MEN WHO KILL WOMEN: SEMANTIC MAPS FOR THE IDENTIKIT OF THE KILLER AND MURDERED WOMEN

Domenica Fioredistella Iezzi, Mario Mastrangelo

1. Introduction

Since 1999, the United Nation have fixed on November 25th as International Day for Elimination of violence against women (Resolution no. 54/99) to remind the assassination of Mirabel sister who were massacred while on their way to visit their husbands in prison. In Italy, only since 2005, women’s refuges celebrate this date. This violence is gender-based, in fact, the acts of violence are committed on women deliberately because they are women. The UN Declaration on the Elimination of Violence Against Women states that: “violence against women is a manifestation of historically unequal power relations between men and women” and that “violence against women is one of the crucial social mechanisms by which women are forced into a subordinate position compared with men”\(^1\). The extreme manifestation of violence against women is called also femicide. According to Russell (2011) femicide is killing of women by males because they are females. Although, in many cases, this expression is used to indicate a series of violence against women including the assassination, in this context, we will use this term with the meaning of murder of women.

Femicide is a social phenomenon present throughout the world, although with different intensity. El Salvador has the sad record of femicide rate (13.2%), but all Latin American countries have a very high prevalence rate (pr), e.g. Brazil in 2008 have had a pr = 5.4 with 4,014 victims, also South Africa also has a very high-risk index with 3,357 women killed in 2008 (pr = 10.1).

In Italy, from 2000 to 2012, there were 2,220 cases of homicide of women (mean of 171 victims per year), 1,570 of which occurred at home (70.7%). In 2012, domestic femicides represented 67.3% of the total femicides (Iezzi, 2013). In recent years, important legislative measures on violence against women have been approved, e.g. the anti-rape and stalking law (low 38/2009) or law against femicide

---

\(^1\) See website: http://www.un-documents.net/a48r104.htm.
(law 119/2013) to curb this phenomenon, but the number of femicide no has registered significant changes.

The aim of this paper is to classify the Italian femicides occurred in 2013.

2. Data and methods

We collect our data from the website of the Corriere della Sera newspaper. In the webpages dedicated to femicide phenomenon, it contains a short summary of 138 violent deaths of women, occurred in 2013. Using these recaps, we build a little textual Corpus, with 22,009 tokens and 4,537 types, the type-token ratio equal to 20.61 and the percentage of hapax equals to about 59%. For every case we also collect a list of structured variables: homicide modalities (with 8 modalities: firearm, cutlass, blunt object, beating, choking and strangling, mixed manner, unknown and others); relationship victim-murderer (with 8 modalities: husband, current partner, former partner or husband, son, relative, extraneous, unknown and other); age and nationalities of victim and murderer.

The adopted strategy is the following:
A) **STEP 1: Collection of documents to structure a corpus**
   A.1) Sourcing: Internet;
   A.2) Formatting: transformation UTF-8 digital encoding;
   A.3) Archiving: structuring of a corpus;
B) **STEP 2: Pre-processing**
   B.1) Tokenize the input text;
   B.2) Case folding;
   B.3) Lemmatization.
C) **STEP 3: Vector Space Model**
   C.1) document Indexing;
   C.2) building a bag of words weighing scheme;
   C.3) selection of the keyword and construction of tfidf scheme;
D) **STEP 4 Extraction of information and construction of patterns**
   D.1) Correspondence Analysis (CA) of Lexical Tables;
   D.2) Clustering on the principal components of CA;
   D.3) Multiple Factorial Analysis on tfidf matrix.

For describing femicide we selected 113 keywords, based on the extraction of the characteristic language. In this way, we obtained a matrix $\mathbf{M}$ with size 113 words x 138 cases: each case is represented by a vector of weighed terms of the form $d_j = (w_{i1}, w_{i2}, \ldots, w_{in})$, a column of matrix $\mathbf{M}$, where $w_i$ represents the
weight for the term $t_i$, which is attached to case $d_j$. We adopt the Term frequency-inverse document frequency (TF-IDF) weighting scheme:

$$w_{ij} = \frac{n_{ij}}{\max n_{ij}} \log \frac{N}{n_i}$$

where $\max n_{ij}$ is the maximum frequency of word $i$ in a corpus. $N$ is the total number of documents, and $n_i$ is the number of documents/cases in which the word $i$ appears.

