

# Sentiment analysis of emotional words in a classical text web corpus

Cite as: AIP Conference Proceedings **2644**, 030027 (2022); <https://doi.org/10.1063/5.0104738>  
Published Online: 07 November 2022

Nurul Norasuwat Rosli, Nordiana Hamzah, Muhamad Fadzllah Zaini, et al.



View Online



Export Citation

## ARTICLES YOU MAY BE INTERESTED IN

[Geometric lexical representative perspectives: The impact of threshold values through #LancsBox software](#)

AIP Conference Proceedings **2644**, 030026 (2022); <https://doi.org/10.1063/5.0104817>

[Diachronic analysis of the f-words in English song lyrics: A computational linguistics perspective](#)

AIP Conference Proceedings **2644**, 030020 (2022); <https://doi.org/10.1063/5.0104823>

[Modal verbs through the years: Computational linguistics perspective](#)

AIP Conference Proceedings **2644**, 030021 (2022); <https://doi.org/10.1063/5.0104825>



## APL Quantum

**CALL FOR APPLICANTS**

### Seeking Editor-in-Chief

# Sentiment Analysis of Emotional Words in a Classical Text WEB Corpus

Nurul Norasuwat Rosli<sup>1, b)</sup>, Nordiana Hamzah<sup>1, a)</sup>, Muhamad Fadzllah Zaini<sup>1, c)</sup>,  
Hasrina Baharum<sup>1, d)</sup>, Farra Humairah Mohd<sup>1, e)</sup>, Nordiana Ab Jabar<sup>2, f)</sup>, Ashrol  
Rahimy Damit<sup>3, g)</sup>, and Rosmani Omar<sup>4, h)</sup>

<sup>1</sup>*Department of Malay Language and Literature, Sultan Idris Education University, 35900 Tanjong Malim, Perak, Malaysia*

<sup>2</sup>*Department of Heritage Studies, Faculty of Creative Technology and Heritage, Universiti Malaysia Kelantan, 16310 Bachok, Kelantan, Malaysia*

<sup>3</sup>*Department of Malay Language and Linguistics, Faculty of Arts and Social Science, Universiti Brunei Darussalam, Jalan Tungku Link, Gadong BE1410, Brunei*

<sup>3</sup> *Department of Publication, Dewan Bahasa dan Pustaka, Jalan Dewan Bahasa, 50460 Kuala Lumpur, Malaysia.*

<sup>a)</sup> Corresponding author: [diana.azmi@fbk.upsi.edu.my](mailto:diana.azmi@fbk.upsi.edu.my)

<sup>b)</sup> [nurulnorasuwatrosli94@gmail.com](mailto:nurulnorasuwatrosli94@gmail.com)

<sup>c)</sup> [mfadzllah@fbk.upsi.edu.my](mailto:mfadzllah@fbk.upsi.edu.my)

<sup>d)</sup> [hasrina@fbk.upsi.edu.my](mailto:hasrina@fbk.upsi.edu.my)

<sup>e)</sup> [farra.humairah@fbk.upsi.edu.my](mailto:farra.humairah@fbk.upsi.edu.my)

<sup>f)</sup> [nordiana.aj@umk.edu.my](mailto:nordiana.aj@umk.edu.my)

<sup>g)</sup> [rahimy.damit@ubd.edu.bn](mailto:rahimy.damit@ubd.edu.bn)

<sup>h)</sup> [ros\\_omar@dbp.gov.my](mailto:ros_omar@dbp.gov.my)

**Abstract.** This study explores the emotional words based on sentiment analysis. The term ‘emotional words’ refers to feeling-related words that explain an individual’s emotions. This study employed data from a corpus site, namely the Malay Concordance Project, a compilation of 165 texts and 5.8 million words, including 140, 000 versions of text. An analysis sentiment model was employed through collocation. This model combined a number of pertinent elements, including the basic corpus approach, basic word approach and basic dictionary approach. Three emotional words were chosen based on their high degree of representativeness in the corpus context. These were anger-related words for the emotion of anger, amok/rage-related words for the emotion of rage, and revenge-related words for the emotion of revenge. The results from the testing of the words anger and rage showed negative sentiments; meanwhile, revenge-based words exhibited positive sentiments. This was because the context of the collocation facilitates sentiment data reading and can be carefully filtered.

