Department of Computer Science
University of Houston

Analysis of Human Summaries for Automatic Summarization

Araly Barrera and Rakesh Verma

Department of Computer Science
University of Houston
Houston, TX, 77204, USA
http://www.cs.uh.edu

Technical Report Number UHCS-10-04

January 17, 2011

Keywords:

Analysis of human summarization, automatic summarization

Abstract

The current "Information Explosion" necessitates methods to reduce the vast amounts of text found from online sites and other sources through automated summarization. As computationally complex as automated text summarization may be, improved Natural Language Processing methods and closer semantic analysis are progressively used for overall summarization improvement. The purpose of this study is to analyze a dataset of human summaries in order to determine the presence of "markers" that humans use for summarization. The relationships found are intended to be used for improvements to a current automatic summarizer program called, WN-SUM, a fusion of semantic and statistical methodologies, and to other automatic summarization systems as well. Both correlation and regression analysis are performed here through scatter plots derived from psychological experimentation data on human summarization. Briefly, our experiments indicate that: raw sentence position score is not at all a good indicator for sentence selection, the low correlation is surprising; however, there is a significant correlation for a certain kind of sentence position and frequency in human summaries, and the relationship we found is not linear. In keyword matching experiments, the correlations are not as strong. The results obtained in this study have implications for both extractive and abstractive summarization.

Research supported in part by NSF grants CCF 0306475 & CNS 0755500

Analysis of Human Summaries for Automatic Summarization

Araly Barrera and Rakesh Verma

Abstract

The current "Information Explosion" necessitates methods to reduce the vast amounts of text found from online sites and other sources through automated summarization. As computationally complex as automated text summarization may be, improved Natural Language Processing methods and closer semantic analysis are progressively used for overall summarization improvement. The purpose of this study is to analyze a dataset of human summaries in order to determine the presence of "markers" that humans use for summarization. The relationships found are intended to be used for improvements to a current automatic summarizer program called, WN-SUM, a fusion of semantic and statistical methodologies, and to other automatic summarization systems as well. Both correlation and regression analysis are performed here through scatter plots derived from psychological experimentation data on human summarization. Briefly, our experiments indicate that: raw sentence position score is not at all a good indicator for sentence selection, the low correlation is surprising; however, there is a significant correlation for a certain kind of sentence position and frequency in human summaries, and the relationship we found is not linear. In keyword matching experiments, the correlations are not as strong. The results obtained in this study have implications for both extractive and abstractive summarization.

Index Terms

Analysis of human summarization, automatic summarization

I. Introduction

The World Wide Web brings forth an alarming rate of text documents (from articles to research papers, etc.) on a daily basis, making it difficult for people to keep up. Thus, as complex as automatic text summarization can be, it has become increasingly essential to combat the information explosion. The goal of automatic summarization is to be able to compress large amounts of text that people encounter everyday into summaries for ease of reading and understanding. There are two main types of summarization: abstractive – the construction of original sentences, and extractive – the concatenation of source sentences. Our focus in this study is to analyze the effects that presence of text *markers* in articles have as humans are asked to construct summaries for them. Here, we analyze a human-generated summary dataset provided to us from prior psychological studies and use it to make a stronger statistical investigation. Besides the intrinsic interest of such analysis, the relationships found through such analysis are intended to be used to improve WN-SUM [4], an extractive summarizer, and benefit future summarizers as well.

WN-SUM is an automated summarizer, which has been developed for the construction of extractive summaries of a given article. WN-SUM selects sentences considered most important (and best representations of an article's theme) are selected from the source text to serve as its summary. The approach to sentence selection is currently based on the computation of a sentence score derived from the implementation of several linguistic analysis tools such as WordNet [15], SenseLearner [13], and TextRank [12]. More details on WN-SUM are found in [4] but are also provided under the WN-SUM section. Although WN-SUM is only extractive, some of our insights would be useful for abstractive summarization as well as when the abstractive summary depends on variables studied here.

The analysis presented in this paper is based on a 1996 psychological experiment performed by Lorch and Lorch, as described in [9]. The experiment consisted of a summarization task assigned to a group of participants, who were each provided a version of an article that either contained headings or not. The main purpose was to determine the effects that article headings, or other *signaling* devices, present within the text had on text recall and on conceptual understanding of the article. This data, in turn, allowed us to perform a more extensive correlation and regression analysis whose findings can help improve the scoring sentence methodology of WN-SUM.

Research supported in part by NSF grants CCF 0306475 & CNS 0755500

The rest of the paper is organized as follows. In the next section, we give a brief introduction to WN-SUM. In Section we give the characteristics of the dataset used for our analysis. Sections presents the results of the correlation and regression analysis perfomed by us. Section discusses related work and Section presents modifications made to WN-SUM. Section concludes the paper and Section provides all additional plots generated in this study.

II. WN-SUM

The WN-SUM system is based on the document map framework used for the construction of a document's extractive summary. We intend to improve the performance of this summarizer by adjusting its current systematic methods to more closely match human tendencies on sentence extraction. The basic premise of the current algorithm is on the selection of a document's most "thematic sentences" as a representation of its summary. This process of sentence selection is based on a final sentence score obtained from three individual scoring methods. Sentence position, popular keyword presence, and semantic relationships are all used to compute a sentence's final score and overall importance to a given article. The following is a basic overview of WN-SUM but refer to [4] for more details:

1) A single article or text document is taken.
2) Sentences are separated and words are lexically and grammatically analyzed. Named entity recognition, parts of speech tagging, and WordNet sense tagging are all performed as a preprocessing step. Named entity tagging is done with Standford's Named Entity Recognizer 1.0 [18], part of speech tagging is performed with Stanford POS tagger [19], and WordNet sense identification is accomplished using SenseLearner 2.0 [13].
3) Sentences are analyzed and assigned a score based on:
    - **Sentence Position -** Sentences at beginning and end of document are weighted more based on the hypothesis that introductory and concluding sentences are better candidates for final summaries.
    - **WordNet hypernymic distance -** The hypernymic distance of words in a sentence and to other sentences are used to measure how close a word is to its root form. The assumption is that sentences containing words closer to their root forms are most likely to be thematic sentences.
    - **TextRank Keyword Rank -** Algorithm that ranks words based on the popularity of them within a text document [12].
4) Scores are combined into a final weighted sentence score.
5) Highest scoring sentences are selected for final summary.

III. The Dataset

Ninety-nine participants in the original 1996 experiment described in [9] were asked to read an article titled, ”Energy Problems and Solutions" (four single-spaced typewritten pages in size) and in turn demonstrate their understanding of the text through a written 15-sentence summary. Though each reader received only one article to summarize, two main versions of the article were distributed, *A* and *B*, each containing a total of 20 topics. Both versions of the article contained identical topics but were intentionally reordered among each other (this to examine the topic selection behavior of readers). Each version was further divided into two sub-versions, representing either the article versions that contained headings, *YA* and *YB*, or the corresponding versions that did not, *NA* and *NB*. Hence, four different article versions containing varying number of sentences were randomly assigned among the readers. The articles' sizes are as follows: *YA* containing 227 sentences, *NA* with 205, *YB* with 236, and *NB* with 214 sentences. Among all ninety-nine distributed articles, 25 articles represented each article version with the exception of *NA*, containing 24.

We use Lorch's original experimental data, in turn, to conduct a more detailed statistical study on various factors including topic headings, sentence positions, popular keyword presence in sentences, and the relationships found to each other. Our intentions are to use these findings as a means of better simulating human-extraction in automated summarization.

The entire article dataset, along with the resulting generated summaries were provided by Lorch and described further below. Figure 1 provided specific article details as well as distribution numbers among participants. The following items composed the dataset used for the analysis presented here.

1) Four versions of the article with distinct conditions (Figure 1).

Article Version	Topic Total	Headings Included (Section and Topic)
YA	20	Yes
NA	20 (YA Topics)	No
YB	20	Yes
NB	20 (YB Topics)	No

Fig. 1. Article Data Description

2) Ninety-nine human-written summaries for their corresponding articles composed of 15 sentences each.

The analysis we present utilized this dataset. It began with the enumeration of all sentences (including the title and the various topic headings present) within each article version as a means of representing article sentence positions. The human-generated summary sentences were then manually compared to the corresponding article sentences for matching. These matches represent the article sentence extractions the participants used as their summaries. The resulting sentence extraction frequencies became the basis of our statistical analysis.

Our investigation is decomposed into the following two tasks:

- ***Correlation Analysis:*** To determine the presence of specific indicators humans use when extracting summaries.
- ***Regression Analysis:*** Correlations found in the first task are then studied further to determine functions based on these indicators for their implementation to WN-Sum and possibly other summarizers.

IV. Correlation and Regression Analysis

Our dataset was used to analyze four different conditions for correlation strength. Correlation was measured with Pearson's correlation coefficient (represented by ρ), with values -1 and 1 representing strongest negative and strongest positive relationships, respectively. This analysis was divided between article sentence and keyword levels.

Figures 6-29 show the strongest correlated conditions that proceeded onto the regression analysis stage to find approximated regression trends. Linear, logarithmic, and quadratic regression lines were the three forms of lines considered. Logarithmic and quadratic regression trends were primarily found as closest relationship approximations. Logarithmic relationships were found superior (resulted with higher R^2 values) to quadratic. Explanation of variance coefficient or, informally speaking, closeness of fit (represented by R^2), and regression line standard error for coefficients are provided as well.

Sentence level tests were based on sentence extraction frequencies. Once *YA*, *NA*, *YB*, and *NB* article sentences were enumerated with position values starting from 0 (for the title), the human-generated summaries were then taken to a matching stage. This matching stage simply consisted on a tally of an article's sentences used in a summary (the tally of sentence positions) to determine sentence extraction frequencies for each article version. Sentence extraction frequencies for all ninety-nine summaries and all four articles were collected and used in this part of the analysis.

Keyword level tests, on the other hand, considered the frequency of the top 100 most popular article keywords present in the summaries. Each article was first processed by the TextRank algorithm [12], for the collection of the top 100 ranked words for residing in each of them. Summary sentence words were then compared to the corresponding article's TextRank popular keywords for matches and frequencies for these were computed as well.

The following were the condition pairs tested for each article and the resulting plots containing the correlation coefficient value obtained from the analysis.

Sentence Level Tests:

Figures 2 - 13 show plots based on the frequencies of human sentence extractions versus conditions such as sentence position alone, positional distance from a previous heading, and positional distance from a nearest heading. Because logarithmic curves proved closest plot trends in regression analysis, only logarithmic regression lines

Matching was based on relative thematic closeness of the two sentences due to the abstractive nature of the human summaries for each article version are provided below. Refer to Appendix (Section) to view all other regression line tests performed.

1) **Frequency vs. Sentence Positions (Figures 2, 3, 4, & 5):**
   Observations: Based on low correlation and diffused positioning of points shown for Figures 2, 3, 4, and 5 in all article tests, no direct relationships were observed for the Sentence Frequency vs. Sentence Position tests.
   Based on the observations made last, we now calculate the sentence distances from a previous heading (that is, the number of sentences away from a proceeding heading). We also consider sentence distance from any nearest heading, whether occurring before or after the sentence, again in terms of number of sentences. For each distance value in each article version, we gather corresponding sentence extraction frequencies for comparisons.

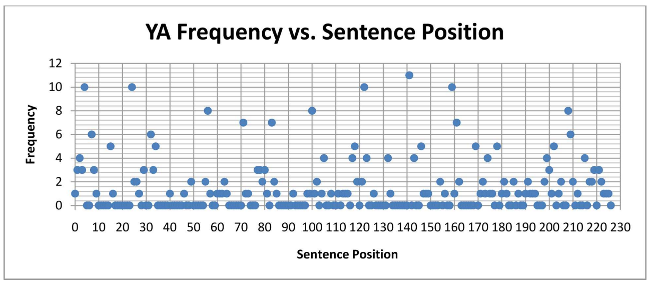

Fig. 2. YA Frequency vs. Sentence Position Plot with ρ= 0.01387

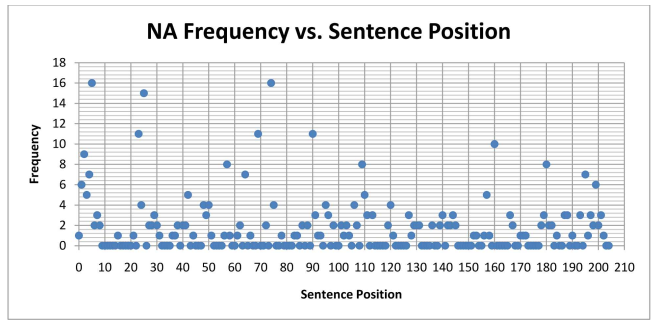

Fig. 3. NA Frequency vs. Sentence Position Plot with ρ= -0.13109

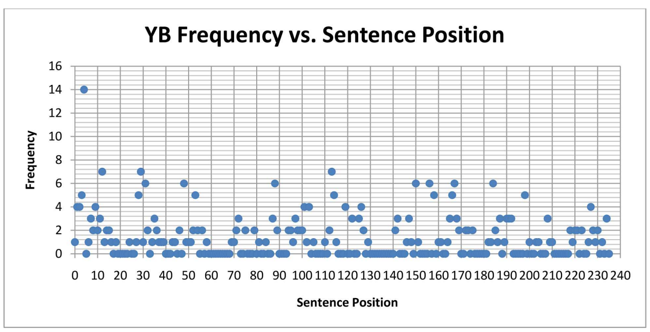

Fig. 4. YB Frequency vs. Sentence Position Plot with ρ= -0.14595

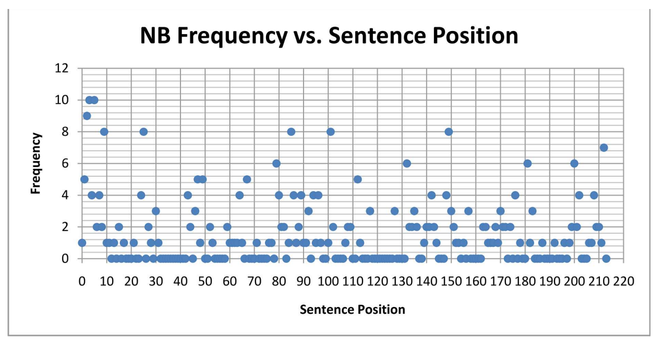

Fig. 5. NB Frequency vs. Sentence Position Plot with ρ= -0.11443

2) **Frequency vs. Distance from Previous Headings (Figures 6, 7, 8, & 9):**
   Observations: Figures 6, 7, 8, and 9 show stronger correlation coefficients. Therefore, the sentence frequency and its distance from a previous heading have a direct relationship for all article tests. This test thus moved on to the regression analysis to find the best fit curve.
3) **Frequency vs. Distance from Nearest Headings (Figures 10, 11, 12, & 13):**
   Observations: Figures 10, 11, 12, and 13 also demonstrate high correlation coefficients for a sentence's frequency and distance from its nearest topic heading. Noticeable curves such as the one for *YB* (Figure 12) with correlation coefficient -0.8726 reinforce the relationship for this test and therefore moved on to the regression analysis.

Keyword Level Tests:

The TextRank algorithm was used here to determine the *top 100* ranked words for each article. Sentences in articles were then ranked based on their frequencies in human sentence extractions. That is, those ranked highest with number 1, were those sentences extracted most often in the human summaries. If two sentences resulted with the same frequency, the ranks were then averaged as is commonly done in statistical analysis of ranks. For

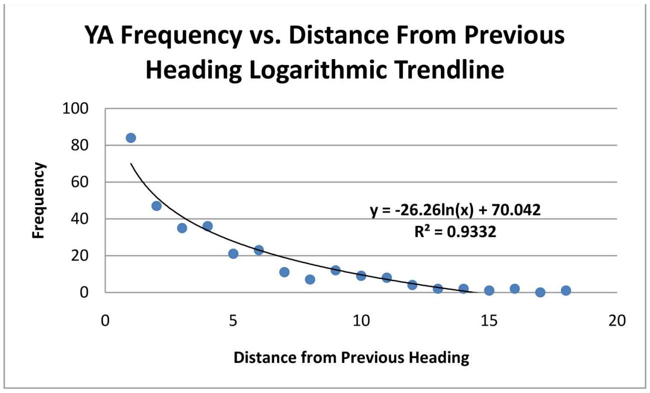

Fig. 6. YA Frequency vs. Distance From Previous Heading with ρ= -0.7167 & closest regression function y= -26.26$ln(x)$ + 70.042 with Standard Error= 12.538 and R^2= 0.9332.

