Language Processes and Related Statistics in the Posts Associated to Disasters on Social Networks

S.C. Bolea

Speranţa Cecilia Bolea
Institute of Computer Science of the Romanian Academy - Iasi Branch
Romania, Iasi, Carol I, 8
cecilia.bolea@iit.academiaromana-is.ro

Abstract: This paper provides a detailed and long-period statistics of the use of synonyms in posts related to specific events on social networks (SNs), an extended analysis of the correlations of the flows of the synonyms in such posts, a study of the applicability of Zipf's law to posts related to specific events on SNs, and an analysis of the dynamics of the fluxes of synonyms in the posts. The paper also introduces the study of the distances in the phase space for the characterization of the dynamics of the word fluxes on social networks. This article is a partial report on recent research performed for a deeper analysis of social networks and of processes developing on social networks, including used lexicon, dynamics of messages related to a specific type of topic, and relationships of the processes on SNs with external events.
Keywords: social networks, disaster, analytics, language statistics, correlative analysis, Zipf's law, dynamics.

1 Introduction

Analytics aimed to social networks have explosively developed during the last few years, with some of them focusing on the rescue from disasters [3], [5], others proposing means for disaster management [6], [10], [26] and others on disaster prevention and mitigation [1, 2], [11], [15], [22], [27].

The main purpose of this paper is fact-finding about the language of the social networks and related posts, for messages connected to dramatic events such (fires and earthquakes). In subsidiary, we are interested in the distributions of the most used words and of the synonyms. A hypothesis we made in connection with the word probability distribution is that it departs from the typical distribution of the language, which is a Zipf's distribution.

This article is a partial report on recent research related to the deeper analysis of social networks and of processes developing on SNs; the analysis refers to the lexicon used, dynamics of messages related to a specific type of topic, and the relationships of the processes on SNs with external events. The general lexicon was studied in [24], and issues related to the use of synonyms in tweets were discussed in [4], [20], and [24].

The main contributions of this paper include the detailed and long-period statistics of the use of synonyms in posts related to specific events on SNs, the extended analysis of the correlations of the flows of the synonyms [29] in those posts, a study of the applicability of Zipf's law to posts related to specific events on SNs, and the detailed analysis of the dynamics of the fluxes of synonyms in the posts. The research was performed in the frame of a larger project, being a part of it (see Acknowledgments).

We advance the study of correlation between time series of the number of words occurrences especially of synonyms in messages related to disasters effects on SNs, and compare the SNs from this point of view. Also, we apply a recently proposed nonlinear technique of analysis of time series for better characterizing, and differentiating specificities of SNs.

One of the purposes of this article is to present a full year data statistics for synonyms usage in posts related to two types of potentially disastrous events, namely earthquakes and fires, and to analyze in detail the dynamics of the words most used in such posts. The utility of this study is multifold and includes the optimization of searches in software applications for disaster monitoring and management, and better understanding the language use under and in relation with such events [18]. The results may also be of interest to psycho-linguistics, beyond descriptive linguistics. A special emphasis is placed on the correlative study of the dynamics of the flows of synonyms in the posts.

This research associates with other analyses of the language used in SNs reported recently in the literature, for example with the vast literature on sentiment analysis on SNs [7], [12], including the time of disaster events [16], [17].

Both the mentioned project and this study relates to previous interests of our research group in computational linguistics and speech technology. The research in the project and partly in this study also reflects the interest in detecting attitudes in messages and to a standing interest in detecting emotions in speech and texts as in the studies [9], [13], [25], [28].

2   Data collection and processing method

The data collection process was described in [24]; data comprises all posts, including articles and blogs pointed to by messages on Twitter and Google+, during the period March 2015 and March 2016, as detected with queries using the keywords *(cutremur OR seism) AND Vrancea*, respectively *incendiu OR foc*.

An important difference between the dataset we used and the data in other papers is that we have not restricted ourselves only to the messages, but we also downloaded the media articles or blogs having addresses in the collected messages. Therefore, the most or at least a large part of the text analyzed comes from the media, not from the respective social network (SN) messages. Thus, a larger lexicon was included in the database. Yet, the most frequent words found beyond stop words have been those intimately related to the search condition, which is an interesting and somewhat unexpected fact. The text was lemmatized before performing the statistics. In doing so, words with various grammatical forms were reduced to their main representative. We used the free lemmatizer for the Romanian language from RACAI Institute, available at [30]. The statistic of words in this study actually represents the statistic of lemmas in the dataset.