## INTRODUCTION

Emotion is a complex multi-dimensional concept that provides an insight into human personality and behaviour. In life, humans express emotions when they encounter various problems, incidents and environments, as well as all minor features surrounding them [1]. A variety of communication methods are used to convey human emotions to others. The most frequent way to do this is through speech and facial expressions. However, with the advancement of technology and communication applications, humans are likely to express their emotions to other individuals using various social networks [2]. Although technological advances allow users of social networking platforms to express their emotions through audio and video channels, text remains the most common form of communication across social networks. Various methods have been used to correctly detect specific emotions in a text. Due to the complexity of

human emotions, detecting emotions accurately from a text poses numerous challenges that must be addressed. Identifying emotions in a text becomes more difficult when they are expressed through one part of the text [3]. Occasionally, emotions in the text are so implicit that automatic emotional detection is almost impossible [4]. Many sarcastic meanings are difficult for other scholars to recognise, let alone for machines to detect properly. Emotions are defined as expressing one's mood or relationship with others [5]. In general, emotions are human reactions to an event that they experience. For example, if a person writes "I am happy", it is understood that their emotion is happiness and the sentiment behind this emotion is positive [5]. In emotional research, the need to investigate historical texts is largely performed by literary or language scholars. Many classical texts have still not been thoroughly explored. For language scholars, several emotions represent the focus of metaphorists, namely those of anger and love. The former emotion is studied using traditional methods, in which the researcher must read the text manually to identify expressions of anger. Such emotions have distinctive levels of depth, and involve anger, rage and revenge.

## METHODOLOGY

### Statistical Corpus Linguistics

This study used a qualitative design to examine emotional scenes in classical Malay prose texts through three words that are regularly used to reflect emotions, namely the words anger, rage and revenge. In this study, the Corpus Linguistic Statistics approach was utilised, as pioneered by [6]. This is a statistical word observation that uses frequency values and scores to measure the relationship between words, known as collostructure [7]. Collostructure serves as a test of the distribution between a word and significant word relationships. Collostructure emphasises the relationship between word and context to indicate the interdependence of existing words. To observe the collostructure accurately, three tests were performed, namely the t-score, MI-score and logDice. The following equations represent the t-score, MI-score and logDice tests:

$$t\text{-score} = \frac{O_{11} - E_{11}}{\sqrt{O_{11}}}$$

[1]

$$MI\text{-score} = \frac{O_{11}}{E_{11}}$$

[2]

$$\logDice = 14 + \frac{2 \times O_{11}}{R_1 + C_1}$$

[3]

In these tests,  $O$  represents the observation frequency, while  $E$  represents the expected frequency. The numbers after  $O$  and  $E$  refer to the relevant rows and columns. For instance,  $O_{12}$  refers to the frequency of observations in the first row and the second column for the first table of contingency. The value of  $R_i$  refers to the frequency of nodes in the whole corpus. This measurement indicates the collostructure and significant relationships. Before choosing a specific union test, it is necessary to determine the type of collocation to use, depending on the research question. All word unions measure focused collocation according to two main dimensions, frequency and exclusivity. Frequency refers to the number of times that nodes and collocations occur together in the corpus. Exclusivity refers to certain aspects of the collocation relationship, in which words occur alone or are dominant in their respective classes. Collostructure has the important function of translating semantics between nodes and word collocations. Moreover, its function is also to examine the representativeness of different levels to explain the semantics in the relationship between words (collocation). In addition, the collostructure formed by word unit testing (collocation) provides representative semantic generalisations across the meaning of collocation items. Another aspect is semantic generalisations about the function of slots in construction, in terms of the role of argument. As such, statistical and semantic relationships are closely related, based on collostructure. This approach allows a more refined form of sentiment analysis.

### Sentiment Analysis

Sentiment analysis (SA), sometimes referred to as opinion polling, is a research space that aims to analyse people's sentiments or opinions about entities related to them, such as topics, events, individuals, issues, services, products, organisations and attributes. Sentiments or opinions are not detached from the emotional words that are used repeatedly in daily life. In psychology, scholars regard emotions and sentiments as different concepts. Emotions are defined as complex psychological conditions and play an important role as operational motivators. Paul Eckman proposed six basic emotions: happiness, sadness, anger, fear, shock, and disgust [8]. Furthermore, these emotional elements are enriched by the emotions of pride, joy, shame, and humiliation. There are three main components of

emotions: experience, response and behaviour. Sentiment, on the other hand, refers to a mental attitude that is formed based on emotions. Sentiment is used to convey an individual's thoughts that originate from his emotions. Compared to emotions, sentiment expresses a deeper concept that is not limited to the psychological dimension. McDougall notes that sentiment usually connects primary emotions with actions [9]. Basic research on sentiment analysis involves the classification of polarisation, which explores whether an opinion expressed in a document or sentence about a particular feature or aspect is positive, negative or natural.