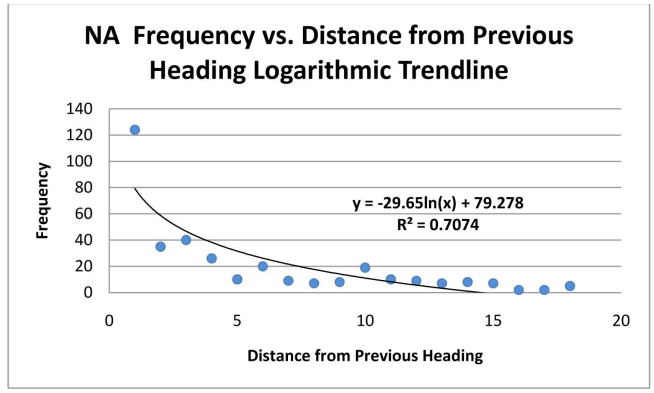

Fig. 7. NA Frequency vs. Distance From Previous Heading with ρ= -0.6489 & closest regression function y= -29.65$ln(x)$ + 79.278 with Standard Error= 21.781 and R^2= 0.7074.

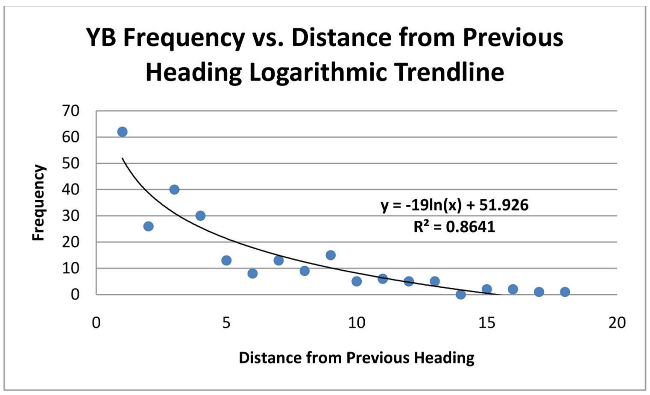

Fig. 8. YB Frequency vs. Distance From Previous Heading with ρ= -0.8268 & closest regression function y= -19$ln(x)$ + 51.926 with Standard Error= 9.9023 and R^2= 0.8641.

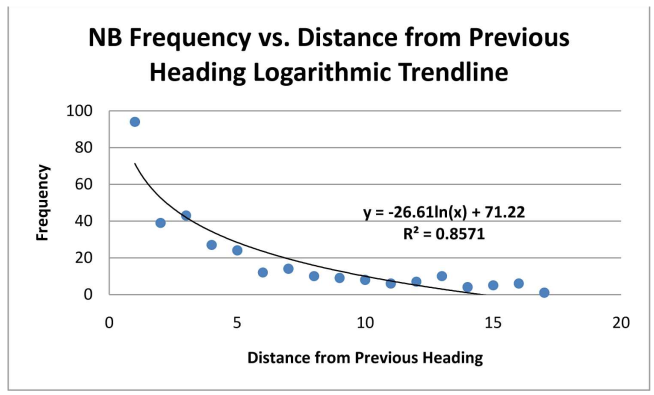

Fig. 9. NB Frequency vs. Distance From Previous Heading with ρ= -0.7627 & closest regression function y= -26.61$ln(x)$ + 71.22 with Standard Error= 15.0847 and R^2= 0.8571.

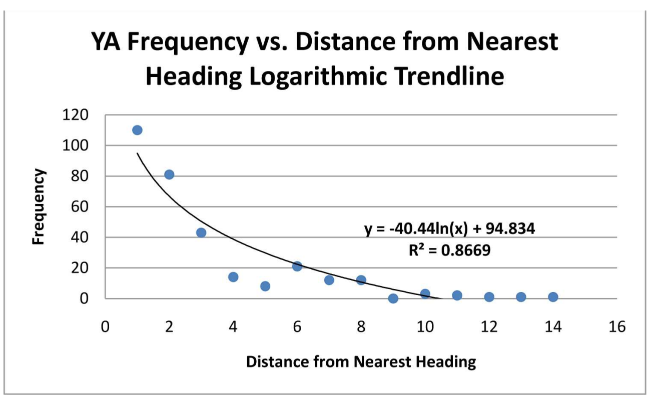

Fig. 10. YA Frequency vs. Distance From Nearest Heading with ρ= -0.7785 & closest regression function y= -40.44$ln(x)$ + 94.834 with Standard Error= 21.9618 and R^2 = 0.8669.

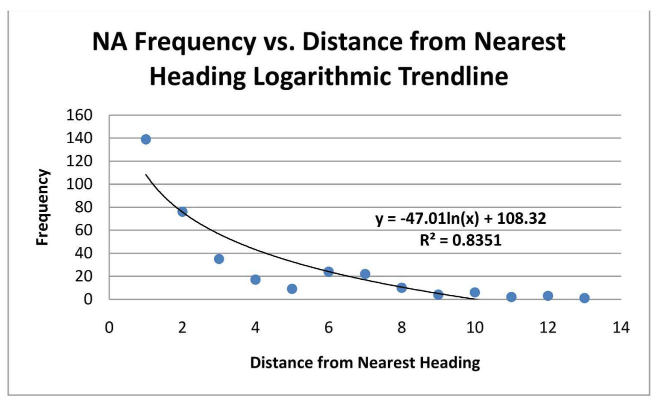

Fig. 11. NA Frequency vs. Distance From Nearest Heading with ρ= -0.7510 & closest regression function y= -47.01$ln(x)$ + 108.32 with Standard Error= 27.150 and R^2= 0.8351.

example, if two sentences were tied for ranks 4 and 5, they were given the average rank of 4.5. The frequency of top 100 words contained in the sentences were totaled and either normalized or left raw as shown in the tests below. Normalization was calculated with the following:

$$KeywordMatchNormalization = \frac{\sum Top100KeywordMatch}{\sum Numberofwordsinsentence} \tag{1}$$

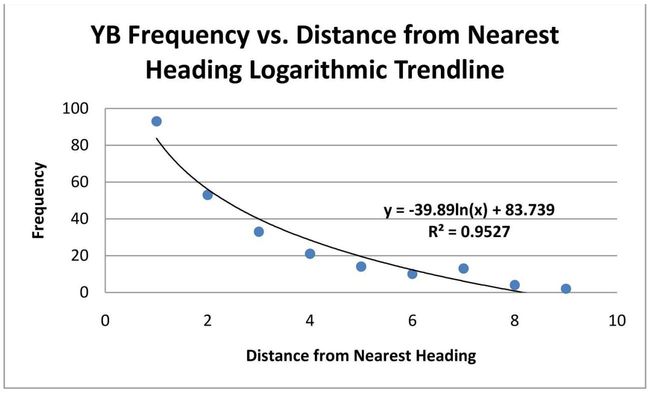

Fig. 12. YB Frequency vs. Distance From Nearest Heading with ρ= -0.8726 & closest regression function y= -39.89$ln(x)$ + 83.739 with Standard Error= 15.343 and R^2= 0.9527.

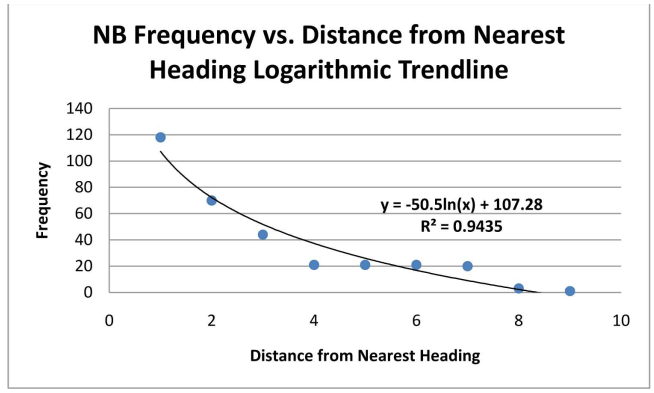

Fig. 13. NB Frequency vs. Distance From Nearest Heading with ρ= -0.8751 & closest regression function y= -50.5$ln(x)$ + 107.28 with Standard Error= 19.38 and R^2= 0.9435.

- Normalization Matches - For these tests, an article's sentences were each ranked based on equation (1) to compute the presence of *top 100* keywords generated by TextRank and normalized with respect to the total sentence keyword count. Duplicate rankings would be resolved by taking corresponding values computed by (1) and averaging those to form a single rank.
- Raw Matches - For these tests, an article's sentences were each ranked based on equation (1) to compute the presence of *top 100* keywords generated by TextRank and left un-normalized. Duplicate rankings would also be resolved by taking corresponding values computed by (1) and averaging those to form a single rank.

The following keyword level tests were analyzed using the scatter plots below. Pearson's correlation coefficients are provided as well (represented with ρ). Figures 14-21 plot the keyword matches against the sentence rank. Next, we present plots shown in Figures 21 through 29 for raw and normalized keyword weights versus the sentence rank. The weights are obtained using the TextRank weight for each keyword and normalization is done using the sum of the weights for the words (those that are in the top 100) in a sentence rather than the length of a sentence. As in the sentence-level tests, only logarithmic regression lines are shown below but all others are provided in the Appendix, Section .

1) **Normalized Keyword Frequency vs. Sentence Rank (Figures 14, 15, 16 & 17):**
   Observations: The normalization of keyword frequency vs. sentence ranks as shown in Figures 14, 15, 16, and 17 demonstrate a reasonably strong relationship in this test. Although most correlation coefficients here result lower than those seen in the sentence-level tests, we nonetheless move these tests up to regression analysis stage.

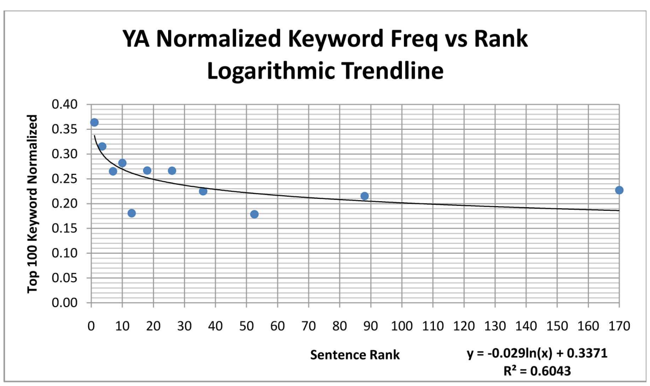

Fig. 14. YA Normalized Keyword Frequency vs. Sentence Rank with ρ= -0.4379 & closest regression function y= -0.029$ln(x)$ + 0.3371 with Standard Error= .0528 and R^2= 0.6043.

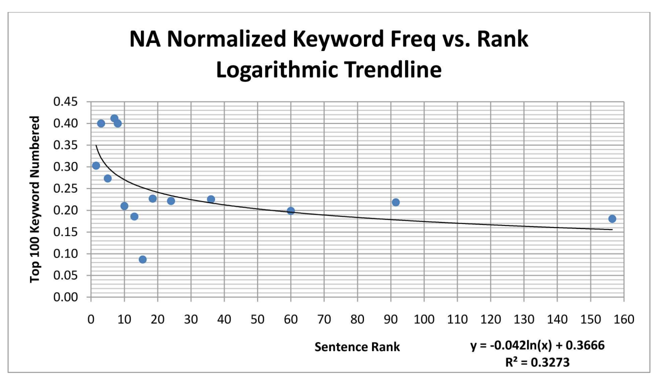

Fig. 15. NA Normalized Keyword Frequency vs. Sentence Rank with ρ= -0.3884 & closest regression function y= -0.042$ln(x)$ + 0.3666 with Standard Error= 0.0910 and R^2= 0.3273.

2) **Raw Keyword Frequency vs. Sentence Rank (Figures 18, 19, 20, & 21):**

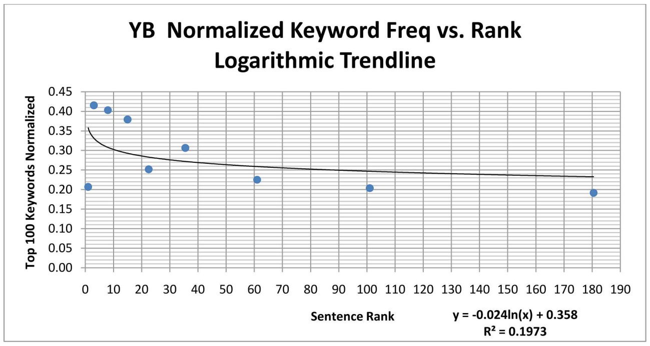

Fig. 16. YB Normalized Keyword Frequency vs. Sentence Rank with ρ= -0.6208 & closest regression function y= -0.024$ln(x)$ + 0.358 with Standard Error: 0.0763 and R^2 = 0.1973.

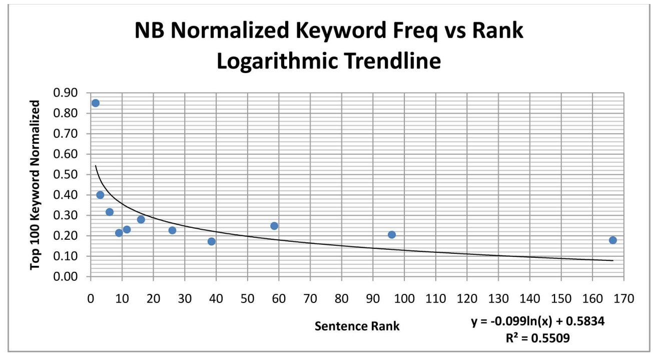

Fig. 17. NB Normalized Keyword Frequency vs. Sentence Rank with ρ= -0.4131 & closest regression function y= -0.099$ln(x)$ + 0.5834 with Standard Error: 0.1855 and R^2 = 0.5509.

Observations: Figure 19 (*NA*) shows a lower correlation coefficient than those resulting in the normalized keyword frequencies (Figures 14, 15, 16, and 17). Correlation strengths for these conditions do not show sufficient relation but moved on to the regression analysis nonetheless.

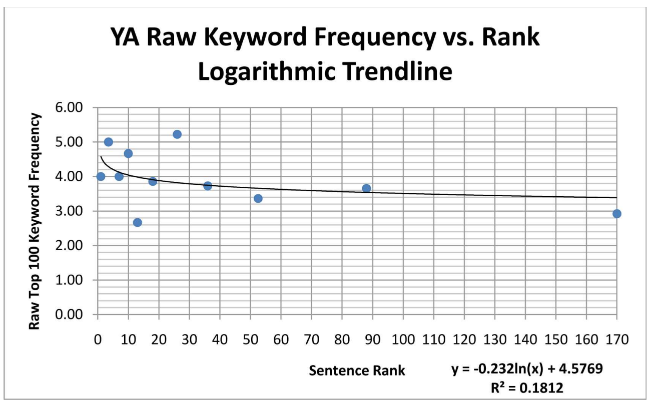

Fig. 18. YA Raw Keyword Frequency vs. Sentence Rank with ρ= -0.4917 & closest regression function y= -0.232$ln(x)$ + 4.5769 with Standard Error= 0.7355 and R^2= 0.1812.

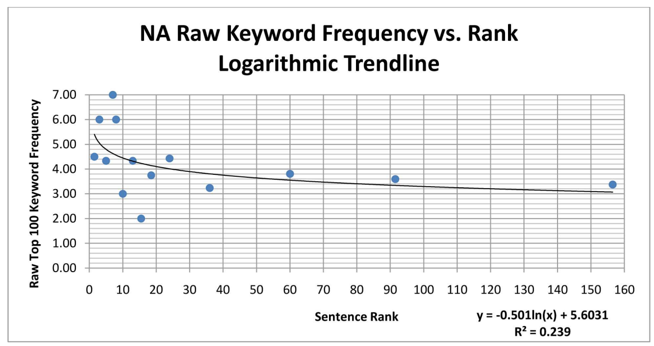

Fig. 19. NA Raw Keyword Frequency vs. Sentence Rank with ρ= -0.3523 & closest regression function y= -0.50$ln(x)$ + 5.6031 with Standard Error= 1.2981 and R^2= 0.239.

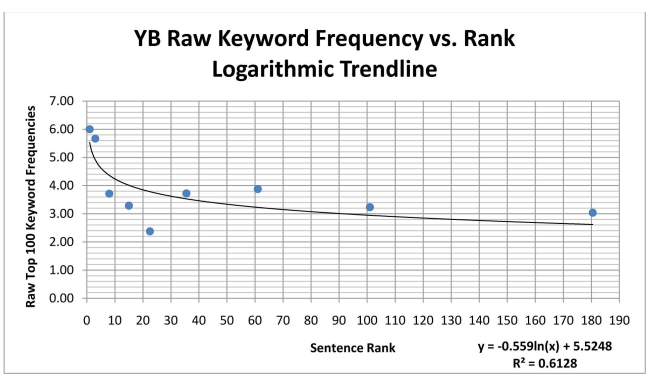

Fig. 20. YB Raw Keyword Frequency vs. Sentence Rank with ρ= -0.4556 & closest regression function y= -0.559$ln(x)$ + 5.5248 with Standard Error= 1.1403 and R^2= 0.6128.

