Analysing occupational safety culture through mass media monitoring

Livia Celardo[1], Rita Vallerotonda[2], Daniele De Santis[2], Claudio Scarici[2], Antonio Leva[2]

[1] Sapienza University of Rome

[2] INAIL Research – Headquarters for Research of the Italian National Institute for Insurance against Accidents at Work

Abstract 1

In the last years, a group of researchers within the National Institute for Insurance against Accidents at Work (INAIL) has launched a pilot project about mass media monitoring in order to find out how the press deal with the culture of safety and health at work. To monitor mass media, the Institute has created a relational database of news concerning occupational injuries and diseases, that was filled with information obtained from the newspaper articles about work-related accidents and incidents, including the text itself of the articles. In keeping with that, the ultimate objective is to identify the major lines for awareness-raising actions on safety and health at work. In a first phase of this project, 1,858 news articles regarding 580 different accidents were collected; for each injury, not only the news texts but also several variables were identified. Our hypothesis is that, for different kind of accidents, a different language is used by journalists to narrate the events. To verify it, a text clustering procedure is implemented on the articles, together with a Lexical Correspondence Analysis; our purpose is to find language distinctions connected to groups of similar injuries. The identification of various ways in reporting the events, in fact, could provide new elements to describe safety knowledge, also establishing collaborations with journalists in order to enhance the communication and raise people attention toward workers' safety.

Abstract 2

Negli ultimi anni un gruppo di ricercatori all'interno dell'Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro e le malattie professionali (INAIL) ha lanciato un progetto pilota riguardante il monitoraggio dei mass media con lo scopo di analizzare come la stampa tratta la salute e la sicurezza sul lavoro. A tal fine, l'Istituto ha istituito un database relazionale delle notizie riguardanti gli infortuni e le malattie, incluso il testo stesso delle notizie. L'obiettivo finale del progetto è dunque quello di identificare le direttrici principali su cui muoversi per azioni di sensibilizzazione su salute e sicurezza sul lavoro. Nella prima fase del progetto, 1,858 articoli di giornale riguardanti 580 infortuni sono stati raccolti; per ogni evento, non solo il testo della notizia ma anche diverse variabili sono state individuate. La nostra ipotesi è che per diversi tipi di infortunio un diverso linguaggio viene usato dai giornalisti per narrare l'accaduto. Per verificare ciò, una procedura di Text Clustering è stata implementata sugli articoli, insieme ad una Analisi delle Corrispondenze Lessicali; il nostro obiettivo è quello di individuare delle differenze nel linguaggio in relazione a diversi gruppi di infortuni. L'identificazione di diversità nel modo in cui viene riportata la notizia al lettore può fornire nuovi elementi per descrivere la cultura della sicurezza, al fine di instaurare delle collaborazioni con i giornalisti stessi per rendere migliore la comunicazione e accrescere l'attenzione del cittadino verso la sicurezza del lavoratore.

1. Introduction

The study described here grew out of the collaboration between the Department of Social Sciences and Economics of Sapienza University of Rome and the Headquarters for Research of INAIL (Italian National Institute for Insurance against Accidents at Work) where, since 2012 a team of researchers has developed the idea of monitoring the mass media in view of prevention against accidents at work (INAIL, 2015).

With this in mind, those researchers achieved the so-called "Repertorio Notizie SSL" (*News Repository on Occupational Safety and Health*), that is a relational database of media news related to occupational injuries and diseases. The objective of this project is to observe the culture of occupational safety and health communicated by mass media agencies in order to identify new elements for increasing prevention against accidents at work. In this study we focus on the hypothesis that there are some asymmetries in the language used to describe the injuries depending on the characteristics of the event. To test it, we performed on the repository data some Automatic Text Analysis procedures.

The article is structured as follow: in section no.2, the News Repository is presented; in section no.3, data are presented and the methodology is exposed; in section no.4, the results of the analyses are shown; in section no.5, conclusions are drawn.

