Creativity Detection in Texts

Costin – Gabriel Chiru
Department of Computer Science and Engineering
Politehnica University of Bucharest
Bucharest, Romania
E-mail: costin.chiru@cs.pub.ro

***Abstract*— In this paper, we present a model that was intended to discriminate creative from non-creative news articles. In order to build the classifier, we have combined nine different measures using a stepwise logistic regression model. The obtained model was tested in two experiments: the first one tried to discriminate between news articles about the US 2012 Elections from different newspapers versus articles taken from The Onion (a website providing satiric news) on the same subject, while the second one evaluated the capacity of the model to generalize over different topics and text genres. The experiments showed that the system achieves 80% accuracy, but the lack of true positives from the second experiment raised the question of whether we really identified creativity or in fact we detected satire (as the assumption for the training corpus was that the satiric news from The Onion were also creative).**

Keywords-Creativity; Satire; Natural Language Processing; Metrics for Creativity Detection

I. INTRODUCTION

According to Zhu et al. [1], the definition of creativity is the ability to transcend traditional ideas, patterns, relationships into meaningful new ideas, interpretations, etc.

The goal of this paper is to identify whether a text is creative or not. To determine this, several steps were undertaken. First of all, we tried to identify the elements that define a creative text. After that, the most important features that explain creativity were chosen. A model for automatic creativity detection was derived as the final result.

In order to do that, nine different measures were explored: Type-to-Token Ratio [2], Word Norms Fraction [2], Google Similarity Distance [3], Explicit Semantic Analysis [4], Number of Named Entities, Named Entities Score, Wordnet Similarity [5], Coherence measure [6], and Latent Semantic Analysis (LSA) measures [7].

The paper continues with a short presentation of the current approaches for creativity detection and after that we describe the nine measures that we have investigated in our experiments for identifying creativity. Section 3 details the architecture of our application and after that we present the two experiments that we undertook and the obtained results. The paper ends with our conclusions based on these experiments.

II. STATE OF THE ART

Renouf [8] describes creativity as the thought of acting or the quality of an unpredictable departure from the rules of regular word formation.

In texts, the creativity measures "new and creative ways of expressing a given idea" and it is called linguistic creativity [9]. Measuring it has been known for its complexity.

A machine learning algorithm has been developed by Zhu et al. [1] to measure creativity by developing subjective creativity metrics. The aim of the algorithm was to use a linear regression model with 17 features derived from computer science and psychology perspectives [1].

Jordanous [10] proposed a Standardized Procedure for Evaluating Creative Systems (SPECS), which follows three steps for determining whether a computational system can be defined as creative or not. The three steps are: creativity identification, the derivation of standards to be used for evaluation of creativity and the system testing according to those standards [10].

Other researches related to linguistic creativity were focused on understanding and using metaphors [11][12] and analogies [12] or on explaining the appearance of new words from already existing ones (e.g., "television"+ "marathon" = "telethon")[13].

Creativity detection in song lyrics has also been carried out by Hu and Yu [2] by comparing three measures, two of them being adapted from the work of Zhu et al. [1]. Those two measures were Word Norms Fraction and Wordnet Similarity. The metrics proved to be able to determine the different aspects of identifying mood and creativity in a lyric. The Word Norms Fraction was used to calculate the lyrics' "usualness", while WordNet Similarity was involved in determining the similarities between concepts [2].

Still, the research for identifying creativity is in its early stages and in this paper we intended to make a step forward by developing a model that is able to discriminate between creative and non-creative texts, using a stepwise logistic regression model built on a corpus of creative and non-creative texts.

III. CREATIVITY MEASURES

In order to decide where creativity occurs, there is a widespread support that two important criteria are *novelty* and *quality*:

- *Novelty*: To what extent an item is different to the existing samples of its genre?
- *Quality*: How good the item really is?

In this paper, we tried to capture these two criteria through nine different measures: some of them were intended to capture novelty by identifying how ordinary a text is, while the others – the semantic ones – were used for detecting the quality of that text. Combining them, we hoped

 ISBN: 978-1-61208-280-6

that we would be able to determine the degree of creativity of a given text. The nine measures that we investigated are described in further detail bellow.