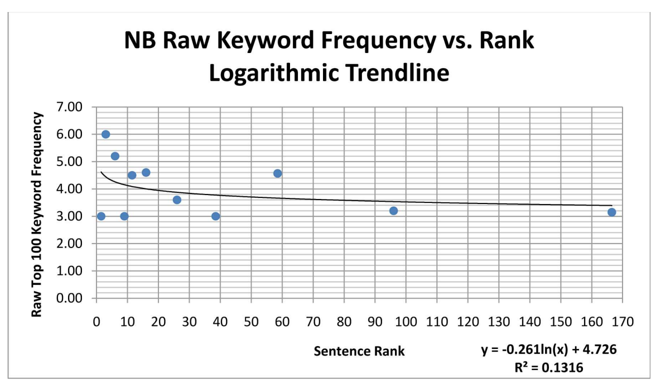

Fig. 21. NB Raw Keyword Frequency vs. Sentence Rank with ρ= -0.393 & closest regression function y= -0.261$ln(x)$ + 4.726 with Standard Error= 1.0118 and R^2 = 0.1316.

3) **Normalized Keyword Weight vs. Sentence Rank (Figures 22, 23, 24, & 25):**
Observations: Figures 22, 23, 24, and 25 show fairly strong correlations such as that found in *YB* (Figure 24) with 0.6601. This therefore indicates a significant pattern in higher popular keyword use within most

frequently extracted sentences.

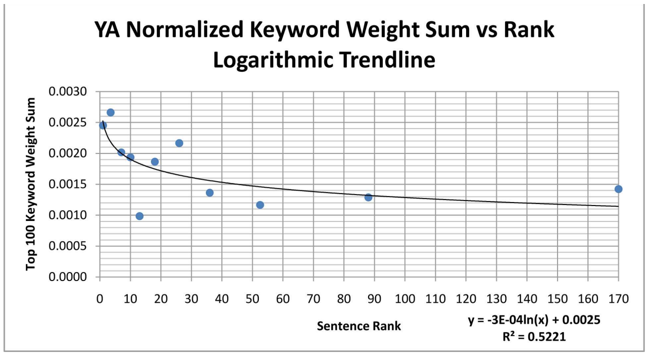

Fig. 22. YA Normalized Keyword Weight Sum vs. Sentence Rank with ρ= -0.4794 & closest regression function y= -3E-04$ln(x)$ + 0.0025 with Standard Error= .000507 and R^2= 0.5221.

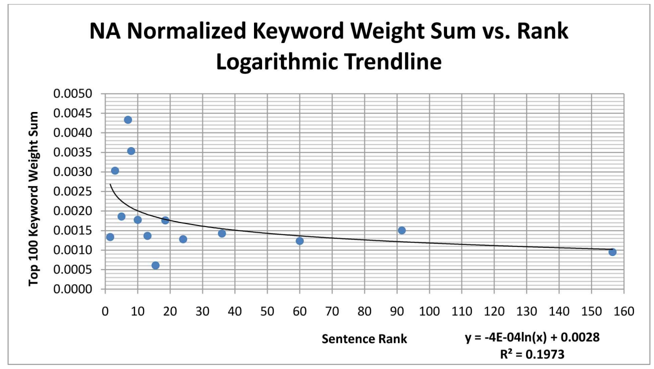

Fig. 23. NA Normalized Keyword Weight Sum vs. Sentence Rank with ρ= 0.3954 & closest regression function y= -4E-04$ln(x)$ + 0.0028 with Standard Error= 0.001002 and R^2= 0.1973.

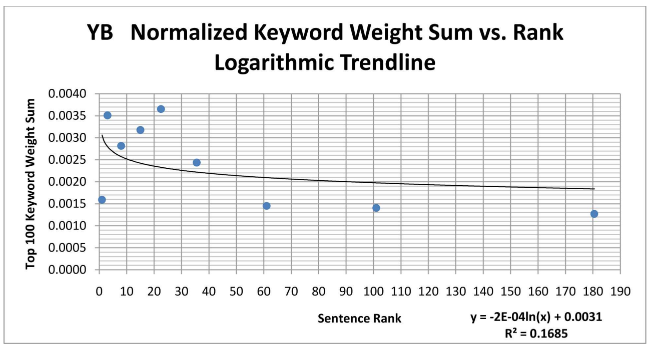

Fig. 24. YB Normalized Keyword Weight Sum vs. Sentence Rank with ρ= 0.6601 & closest regression function y= -2E-04$ln(x)$ + 0.0031 with Standard Error= 0.00077 and R^2= 0.1685.

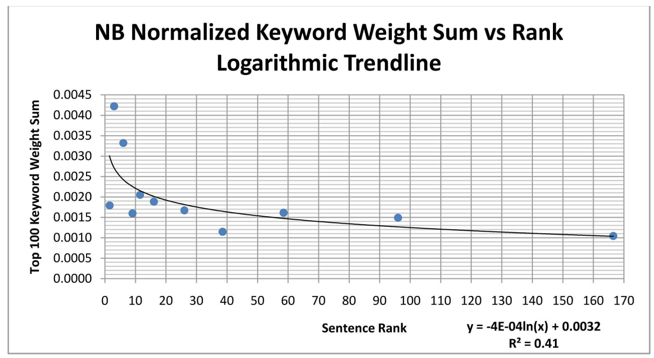

Fig. 25. NB Normalized Keyword Weight Sum vs. Sentence Rank with ρ= 0.5276 & closest regression function y= -4E-04$ln(x)$ + 0.0032 with Standard Error= 0.000852 and R^2 = 0.41.

4) **Raw Keyword Weight vs. Sentence Rank (Figures 26, 27, 28 & 29):**
Observations: Figures 26, 27, 28, and 29 for raw keyword weight sum versions also show strong correlations in *YA* (Figure 26) with -0.5128 and *YB* (Figure 28) with -0.6726. When comparing these to the normalization of keyword weights previously mentioned, both sets of plots indeed demonstrate a stable relationship when the popular keywords are found in highly ranked sentences.

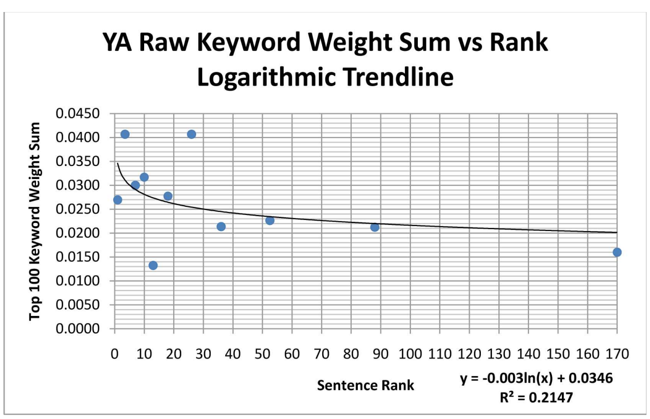

Fig. 26. YA Raw Keyword Weight Sum vs. Sentence Rank with ρ= -0.5128 & closest regression function y= -0.003$ln(x)$ + 0.0346 with Standard Error= 0.008085 and R^2= 0.2147.

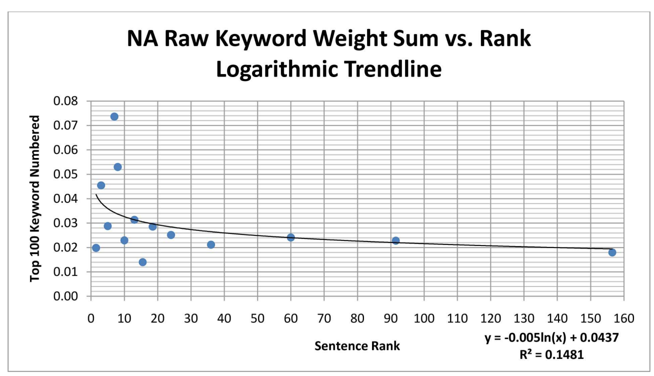

Fig. 27. NA Raw Keyword Weight Sum vs. Sentence Rank with ρ= -0.3707 & closest regression function y= -0.005$ln(x)$ + 0.0437 with Standard Error= 0.01567 and R^2 = 0.1481.

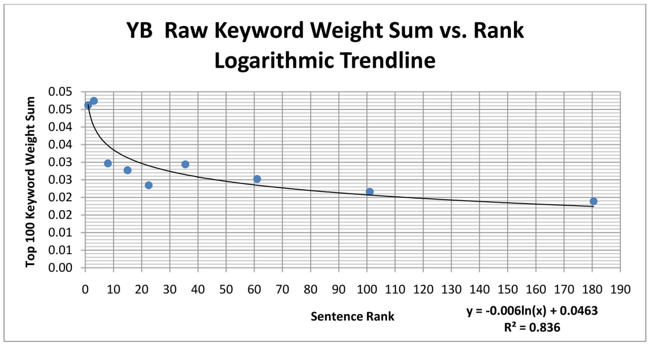

Fig. 28. YB Raw Keyword Weight Sum vs. Sentence Rank with ρ= -0.6726 & closest regression function y= -0.006$ln(x)$ + 0.0463 with Standard Error= 0.00806 and R^2= 0.836.

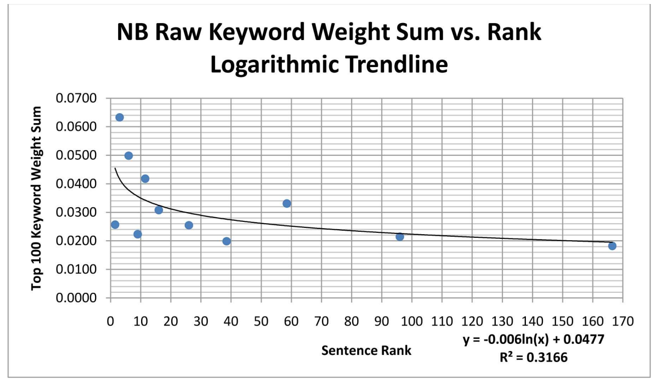

Fig. 29. NB Raw Keyword Weight Sum vs. Sentence Rank with ρ= 0.5071 & closest regression function y= -0.006$ln(x)$ + 0.0477 with Standard Error= 0.01291 and R^2= 0.3166.

Discussion: Overall, we see that sentence position with respect to distances from nearest and previous headings showed the strongest correlation. The high correlations observed for *NA* and *NB* (those article versions that did not contain headings) for this variable also show that human readers can pick up on topic changes easily, even without the use of topic headings. For this, Lorch and Lorch suggest that readers have strong abilities in distinguishing article thematic changes occurring throughout a document and are equally likely to prioritize thematic-driven content as with those that did see topic headings.

The next strongest correlations were observed for keyword weights normalized and raw (un-normalized), rather than keyword matches. This indicates that keyword weighting schemes such as TextRank are likely to strengthen automated summarizers.

V. Related Work

Sentence position has been considered important to summarization and information extraction ever since the late 1950s [1]. Many researchers have proposed using it for automatic summarization, e.g., see [3], [14], [7], [6] and [10]. The importance of sentence position in *book length* documents was studied by [11], which are outside the scope of our study. Most researchers use sentence position based on their subjective knowledge of the language in which the document is written. Many use a linear function of the sentence position [7], [6] or sentence position with respect to a centroid sentence [14], others use either the first few sentences in a paragraph or the document. But, to our knowledge, this is the first objective study that attempts to analyze human summary data for a "newspaper-length" article. Moreover, our work shows the importance of considering derived variables from the sentence position, not just the raw sentence position, and we observe a logarithmic relationship rather than a linear relationship.

The importance of keywords or key phrases for summarization is also well-recognized since at least Edmundson's work [3]. Again, many researchers have proposed using it for automatic summarization, e.g., [12], [4], etc. Again, our work represents the first objective study, to our knowledge, of actual human summary data. Our work shows that weighted keyword schemes such as [12] are likely to perform better in automatic summarization rather than raw counts of keywords [4].

VI. Keyword Match and Nearest Distance Analysis

The following plots represent a deeper look at whether sentence position is a better predictor than keyword matches for inclusion in human summaries. In the plots for fixed number of keyword matches, we plot the sentence distance from nearest heading versus the frequency. Since there are only 25 summaries per document, to get a reasonable number of sentences with the same number of keyword matches, we considered all four articles together. From the correlation analysis performed here, results demonstrated stronger correlation values for higher selection of sentence positions closer to topic headings than the presence of most popular keywords among the sentences. The lower correlation values and observable inconsistencies among the keyword matching tests demonstrate that the extraction of sentences does not have sufficient correlation to the ranking of keyword popularity among an article.

Nearest sentence distances to topic headings were compared against sentence frequencies for all article versions combined (Figures 30 and 32). Keyword match numbers were left fixed for closer behavior analysis and reinforced the assumptions that sentence positions closer to topic headings had higher selection frequencies.

Keyword matching was also compared against sentence frequencies for all article versions combined (Figures 31 and 33). Matching consisted of raw counts among TextRank-generated and topic heading keywords within the article sentences. Nearest Distances were left fixed and showed insufficient relationship to selection of sentence frequencies.

- *Topic Heading Matching:*

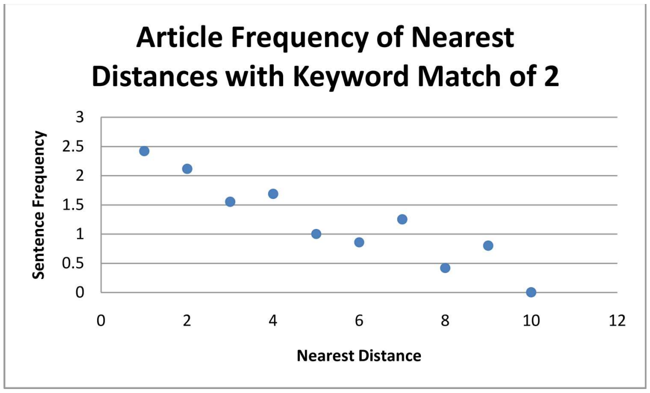

Fig. 30. Nearest Distance vs. Sentence Frequencies with Topic Keyword Match of 2 fixed with ρ= -0.93279

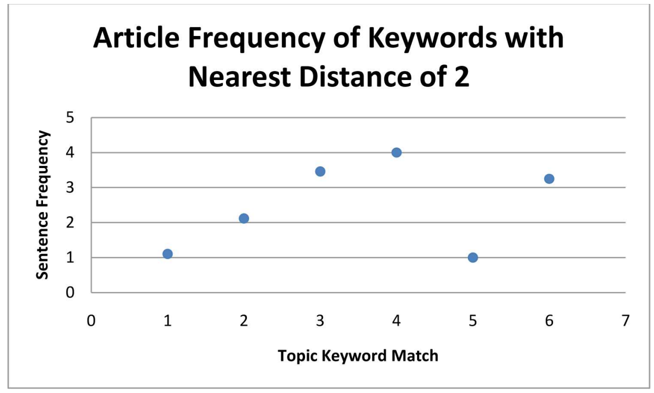

Fig. 31. Topic Keyword Match vs. Sentence Frequencies with Nearest Distance of 2 fixed with ρ= 0.333408

- *TextRank Keyword Matching:*

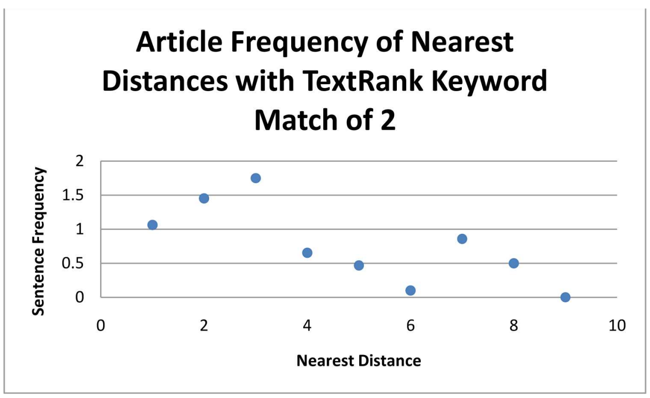

Fig. 32. Nearest Distance vs. Sentence Frequencies with TextRank Keyword Match of 2 fixed with ρ= -0.31398

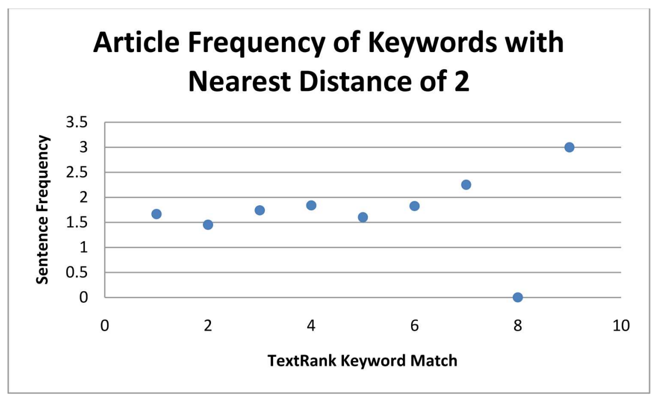

Fig. 33. TextRank Keyword Match vs. Sentence Frequencies with Nearest Distance of 2 fixed with ρ= 0.382985

VII. WN-SUM MODIFICATIONS

The following were changes and modifications made to WN-SUM. None of these reflect any of the findings presented here, but were minor changes made to further improve the summarizer prior to making extensive changes to its current methodologies. These include making changes to WN-SUM's TextRank module and removing topic headings at certain points in the program. Changes reflecting findings from our statistical analysis and their performance effects on sentence extraction are left for future work.