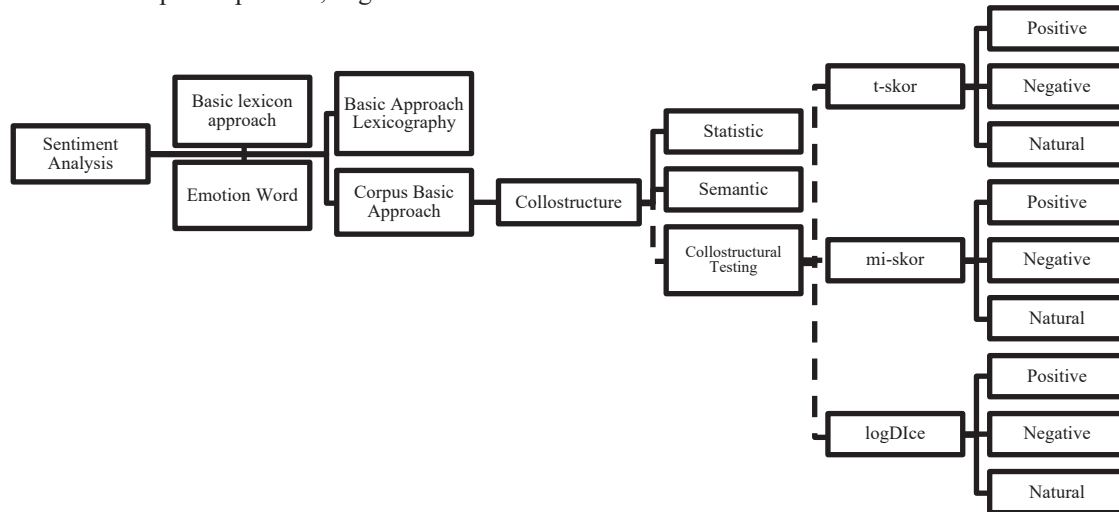


FIGURE 1. Analysis model of sentiment analysis via collocation

Figure 1 shows the sentiment analysis model that follows the collocation sub approach. In this model, sentiment analysis uses a word-based system consisting of a basic dictionary approach and a corpus-based approach. The former refers to the procedure of considering a word's meaning as well as the background of each word's information, as used in the analysed sentiments. The basic corpus approach refers to the linguistic corpus statistical approach, which describes collocation (collocation structure) as consisting of the value of the relationship between a word and the semantic generalisation collocation. This study used semantics by observing keywords in context to obtain an overall collection of positive, negative and natural target plots. The focus of this study was words related to emotions, namely anger, rage and revenge. This study used word data from the Malay Concordance Project (MCP), a compilation of 165 texts and 5.8 million words, including 140,000 versions (<http://mcp.anu.edu.au/Q/mcp.html>). The Malay Concordance Project is a collection of classical Malay texts gathered over an extensive time period. This study used prose text and brought together 70 classical texts consisting of Malay tales. Through this study, three words - anger, rage (*amok*) and revenge - were generated to identify the emotional scenarios highlighted in the classical texts. The frequency of the word anger in the classical prose texts. There are 441 instances of this word, involving 60 classical prose texts in the Malay Concordance Project. The classical prose text with the highest frequency of the representation of the word anger is The Tale of Sang Boma (Hikayat Sang Boma), with 83 instances. Meanwhile, the classical prose texts with the lowest frequency value of anger words, with one instance each, are The Origin of the Jinn and Gods (Asal Bangsa Jin & Dewa-Dewa), Stories Composed by Abdullah (Ceritera2 karangan Abdullah), Abdullah's Voyage to Kelantan (Pelayaran Abdullah ke Kelantan), Abdullah's Voyage to Mecca (Pelayaran Abdullah ke Mekah), Surat al-Anbiya', Takzir Chapter (Bab Takzir), The Tale of Banjar and Waringin City (Hikayat Banjar dan Kota Waringin), Jokes (Cerita Jenaka), The Tale of Tanah Hitu (Hikayat Tanah Hitu), The Tale of Hassanuddin (Hikayat Hassanuddin), Malay and Bugis Pedigree (Salasilah Melayu dan Bugis), The Tale of Bengkulu State Order (Hikayat Perintah Negeri Bengkulu), The Tale of Pahang (Hikayat Pahang), The Tale of Panji Kuda Semirang (Hikayat Panji Kuda Semirang), Shihabuddin's Tract (Risalah Shihabuddin), The Tale of Seri Kelantan (Hikayat Seri Kelantan), The Tale of Siak (Hikayat Siak), and The Tale of Tambo Barus Hilir (Hikayat Tambo Barus Hilir). The frequency of the word rage in classical prose texts from the Malay Concordance Project. As this diagram indicates, the word rage occurs most often in the texts of The Tale of Sang Boma (34 times) and The Tale of Siak (33 times). Texts with the lowest frequency of the word amok, with a value of 1, are Abdullah's Voyage to Mecca, The Story of Bangka, The Tale of Bakhtiar, The Tale of Tanah Hitu, Misa Melayu, The Tale of Merpati Mas and Merpati Perak, The Tale of Nakhoda Muda, The Tale of Bengkulu State Order, The Tale of Raja Pasai, The Tale of Pelanduk Jenaka, The Tale of Seri Rama, The Tale of Seri Kelantan, and The Tale of Sikka Kingdom. In total, 49 classical prose texts show consistent word frequency,