A. Type-To-Token Ratio

Type-to-Token Ratio is defined as the number of unique terms in a text divided by the total number of terms. It is often used to measure the vocabulary richness of a text [2].

$$m_1 = \frac{C_{unique}(x)}{n} \quad (1)$$

where C_{unique} is the number of unique words in a text and n is the total number of words.

B. Word Norms Fraction

Word norms represent associations between words, while Word Norms Fraction measures the "usualness" of a text [2]. According to Hu and Yu [2], texts with high occurrences of word norms should indicate high "usualness" and thus low creativity since creativity often corresponds to unusual patterns. In order to compute the "usualness" of word pairs, we have used the 72,176 pairs of word pairs offered by Free Association Norms [14].

$$m_2 = \frac{C_{norm}(x,y)}{n} \quad (2)$$

where $C_{norm}(x,y)$ is the number of word pairs that appear in Free Association Norms and n is the total number of words in the text.

C. Google Similarity Distance

Google Similarity Distance [3] measures similarity of words and phrases from the World Wide Web using Google page counts. It is based on the concept that the probabilities of Google search terms (conceived as the frequencies of page counts returned by Google divided by the number of pages indexed by Google), approximate the relative frequencies of those search terms as actually used in society. The Google Similarity Distance is given by:

$$m_3 = \frac{\max\{log(f(x)),log(f(y))\} - \log(f(x,y))}{\log(M) - \min\{log(f(x)),log(f(y))\}} \quad (3)$$

where f(x) denotes the number of pages containing x, and f(x, y) denotes the number of pages containing both x and y. M is the total number of web pages indexed by Google and during their experiments (in 2007), it was shown to have a value of 8,058,044,651. Nowadays, M is considered to have a value of 50 billion [15]. This value was used for the calculation of the Google Similarity Distance.

D. Explicit Semantic Analysis

The aim of the Explicit Semantic Analysis (ESA) [4] is to compute the semantic relatedness between the vectors of words using Wikipedia as the knowledge base. Wikipedia has been known as the largest online knowledge repository and it has been proven to be highly organized and regularly maintained, thus ensuring its consistency. This method uses Wikipedia's concepts and explicitly represents the meaning of a given text in terms of the concepts in Wikipedia. ESA manipulates concepts based on human cognition, which is why it is explicit in a sense compared to the Latent Semantic Analysis approach.

The input of this method is a plain text with concepts represented by the Wikipedia articles ranked according to their relevance using classic text classification algorithms. Each concept is represented as an attribute vector with assigned weights using Term Frequency–Inverse Document Frequency (TFIDF) and afterwards an inverted index is built. Once the text is represented by a semantic interpretation vector, simple cosine similarity is used to compute the semantic relatedness.

E. Number of Named Entities

This measure gives the total number of named entities found in the text, in order to check if the creativity of a text is related to the number of named entities used in the text.

m_5 = number of named entities in a text

F. Named Entities Score

This measure gives the proportion of distinct named entities used in the text. It is computed by dividing the number of distinct named entities by the total number of named entities.

$$m_6 = \frac{Number\ of\ distinct\ named\ entities}{Total\ number\ of\ named\ entities} \quad (4)$$

G. WordNet Similarity

The WordNet Similarity measure is based on the lexical database Wordnet [5]. It returns a value denoting how similar two word senses are, based on the shortest path that connects their senses in the WordNet lexical ontology.

H. Coherence Measure

Coherence can be thought of as how meanings and sequences of ideas relate to each other in a text. One approach of measuring coherence in a text is to compare sentences and check how similar they are. The coherence measure we propose is computed from the pair-wise sentence similarity. This measure is based on the coherence score proposed by He et al [6].

Given a set of documents D = $\{d_i\}$, i = 1...M, we define the coherence score as the proportion of "coherent" pairs of documents with respect to the total number of document pairs within D. The criterion of being a "coherent" pair is that the similarity between the two documents in the pair should meet or exceed a given threshold. Formally, given the document set D and a threshold τ, we have:

$$\delta(d_i, d_j) = \begin{cases} 1, & if\ sim(d_i, d_j) \geq \tau \\ 0, & otherwise \end{cases} \quad with\ i \neq j \in \{1..M\} \quad (5)$$

where cosine similarity is taken as the similarity between documents d_i and d_j and the threshold is set to 0.05. Then, coherence score of the document set D is defined as:

$$m_8 = \frac{\sum_{i \neq j \in \{1..M\}} \delta(d_i, d_j)}{\frac{1}{2}M(M-1)} \quad (6)$$

I. LSA Measures

Latent Semantic Analysis (LSA) is the best known and most widely used vector-space method for computing semantic similarity using dimensionality reduction [7]. It involves the application of Singular Value Decomposition (SVD) to a document-by-term matrix to reduce its rank. We used LSA to analyze sentence-to-sentence similarity of texts. Each sentence is treated as a document and LSA is performed on the document-by-term matrix. From the resulting matrix with reduced dimension, four different measures are computed.