A. TextRank Module Changes

Sentences containing most of a document's popular keywords have been best known to be used for its summary. Therefore, the TextRank implementation (based on Mihalcea's algorithm [12]) integrated in WN-SUM works to extract a document's major keywords to determine best candidate sentences and the composure of the final summary. This method essentially assigns a score to words and sentences within the document to represent the relevance to the topic of the text. It also ranks these words and sentences based on the weight of the scores from highest to lowest.

After a number of test runs and careful observations made on the current TextRank program, it was determined that letter case was an issue that would give erroneous word ranking results. For instance, words such as "oil" and "Oil" may have the same word sense and meaning, but would result with different TextRank scores and ranking values due to the program's failure to match lower and uppercase letters as in the letter "o" in words "oil" and "Oil." To resolve this case-sensitive flaw, the program was modified to lower-case all words that were not tagged as named entities and began a sentence. All other words would be left as is for proper word matching.

Another observation made on TextRank was on its use on parentheses tagging as valid words. Prior to named entity and part of speech tagging of words, the program eliminates punctuations and other non-letter characters from further processing. However, parentheses characters such as '(' and ')' were tagged as "LRB" to represent a left parentheses and "RRB" to represent a right parentheses and were not eliminated from the valid word set. Thus, the program would consider these tags as acceptable working words rather than garbage ones and would then proceed onto the calculations of scores and rankings. To resolve this flaw, the program was modified so that words identified as "LRB" and "RRB" were eliminated before their storage in the valid word set prior to continued TextRank processing.

B. Heading Filtering

A major observation made in WN-SUM was on its lack of topic heading removal at times when the presence of headings was unwanted. For instance, numerous initial WN-Sum executions resulted with topic heading extractions for summaries when articles contained topic headings. Topic heading presence in a summary certainly diminishes the quality of the summary and is thus a very important thing to avoid. To resolve this problem, a parser used for the identification incomplete sentences, as in those of headings, would be required to be implemented for their elimination. The LinkGrammar parser [16] was therefore used to parse the articles before their implementation into WN-SUM so that all heading sentences such as "Greenhouse Effect" would be extracted out of the document before it proceeded onto the computation of its final summary. Lingua LinkParser [17], LinkGrammar's API written in Perl, was the package used to accomplish the heading elimination task.

VIII. CONCLUSIONS

The analysis performed on human summaries demonstrated that sentences closer to topic headings and other related "signaling" devices present in an article are more frequently used for summary extraction. Sentences containing an article's most popular words are also prioritized by readers for sentence selection, those of which used to represent the article's final summary. The observations of these direct relationships and the functions obtained from the regression analysis here will be implemented to the WN-SUM summarizer for future evaluations. We, for instance, find especially valuable to consider the prioritization of sentences closer to headings, instead of those found only in introductory and concluding sections as implemented in the current version of WN-SUM. Additionally, thematic content and semantic closeness to thematic content are also important article aspects to consider for the extraction of better quality sentences. Overall, the intention of this analysis has been take greater into account important human extraction tendencies for high quality automated summarization.

IX. ACKNOWLEDGEMENTS

This work was supported by the University of Houston's Computer Science NSF REU Program with NSF grants CCF 0306475 & CNS 0755500, University of Houston's LSAMP Bridge To Doctorate Program, and through Grant no. DUE0737404. The authors would also like to thank Dr. Lorch's for his contributions and for providing the data used for this analysis.

References

[1] Baxendale, P. (1958). Machine-made index for technical literature - an experiment. IBM Journal of Research Development, 2(4):354-361.
[2] Cheung, Jackie CK. (2000). Comparing Abstractive and Extractive Summarization of Evaluative Text: Controversiality and Content Selection, 133.
[3] Edmundson, H. P. (1969). New methods in automatic extracting. Journal of the ACM, 16(2):264-285.
[4] Filozov, F. and Verma, R.M. (2010). Document Map and WN-SUM: A New Framework for Automatic Text Summarization and a First Implementation. UH-CS-10-03.
[5] Fox, John. Nonparametric Simple Regression: Smoothing Scatterplots. Iowa City: Sage Publications, 2000.
[6] Ishikawa, K. et al. (2003). Trainable Automatic Text Summarization Using Segmentation of Sentence. Proceedings of the Third NTCIR Workshop.
[7] Li, S., Wang, W., and Wang, C. (2008). TAC 2008 Update Summarization Task of ICL. Proceedings of TAC.
[8] Lin, C.-Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. In Proceedings of Workshop on Text Summarization Post-Conference Workshop (ACL2004), Barcelona, Spain.
[9] Lorch, R.F., Jr. & Lorch, E. P. (1996a). Effects of Headings on Text Recall and Summarization. Contemporary Educational Psychology, 21, 261-278.
[10] Mani, I., and Maybury, M. Advances in Automatic Text Summarization, MIT Press, 1999.
[11] Mihalcea, R., and and Ceylan, H. (2007). Explorations in Automatic Book Summarization, in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2007), Prague, June 2007.
[12] Mihalcea, R., and Tarau, P. (2004). Textrank: Bringing order into texts. In *EMNLP 2004: Proceedings of the Conference on Empirical Methods in Natural Language Processing*.
[13] Mihalcea, R., and Csomai, A. (2005). Senselearner: Word sense disambiguation for all words in unrestricted text. In *ACL*.
[14] Radev, D. R., Jing, H., Stys, M., and Tam, D. (2004). Centroid-based summarization of multiple documents. Information Processing and Management 40 (2004), 40:919-938.
[15] WordNet: An Electronic Lexical Database, Edited by Christiane Fellbaum, MIT Press, 1998.
[16] Temperley, D., Sleator, D., and Lafferty, J. LinkGrammar. `http://www.link.cs.cmu.edu/link/`
[17] Brian, D. Lingua LinkParser. `http://search.cpan.org/~dbrian/Lingua-LinkParser/`
[18] Standford Natural Language Processing Group. Stanford Named Entity Recognizer (NER). `http://nlp.stanford.edu/software/CRF-NER.shtml`
[19] Stanford Natural Language Processing Group. Stanford Part-of-speech Tagger (POS). `http://nlp.stanford.edu/software/tagger.shtml`

X. Appendix

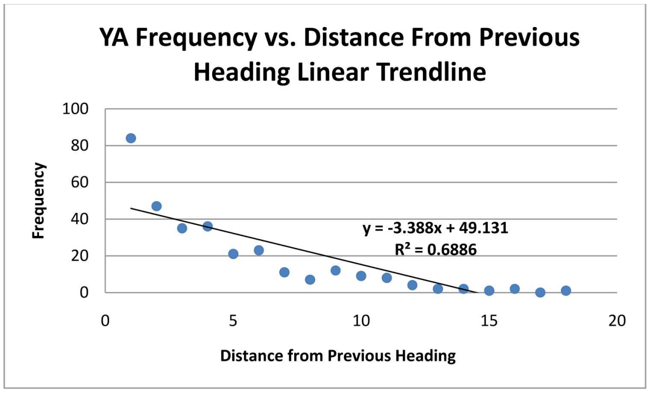

Fig. 34. YA Frequency vs. Distance From Previous Heading with ρ= -0.71673 & closest regression function y= -3.388(x) + 49.131 with Standard Error: 12.5387 and R^2 = 0.6886.

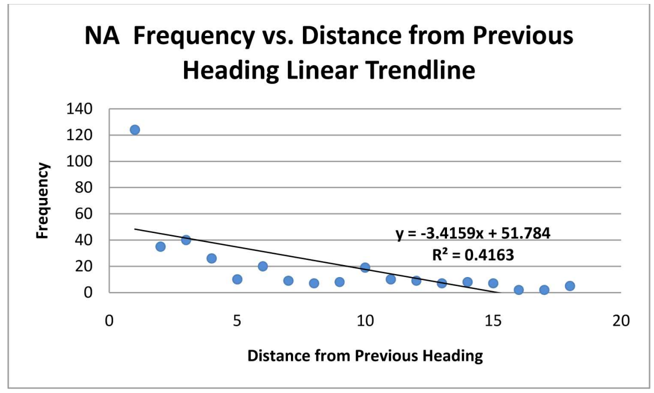

Fig. 35. NA Frequency vs. Distance From Previous Heading with ρ= -0.6489 & closest regression function y= -3.4159(x) + 51.784 with Standard Error= 21.781 and R^2 = 0.4163.

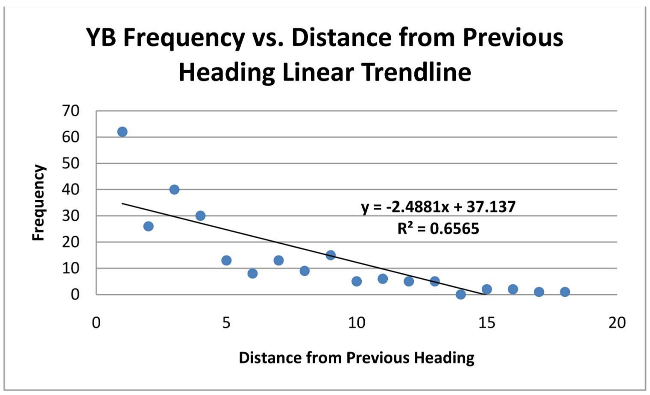

Fig. 36. YB Frequency vs. Distance From Previous Heading with ρ= -0.8267 & closest regression function y= -2.4881(x) + 37.137 with Standard Error= 9.9028 and R^2 = 0.6565.

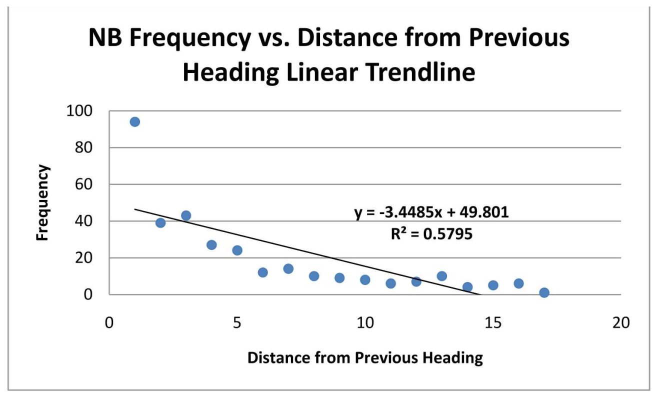

Fig. 37. NB Frequency vs. Distance From Previous Heading with ρ= -0.7627 & closest regression function y= -3.4485(x) + 49.801 with Standard Error= 15.084 and R^2 = 0.5795.

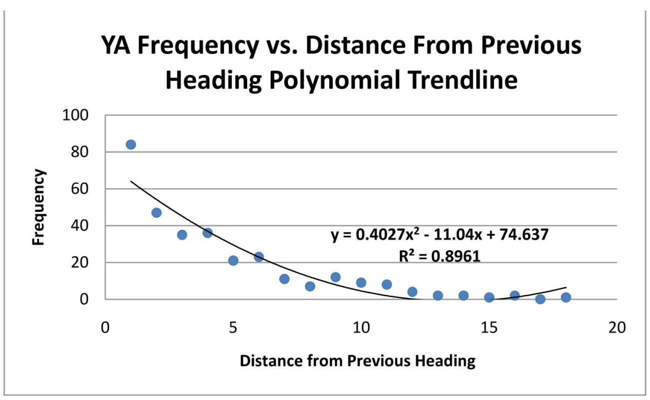

Fig. 38. YA Frequency vs. Distance From Previous Heading with ρ= -0.7167 & closest regression function y= 0.4027(x^2) - 11.04(x) + 74.637 with Standard Error= 12.538 and R^2= 0.8961.

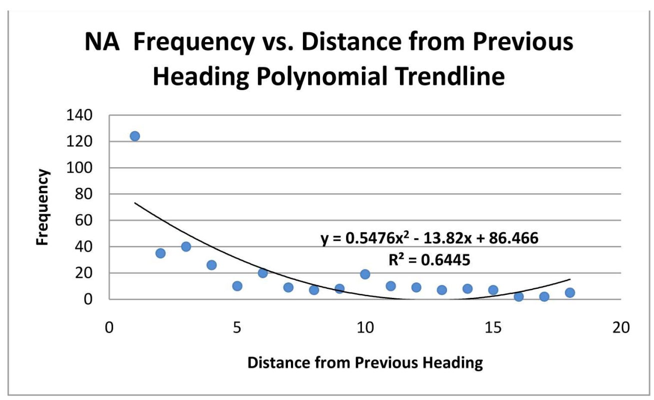

Fig. 39. NA Frequency vs. Distance From Previous Heading with ρ= -0.6489 closest regression function y= 0.5476(x^2) - 13.82(x) + 86.466 with Standard Error= 21.781 and R^2= 0.6445.

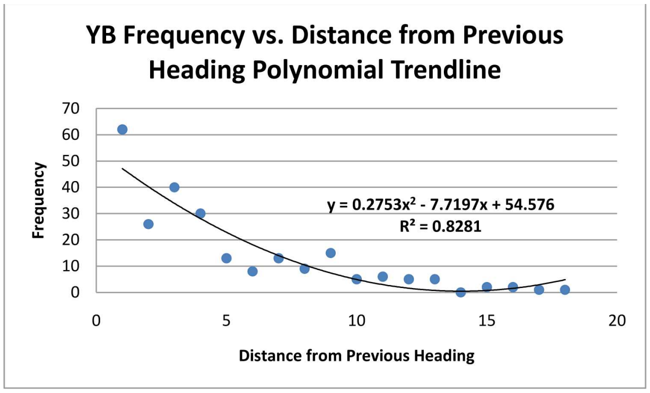

Fig. 40. YB Frequency vs. Distance From Previous Heading with ρ= -0.8267 closest regression function y= 0.2753(x^2) - 7.7197(x) + 54.576 with Standard Error= 9.9028 and R^2= 0.8281.

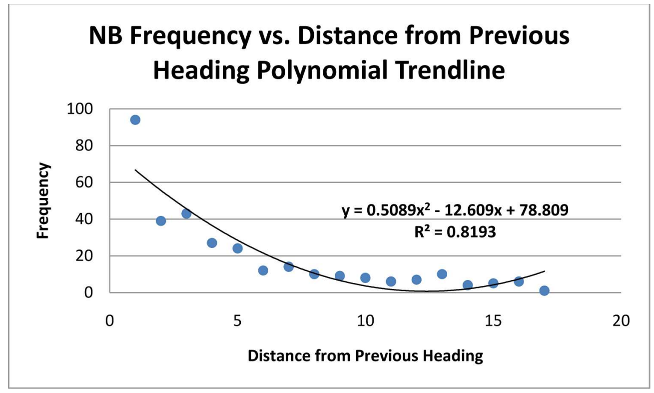

Fig. 41. NB Frequency vs. Distance From Previous Heading with ρ= -0.7627 & closest regression function y= 0.5089(x^2) - 12.609(x) + 78.809 with Standard Error= 15.084 and R^2= 0.8193.

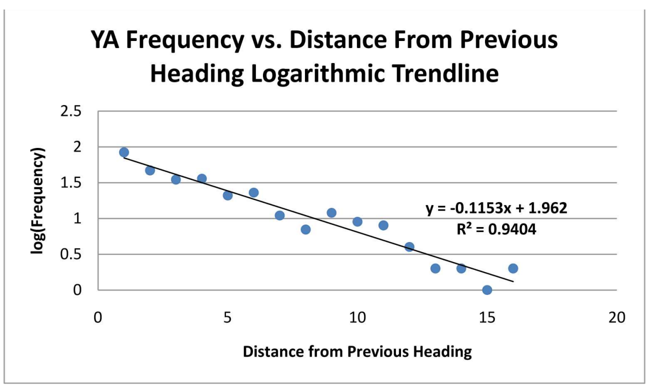

Fig. 42. YA Frequency vs. Distance From Previous Heading with ρ= -0.9697 & closest regression function y= -0.1153(x) + 1.962 with Standard Error= 0.1431 and R^2 = 0.9404.

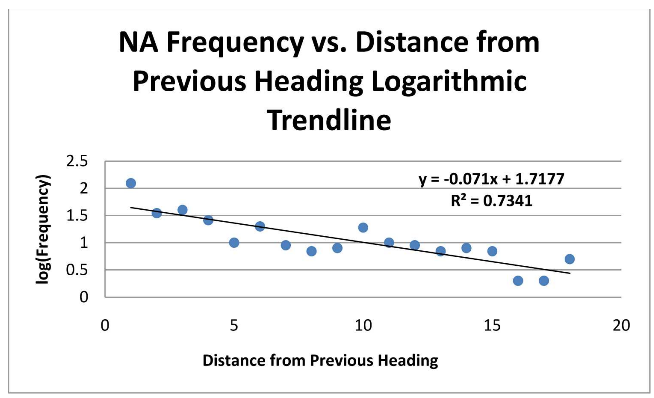

Fig. 43. NA Frequency vs. Distance From Previous Heading with ρ= -0.8568 & closest regression function y= -0.0771(x) + 1.7177 with Standard Error= 0.235 and R^2 = 0.7341.