with a total of 283 instances of the word *rage*. The word frequency of the word *revenge* was generated, with an overall representative value of 79 times. The frequency of the word *revenge* is the highest in the classical prose text of The Tale of Hang Tuah, with a frequency value of 9 times. The prose texts with the lowest frequency of the word *revenge* include The Tale of Amir Hamzah, The Tale of Abdullah bin Abdul Kadir, Abdullah's Voyage to Kelantan, The Tale of Aceh, Surat Al-Anbiya', The Story of Bangka, The Tale of Banjar and Kota Waringin, The Tale of Bayan Budiman, The Tale of Raja Bikrama Sakti, The Tale of Sang Boma, Jokes (*Cerita Jenaka*), The Tale of the Devil, The Tale of Inderaputera, The Tale of Johor and Pahang, The Holy Book (*Kitab Suci*), The Malay and Bugis Tract, The Tale of Marakarma (The Poor), The Tale of Nakhoda Asik, The Tale of Pandawa Lima, The Tale of Raja Pasai, The Tale of Purasara, The Tale of Seri Rama, The Syekh Jalaluddin Explanation Letter, The Malay History, The Tale of Siak, *Tuhfat al-Nafis* and *Taj al-Salatin* (Roorda). There are 40 classical prose texts with instances of the word *revenge*. Therefore, based on these three emotional words, sentiment analysis was conducted through t-score, MI-score and logDice testing. These three emotion-based words were found in classical prose texts with high frequency values, especially the word *anger*, which occurs in 60 out of the 70 classical prose texts in the MCP. Meanwhile, the words *revenge* (in 40 classical prose texts) and *rage* (in 49 classical prose texts) were found in the 70 texts in the MCP.

## RESULT AND DISCUSSION

As a result of data generated for the emotions of anger, rage and revenge, three words provide further expression of those emotions. Among these are the words *anger*, *rage* and *revenge*, which express emotion in the classical prose texts. Through the prose texts' word frequency list, the frequency value of each of these emotional lexicons is referenced: the word *anger* = [120; 53.961687], the word *rage* = [246; 158.382694], and the word *revenge* = [63; 139.565795]. This word selection was based on a list of word frequencies with a high number of appearances. Thus, the expression of emotions was observable through the high frequency value of their respective usages.

### The Emotion Of Anger Revealed Through Words Expressing Anger

Earlier findings of this study used a word that represents angry emotions, namely *anger*. This word is often used in classical prose and has the same meaning as *anger*. The following figure shows the generated results for words expressing anger.

Collostructional [amarah]