2. The tool

News Repository on Occupational safety and health (NeRO) is a tool created to allow analyses of news contents and texts related to occupational diseases and injuries. In fact, our strategic objective is to increase public awareness and safety culture through a different approach, which will be also based on the study of news articles, their composition and communication dynamics. So, the first operational purpose is to understand:
- which kind of terms are used in news articles about accidents at work or occupational diseases;
- what inspires a title;
- how the same news is treated by different sources/media;
- how the news text could be interpreted in different ways due to who communicates the news itself;
- whether or not some specific aspects of the events are considered by media.

Our study plans to analyze the cultural characteristics of mass media communication regarding occupational safety and health (OSH), observing the attitude of mass media (and journalists) towards the subject and the way users perceive the news depending on which words are used.

As mentioned before, NeRO is an *ad hoc* relational database, centred on the gathering of newspaper articles regarding accidents at work, but it is also arranged to gather news on near misses, occupational diseases and incidents from all kind of sources (press, television or radio). It involves several digital interconnected *tables*, which contain structured – i.e. based on appropriate classifications – and unstructured – i.e. textual – information. Information retrieval regards events happened in Italy and it could contain both online and directly consulting newspapers, since we exploited Google Alert Service (using some suitable keywords) and a daily-newspaper subscription ("la Repubblica"). The reference unit is the event (right now, we are restricting events to accidents) and different aspects and information are linked to it: one or more articles about it, one or more workers injured, and so on. The data-entry interface consists of a series of thematic screens, starting from the opening one,

which covers the list of already recorded events. These screens allow to enter the following data, step by step:

– [Screen "Event"] Text containing event description, date of the event, venue, company where accident occurred (if appropriate), economic activity;
– [Screens "News"] Texts of each article related to the event, newspaper name (or press affiliation), news title, web url, date of the article;
– [Screens "Worker" and Sub-screens "Accident" and "Harms, disorders or diseases"] Injured worker's biographical data, information about accident, type of injury, physical implication or resulting disease.

3. Methodology and data

The repository, at the end of data collection, was composed of 1,858 news, related to almost six hundreds different accidents. In order to analyse the content of the news texts in connection with the characteristics of the different events, we performed a content analysis using the Reinert's method (Reinert, 1983) for a descendant hierarchical partition. This algorithm, starting from the co-occurences matrix, generates groups of lexical units – i.e. words – that more co-occur in the texts. Then, the lexical groups were projected on the factorial axes, together with the variables modalities, using the Lexical Correspondence Analysis (Lebart, Salem and Berry, 1997); in this way, we could observe how the language is connected to the accidents features. Finally, to better understand the differences between news texts we analysed the specificities related to the modalities of the variables.

4. Main results and discussion

The cluster analysis made on news texts using the Reinert's method– choosing as segments the articles – produced three lexical groups (in order, the red, the blue and the green ones, in Figure 1):

- Cluster 1 (56.5%): in this group are included the words related to the description of the events, in terms of what happened;
- Cluster 2 (26.5%): here we have the terms connected to the road accidents;
- Cluster 3 (17%): this group is about the emotional aspects connected to the events.

We projected the lexical groups (Figure 1) and the modalities of the variables related to the events (Figure 2) on the first two factors obtained using the lexical correspondence analysis.
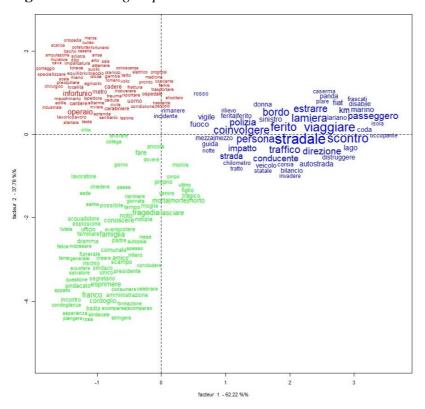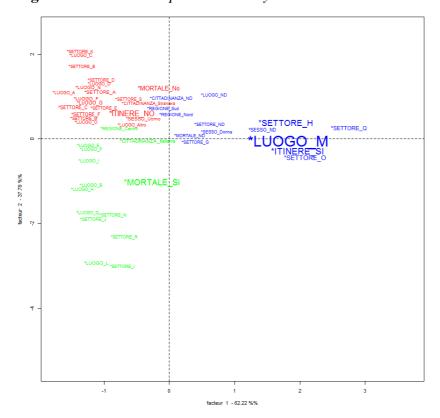
Figure 1 *Lexical groups*