1) Average similarity between adjacent sentences

From the reduced dimensionality matrix $\hat{X}$, the sentence similarity matrix S = [s_{ij}] = $\hat{X}^T\hat{X}$ is computed. The matrix S gives the similarity of all pairs of sentences. Average similarity between adjacent sentences is computed as follows:

$$m_{9a} = \frac{\sum_{i=1,j=i+1}^{n-1} s_{ij}}{n-1} \quad (7)$$

2) Average similarity between sentences

From the sentence matrix S, average similarity between pairs of sentences is given by:

$$m_{9b} = \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} s_{ij}}{\frac{1}{2}n(n-1)} \quad (8)$$

3) Average cosine similarity between adjacent sentences

This measure is similar to m_{9a} and gives the average of similarity of all pairs of adjacent sentences. However, this measure uses cosine similarity instead of the sentence similarity matrix S. Cosine between the sentence-vectors obtained from SVD is computed for all pairs of adjacent sentences and their average is taken.

$$m_{9c} = \frac{\sum_{i=1,}^{n-1} Cosine(\hat{x}_i, \hat{x}_{i+1})}{n-1} \quad (9)$$

4) Average cosine similarity between sentences

This measure is similar to m_{9b} but like in m_{9c}, cosine similarity is used to compute this measure. This measure is the average of cosine between the sentence-vectors obtained from SVD computed for all pairs of sentences.

$$m_{9d} = \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Cosine(\hat{x}_i, \hat{x}_j)}{\frac{1}{2}n(n-1)} \quad (10)$$

IV. Experiment's Architecture

The main experiment architecture consisted of three main modules: web crawling, corpus building and creativity assessment using the creativity measures presented above. The first two modules are shown in Figure 1, while the third one is detailed in Figure 2. Each process will be further detailed.

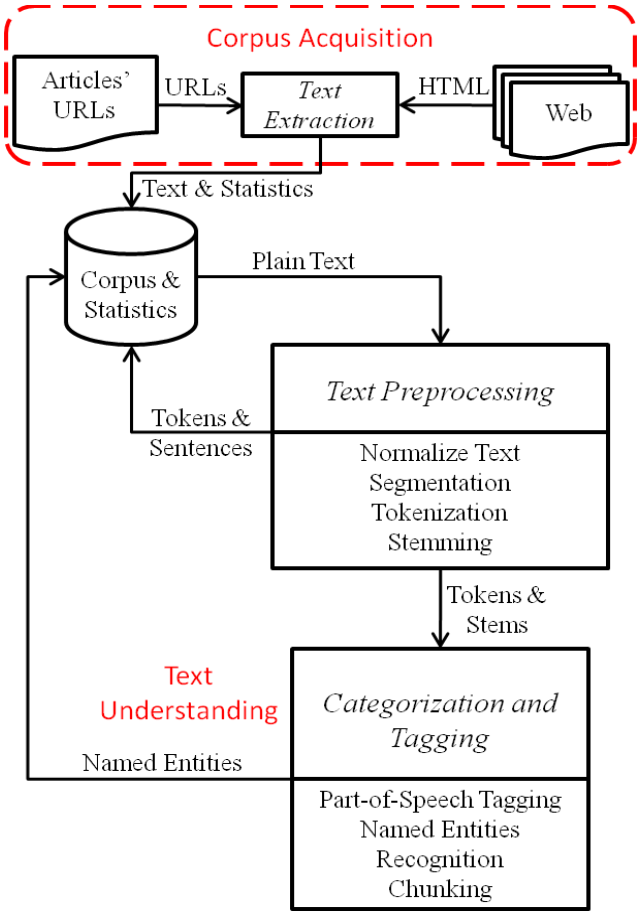

Figure 1. First two modules of the experiment: Web Crawling and Corpus Building involving Text Extraction, Preprocessing, Categorization and Tagging.