The dataset was conventionally split by SN and by months and is reported as such in the paper, for example in the correlative analysis of synonyms and in the analysis of their dynamics of occurrences. The descriptive statistics, including Zipf's law detection, refers to the whole database, on each SN, thus extending the results in [24].

The dataset includes 3.715 posts for the *earthquake* type of event (database named *CutremurDB*) and 1.106 for *fire* type (database named *FocDB*). We reiterate that the only events of interest here are the earthquakes and the fires. The total number of posts, words (without stopwords), and the number of characters (with spaces) are given in Table 1 for the two classes of posts. The stopwords are the words such as "0 retweets 0 likes", "0 retweets 2 like", "view summary" and "newline", and their form depends on the type of browser used (Google Chrome, Mozilla Firefox).

The *FocDB* database is larger than *CutremurDB* because fires are much more frequent events, thus produce more posts on SNs. Interestingly, the posts related to earthquakes have a small number of lemmas (Table 1). The average number of words per post are: 126 words for Google+ and 124 words for Twitter in *CutremurDB*, and 317 words for Google+ and 131 words for Twitter in *FocDB*.

Table 1: The number of posts and lemmas

	Posts	Lemma	No. of lemmas	Char with spaces
Google+(*CutremurDB*)	318	6.872	40.196	174.030
Twitter(*CutremurDB*)	941	12.229	116.667	632.979
Google+(*FocDB*)	245	15.041	77.909	466.618
Twitter(*FocDB*)	279	10.396	36.665	231.293

Both databases, *CutremurDB* and *FocDB* have the same structure and contain the fields: *year*, *month*, *posts* (number of posts from the month), *lemma* (number of lemmas), *no. of lemmas* (number of lemmas occurrences), *char with spaces* (number of characters including spaces) and $lemma_1, lemma_2, \cdots, lemma_n$, where $lemma_i, i \in \{1, n\}$ represents the significant words referring earthquake and fire (nouns, verbs, adjectives, adverbs). The other words such as pronouns, prepositions, conjunctions, numerals and punctuations have no relevance in our study, and they do not appear in our databases.

3 Main findings

3.1 Descriptive statistics of synonyms used in relation to specific events

This study contributes to the descriptive statistics of the SN language by determining the probabilities of the most frequent words and synonyms in posts related to earthquakes and fires on social networks, on extended periods. The main results are summarized in Table 2, where *lemma* denotes the number of lemmas and *No. of lemmas* the number of lemma's occurrences.

Table 2: The probabilities of the most important lemmas for CutremurDB and FocDB

CutremurDB	cutremur	seism	seismic	magnitudine	Richter	Vrancea	epicentru	adâncime	produce
p for Google+	0,0310	0,0123	0,0069	0,0158	0,0149	0,0244	0,0041	0,0079	0,0133
p for Twitter	0,0311	0,0130	0,0064	0,0165	0,0137	0,0244	0,0054	0,0085	0,0153

FocDB	incendiu	foc	izbucni	pompier	suferi	stinge	flacără	pericol	arde
p for Google+	0,0092	0,0069	0,0018	0,0027	0,0003	0,0006	0,0012	0,0002	0,0012
p for Twitter	0,0260	0,0164	0,0055	0,0098	0,0005	0,0024	0,0047	0,0007	0,0027

The best correlation found (in CutremurDB) is for the lemmas: *cutremur - Vrancea* and *cutremur - produce* (see Table 3). This is because the messages about earthquakes found on SNs have a very similar "Un cutremur cu magnitudinea de 3, 5 grade pe scara Richter s-a produs luni seara în zona seismică Vrancea...", "An earthquake with a magnitude of 3.5 on the Richter scale occurred Monday evening in Vrancea seismic zone...". The correlations between *cutremur* and *seism* synonyms is a little smaller than the above. The synonyms *incendiu* and *foc* have a higher value of correlation than other pairs of words (see Table 3).