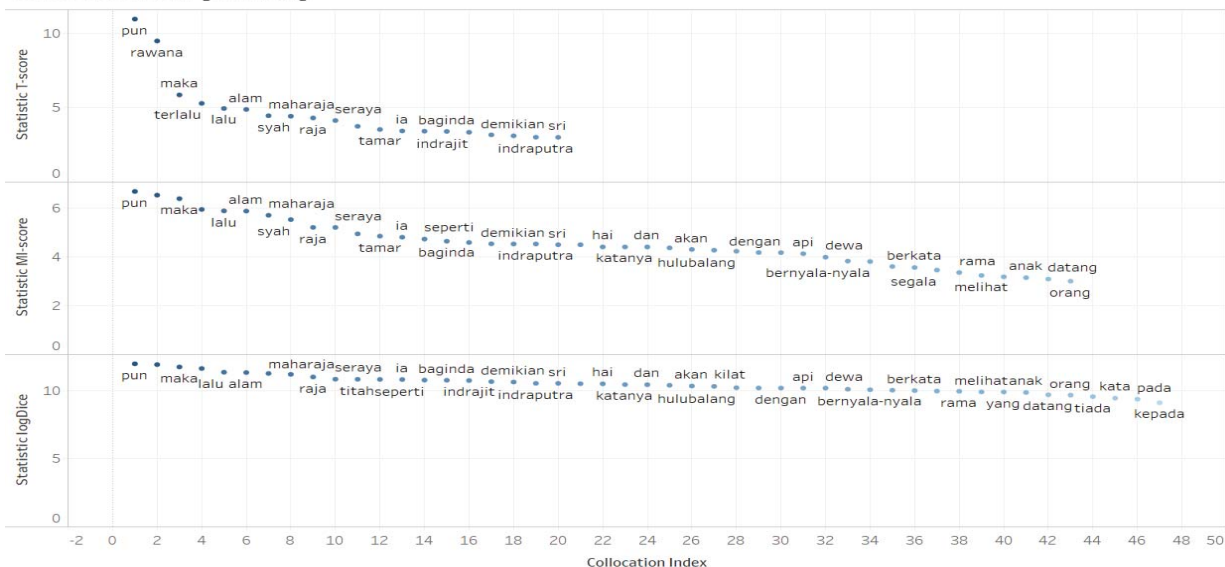


FIGURE 2. The word anger in t-score, MI-score and logDice tests

Figure 2 shows a word representation of anger derived from the t-score collocation testing. The representative results of this collocation reveal 20 words associated with the word *rage*. In this test, word content was represented less, with only 10 lexicons collocated with the word *anger*. Several word contents were represented, those of emperor

(*maharaja*) = [4.94451771], king (*raja*) = [4.422059311], *rawana* = [4.306622277], nature (*alam*) = [3.536332878], while (*seraya*) = [3.435514775], and he (*baginda*) = [3.172671459]. Examining this representative list, the test results provide a word collective structure of anger that is dominated by words referring to rulers, such as emperor and king. This level of testing provided a joint word consumption impact between the word anger and ‘government’ or ‘leaders’. Next, testing resumed with the word anger being assessed using MI-score, as shown in Figure 5. The MI-score results show 43 collostructure words related to anger. In this collostructure, a number of words that refer to content words were discovered, such as thunder (*kilat*) = [6.678045816], order (*titah*) = [5.941080177], date (*tamar*) = [5.882186547], his sword (*pedangnya*) = [5.526042705], *indraputra* = [5.204114587], flaming (*bernyala-nyala*) = [4.803576666], warload (*hulubalang*) = [4.587443298], he (*baginda*) = [4.526042726], fire (*api*) = [4.496295365], god (*dewa*) = [4.496295365], emperor (*maharaja*) = [4.36804089], see (*melihat*) = [4.175545454], he said (*katanya*) = [4.172405764], king (*raja*) = [4.127493347], *rama* = [3.807224479], son (*anak*) = [3.603210599], datang (*come*) = [3.562568585], and people (*orang*) = [3.180267891]. Through all the words pertaining to the collostructure of the word anger, three words can also be interpreted as expressing anger, including the words lightning, his sword and fire. These words connote negative representations and are closely related to the lexical item anger. The words fire and lightning are two implicit representations of anger. Furthermore, the words ‘his sword’ (*pedangnya*) form an image closely related to the word anger, as they represent the emotion of anger and instability. This section presents the results for the word anger in the logDice test (refer to Figure 2). Based on Figure 2, the word anger was tested and 47 words in the collostructure were generated. Through this collective word representation of anger, three words remain that are similar to the MI-score results: lightning, sword and fire. Among the represented words include *jalis* [10.6576078], *hanuman* [10.02730735], and flaming (*bernyala-nyala*). The word *jalis* (جليس) means a friend or companion [10]. Meanwhile, *hanuman* refers to the noun for *keris* (Malay dagger) which is *keris hanuman*, a Javanese type of *keris* that is a weapon shaped like a monkey’s head [11]. The word flaming (*bernyala-nyala*) is representative of burning or emitting flames and is specific to light. This word is often used in conjunction with the word fire to represent angry emotions [12].

### The Emotion Of Rage Expressed Through Rage-Related Words

According to the results for the word rage, emotions of rage have been represented by rage-related verbs, and refer to acts of attacking blindly due to being excessively angry. The emotion rage is the second stage that represents anger. Through the tests conducted on the word rage, three sets of results were generated, from the t-score, MI-score and logDice (refer to Figure 3).

Collocstructional [mengamuk]