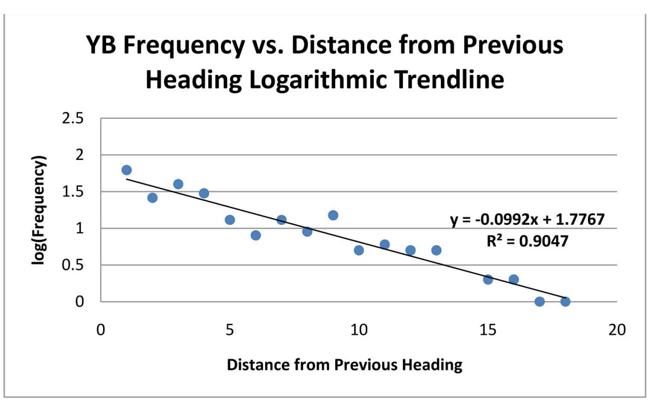

Fig. 44. YB Frequency vs. Distance From Previous Heading with ρ= -0.95414 & closest regression function y= -0.0992(x) + 1.7767 with Standard Error= 1.5103 and R^2 = 0.9047.

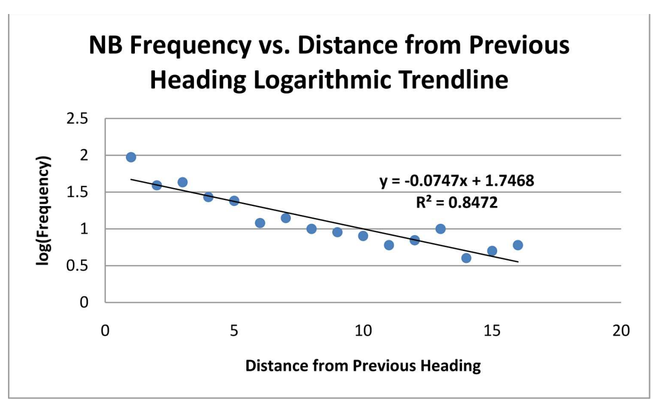

Fig. 45. NB Frequency vs. Distance From Previous Heading with ρ= -0.9204 closest regression function y= -0.0747(x) + 1.7468 with Standard Error= 0.15626 and $R^2 = 0.8472$.

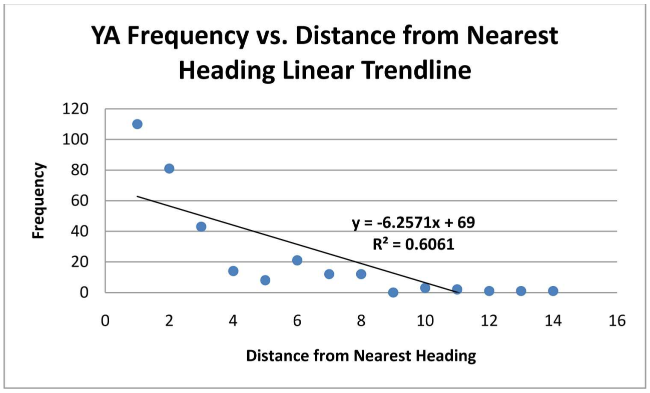

Fig. 46. YA Frequency vs. Distance From Nearest Heading with ρ= -0.7785 & closest regression function y= -6.2571(x) + 69 with Standard Error= 21.9619 and R^2 = 0.6061.

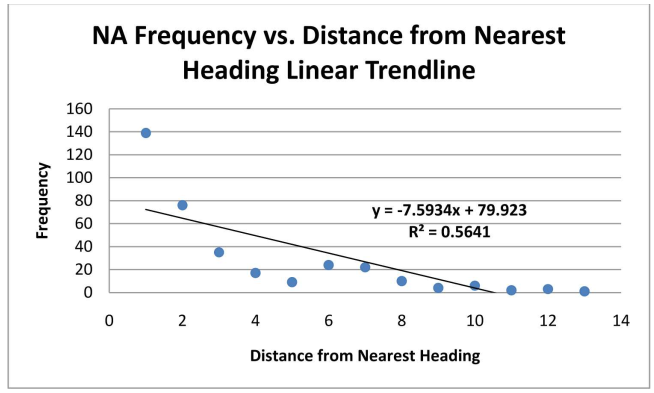

Fig. 47. NA Frequency vs. Distance From Nearest Heading with ρ= -0.75108 & closest regression function y= -7.5934(x) + 79.923 with Standard Error= 27.149 and R^2 = 0.5641.

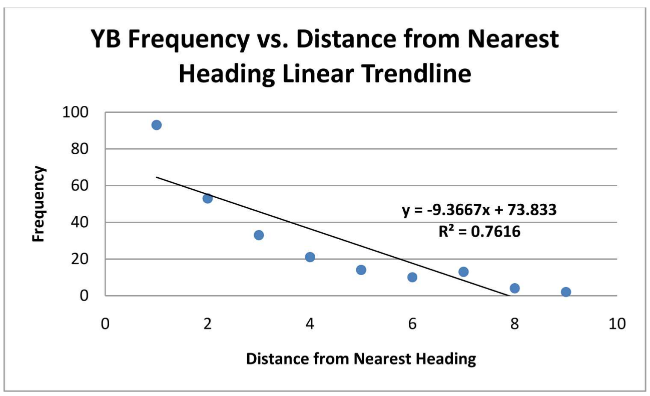

Fig. 48. YB Frequency vs. Distance From Nearest Heading with ρ= -0.8726 & closest regression function y= -9.3667(x) + 73.833 with Standard Error= 15.3433 and R^2= 0.7616.

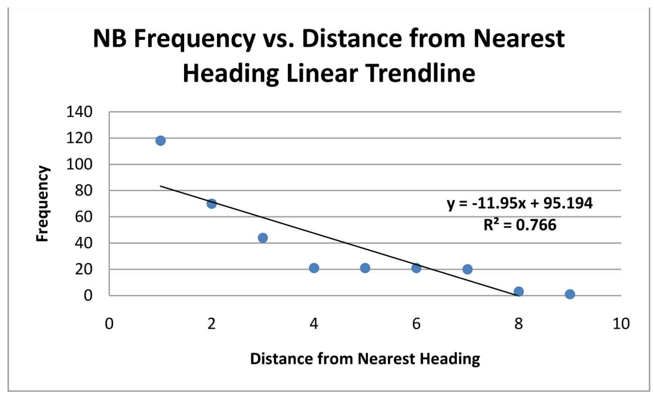

Fig. 49. NB Frequency vs. Distance From Nearest Heading with ρ= -0.8752 & closest regression function y= -11.95(x)+ 95.194 with Standard Error= 19.3803 and R^2= 0.766.

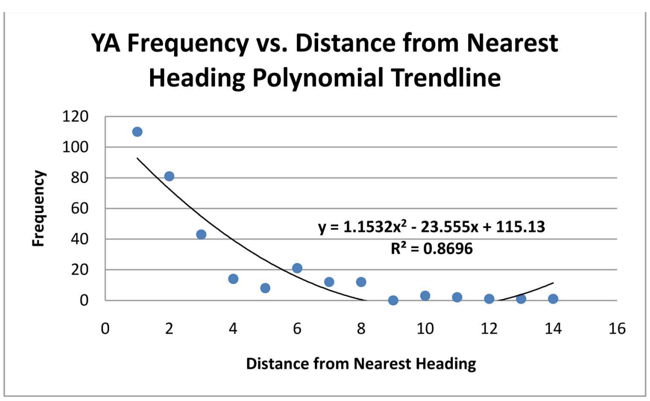

Fig. 50. YA Frequency vs. Distance From Nearest Heading with ρ= -0.7785 & closest regression function y= 1.1532(x^2) - 23.555(x) + 115.13 with Standard Error= 21.9619 and R^2= 0.8696.

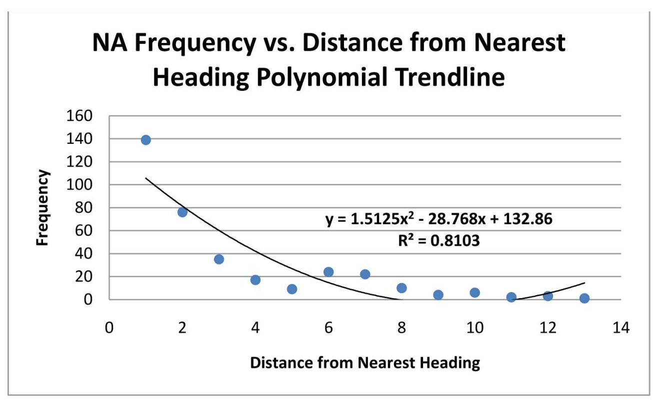

Fig. 51. NA Frequency vs. Distance From Nearest Heading with ρ= -0.751 & closest regression function y= 1.5125(x^2) - 28.768(x) + 132.86 with Standard Error= 27.149 and R^2= 0.8103.

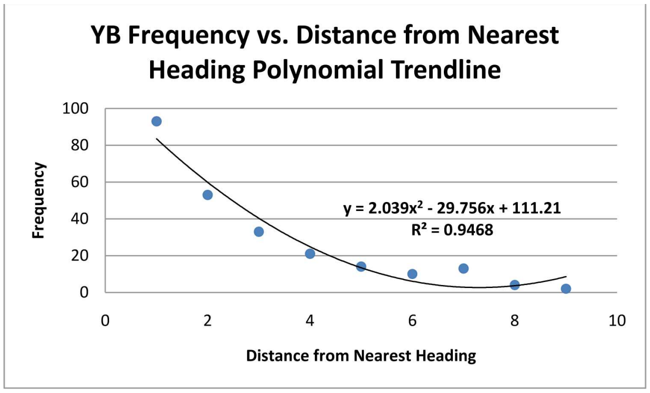

Fig. 52. YB Frequency vs. Distance From Nearest Heading with ρ= -0.8727 & closest regression function y= 2.039(x^2) - 29.756(x) + 111.21 with Standard Error= 15.343 and R^2= 0.9468.

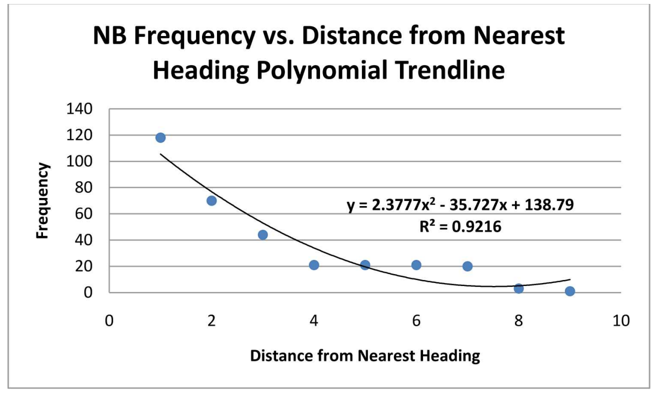

Fig. 53. NB Frequency vs. Distance From Nearest Heading with ρ= -0.8752 & closest regression function y= 2.3777(x^2) - 35.727(x) + 138.79 with Standard Error= 19.38 and R^2= 0.9216.

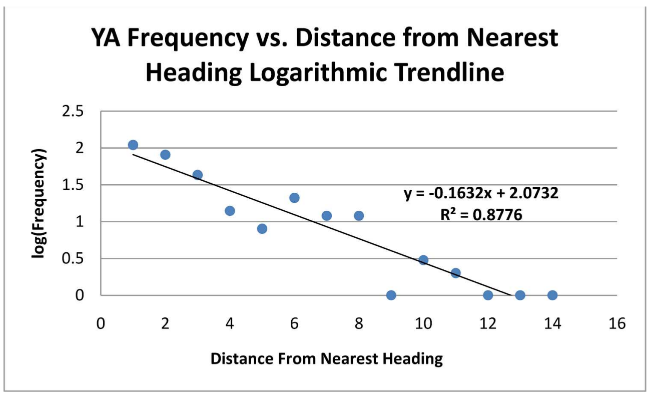

Fig. 54. YA Frequency vs. Distance From Nearest Heading with ρ= -0.9628 & closest regression function y= -0.1632(x) + 2.0732 with Standard Error= 0.2016 and R^2= 0.8776.

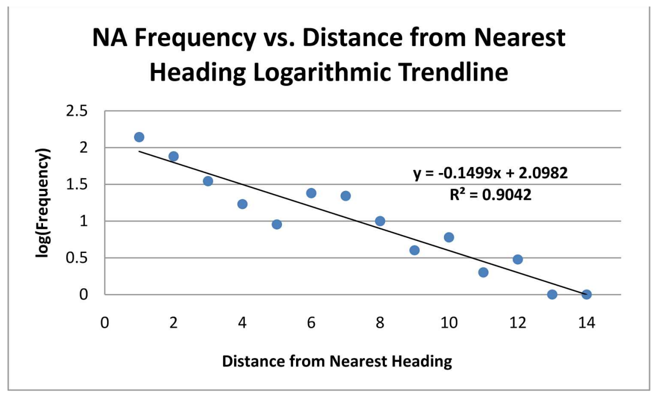

Fig. 55. NA Frequency vs. Distance From Nearest Heading with ρ= -0.9397 & closest regression function y= -0.1499(x) + 2.0982 with Standard Error= 0.2219 and R^2 = 0.9042.

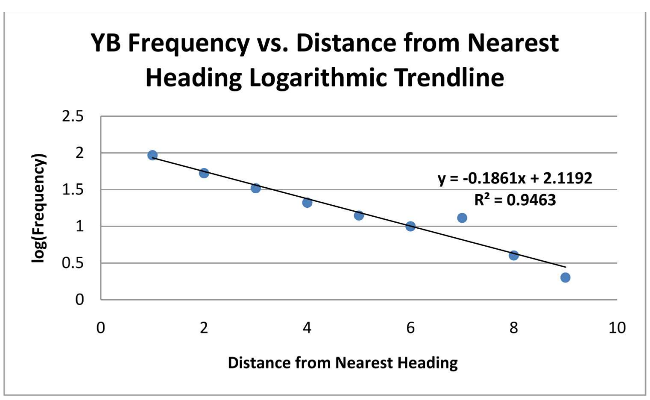

Fig. 56. YB Frequency vs. Distance From Nearest Heading with ρ= -0.9728 & closest regression function y= -0.1861(x) + 2.1192 with Standard Error= 0.1298 and R^2 = 0.9463.

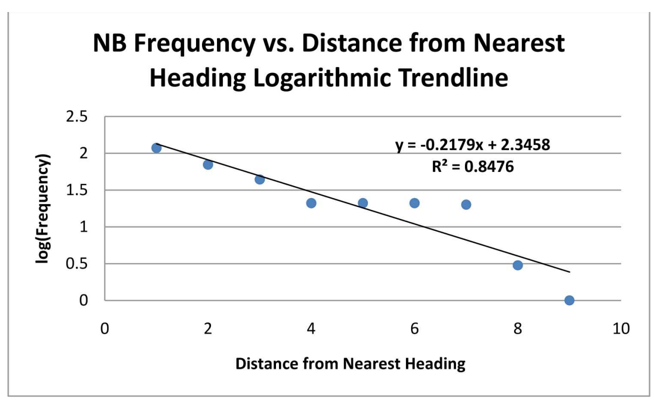

Fig. 57. NB Frequency vs. Distance From Nearest Heading with ρ= -0.9207 & closest regression function y= -0.2179(x) + 2.3458 with Standard Error= 0.2705 and R^2 = 0.8476.

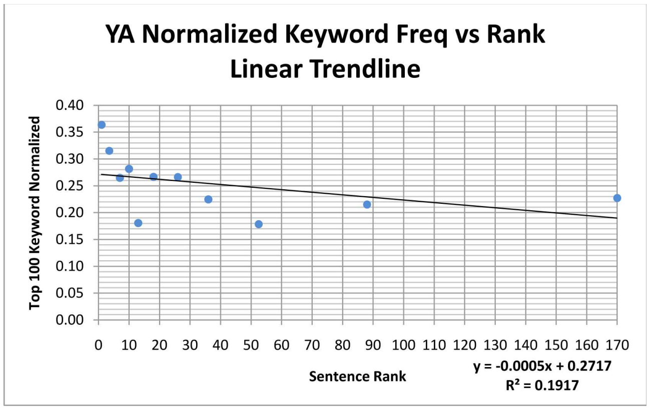

Fig. 58. YA Normalized Keyword Frequency vs. Sentence Rank with ρ= -0.4379 & closest regression function y= -0.0005(x) + 0.2717 with Standard Error= .0528 and R^2= 0.1917.