A. News articles extraction

During the experiment, we selected 118 articles that were debating about the US 2012 Elections. The articles were collected from 12 news sites from six different countries:

- *UK*: BBC, Wired, The Independent, The Sun;
- *Canada*: CBC;
- *Australia*: News.com.au, The Australian, Sydney Morning Herald;
- *USA*: Foxnews and Huffington Post;
- *South Africa*: News24;
- *New Zealand*: The NZ Herald.

In addition, 67 articles on the same topic were also extracted from The Onion [16]. As The Onion is a satire news organization, these articles were assumed to be more creative than the news. This assumption is based on the fact that while the news articles only present the facts/events, the ones from The Onion should involve either additional feelings towards these facts/events that would transform the articles into satires or satiric parallelisms with other facts/events (otherwise not being published). And both these actions could be triggers for creativity.

 ISBN: 978-1-61208-280-6

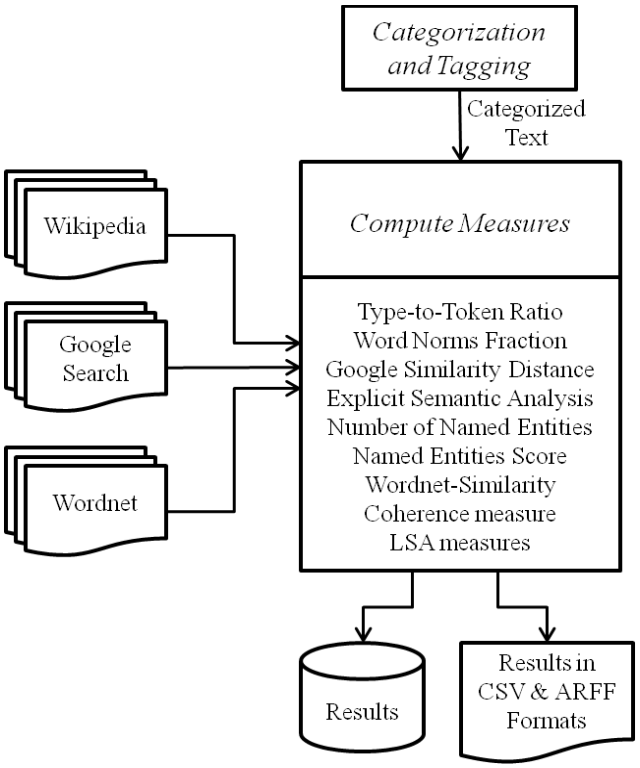

Figure 2. Measures Calculation modules.

B. Preprocessing, Categorization and Tagging

Once the articles' content was saved, preprocessing techniques (from the field of Natural Language Processing) were applied to the text (such as: tokenizing, sentence segmentation, stemming, and stop words removal). The next step was the part-of-speech tagging of the text and after that the assignment of categories (creative or not) based on being a The Onion article or not.

C. Computing the measures for each text

The next process consisted of computing the nine measures described above for each of the gathered documents. The process used information from Wikipedia, Google search and Wordnet (see Figure 2). The obtained results were min-max normalized to set the values in a range between 0 and 1 and then they were saved in order to be used as input files for the predictive models that are built in the next step.

D. Stepwise Logistic Regression

To evaluate the performance of each measure and to obtain a model able to predict the creativity or non creativity of an input text, a stepwise logistic regression model was built. This model gives a weight for each of the used features (which can be interpreted as an importance coefficient), helping to identify those that are most relevant. These weights show how strong is the correlation of the corresponding metric with creativity. The larger the absolute value of the coefficient is, the more important is the feature in determining whether the text is creative or not. This part was done with the help of Orange [17] and Tanagra [18], both open source data mining tools. The diagram for the workflow designed in Orange is shown in Figure 3.

V. Experiments and Results

In order to evaluate our work, we did two experiments: the first one was done for assessing the value of our built classifier and was tested versus the news articles that we have extracted, while the second one was intended to measure the capacity of the classifier to adapt (to what degree the classifier can be generalized in order to be applied to any kind of text?).

A. Assessment Experiment

The first thing that we had to do was to determine the parameters of the logistic regression model based on the values that we obtained for the set of news articles that we extracted from the web. A graphical representation of the values obtained for these documents for each of the measures described in Section 3 is provided in Figure 4.

