temporality provided by the framework of finite state temporality [25], [26].

These findings are useful within the wider research area of cognitive infocommunications, which examines technology that extends human capabilities [27], [28], [29], [30]. It has been argued within this field that in order to understand the nature of technology that extends human capabilities and which is accepted to the extent that it is assimilated into the concept of humanity, it is necessary to understand both human capabilities without that assimilation and the new technology, each in isolation [31]. They argue further that natural language is an example of such a technology (alongside clothing, money and more recent advances, like calculators), and that, notwithstanding all the advances of linguistics and cognitive science, there remains much to learn about human use of language. Our work is in that spirit, and not without relevance to the possibility of other technologies that build on this. For example, if the findings that we present about events English newspaper reporting generalize to other languages, and other text genres, then topic and emotional intensity classifiers may be informed by relative distributions of nouns and verbs.

ACKNOWLEDGMENT

This research is supported by the Science Foundation Ireland (grants 12/CE/I2267 and 13/RC/2106) through the CNGl programme and the ADAPT Centre.

REFERENCES