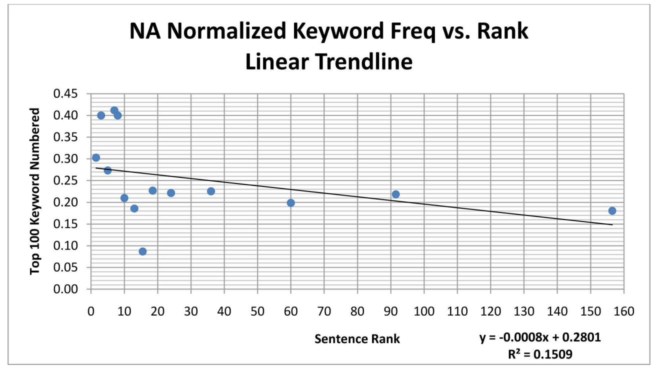

Fig. 59. NA Normalized Keyword Frequency vs. Sentence Rank with ρ= -0.3884 & closest regression function y= -0.0008(x) + 0.2801 with Standard Error= 0.0910 and R^2= 0.1509.

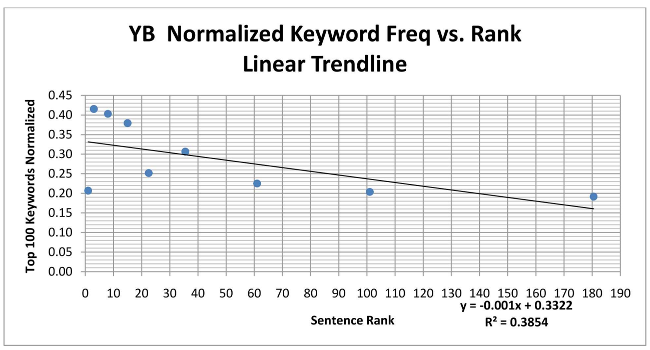

Fig. 60. YB Normalized Keyword Frequency vs. Sentence Rank with ρ= -0.6208 & closest regression function y= -0.001(x) + 0.3322 with Standard Error= 0.0763 and R^2= 0.3854.

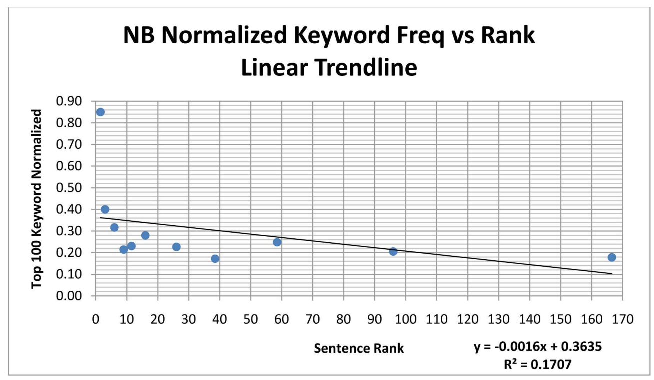

Fig. 61. NB Normalized Keyword Frequency vs. Sentence Rank with ρ= -0.4131 & closest regression function y= -0.0016(x) + 0.3635 with Standard Error= 0.0763 and R^2= 0.1707.

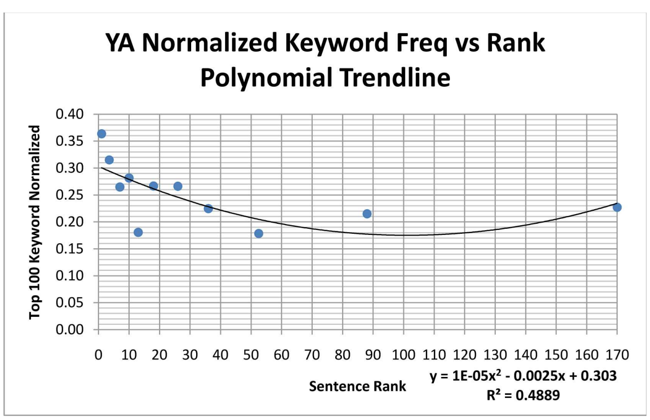

Fig. 62. YA Normalized Keyword Frequency vs. Sentence Rank with ρ= -0.4379 & closest regression function y= -1E-05(x^2) - 0.0025(x) + 0.303 with Standard Error= .0528 and R^2= 0.4889.

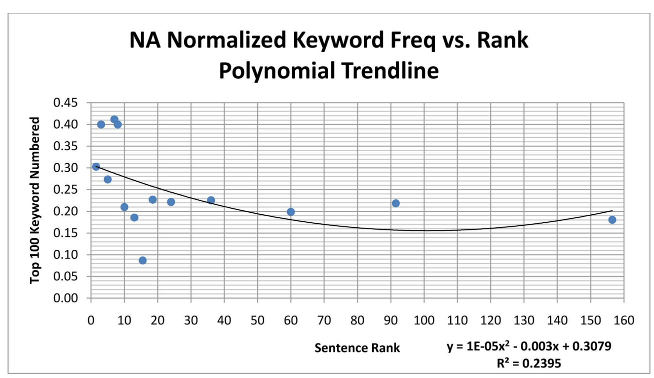

Fig. 63. NA Normalized Keyword Frequency vs. Sentence Rank with ρ= -0.3884 & closest regression function y= 1E-05(x^2) - 0.003(x) + 0.3079 with Standard Error= 0.0910 and R^2= 0.2395.

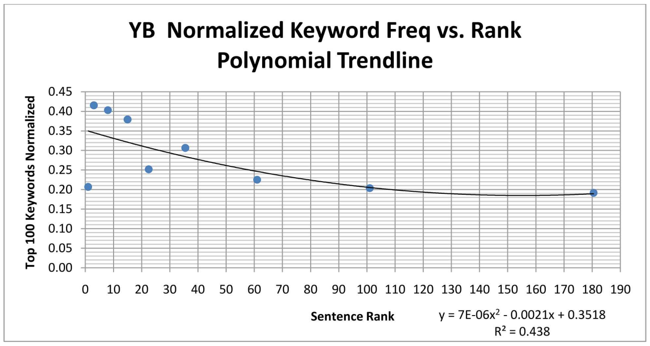

Fig. 64. YB Normalized Keyword Frequency vs. Sentence Rank with ρ= -0.6208 & closest regression function y= -7E-06(x^2) - 0.0021(x) + 0.3518 with Standard Error= 0.0763 and R^2= 0.438.

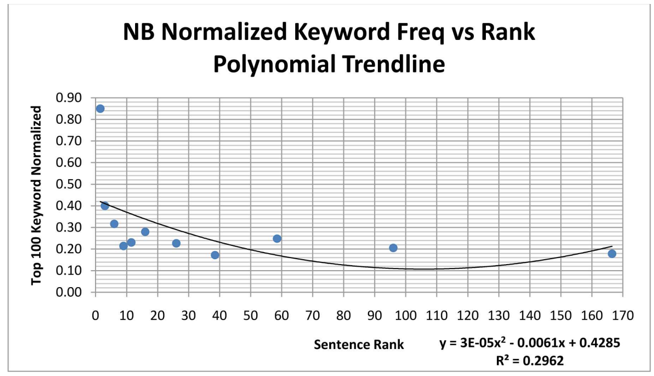

Fig. 65. NB Normalized Keyword Frequency vs. Sentence Rank with ρ= -0.4131 & closest regression function y= 3E-05(x^2) - 0.0061(x) + 0.4285 with Standard Error= 0.1855 and R^2 = 0.2962.

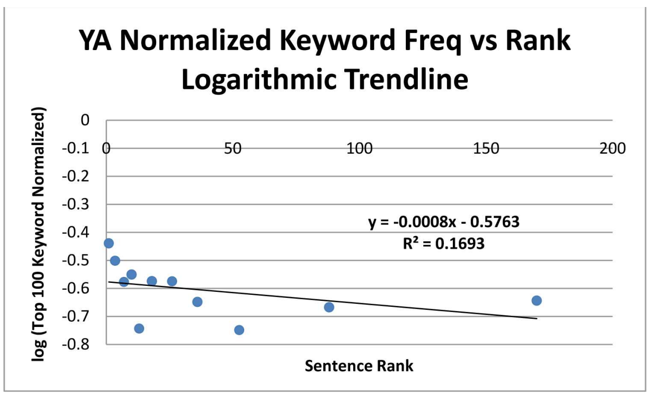

Fig. 66. YA Normalized Keyword Frequency vs. Sentence Rank with ρ= -0.4379 & closest regression function y= -0.0008(x) - 0.5763 with Standard Error= .0915 and R^2 = 0.1693.

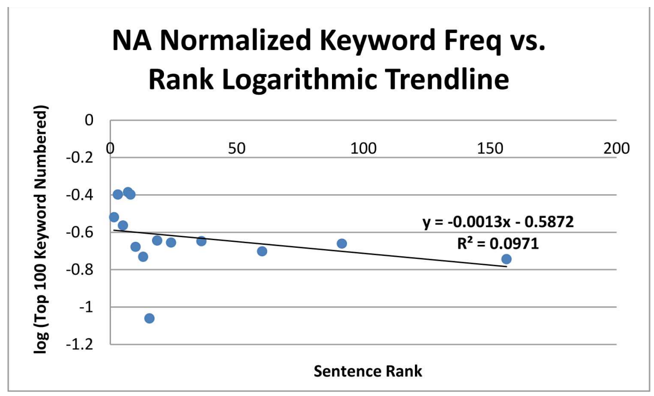

Fig. 67. NA Normalized Keyword Frequency vs. Sentence Rank with ρ= -0.3115 & closest regression function y= -0.0013(x) - 0.5872 with Standard Error= 0.174 and R^2= 0.0971.

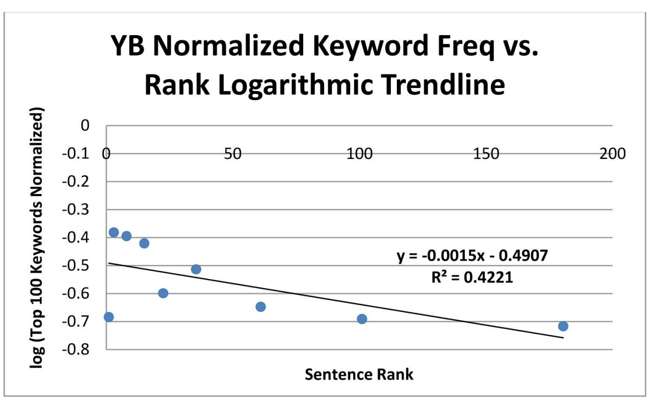

Fig. 68. YB Normalized Keyword Frequency vs. Sentence Rank with ρ= -0.6497 & closest regression function y= -0.0015(x) - 0.4907 with Standard Error= 0.1102 and R^2= 0.4221.

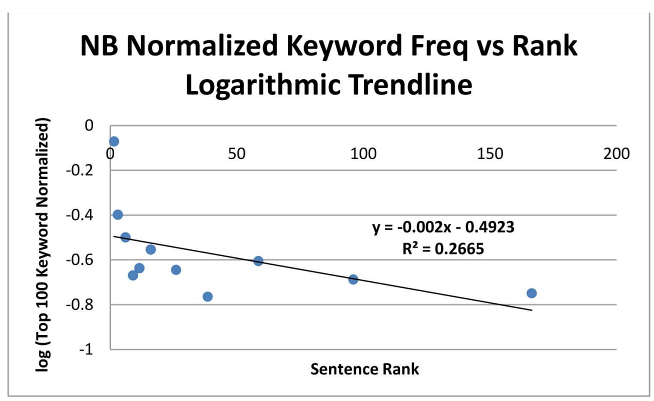

Fig. 69. NB Normalized Keyword Frequency vs. Sentence Rank with ρ= -0.4131 closest regression function y= -0.002(x) - 0.4923 with Standard Error= 0.178 and R^2 = 0.2665.

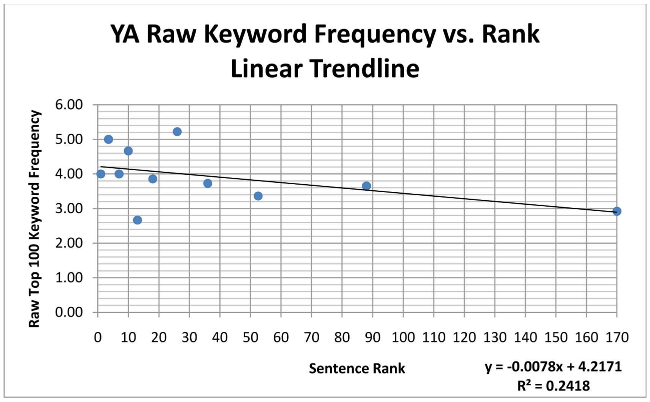

Fig. 70. YA Raw Keyword Frequency vs. Sentence Rank with ρ= -0.4917 & closest regression function y= -0.0078(x) + 4.2171 with Standard Error= 0.7355 and R^2= 0.2418.

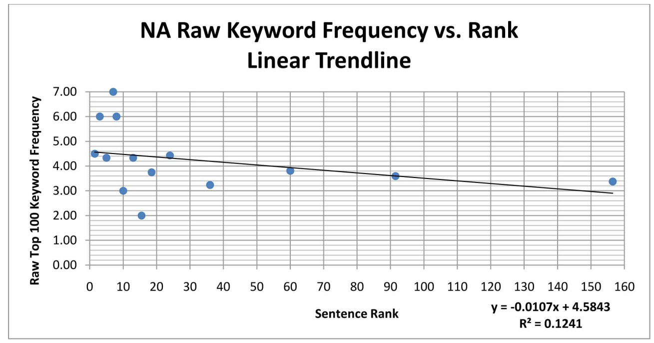

Fig. 71. NA Raw Keyword Frequency vs. Sentence Rank with ρ= -0.3523 & closest regression function y= -0.0107(x) + 4.5843 with Standard Error= 1.2981 and R^2= 0.1241.

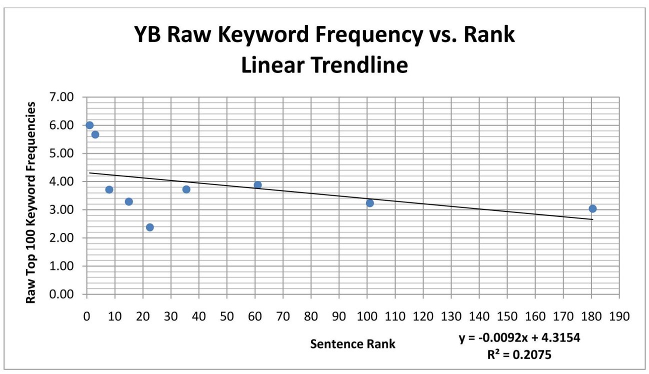

Fig. 72. YB Raw Keyword Frequency vs. Sentence Rank with ρ= -0.4556 & closest regression function y= -0.0092(x) + 4.3154 with Standard Error= 1.1403 and R^2= 0.2075.

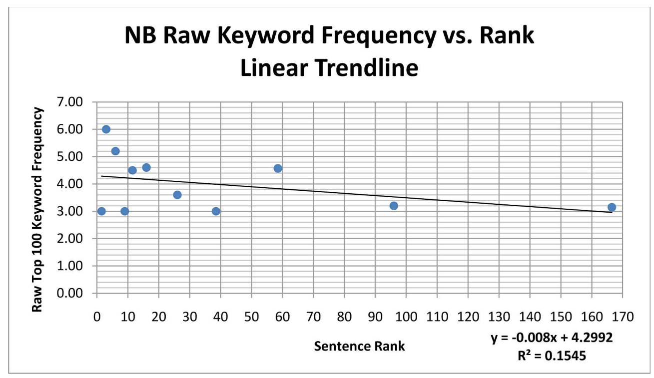

Fig. 73. NB Raw Keyword Frequency vs. Sentence Rank with ρ= -0.393 & closest regression function y=-0.008(x) + 4.2992 with Standard Error= 1.0118 and R^2= 0.1545.

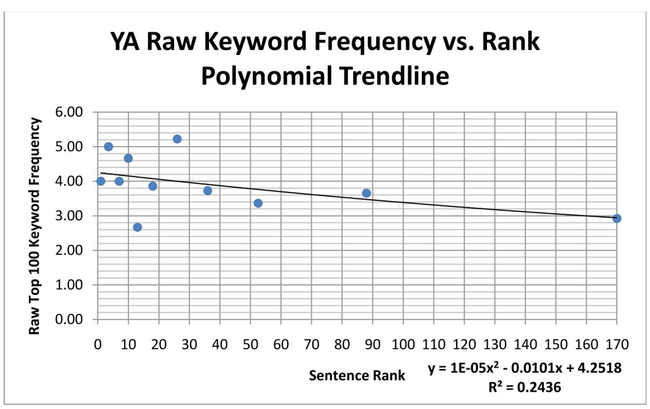

Fig. 74. YA Raw Keyword Frequency vs. Sentence Rank with ρ= -0.4917 & closest regression function y= 1E-05(x^2) - 0.0101(x) + 4.2518 with Standard Error= 0.7355 and R^2= 0.2436.