It was shown in [19], [23] that there is a strong correlation between the number of posts related to potentially or actually disastrous events and the intensity of the event. Precisely, in the cited papers it was found that the peak of the number of posts per day (post flux) related to the event correlates with the number of victims in several types of disasters; when the event is only potentially disastrous, such as a small seism, the peak of the post flux was found to correlate with the magnitude of the earthquakes. In addition, it was shown in [19], [23] that the flux of posts related to such events is well represented by a pulse with exponential growth and decay, in the form

Table 3: Examples of correlations in the use of words: correlations of the time series of the monthly number of occurrences of the lemmas

Word pairs CutremurDB	Correlation value		Word pairs FocDB	Correlation value	
	Google+	Twitter		Google+	Twitter
cutremur - seism	0,87	0,96	incendiu - foc	0,99	0,95
cutremur - seismic	0,78	0,90	incendiu - izbucni	0,90	0,66
cutremur - magnitudine	0,89	0,96	incendiu - pompier	0,69	0,92
cutremur - Richter	0,86	0,93	incendiu - suferi	0,77	0,69
cutremur - Vrancea	0,97	0,97	incendiu - stinge	0,82	0,26
cutremur - epicentru	0,81	0,86	incendiu - flacără	0,87	0,36
cutremur - adâncime	0,82	0,99	incendiu - pericol	0,61	0,37
cutremur - produce	0,96	0,98	incendiu - arde	0,83	0,75
cutremur - grad	0,87	0,98	incendiu - distruge	0,79	0,08
cutremur - intensitate	0,61	0,85	incendiu - răni	0,93	0,04
cutremur - România	0,83	0,92	incendiu - România	0,81	0,80

$$n(t) = \{ \begin{matrix} Ae^{\lambda(t-t_0)} \\ Ae^{\lambda(t_m-t_0)}e^{-\kappa(t-t_m)} \end{matrix} \tag{1}$$

Above, t is time, $n(t)$ is the number of posts per day (flux) at time t, t_0 is the moment of the event, t_m is the moment when the peak occurs, and λ and κ are coefficients. A similar evolution was verified for the words *cutremur* and *foc* in our database, see Figure 1.

Notice that during the months of autumn and winter the number of fires increases compared with the other months of the year. Yet the summer heat may cause vegetation fires. These facts are mirrored by the number of related messages on SNs, see *foc* in Figure 1.

3.2 Zipf's law and the lexicons of posts on SNs

Recall that Zipf's law [14] states that the logarithm of the probability of a word in a language is related to the rank of the word,

$$log(p(w)) = \alpha \times rank(w) + \beta \tag{2}$$

Zipf-like laws, that is, power laws, are well known in statistical linguistics, see for example [14], and in many other domains such as populations of cities, economy, and biology. Explained in more detail, this law says that, when ordering the words of any language, including Romanian, according to their frequency of apparition in a large corpus, the logarithm of their apparition probability decreases linearly with their rank. This means that selecting at random words in a specified language, the law should not apply. We have been interested, see [24], to find if Zipf's law is relevant to the database related to disasters. Our initial hypothesis was that the law would not be valid because of the very restrictive set of words used in such posts, their low rank in the language, and their supposedly almost random choice from the lexicon of the global lexicon of the language.

Yet, our hypothesis was overturned [24], in that at least the most frequently used words obeyed Zipf's law, while words with higher rank confirmed the hypothesis. The results of log-lin representation of the probabilities of the most frequent 8 words, topic *seism* for Google+ (3), and Twitter (4) are:

$$Y = -0,1969X - 3,3642; R^2 = 0,9555 \tag{3}$$
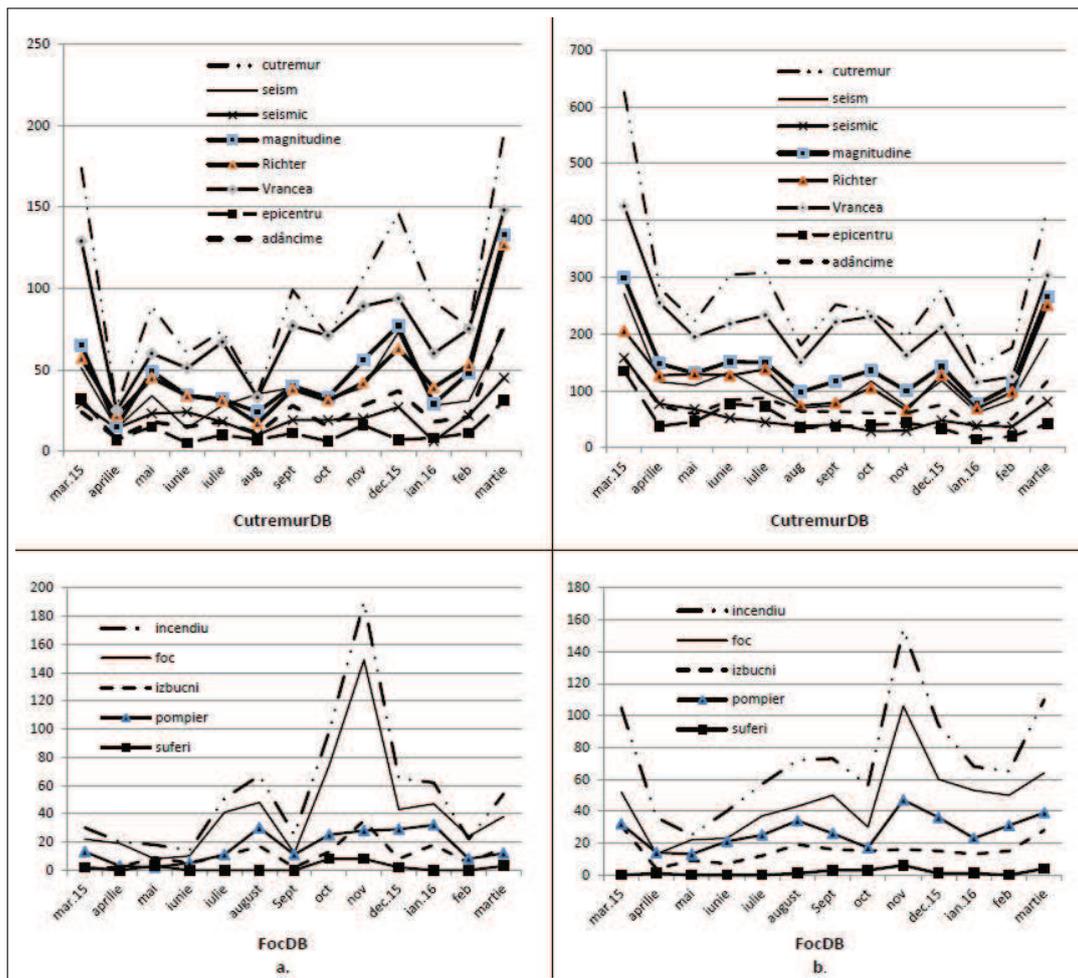
$$Y = -0,1592X - 3,0524; R^2 = 0,9494 \tag{4}$$

Figure 1: Time series of keywords occurrences: a. Google+, b. Twitter

The main results are shown in Fig. 2, indicating a good agreement ($R^2 > 0.7$) with Zipf's law. Notice that on Google+ the slope slightly differs from that on Twitter, indicating a lexicon-based personality of the two SNs. Additional results are given on the SRoL website described in [9].
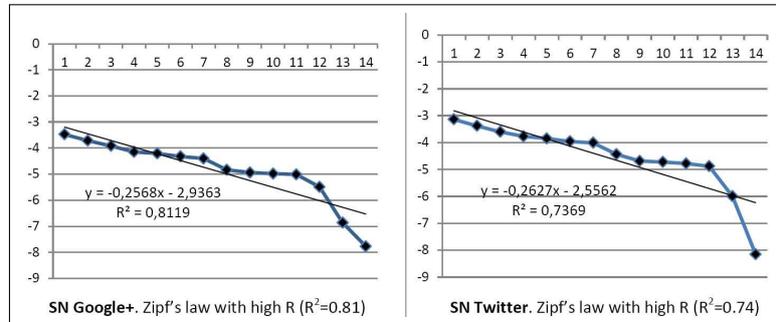

Figure 2: Log-lin representation of the probabilities of the most frequent 14 words, topic *seism*