After applying the logistic regression to this data set augmented with the obtained measures for each of the news articles, the equation defining the creativity (the negative class) or non-creativity (the positive class) was given by the formula:

$$\Pr(Y = 1 | X1, \dots, X9) = F(Bi * Xi),\ i = 0..9 \qquad (11)$$

, where: $X_0 = 1$, $X_1 - X_9$ represent the measures (Type-To-Token Ratio, Word Norms Fraction, Google Similarity Distance, ESA, Number of Named Entities, Named Entities Score, Wordnet Similarity, Coherence Measure, LSA), while B_0 is the bias factor and $B_1 - B_9$ are the parameters associated to each of the measures. The values obtained from the model were: $B_0 = 1.83$, $B_1 = 0$, $B_2 = 3.585$, $B_3 = 3.255$, $B_4 = 0$, $B_5 = -2.897$, $B_6 = 2.485$, $B_7 = -9.799$, $B_8 = 0$, $B_9 = 3.445$, resulting in a classifier for being creative or not, given by (12).

$$\Pr(Y = 1 \mid X_1, \dots, X_9) = F(1.83 + 3.585 * X_2 + 3.255 * X_3 - 2.897 * X_5 + 2.485 * X_6 - 9.779 * X_7 + 3.445 * X_9) \qquad (12)$$

A couple of observations should be drawn based on the model represented by (12). First of all, one can see that the bias factor (B_0) is positive, reflecting the fact that most of the texts are non-creative. Secondly, we saw that the Type-To-Token Ratio had no influence against the creativity of a text. This implies that both creative and non-creative texts had similar ratios of unique terms. On the other hand, Word Norms Fraction had the highest positive influence (showing evidence of a non-creative text), which was confirming our expectations since high values for this measure witnessed high "usualness" of the text, which contradicts the definition of creativity. More than that, the high value received by the parameter of the Google Similarity Distance comes to augment the drive towards the text "usualness".

Regarding the analysis of named entities, the classifier considered that the use of named entities is a sign of creativity (the parameter for the Number of Named Entities

is negative), but in the same time it regards the use of distinct such entities as being non-creative (positive Named Entities Score) which was at least confusing at the beginning. After a deeper analysis of the texts, we have reached the conclusion that this fact was correct, since the more distinct named entities were found in text, the less space was dedicated to expressing the author's sentiments related to the events described (which we consider to offer the opportunity for creativity) because that space was filled with the facts expressed by the named entities.