Figure 2 *Lexical correspondence analysis*

As shown in the figure no. 2, there are some interesting characterizations of the language used in newspapers. Some variables, like the economic activity and the accident site, present a strong lexical differentiation among the modalities; this means that who is narrating the event - i.e. the journalist - uses a specific language to describe the accident, on the basis of these characteristics. The other variables presented no particular specificities, except for the one related to the mortality of the accident. In fact, as shown in the figure no. 2, on the second factor the variable "accident mortality" is best represented because of the position and the distance of the modalities "yes" and "no" from the origin. To better understand the lexical differences, we analysed also the specificities (Bolasco and De Mauro, 2013; Lafon, 1980; Lebart, Salem and Berry, 1997) for this particular variable.

Table 1 *Analysis of the specificities – Variable: "accident mortality"*

Fatal accident - No	*z = test-value*	Fatal accident - Yes	*z = test-value*
Hospital	59.17	Tragedy	35.68
Serious	58.84	Family	27.17
To transfer	54.90	Useless	23.62
Dangerous	28.38	To leave	19.84
Rescue	24.13	Victim	18.68
Ambulance	24.09	Tragic	17.71
Leg	23.12	Friend	14.95
Injury	22.06	Band	14.89
Trauma	20.55	Condolence	12.65
Hand	18.84	Province	12.15
Fracture	16.70	Son	11.49
Helicopter	13.70	Wife	11.48
Bus	12.23	Escape	10.63
Crossroad	10.20	Mayor	9.11

Starting from the results showed in table no.1, we can observe that there is a significant difference in the language utilized when the accident is fatal or not. The terms used in the case of a non-fatal event are related to the description of the injury, while in the case of a mortal accident the situation is completely different: the words utilized refer to the emotional sphere of the event, so concepts like the family or the unpredictability are very often used to describe what was happened.

5. Conclusions

The project here presented showed how News Repository on OSH (NeRO) can contribute to analyse occupational safety and health, although in some institutions there are already databases dedicated to newspaper articles dealing with OSH. Actually, in addition to news texts, NeRO provides several systematized information, enabling to filter news according to various search criteria and, above all, to carry out a number of studies and organized analysis on textual data, too. In this paper we showed one of the study we implemented on Repository data using Automatic Text Analysis. The results revealed that a large amount of information is contained within these data; anyway, some information asymmetries are present. For that

reason, it will be essential to set up a discussion with a network of journalists and other experts, in order to improve and enhance the media communication. The challenge is to get out from the inner circle of prevention practitioners and build a bridge that could connect the Institution to a more general public, also contemplating liaison organizations (such as trade unions and employers' associations).

References

Bolasco S. and De Mauro T. (2013). *L'analisi automatica dei testi: fare ricerca con il text mining*. Carocci Editore.

Iezzi D. F. (2012). Centrality measures for text clustering. *Communications in Statistics-Theory and Methods*, *41*(16-17), 3179-3197.

INAIL. (2015). *Il monitoraggio dei mass media in materia di salute e sicurezza: Strumenti per la raccolta e l'analisi delle informazioni*.

Lafon P. (1980). Sur la variabilité de la fréquence des formes dans un corpus. *Mots*, *1*(1), 127-165.

Lebart L., Salem A. and Berry L. (1997). *Exploring textual data*(Vol. 4). Springer Science & Business Media.

Reinert M. (1983). Une méthode de classification descendante hiérarchique: application à l'analyse lexicale par contexte. *Les cahiers de l'analyse des données*, *8*(2), 187-198.