Analyzing the equations 3, 4 and Figure 2, the first ranked words (up to 8) obey the Zipfs law. In case of *FocDB*, the first ranked words(up to 8 or 14) obey Zipf's law (see (5), (6) and Figure 3).
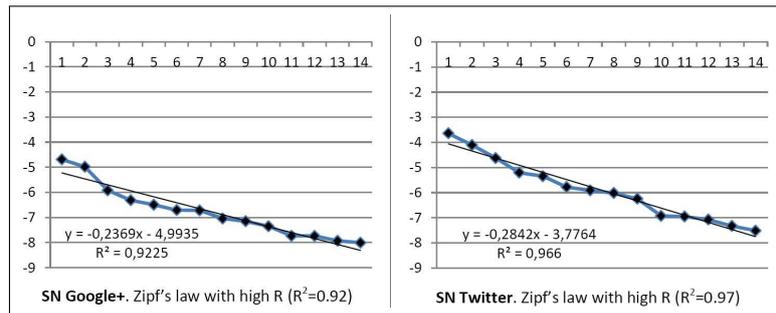

Figure 3: Log-lin representation of the probabilities of the most frequent 14 words, topic *incendiu*

$$Y = -0,33X - 4,6241; R^2 = 0,8933; in\ Google+\ SN \tag{5}$$

$$Y = -0,3481X - 3,5163; R^2 = 0,9489; in\ Twitter\ SN \tag{6}$$

3.3 Dynamics of the fluxes of synonyms

The notion of dynamics of SNs originates from the papers [19], [23], where it was shown that the flow of messages streaming from an event on SNs is predictable and has a dynamics that has characteristics of nonlinearity. It was shown that the time series of the number of messages related to events correlate with the level of seismic activity [19], [23].

One of the issues only slightly addressed in [24] is that of the dynamics of the use of keywords and synonyms. In many respects, this dynamics resembles with the dynamics of the number of messages addressing a certain disaster event, as dealt in [19], [23]. The interest in the dynamics is multifold and includes the expectation of correlation of the query based on keywords and the number of returned messages related to a specified event (probability of finding messages related to an event when the search uses a given keyword), and the dynamics of the joint use of synonyms in messages. On the other hand, the study [24] proved that, while the time series of the number

of occurrences of the words "cutremur" and "seism" are strongly correlated, they have dynamics that are clearly differentiated in the phase space.

In this sub-section, we study the difference in the dynamics of synonyms, continuing [24], by generating the phase-space plots (maps) of the time series of the synonyms and then by applying the method of *windows* in the time space, as proposed in [21]. That method is essentially an approximate way of determining the probability that the attractor of the time series has a point in a specified rectangular region of the phase space. Four windows were defined in the rectangular coverage of the attractors, as suggested in [24], and the counts for the respective four windows was performed, producing a vector with four components for each time series. Then, the distance between the vectors was determined (method due to HN Teodorescu). We found that the Euclidean distance is low only for the pair of synonyms *cutremur* and *seism* on the Twitter SN and for Google+ SN, this distance is a little larger.

The phase space, in a simplified case of the phase plane for the dynamics $x(t)$, is simply defined as the plane $(x, \dot{x})$, $\dfrac{dx}{dt} = \dot{x}$, with the corresponding graphical representation. The characterization method introduced in [21], consists in determining the frequency of a point in the phase space that lies in the 2D interval (rectangle) $[a_1, a_2] \times [b_1, b_2]$, for a set of non-overlapping rectangles defined in the phase space. Let us consider the attractor of the time series of the flux of a specified word, $n_w(t_k)$, for a given time duration $t_1, \cdots, t_p$ framed in the rectangle $[min_{j \in \{1,p\}} n_w(t_k), max_{j \in \{1,p\}} n_w(t_k)] \times [min_{j \in \{1,p\}} dn_w(t_k), max_{j \in \{1,p\}} dn_w(t_k)]$, where $dn_w(t_k) = n_w(t_k) - n_w(t_{k-1})$. Divide this rectangular subspace of the phase space into four equal rectangles, overlapping only on one of their edges,

$$R_1 = [min_{j \in \{1,p\}} n_w, \frac{min_j n_w + max_j n_w)}{2} \times [min_j dn_w, \frac{min_j dn_w + max_j dn_w}{2}] \qquad (7)$$