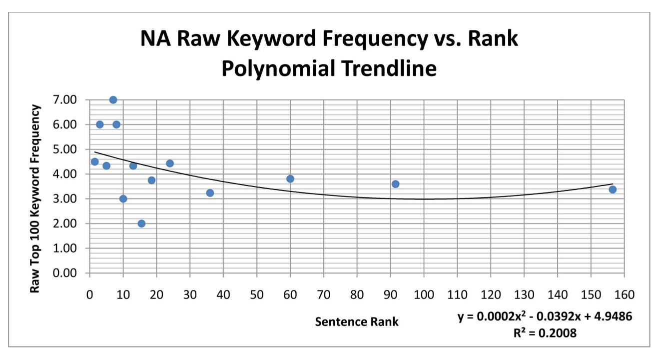

Fig. 75. NA Raw Keyword Frequency vs. Sentence Rank with ρ= -0.3523 & closest regression function y= 0.0002(x^2) - 0.0392(x) + 4.9486 with Standard Error= 1.2981 and R^2= 0.2008.

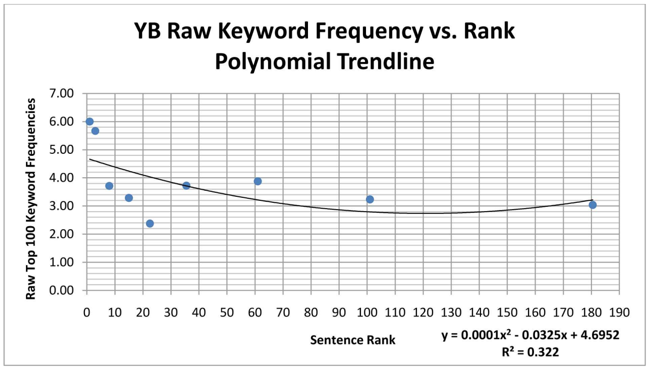

Fig. 76. YB Raw Keyword Frequency vs. Sentence Rank with ρ= -0.4556 & closest regression function y= 0.0001(x^2) - 0.0325(x) + 4.6952 with Standard Error= 1.1403 and R^2= 0.322.

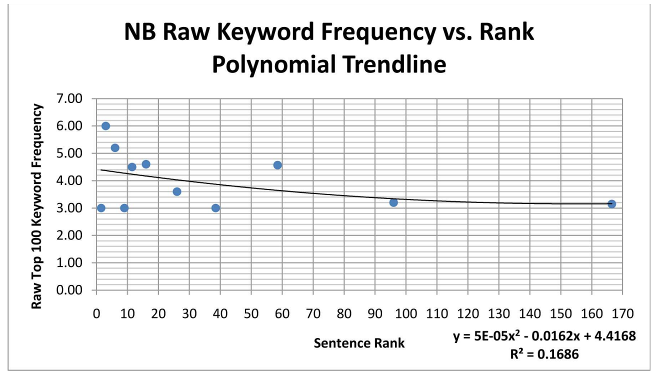

Fig. 77. NB Raw Keyword Frequency vs. Sentence Rank with ρ= -0.393 & closest regression function y= 5E-05(x^2) - 0.0162(x) + 4.4168 with Standard Error= 1.0118 and R^2= 0.1686.

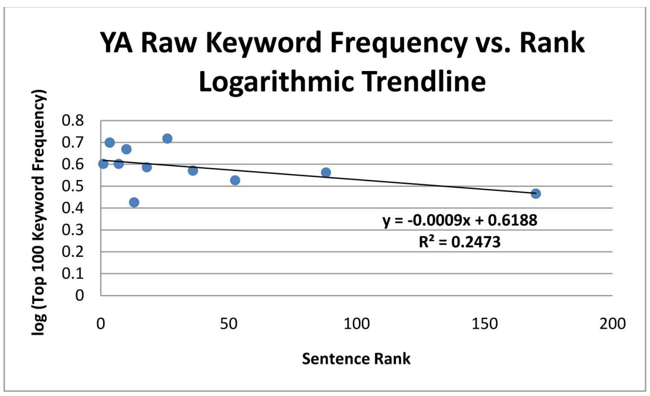

Fig. 78. YA Raw Keyword Frequency vs. Sentence Rank with ρ= -0.4973 & closest regression function y= -0.0009(x) + 0.6188 with Standard Error= 0.0827 and R^2= 0.2473.

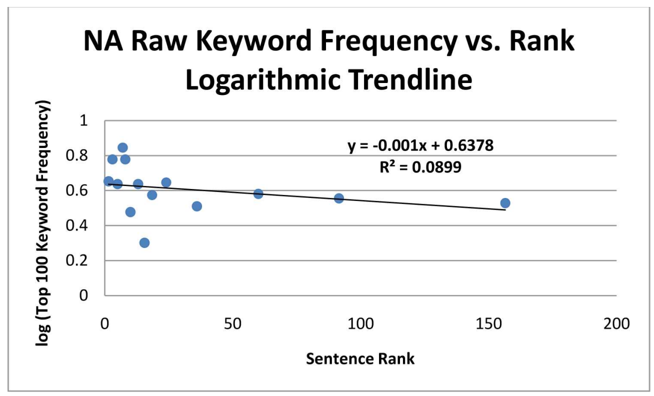

Fig. 79. NA Raw Keyword Frequency vs. Sentence Rank with ρ= -0.2999 & closest regression function y= -0.001(x) + 0.6378 with Standard Error= 0.138 and R^2= 0.0899.

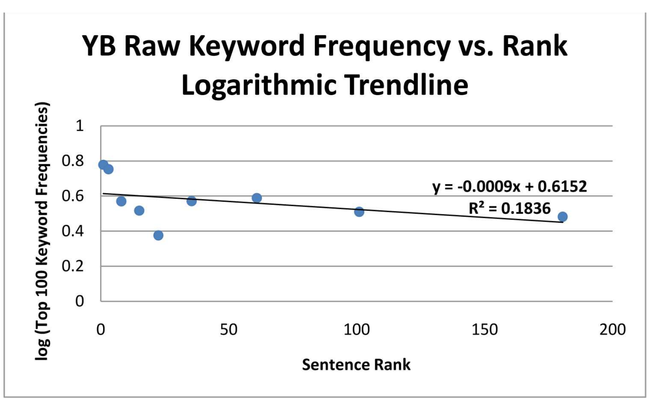

Fig. 80. YB Raw Keyword Frequency vs. Sentence Rank with ρ= -0.4284 & closest regression function y= -0.0009(x) + 0.6152 with Standard Error= 0.1228 and R^2= 0.1836.

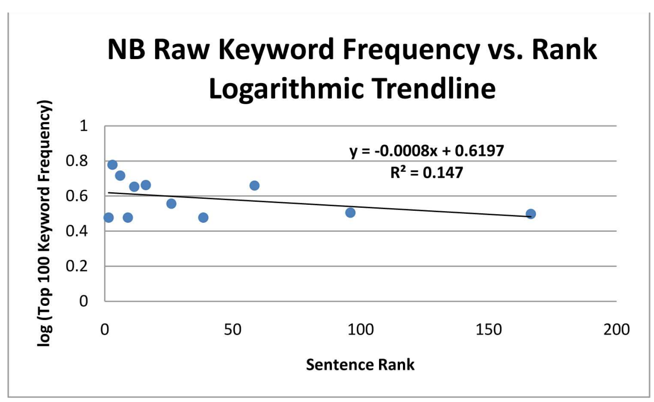

Fig. 81. NB Raw Keyword Frequency vs. Sentence Rank with ρ= -0.393 & closest regression function y= -0.0008(x) + 0.6197 with Standard Error= 0.1068 and R^2= 0.147.

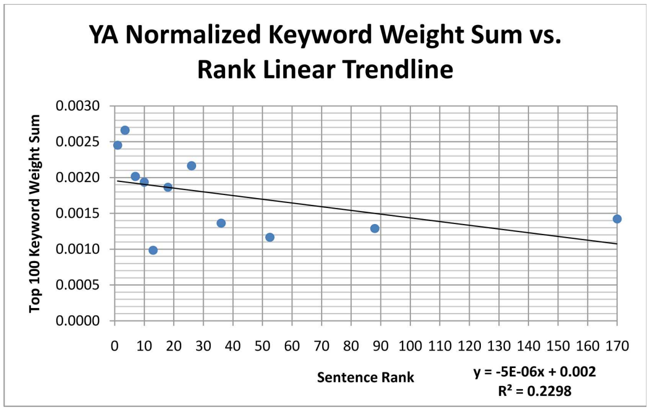

Fig. 82. YA Normalized Keyword Weight Sum vs. Sentence Rank with ρ= -0.4794 & closest regression function y= -5E-06(x) + 0.002 with Standard Error= .0005 and R^2= 0.2298.

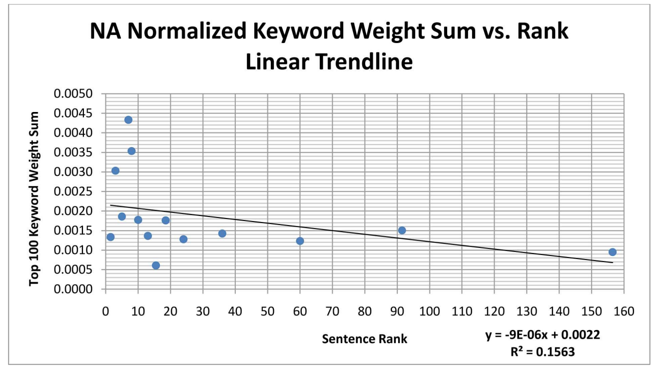

Fig. 83. NA Normalized Keyword Weight Sum vs. Sentence Rank with ρ= 0.3954 & closest regression function y= -9E-06(x) + 0.0022 with Standard Error= 0.001 and R^2= 0.1563.

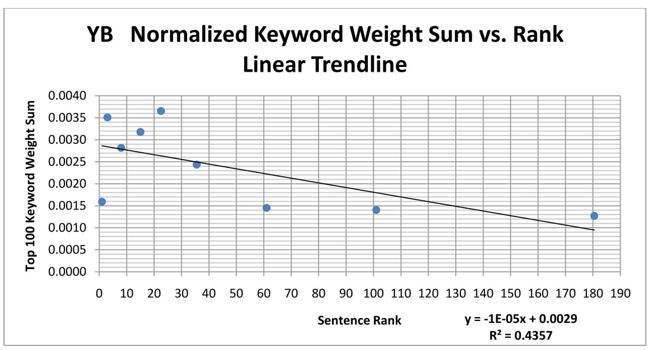

Fig. 84. YB Normalized Keyword Weight Sum vs. Sentence Rank with ρ= 0.6601 & closest regression function y= -1E-05(x) + 0.0029 with Standard Error= 0.00077 and R^2= 0.4357.

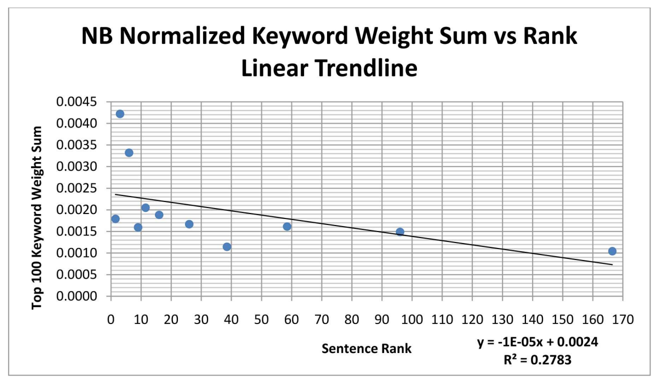

Fig. 85. NB Normalized Keyword Weight Sum vs. Sentence Rank with ρ= 0.5276 & closest regression function y= -1E-05(x) + 0.0024 with Standard Error= 0.00085 and R^2= 0.2783.

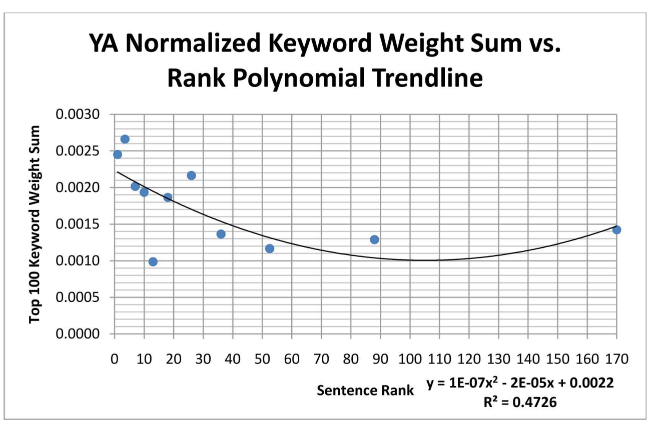

Fig. 86. YA Normalized Keyword Weight Sum vs. Sentence Rank with ρ= -0.4794 & closest regression function y= 1E-07(x^2) - 2E-05(x) + 0.0022 with Standard Error= .0005 and R^2= 0.4726.

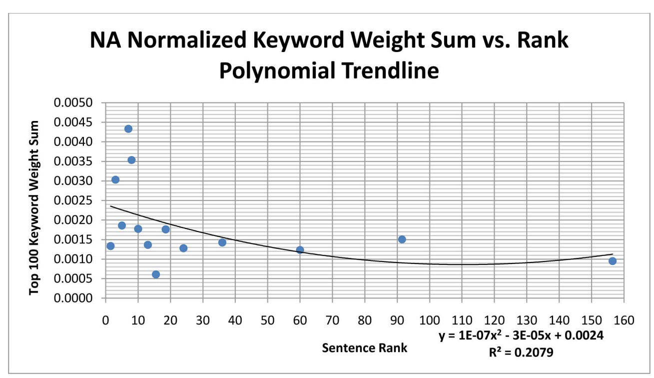

Fig. 87. NA Normalized Keyword Weight Sum vs. Sentence Rank with ρ= 0.3954 & closest regression function y= 1E-07(x^2) - 3E-05(x) + 0.0024 with Standard Error: 0.001 and R^2= 0.2079.

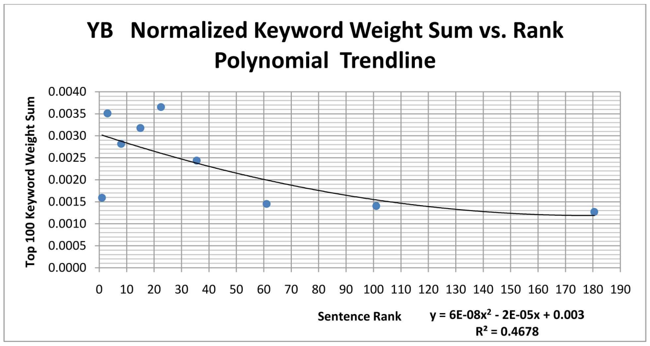

Fig. 88. YB Normalized Keyword Weight Sum vs. Sentence Rank with ρ= 0.6601 & closest regression function y= 6E-08(x^2) - 2E-05(x) + 0.003 with Standard Error= 0.00077 and R^2= 0.4678.

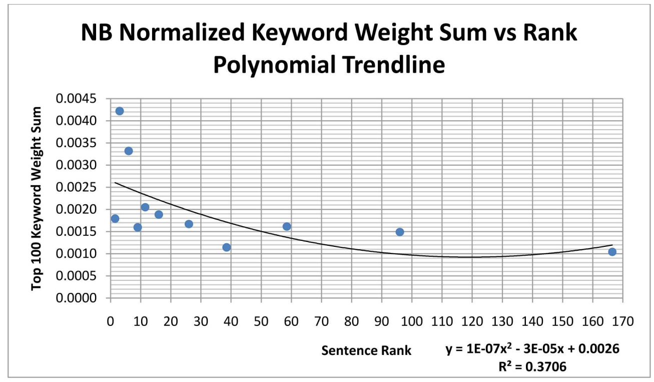

Fig. 89. NB Normalized Keyword Weight Sum vs. Sentence Rank with ρ= 0.5276 & closest regression function y= 1E-07(x^2) - 3E-05(x) + 0.0026 with Standard Error= 0.00085 and R^2= 0.3706.