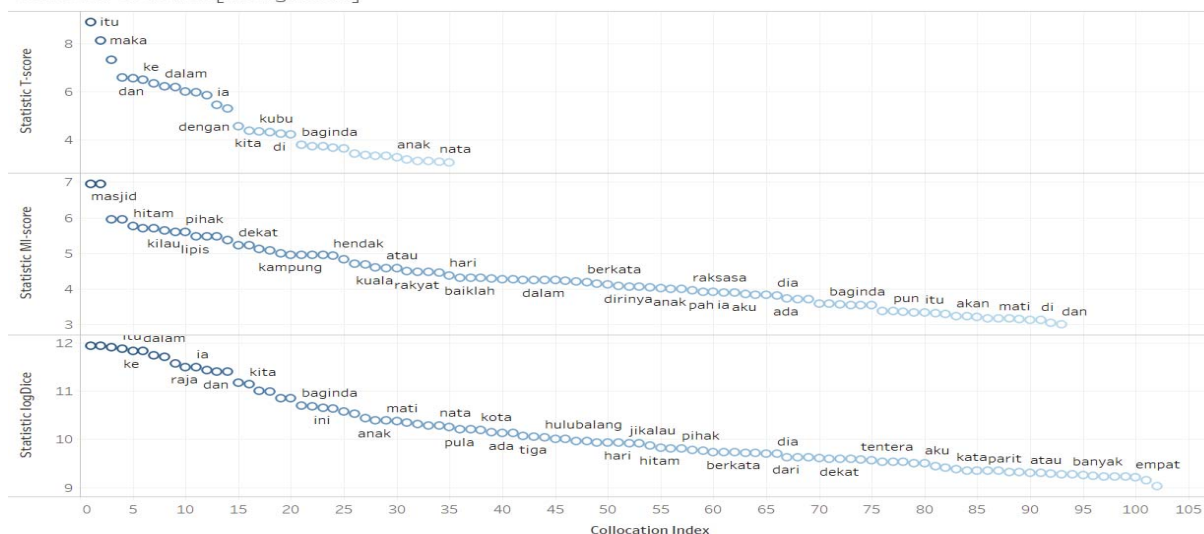


FIGURE 3. Mengamuk (raging) words according to the t-score, MI-score and logDice tests

Figure 3 shows the results for the word *mengamuk* (raging), with 35 words involved in the collostructure. Among the words involved are people (*orang*) = [6.591576198], citizen (*rakyat*) = [6.334835009], king (*raja*) = [5.855218538], fort (*kubu*) = [4.33179292], warrior (*panglima*) = [4.213041125], he (*baginda*) = [3.775724035],

emperor (*maharaja*) = [3.413623413], die (*mati*) = [3.320151826], son (*anak*) [3.245753982], monster (*raksasa*) = [3.103124932], and me (*patik*) = [3.050317394]. Through this representation of collocation, several words can be examined, i.e., the word correlation of rage with the words dead and monster. These two words refer to the context of consumption in the collocation of the word rage by showing a negative scenario involving death and mythology, and often including monsters. Giants are mythical subjects that often appear in classical texts and provide interpretations of destruction and rage [13]. The word death indicates the impact of the outcome of rage-related incidents. Based on the psychological context, the act of rage is a mental disorder in which a person acts on some event in a manner that is outside normal limits of behaviour [14]. This word is often used in classical Malay texts to illustrate scenarios of rage [15]. Thus, the word death is the result of actions involving rage. Next, Figure 3 reveals the results of the word raging based on MI-score testing. The results showed that 93 words were collocated with the word raging. Among the words studied were mosque (*masjid*) [6.957408912], black (*hitam*) [5.764763742], army (*tentera*) [4.957408728], raid (*menyerbukan*) [4.677300949], enemy (*musuh*) [4.24120175], war (*perang*) [4.234942754] and run (*lari*) [3.294443783]. All these word representatives indicate the context of consumption along with the word rage. Rage words are used based on places, such as the words mosque and fort (representative in the t-score and MI-score testing). In addition, a representative colour image, the word black, is used in relation to the word raging. The context of the act of rage also involves the presence of the word togetherness of the army and the enemy. Furthermore, certain words provide representations of rage with the verbs pollen, war and running. Lexicons involving the words dead and monsters also appear in representative MI-score results. Figure 3 shows the word rage through the logDice testing, the results of which reveal 102 word collocations. Among the words that appear in this representation are those which indicate human entities, such as citizen (*rakyat*)= [11.93195461], king (*raja*)= [11.49957101], warrior (*panglima*)= [11.00904514], emperor (*maharaja*)= [10.52551404], prince (*raden*)= [10.34568945], warlord (*hulubalang*)= [10.03190925], army (*tentera*)= [9.573735245], and slave (*hamba*)= [9.535113952]. These all reflect the strata of historical societies. The word rage affects the strata of a society or government that triggers war or chaos. Most such texts describe the rampant emotions that occur between one party and another when they encounter each other [16], [17]. In addition, the results are representative of areas or places, such as the words fort (*kubu*)= [11.16710999], town (*kota*)= [10.19264508], village (*kampong*)= [9.966576999], and mosque (*masjid*)= [9.624960569]. These words, illustrating places or areas, are used with the word raging.