By joining the $M$ matrix and the structured variables previously described, we obtained a new matrix of size 120 variables (keywords plus structured variables) x 138 cases, that is a dataset composed by several sets of variables, both continuous and categorical, which we want to study simultaneously.

For this aim and in order to reduce the number of attributes (dimensions) of our matrix, we used the Multiple Factor analysis (MFA) (Escofier & Pages, 1988), that is based on repeated implementations and recaps of Principal Components Analysis (PCA) for continuous variables and Multiple Correspondence Analysis (MCA) for categorical variables, and, among other things, allows to balance the influence of each set of variables and to study the links between sets.

We considered the keywords, the homicide modalities and the relationship victim-murderer as active variables, and the others as illustrative variables.

Starting from the first 5 global factorial coordinates obtained with the MFA, our goal is to classify femicide, using our suit of structured and unstructured data.

We adopted a fuzzy approach that allows to deal with the usual lexical ambiguity of textual data better than a hard approach (Iezzi & Mastrangelo, 2014). Therefore, we applied the $c$-means algorithm (Dunn, 1973; Bezdek, 1973), the soft version of the k-means clustering, and the Xie Beni index (Xie-Beni, 1991) to identify the best fuzzy-partition, according to the overall compactness and the separation of the clusters.

3. Results

In 2013, 103 Italian women have been killed and 36 foreign (8 Romanian, 5 Ukrainian, 3 Albanians, and 3 of Kosovo). Figure 1 shows that the women murdered were aged from 16 and 93 with an average of 51.96 (SD 20.29), the authors were aged from 16 and 90, with an average of 43 years old (SD 16.05), but 90 years old for the authors is an outlier. The Pearson correlation points out that the variables "victim age" and "author age" is weakly correlated ($r=0.43$).
Figure 1 *Box-plot on victim and author age*

Figure 2 highlights that the murder weapons most used were cutlass (25.4% of victims), firearm (24.6%), strangulation (18.9%). The chi-square test shows that there is not an association between the weapon used and the territory (*p*-value = 0.553).

**Figure 2** – *Murder weapon used to murdered women in 2013*

Sicily (femicides no.17), Lazio (femicides no.16), and Campania (femicides no.14) have had the highest number of femicides, but the regions with the highest incidence are Umbria (r=0.86), Calabria (0.69), and Sicily (r=0.65). These results
differ from the analysis conducted in the period 2000-2012 in which the North of Italy, particularly Emilia Romagna, recorded a higher incidence (Iezzi, 2013).

Table 1 – Regional distribution of femicide

<table>
<thead>
<tr>
<th>Regions</th>
<th>Number of femicide</th>
<th>Rate for 100,000 woman (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Piemonte</td>
<td>11</td>
<td>0.48</td>
</tr>
<tr>
<td>Lombardia</td>
<td>13</td>
<td>0.25</td>
</tr>
<tr>
<td>Liguria</td>
<td>4</td>
<td>0.48</td>
</tr>
<tr>
<td>Veneto</td>
<td>4</td>
<td>0.16</td>
</tr>
<tr>
<td>Trentino A.A.</td>
<td>3</td>
<td>0.56</td>
</tr>
<tr>
<td>Friuli V. G.</td>
<td>4</td>
<td>0.63</td>
</tr>
<tr>
<td>Emilia Romagna</td>
<td>8</td>
<td>0.35</td>
</tr>
<tr>
<td>Toscana</td>
<td>9</td>
<td>0.46</td>
</tr>
<tr>
<td>Marche</td>
<td>5</td>
<td>0.62</td>
</tr>
<tr>
<td>Umbria</td>
<td>4</td>
<td>0.86</td>
</tr>
<tr>
<td>Abruzzo</td>
<td>4</td>
<td>0.59</td>
</tr>
<tr>
<td>Lazio</td>
<td>16</td>
<td>0.54</td>
</tr>
<tr>
<td>Campania</td>
<td>14</td>
<td>0.47</td>
</tr>
<tr>
<td>Molise</td>
<td>1</td>
<td>0.62</td>
</tr>
<tr>
<td>Calabria</td>
<td>7</td>
<td>0.69</td>
</tr>
<tr>
<td>Puglia</td>
<td>10</td>
<td>0.48</td>
</tr>
<tr>
<td>Sicilia</td>
<td>17</td>
<td>0.65</td>
</tr>
<tr>
<td>Sardegna</td>
<td>4</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Correspondence analysis on lexical table (terms x murder of weapon) detects six overlapped classes (Figure 1). On the fist five components, we apply k-means algorithm that detected three groups:

Cluster no.1: Murders of crime: where the victims are young women, in most cases of prostitutes savagely beaten with sticks, kicked, gagged .. The killers are unknown;

Cluster no.2: family homicides, where the authors are current or former husbands. Women are generally stabbed or killed with a firearm;

Cluster no.3: Multiple murderers, which occur with mixed methods, the body or the face could be were both affected.

Using MFA method, we detect 5 groups. The first group includes 2 types of femicide: murders carried out by robbers, and by victim’s sons with psychosis or trying themselves to rob their mothers. In all cases the victim’s age is advanced and the prevalent modality is choking/strangling, although the membership value is high also for beating. Following keywords belong to this group: rapina, pensionata, anziana, soqquadro, imbavagliata, ecchimosi, cuscino, etc.

The second group includes murders occurred mainly because of strokes and strikes with blunt objects and due to partners (the husband modality for relationship
victim-murderer variable has also high membership value). In this group we find keywords like aggredire, litigare, pugni, bottigliata, martellata, malmenare, mattarello, bastonate, pestaggio, and gelosia as mainly motive of the murders.

**Figure 3** Plane of the first two axes from the Correspondence Analysis of Lexical table

The third group is characterized by the use of fire-arms, mostly employed by husbands and ex partners, although the membership value is high also for the current partner modality. In these cases homicides are often followed by murderer’s suicide. We detect keywords like pistola, fucile, freddare, tempia, nuca, bruciapelo, suicidarsi.

The fourth group includes murders carried out by relatives, mainly fathers and brothers and secondly sons and husbands, in a mixed manner—for example strokes and strangling- or in other ways, like poisoning or burning. It also includes some events of euthanasia, when the victims have a case of degenerative disease. In this
The last group includes murders committed using cutlass, carried out by husbands, partners and particularly strangers, in the latter case often due to rejected sexual advances. Following keywords belong to this group: sangue, lama, coltello, ascia, roncola, sgozzare, fendenti, gola, addome, lite, avance.

Figure 4 shows the 5 groups partition previously described in the first 10 factorial planes.

These analyses showed that a lemmatization when you have short texts, as in this case, allows of reducing the sparseness of matrices. The use of the matrix of weights bag of words, with the choice of frequency thresholds, may result in a loss of valuable information. The results of the analysis
of lexical correspondences tables brought results extremely confused. In contrast, the system of weights tfidf on keywords allows of selecting well-defined profiles. The results showed that, over the years, the numbers of femicide and its profiles have remained unchanged, despite the extended to the public awareness campaign and the many innovations at the regulatory level.

References


SUMMARY

In recent years, several laws have been passed to curb violence against women, but the number of femicide has not decreased.
In this paper, we examine the stories of the women killed in Italy in 2013. For this purpose, we will use the unstructured information from the website of the Corriere della Sera newspaper, with metadata, to build semantic maps and locate the identikit of the killer and the profiles of the women killed.

Domenica Fioredistella IEZZI, University of Tor Vergata, stella.iezzi@uniroma2.it,
Mario MASTRANGELO, University of Tor Vergata, mario.mastrangelo@uniroma2.it