where everywhere above *min* and *max* are for $j \in \{1, p\}$; similarly one defines R_2, R_3, R_4, see Fig. 4. Denote the number of points that fall in these rectangles by $n_1(w_1), \cdots, n_4(w_1)$. Form the vector $\overrightarrow{\omega_1} = (n_1(w_1), n_2(w_1), n_3(w_1), n_4(w_1))$. Perform the same operations for the specified word, w_2, whose dynamics we wish to compare with $\overrightarrow{\omega_1}$, and find $\overrightarrow{\omega_2}$. Then, the distance $d(\overrightarrow{\omega_1}, \overrightarrow{\omega_2})$ between the two dynamics is the distance between $\overrightarrow{\omega_1}$ and $\overrightarrow{\omega_2}$ (personal communication HNT):

$$d(\overrightarrow{\omega_1}, \overrightarrow{\omega_2}) = \sum_{k=1-4} |n_k(w_1) - n_k(w_2)|,$$

or the Euclidean distance would be $d(w_1, w_2) = \sqrt{\sum_{k=1-4} (n_k(w_1) - n_k(w_2))^2}$. A further improvement is to use relative frequencies in the distances, dividing by the total number of points in the phase diagrams, $\nu_k = n_k/p$, where $p = \sum_k n_k$.

The distances between the attractors in the phase plane provides a quantitative, simple characterization of the similarity of the dynamics of the two time series compared, supplementing the information given by the correlation. The phase diagrams for four of the most frequent words found on *Google+* are shown in Figure 4; the corresponding Euclidean distances between the dynamics are given in Table 4, which stands for the matrix $[d_{12}]$, where d_{12} denotes the distance between words.

The results presented in Table 4 show that the larger Euclidean distance is between the dynamics of the words *magnitudine* and *Vrancea*, and the smallest distance is between *seism* and *Vrancea* respectively *cutremur* and *magnitudine*, on Google + (both distances 1.41). Comparatively, the time series for the word pair *cutremur* and *magnitude* have the third higher correlation (see Table 3), only slightly higher than the correlation for the pair *cutremur - seism*. Even less favorably compares the value of the distance between the dynamics of *cutremur* and *Vrancea* $d = 3.75$, the third largest, with the correlation of their time series, which is the highest,
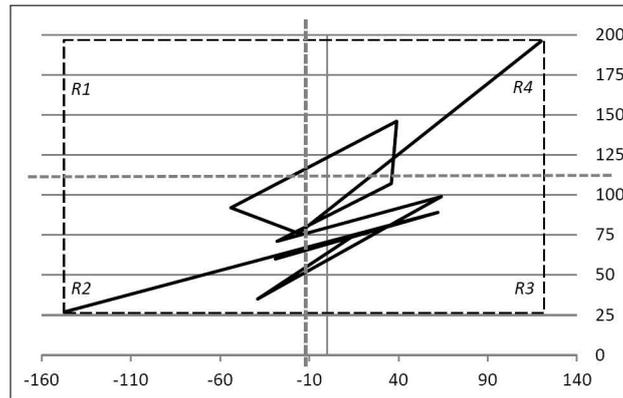
Figure 4: Example of dynamics in the phase plane of the series for *cutremur*

Table 4: Euclidean distances between the dynamics of the words *cutremur*, *seism*, *Vrancea*, and *magnitudine* (relative frequencies)

Words	cutremur	seism	Vrancea	magnitudine
cutremur	0	2,83	3,75	1,41
seism		0	1,41	4,24
Vrancea			0	5,10
magnitudine				0

$C = 0.97$. This shows that distances between the dynamics and the value of the correlation of the respective time series carry at least partly different information and help differentiate between the processes represented by the time series.

4 Discussion

The use of synonyms has also a psycho-linguistic background [24]; as such, it relates to emotions and attitudes expressed in the posts. It would be interesting to follow this study by a research of the uttering of the synonyms on voice-enabled SNs in view of detecting their emotional charge using various methods of characterization, such as those in [8], [9], [13], [28] that report on emotion recognition tools specifically for the Romanian language.

Not all messages posted on SNs contain diacritics. This can present a problem in recognizing lemmas. In this study, all the words with and without diacritics were taken into account.

There is a difference between the same messages on different browsers and SNs. So, the stopwords for TwitterSN are (in Google Chrome): "*no* retweets *no* likes", "Reply", "Retweet", "More", "Like" "newline", and in case of Mozilla's browser the stopwords are "*no* retweets *no* likes" and "newline", where *no* is a number. The stopwords in Google+SN are: "Adauga un comentariu...", "newline". This posts were saved first in .DOCX, than transformed in .TXT.