### The Emotion Of Revenge Expressed Through Revenge-Related Words

The final results refer to expressions of the emotion of revenge through the use of the words vengeance or revenge in classical prose texts. The word revenge was generated and results were obtained for the same three tests. The following are the results of the first test, the word revenge through t-score testing:

Collostructional [dendam]

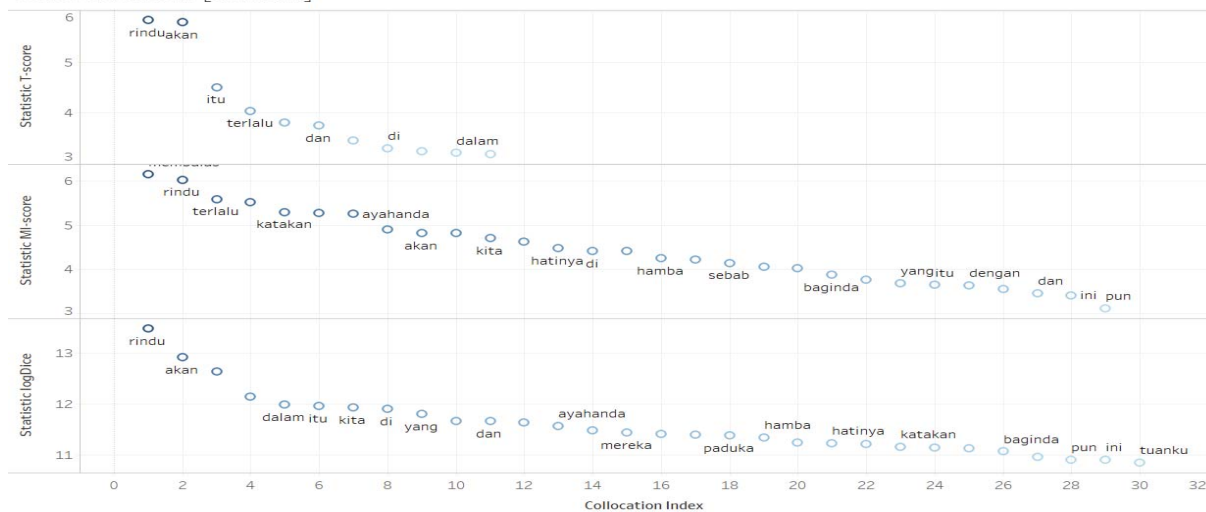


FIGURE 4. The word revenge (*dendam*) from t-score, MI-score and logDice testing

The word revenge has various meanings that highlight the act of holding feelings within the heart, including anger, love and others. The results reveal only 11 word collocations. Among the words involved were miss/ longing

(*rindu*)= [5.825165414], really (*terlalu*)= [4.037283652], deep (*dalam*)= [3.205666447] and we (*kita*) = [3.188595934]. The word longing (*rindu*) has the opposite impact to the interpretation of the word revenge (*dendam*). In the Malay language, there is a phrase ‘longing for revenge’ (*rindu dendam*) which refers to the extreme feeling of wishing to take revenge [18]. The interpretation of revenge is strongly linked with the word longing because the action represents a turbulent human psychology [10]. On the other hand, longing for revenge (*rindu dendam*) also connotes a negative interpretation, that of an extreme addiction [10]. The words very (*terlalu*) and deep (*dalam*) indicate the level of the word revenge because the representations of these two words are significant and positioned to the left and right of the word revenge: [very + revenge] and [revenge + deep] (refer to the discussion section). The pronoun we (*kita*) appears in the t-score test, as its word value is significant with the word revenge. Next, a MI-score test was performed to identify the word representations of collocations with the word revenge (refer to Figure 4). The results of the MI-score test indicated 29 word collocations with the word revenge. Figure 4 shows the collocation word representations of the word revenge. Among the represented words were pay (*membalas*) = [6.142633844], majesty (*paduka*) = [5.268164716], father (*ayahanda*) = [5.249549056], his heart (*hatinya*)= [4.479668836], slave (*hamba*) = [4.249549056], he (*baginda*) = [3.869615368], highness (*tuanku*) = [3.764122228], and king (*raja*) = [3.557671361]. The words longing (*rindu*), intense (*terlalu*) and in (*dalam*) had the highest word values, between 6.02398937 and 4.901625752. Other words represented modifiers like the word very (*sangat*) [5.514602652]. Figure 4 displays the results from the logDice test on the word revenge, from which 30 represented collocation words were observed, none of which revealed any distinction when tested using the MI-score. However, there were differences in the index position and there was an increase in the number of functional words. Meanwhile, the word content remains with no addition. The word miss (*rindu*) is the first index representative, with a value of 13.47107153. The words that represent humans include father (*ayahanda*), sir (*tuan*), Your Majesty (*paduka*), slave (*hamba*), His Majesty (*baginda*), king (*raja*) and my lord (*tuanku*). They carry prudent values to the word revenge (*dendam*). The representative values of words for human entities ranged between 11.55942741 and 10.84619466. The representatives of the t-score, MI-score and logDice test results were assessed based on their significant threshold values. This representativeness is a word collocation trend that appears in the collostructure. This word-based statistical requirement explains the position and relationship of the prudent value between the emotional words, namely anger, rage and revenge, and the collocations with these words. Such an impact facilitates the interpretation of automatic readings to gain a significant understanding of the word evidence and the main emotional words.