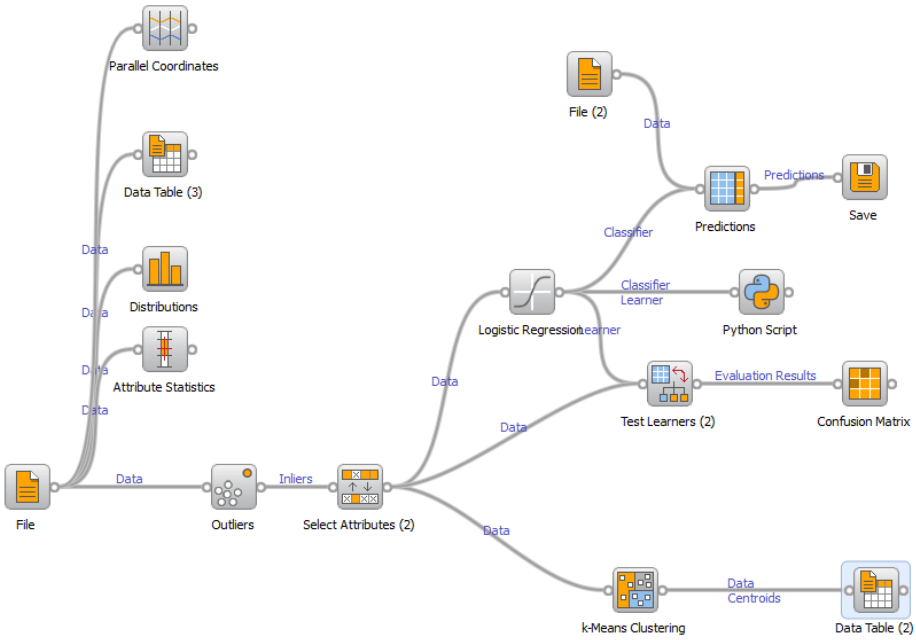

Figure 3. Orange data analysis workflow

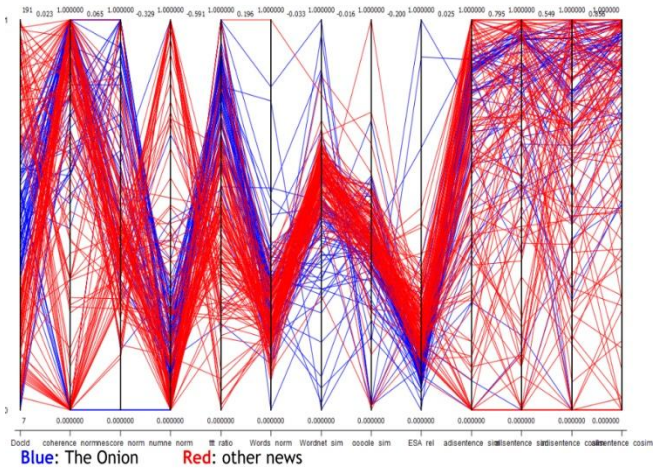

Figure 4. Orange data analysis workflow.

Another interesting result was provided by the investigation of semantic similarities: while ESA proved to have no influence, Wordnet Similarity proved to be the best evidence for creative texts, showing the fact that the concepts are highly connected. This fact gives credit to the definition of creativity in the sense that the more semantic similarity exists between the words, the better qualitative the text is. From the four different options for LSA, the one that proved to be the most correlated with creativity was the average cosine similarity for all sentences. High values for this measure witnessed for non-creative texts, which is natural since LSA reflects the words connections that could be seen in the training corpus (showing a higher "usualness" of the text than the word pairs with smaller LSA scores).

Finally, the analysis of Coherence Measure did not bring anything new, proving that no matter how creative a text is, it should be coherent.

The obtained model was tested in a 10-fold cross validation setup, starting from the assumption that the news taken from the The Onion were creative, while the others were not. The results are presented in Table 1:

TABLE I. EXPERIMENT RESULTS

Values prediction				Confusion matrix	
Real / Predicted	Creative	Non-creative	Sum	Precision	Recall
Creative	46	15	61	0.754	0.6866
Non-creative	21	103	124	0.8306	0.8729
All	67	118	185		

The accuracy for this experiment was 80.54%, which is quite high, considering the difficulty of this task.

B. *Adaptability Experiment*

In order to evaluate the adaptability capacity of our classifier, we tried to evaluate a different type of texts (book reviews taken from [19]) using the same classifier as for the Assessment Experiment. Therefore three masters' students individually evaluated 20 different book reviews, assessing

to each of them a rank between 1 and 3. One was assigned to creative texts, two was assigned to mildly creative texts, while three was assigned to non-creative texts (see Table 2). Unfortunately, the inter-rater agreement Kappa Statistic [20][21] was low (perceived agreement was Po = 0.45), which according to the Kappa interpretation done by Altman [22], was not enough to further consider this ranking. Therefore, in order to improve this situation, we considered instead a binary classification. In order to decide what to do with the mildly creative texts (the ones evaluated with the rank 2), we tested two different situations (evaluating these problematic documents in both possible ways). In the first one, we considered that they were creative, so we evaluated the text as being creative if they formerly received the rank 1 or 2, and non-creative if they received rank 3. Here, the value for the Kappa Statistic was Po = 0.633. In the second situation, we considered that only the texts evaluated with rank 1 were creative and the rest were classified as non-creative (see Table 3). This time, the Kappa Statistic was Po = 0.733. The higher Kappa Statistic score from the second situation gave us a hint that this should be the correct binary classification of the reviews. This decision was also enforced by the fact that, using the majority class (creative/non-creative) amongst the reviewers as the gold standard, in the first situation we ended up with 12 creative texts (out of 20), while in the second we had only 4 texts that were considered to be creative. Since our hypothesis was that there are more non-creative texts than creative ones, the second decision augments the decision made starting from the inter-rater agreement.

After deciding how to consider the reviews initially evaluated with rank 2 and computing the inter-rater agreement, we tried to correlate the output provided by the previously built model with the classes obtained from the reviewers' gold standard evaluations.