The earthquake's posts contain information about the earthquake from a specific day. In almost all messages there are dates about the biggest earthquake from that year, or the previous years. This is the reason for the *CutremurDB* contain a small numbers of words than *FocDB*. The data bases include only the following words: nouns, verbs, adjectives, adverbs; we deleted the conjunctions, prepositions, numerals.

5 Conclusions

One of the interesting findings first detected in the preliminary study [24] and confirmed by this ampler database is that Zipf's law is still valid for the most important most frequent 6-7 words used in the posts related to disaster. This is surprising because the words are not among the most used in a language, that is, they are among the rare words in common language; moreover, they are not selected in any specific way preserving the distribution in the common language, so there is no reason that these words still obey a Zipf-type law. However, the law breaks down when we try to expand the set of the most frequent words beyond the number of eight. These facts may indicate that the word selection mechanism in the common language scales down to the lexicon used when dealing specific topics. This scaling property of language statistics (Zipf's law) was never been pointed out before [24], according to our knowledge.

Social networks have their own specificity, with features that could be likened to the personality of humans; the specific use of synonyms is one of that. The parameters in Zipf's law for the lexicons used in posts related to various types of events may help to automatically identify the SN from a set of posts, only knowing the related type of event eliciting the messages.

The posts which have been studied for one year are mostly messages sent by newspapers and TV's and have an informal structure of mass-media type.

Acknowledgments

Bibliography

[1] Abbasi, M.A. et al. (2012), Lessons Learned in Using Social Media for Disaster Relief - ASU Crisis Response Game, *Social Computing, Behavioral - Cultural Modeling and Prediction*, Springer, LNCS 7227:282-289.

[2] Acar, A.; Muraki, Y.(2011); Twitter for Crisis Communication: Lessons Learned from Japan's Tsunami Disaster, *Int. J. Web Based Communities*, 7(3): 392-402.

[3] Anderson, K.M.; Schram, A. (2011); Design and Implementation of a Data Analytics Infrastructure in Support of Crisis Informatics Research (NIER Track), *The 33rd Int. Conf. Software Engineering, ICSE'11*, Waikiki, Honolulu, USA, May 21-28, ACM, 844-847.

[4] Bolea, S.C. (2015); Vocabulary, Synonyms and Sentiments of Hazard-Related Posts on Social Networks, *8th IEEE Int. Conf. on Speech Technology and Human-Computer Dialogue, SPED*, Bucharest, Romania, October 14-17.

[5] Boulos, M.N.K. et al. (2010), Social Web Mining and Exploitation for Serious Alications: Technosocial Predictive Analytics and Related Technologies for Public Health, Environmental and National Security Surveillance, *Computer Methods and Programs in Biomedicine*, 100(1): 16-23.

[6] Bruns, A.; Burgess, J.E. (2012); Local and Global Responses to Disaster: #eqnz and the Christchurch Earthquake, *Disaster and Emergency Management Conf., Proc., AST Management Pty Ltd*, pp. 86-103.

[7] Cambria, E. et al. (2013), New Avenues in Opinion Mining and Sentiment Analysis, *IEEE Intelligent Systems*, doi:10.1109/MIS.2013.30, 28(2): 15-21.

[8] Ciobanu, A. et al. (2014), Automatic Fury Recognition in Audio Records, *Proc. 12th Int. Conference on Development and Application Systems (DAS)*, Suceava, Romania, May 15-17, 176-179.

[9] Feraru, S.M. et al. (2010), SRoL - Web-Based Resources for Languages and Language Technology e-Learning, *International Journal Computers Communications & Control*, 5(3): 301-313.

[10] Kryvasheyeu, Y. et al. (2016), Rapid Assessment of Disaster Damage Using Social Media Activity, *Science Advances*, DOI: 10.1126/sciadv.1500779, 2(3):e1500779.

[11] Merchant, R.M. et al. (2011), Integrating Social Media into Emergency-Preparedness Efforts, *The New England Journal of Medicine*, 365:289-291.

[12] Nakov, P. et al. (2013), SemEval-2013 Task 2: Sentiment Analysis in Twitter, *2nd Joint Conf. Lexical and Computational Semantics (*SEM), 7th Int. Workshop SemEval*,Atlanta, June 14-15, 2:312-320.