## CONCLUSION

The above results indicate that the words anger (*amarah*) and rage (*amok*) can be classified as negative sentiments. The word revenge (*dendam*), on the other hand, can be classified as a positive sentiment. This word classification was determined based on the evidence between word occurrences in the collostructure. The word anger expresses angry emotions that occur with negative sentiments, indicated by the presence of word collocations such as lightning (*kilat*), his sword (*pedangnya*) and his fire (*apinya*). Such contexts occur predominantly with significantly occurring word frequency. Most rage-related words are used to reflect angry emotions expressed by leaders. This is clearly based on the presence of nouns or pronouns that collocate with the word anger, namely the words the king (*baginda*), king (*raja*), emperor (*maharaja*), and *indraputra*. As for the word rage, emotional expressions of rage are used with negative sentiments when certain negative words are involved. Among the negative words collocated with such emotions is the word dead (*mati*), which is representative with significant frequency. Similarly, representatives of the word leader are also present in the collocations with the word rage. Among the words that express the term leader are king (*raja*), emperor (*maharaja*), commander (*panglima*) and *raden*. Next, the word revenge is an emotional expression of emotions involving revenge, with a classification as a positive sentiment. This was because the collostructure of the word revenge predominantly shows the presence of the word longing (*rindu*). The words longing (*rindu*) + revenge (*dendam*) are present and express the emotion of wishing to take revenge against someone who has left an impression of affection on the individual who expresses those emotions. The word leader also exists in expressions of the emotion of revenge. Among the words referring to leaders are Your Majesty (*paduka*), His Majesty (*baginda*), king (*raja*) and my lord (*tuanku*). Thus, it can be concluded that the emotions of anger, rage and revenge occur with positive and negative classifications through the involvement of historical leaders. The suggestion of future research is that leadership emotions in modern texts can be revisited through analytical sentiment and modern corpus data.



## ACKNOWLEDGMENTS

This paper is based on a research project entitled: Development Of Emotional Intelligence Model Framework Malay Leadership Then And Now According To The Thoughts Of Authors Art For The Society; Grant (FRGS/1/2018/SS102/UPSI/02/1) prepared by the Ministry of Education Malaysia. Acknowledgment is also given to the Malay Concordance Project as a secondary data source.

## REFERENCES

1. K. Sailunaz and R. Alhaji, *Journal of Computational Science* **36**, 101003 (2019).
2. Z. Feng, *Personal and Ubiquitous Computing* **23**, 373 (2019).
3. M. Kaity and V. Balakrishnan, *Knowledge and Information Systems* **62**, 4445 (2020).
4. S.M. Mohammad, *Emotion Measurement* 201 (2016).
5. A. Pak and P. Paroubek, *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010* 1320 (2010).
6. V. Brezina, *Statistics in Corpus Linguistics*, First (Cambridge University Press, United Kingdom, 2018).
7. T. Defanti, A. Grafton, T.E. Levy, L. Manovich, and A. Rockwood, *Lexical Collocation Analysis: Advances and Application*, First (Springer Nature Switzerland AG, Cham, 2018).
8. L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, *Knowledge and Information Systems* **60**, 617 (2019).
9. H.C. Link, *The American Journal of Psychology* **32**, 134 (1921).
10. F. Didonna, *Clinical Handbook of Mindfulness* (2009).
11. Mahyudin, *SARI: Jurnal Alam Dan Tamadun Melayu* **27**, 27 (2009).
12. R. Rashidin, in *Seminar Linguistik Kebangsaan (SLiK2015)* (2015), pp. 29–47.
13. J.H. Ruiz, *Interlingüística* **17**, 475 (2007).
14. L. Adler, D. Marx, H. Apel, M. Wolfersdorf, and G. Hajak, *NCRM Working Paper 1* (2006).
15. J.H. Flaskerud, *Issues in Mental Health Nursing* **33**, 898 (2012).
16. R. Rashidin, *E-BANGI: Jurnal Sains Sosial Dan Kemanusiaan* **3**, 32 (2008).
17. R. Rashidin and N.H. Jalaluddin, *Procedia - Social and Behavioral Sciences* **118**, 412 (2014).
18. B. Noresah, editor, *Kamus Dewan*, 4th ed. (Dewan Bahasa dan Pustaka, Kuala Lumpur, 2013).