Unfortunately, all the reviews were classified as being non-creative, missing 4 creative texts – R5, R6, R9 and R13 – (see Table 3). This might be due to the fact that the built classifier is too specific for The Onion news, and does not find book reviews as creative. However, it should be noted that from these four misclassified reviews, only one was considered creative by all three reviewers. The experiment's accuracy was 80%.

TABLE II. THE RANKS PROVIDED BY THE 2 REVIEWERS FOR THE 20 BOOK REVIEWS CONSIDERING A SCALE WITH 3 VALUES:1 FOR CREATIVE, 2 FOR MILDLY CREATIVE AND 3 FOR NON-CREATIVE

Filename / Reviewer	R1	R2	R3	R4	R5	R6	R7	R8	R9	R 10	R 11	R 12	R 13	R 14	R 15	R 16	R 17	R 18	R 19	R 20
Reviewer1	3	3	2	2	2	2	1	3	2	2	2	2	1	3	3	3	2	1	3	3
Reviewer2	3	3	2	3	1	1	3	3	1	1	2	3	1	1	3	1	2	3	3	2
Reviewer3	2	3	3	2	1	1	3	3	1	2	3	3	1	2	3	2	2	2	2	3

TABLE III. THE FINAL EVALUATION OF THE 20 BOOK REVIEWS

Filename / Evaluation	R1	R2	R3	R4	R5	R6	R7	R8	R9	R 10	R 11	R 12	R 13	R 14	R 15	R 16	R 17	R 18	R 19	R 20
Non-creative	3	3	3	3	1	1	2	3	1	2	3	3	0	2	3	2	3	2	3	3
Creative	0	0	0	0	2	2	1	0	2	1	0	0	3	1	0	1	0	1	0	0

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we presented a model that was intended to discriminate creative from non-creative news articles that was built combining nine different measures. The model could be improved by removing or changing the important assumptions done during the course of this work, such as the The Onion articles being always creative (while the news articles are not) and using just a binary creativity scale: creative and non-creative.

The first part of the experiment presented in this paper delivered the following specific conclusions:

- Word Norms Fraction was the measure that was best correlated with the lack of creativity, which was expected considering the definitions of creativity and of Word Norms Fraction. Google Similarity Distance was in the same situation;
- Named Entities analysis showed that they are signs of a creative text as long as not too many distinct such entities are used;
- Wordnet Similarity proved to be the best evidence for creative texts, while LSA was similar to the measures of Word Norms Fraction and Google Similarity Distance in providing a measure for text "usualness" and therefore giving evidence of non-creative texts. They also have similar weights in the final classifier. ESA had no influence in the built classifier;
- Less coherent texts were expected to be more creative but coherence score was found to have no influence in identifying creativity.

The second part of our experiment investigated the possibility of generalizing the built classifier so that it can be applied to different kinds of texts and/or topics. The difficulty of this task was observed in the very low inter-rater agreement – we believe that more judges are needed to obtain better agreement and build a more robust data set. Also, a finer scale would be useful to cope with the problematic of "how creative" means creative, and to give a better idea of creativity than plain binary values.

 ISBN: 978-1-61208-280-6

The fact that the model built during the first experiment did not consider any review as being creative might be due to the fact that it is tested on a different corpus. It seems also that there are "levels" of creativity according to the analyzed texts: a satire news articles domain may be more creative than books reviews, in general. Thus a bigger data set, comprising different text sources, may achieve better results.

Even though the classifier did not detect creative reviews, the results of both experiments were around 80%, showing that there might be a possibility that the classifier adapts well to different domains and kinds of texts. However, the lack of true positive examples from the second experiment makes us be a little cautious in clearly stating this fact.

These results made us question whether we really identified creativity or we identified a solution to another very difficult problem: satire detection in texts.

The classifier performed reasonably well at differentiating articles from The Onion and from other serious news websites. We believe that increasing the size of the data set, and testing it further, could confirm our assumption. It also shows that satire and creativity are related, since we were searching for creativity but we may have ended up in identifying satire. Previous work has been done about satire detection [23], but increasing the emphasis on semantic similarity, as this work does, could yield better results than those in the referred experiment.

As future work, we plan to verify our assumption related to what make the The Onion articles special (are they expressing creativity, satire, or have we made a wrong assumption considering them to be special?). We intend to do this by using manually classified texts to train the model and then to use it in order to decide whether any of the two assumptions stands and which of them is more adequate.