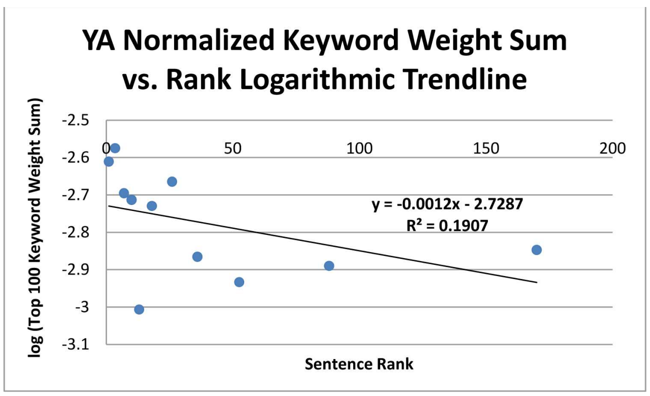

Fig. 90. YA Normalized Keyword Weight Sum vs. Sentence Rank with ρ= -0.4366 & closest regression function y= -0.0012(x) - 2.7287 with Standard Error= 0.1307 and R^2= 0.1907.

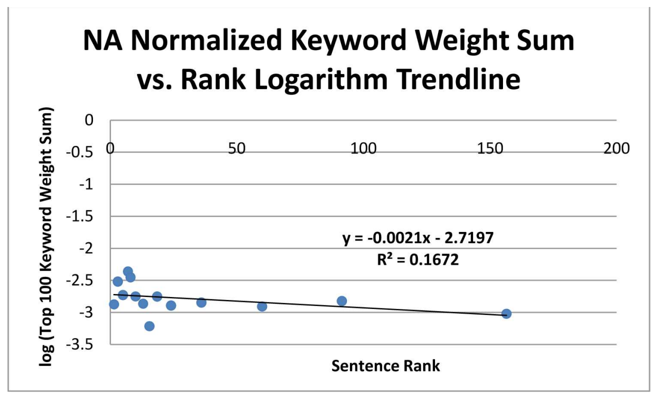

Fig. 91. NA Normalized Keyword Weight Sum vs. Sentence Rank with ρ= -0.4089 & closest regression function y= -0.0021(x) - 2.7197 with Standard Error= 0.2129 and R^2= 0.1672.

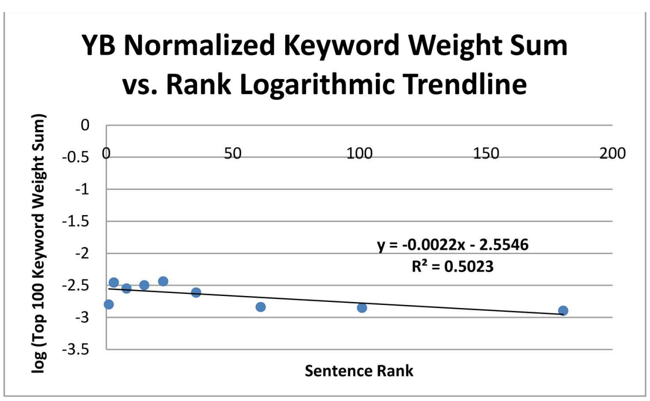

Fig. 92. YB Normalized Keyword Weight Sum vs. Sentence Rank with ρ= -0.7087 & closest regression function y= -0.0022(x) - 2.5546 with Standard Error= 0.1399 and R^2= 0.5023.

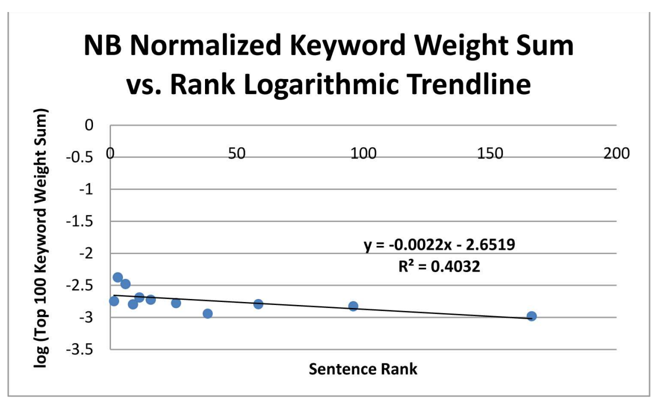

Fig. 93. NB Normalized Keyword Weight Sum vs. Sentence Rank with ρ= -0.6349 & closest regression function y= -0.0022(x) - 2.6519 with Standard Error= 0.1463 and R^2= 0.4032.

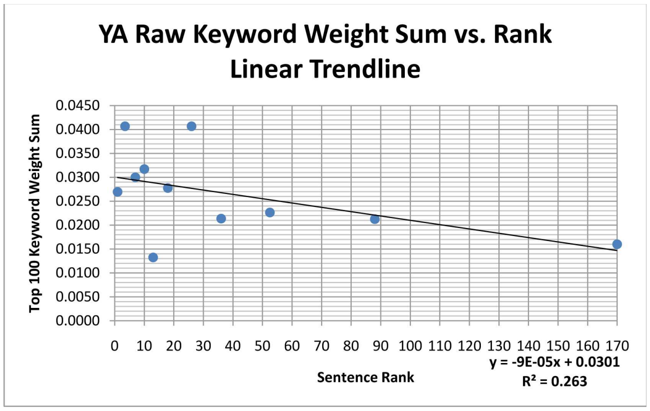

Fig. 94. YA Raw Keyword Weight Sum vs. Sentence Rank with ρ= -0.5128 & closest regression function y= -9E-05(x) + 0.0301 with Standard Error= 0.008 and R^2= 0.263.

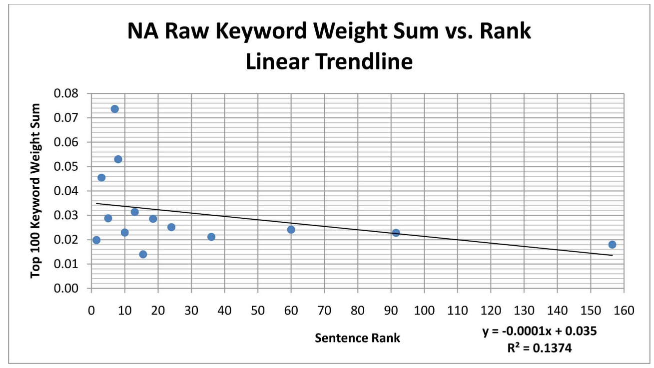

Fig. 95. NA Raw Keyword Weight Sum vs. Sentence Rank with ρ= -0.3707 & closest regression function y= -0.0001(x) + 0.035 with Standard Error= 0.0157 and R^2= 0.1374.

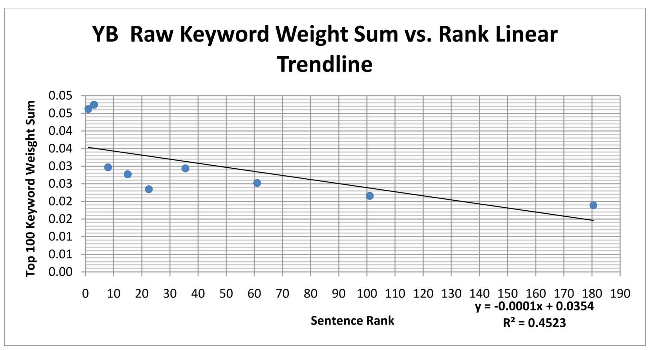

Fig. 96. YB Raw Keyword Weight Sum vs. Sentence Rank with ρ= -0.6726 & closest regression function y= -0.0001(x) +0.0354 with Standard Error= 0.008 and R^2= 0.4523.

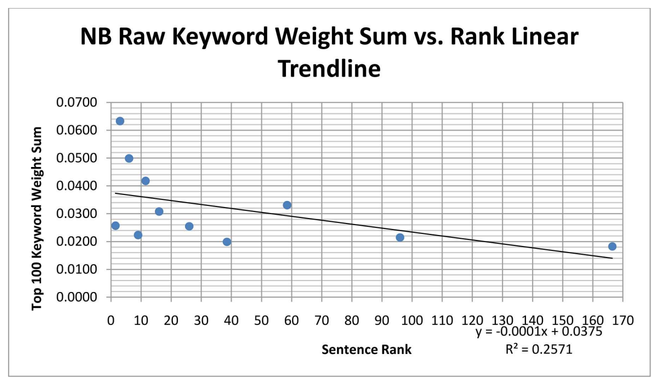

Fig. 97. NB Raw Keyword Weight Sum vs. Sentence Rank with ρ= 0.5071 & closest regression function y= -0.0001(x) + 0.0375 with Standard Error: 0.0129 and R^2 = 0.2571.

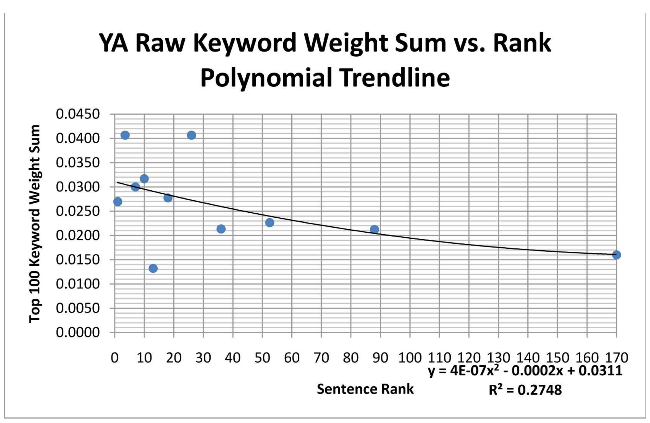

Fig. 98. YA Raw Keyword Weight Sum vs. Sentence Rank with ρ= -0.5128 & closest regression function y= 4E-07(x^2) - 0.0002(x) + 0.0311 with Standard Error= 0.008 and R^2= 0.2748.

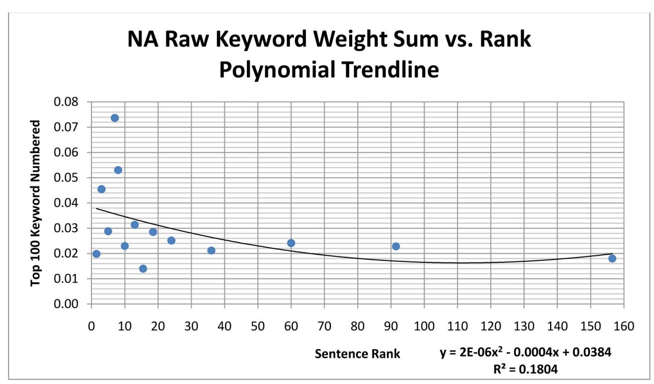

Fig. 99. NA Raw Keyword Weight Sum vs. Sentence Rank with ρ= -0.3707 & closest regression function y= 2E-06(x^2) - 0.0004(x) + 0.0384 with Standard Error= 0.0157 and R^2= 0.1804.

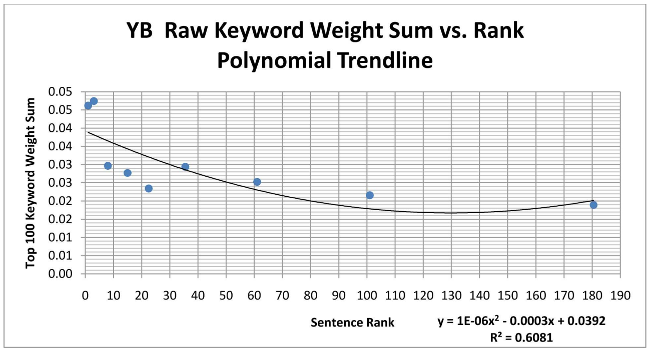

Fig. 100. YB Raw Keyword Weight Sum vs. Sentence Rank with ρ= -0.6726 & closest regression function y= 1E-06(x^2) - 0.0003(x) + 0.0392 with Standard Error= 0.00806 and R^2= 0.6081.

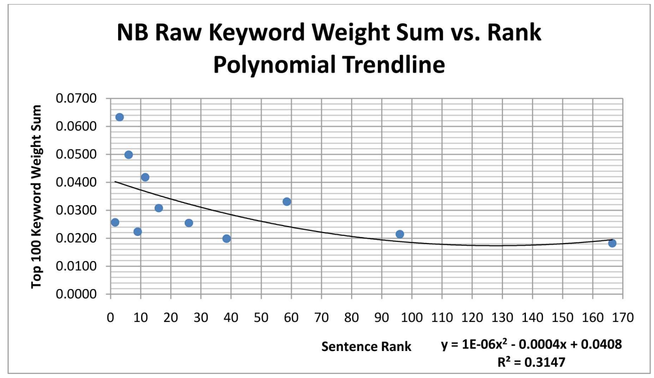

Fig. 101. NB Raw Keyword Weight Sum vs. Sentence Rank with ρ= .5071 & closest regression function y= 1E-06(x^2) - 0.0004(x) + 0.0408 with Standard Error= 0.01291 and R^2 = 0.3147.

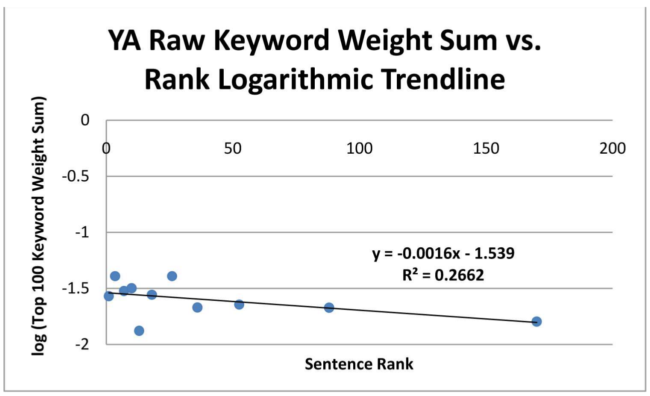

Fig. 102. YA Raw Keyword Weight Sum vs. Sentence Rank with ρ= -0.5159 & closest regression function y= -0.0016(x) - 1.539 with Standard Error= 0.1387 and R^2= 0.2662.

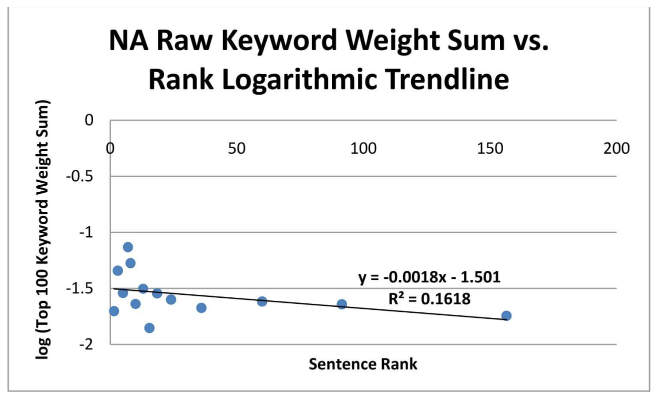

Fig. 103. NA Raw Keyword Weight Sum vs. Sentence Rank with ρ= -0.4023 & closest regression function y= -0.0018(x) - 1.501 with Standard Error= 0.1843 and R^2 = 0.1618.

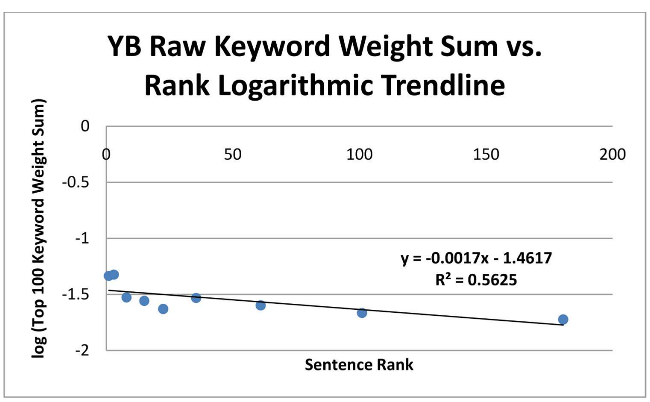

Fig. 104. YB Raw Keyword Weight Sum vs. Sentence Rank with ρ= -0.75 & closest regression function y= -0.0017(x) - 1.4617 with Standard Error= 0.0968 and R^2= 0.5625.

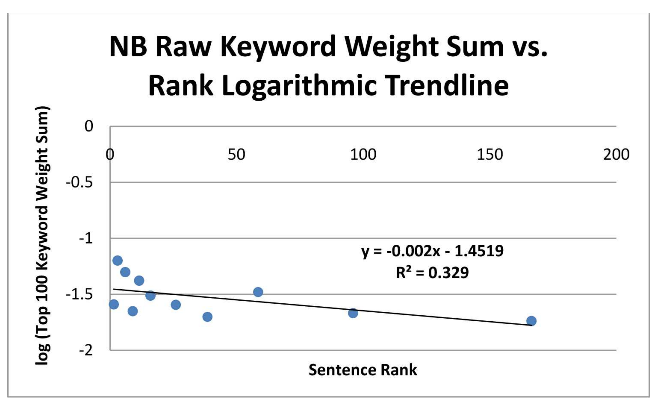

Fig. 105. NB Raw Keyword Weight Sum vs. Sentence Rank with ρ= -0.5736 & closest regression function y= -0.002(x) - 1.4519 with Standard Error= 0.1501 and R^2= 0.329.