[13] Pavaloi, I. et al. (2013), Acoustic Analysis Methodology on Romanian Language Vowels for Different Emotional States, *Proc. Int. Symposium on Signals, Circuits and Systems (ISSCS)*, Iasi, Romania, July 11-12.

[14] Piantadosi, S. T. (2014); Zipf's Word Frequency Law in Natural Language: A Critical Review and Future Directions, *Psychon Bull Rev.*, DOI: 10.3758/s13423-014-0585-6, 5:1112-1130.

[15] Pirnau, M. (2015); Tool for Monitoring Web Sites for Emergency-Related Posts and Post Analysis, *8th IEEE Int. Conf. on Speech Technology and Human-Computer Dialogue, SPED*, Bucharest, Romania, October 14-17.

[16] Saharia, N. (2015); Detecting Emotion from Short Messages on Nepal Earthquake, *Proceedings of the 8th International Conference on Speech Technology and Human-Computer Dialogue, SPED*, Bucharest, Romania, October 14-17.

[17] Saif, H. et al.(2012), Semantic Sentiment Analysis of Twitter, *Proceedings of the 11th Int. Conference The Semantic Web, Part I*, Boston, USA, November 11-15, pp. 508-524.

[18] Temnikova, E. et al. (2015), EMterms 1.0: A teminological Resource for Crisis Tweets, *Proceedings ISCRAM 2015 Conference*, Kristiansand, Norway, May 24-27.

[19] Teodorescu, H.N.L. (2015); On the Responses of Social Networks to External Events, *Proc. ECAI 2015 7th IEEE Int. Conf. on Electronics, Computers and Artificial Intelligence*, Bucharest, Romania, June 25-27, DOI: 10.1109/ECAI.2015.7301138, 13-18.

[20] Teodorescu, H.N.; Bolea S. C. (2016); Analysis of Probabilities of Specified Words' Occurrences in SN Messages related to Catastrophes, *18-th Int. Conf. System Analysis and Information Technology, SAIT*, Kyiv, Ukraine, May 30-June 2.

[21] Teodorescu, H.N. (2012); Characterization of Nonlinear Dynamic Systems for Engineering Purposes - A partial Review, *International Journal of General Systems*, 41(8):805-825.

[22] Teodorescu, H.N. (2013); SN Voice and Text Analysis as a Tool for Disaster Effects Estimation - A Preliminary Exploration, *Proceedings 7th IEEE Conference on Speech Technology and Human - Computer Dialogue*, Cluj Napoca, Romania, October 16-19.

[23] Teodorescu, H.N. (2016); Emergency-Related, Social Network Time Series: Description and Analysis, *Time Series Analysis and Forecasting, Contributions to Statistics*, Springer International Publishing Switzerland, pp.205-215.

[24] Teodorescu, H.N.L.; Bolea S.C. (2016); On the Algorithmic Role of Synonyms and Keywords in Analytics for Catastrophic Events, *Proceedings ECAI-2016, 8th International Conference on Electronics, Computers and Artificial Intelligence*, Ploiesti, Romania, June 30-July 2.

[25] Teodorescu, H.N.; Feraru, S.M. (2007); A Study on Speech with Manifest Emotions, *Lecture Notes In Artificial Intelligence*, Springer Berlin Heidelberg, 4629: 254-261.

[26] Verma, S. et al. (2011), Natural Language Processing to the Rescue?: Extracting Situational Awareness Tweets During Mass Emergency, *Proc. Fifth Int. AAAI Conf. on Weblogs and Social Media*, North America, July, 385-392.

[27] Yang M. et al. (2011), Social Media Analytics for Radical Opinion Mining in Hate Group Web Forums, *J. Homeland Security and Emergency Management*, August 11, 8(1), DOI: 10.2202/1547-7355.1801.

[28] Zbancioc, M.; Feraru, M. (2012); Emotion Recognition of the SROL Romanian Database Using Fuzzy KNN Algorithm, *Int. Symposium on Electronics and Telecommunications IEEE-ISETC 2012 Tenth Edition*, Timisoara, Romania, 347-350.

[29] http://www.dictionardesinonime.ro/.

[30] http://www.racai.ro/en/tools/text/