ACKNOWLEDGMENT

The research presented in this paper was supported by project No. 264207, ERRIC-Empowering Romanian Research on Intelligent Information Technologies/FP7-REGPOT-2010-1.

We would like to thank to the three students that helped us in our experiment (Rajani, Agata and Pavel) and to the reviewers that pointed out an important error that was present in the initial submission. Fixing it greatly improved our results and (we believe that) the paper looks much better now.

REFERENCES

[1] X. Zhu, Z. Xu, and T. Khot, "How creative is your writing? a linguistic creativity measure from computer science and cognitive psychology perspectives," in Proceedings of the Workshop on Computational Approaches to Linguistic Creativity, ACL, 2009, pp. 87-93.

[2] X. Hu and B. Yu, "Exploring the relationship between mood and creativity in rock lyrics," 12th International Society for Music Information Retrieval Conference (ISMIR 2011), 2011, pp. 789-794.

[3] R.L. Cilibrasi and P.M.B. Vitanyi, "The google similarity distance," IEEE Transactions on Knowledge and Data Engineering, 19(3), 2007, pp. 370-383.

[4] E. Gabrilovich and S. Markovitch, "Wikipedia-based semantic interpretation for natural language processing," Journal of Artificial Intelligence Research, 34(2), 2009, pp. 443-498.

[5] About WordNet - WordNet - About Wordnet [online] http://wordnet.princeton.edu/ [retrieved; May, 8, 2013]

[6] J. He, W. Weerkamp, M. Larson, and M. de Rijke, "An effective coherence measure to determine topical consistency in user-generated content," International journal on document analysis and recognition, 12(3), 2009, pp 185-203.

[7] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," Journal of the American society for information science, 41(6), 1990, pp. 391-407.

[8] A. Renouf, "Tracing lexical productivity and creativity in the british media: the chavs and the chav-nots," in Lexical Creativity, Texts and Contexts, John Benjamins Publishing Company, Amsterdam, 2007, pp. 61-89.

[9] T. Veale, "Creative language retrieval: A robust hybrid of information retrieval and linguistic creativity," in Proceedings of ACL 2011, 2011, pp. 278-287.

[10] A. Jordanous, "A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative," Cognitive Computation, 4, 2012, pp. 246-279, ISSN 1866-9956.

[11] Z. Kovecses, "A new look at metaphorical creativity in cognitive linguistics," in Cognitive Linguistics 21(4), 2010, pp. 663-697.

[12] T. Veale, "An analogy-oriented type hierarchy for linguistic creativity," in Knowledge-Based Systems, 19, 2006, pp. 471-479.

[13] A. Lehrer, "Blendalicious," in Munat, J. (Ed.) Lexical creativity, texts and contexts. John Benjamins, Amsterdam, 2007, 115-136.

[14] USF Free Association Norms: Introduction [online] http://web.usf.edu/FreeAssociation/Intro.html [retrieved; May, 8, 2013]

[15] Total Number of Pages Indexed by Google | Statistic Brain [online] http://www.statisticbrain.com/total-number-of-pages-indexed-by-google/ [retrieved; May, 8, 2013]

[16] The Onion - America's Finest News Source [online] http://www.theonion.com/ [retrieved; May, 8, 2013]

[17] Orange – Data Mining Fruitful & Fun [online] http://orange.biolab.si/ [retrieved; May, 8, 2013]

[18] TANAGRA - A free DATA MINING software for teaching and research [online] http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html [retrieved; May, 8, 2013]

[19] Maite Taboada [online] http://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html [retrieved; May, 8, 2013]

[20] J. Cohen, "A Coefficient of Agreement for Nominal Scales" in Educational and Psychological Measurement 20, 1960, pp. 37-46, DOI: 10.1177/001316446002000104.

[21] J. Carletta, "Assessing agreement on classiffication tasks: the kappa statistic," Computational Linguistics, 22(2), 1996, pp. 249-254, ISSN 0891-2017.

[22] D.G. Altman, Practical Statistics For Medical Research. Chapman & Hall, 1991, ISBN 9780412276309.

[23] C. Burfoot and T. Baldwin, "Automatic satire detection: are you having a laugh?," in Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, ACLShort '09, 2009, pp. 161-164.

 ISBN: 978-1-61208-280-6