

The CALA 2020 Proceedings Paper 11 - 1

*Cognitive Anthropology and Language, Paper 1*

**A Study of Malay Manuscripts (Petua Membina  
Rumah): Geometric Pattern Lexical Density**

Muhamad Fadzllah Zaini  
*Universiti Pendidikan Sultan Idris, Malaysia*

Anida Saruddin  
*Universiti Pendidikan Sultan Idris, Malaysia*

Mazura Mastura Muhammad  
*Universiti Pendidikan Sultan Idris, Malaysia*

# A Study of Malay Manuscripts (Petua Membina Rumah): Geometric Pattern Lexical Density

Muhamad Fadzllah Zaini<sup>a</sup>

<sup>a</sup>*Universiti Pendidikan Sultan Idris, Malaysia*

Anida Saruddin<sup>b</sup>

<sup>b</sup>*Universiti Pendidikan Sultan Idris, Malaysia*

Mazura Mastura Muhammad<sup>c</sup>

<sup>c</sup>*Universiti Pendidikan Sultan Idris, Malaysia*

---

## Abstract

This study examines the lexical density of classical Malay texts. This study is divided into three main parts, namely, (1) the form of the keyword list in a classic text compared to a current text, (2) 10 keywords that appear as a result of the Simple Math Procedure (SMP) test, and (3) geometrical patterns that are produced to examine the lexical representations of classical texts according to their SMP values. This study uses the SMP as is pre-defined in the LancsBox 4.5 software. This study was conducted on seven classical texts that have been developed in the *Korpus Petua Membina Rumah* (KPMR) (Tips for Building a Home Corpus). The rationale of this study is to test lexical density as a complex lexical representation in classical texts. In addition, the relation between text and community was also highlighted as a phenomenon of traditional knowledge and feudal thought.

**Keywords:** Language, culture, classic text, corpus linguistics

---

## Introduction

Lexical density is a statistical measurement of the richness of a text. This statistical measurement refers to the lexical frequency that appears when representing a type of word (Gregori-Signes and Clavel-Arroitia 2015). The current study of corpus linguistics contains specifically witnessed lexicons that shape grammar according to the needs of language morphology. This is owing to that lexical wealth can be measured in a corpus. Over the years, classical textual studies have been carried out by incorporating corpus linguistics as an approach (Abainia 2019; Hammo, Yagi, Ismail and AbuShariah 2016; Norri, Junkkari and Poranen 2019; Rubinstein 2019; Timmis 2018). Classical texts contribute significantly to lexical representations according to textual content either chronologically or synchronously. As such, this study tested the lexical density to examine lexical richness and the structural pattern of classical Malay texts in applying the corpus linguistics approach. This study used the KPMR, which is a specialized corpus. This corpus consists of seven classic Malay texts, which are MSS741, MSS1415, MSS1521, MSS1849, MSS2001, Tajul Muluk and Kitab Abu Masyar. The token for this corpus is 14,854 and the value of type is 2,027 (13.67 TTR). The lexical test uses the Simple Math Procedure (SMP) generated through LancsBox 4.5 as a research instrument. This study uses the corpus referred to by the Berita

Harian corpus (KBH) as a lexical comparison of representative keyword lists. The rationale for this study is to examine the lexical structure form of classical Malay texts using PMR. This is significant to identify the richness of Malay lexicons in classical texts.

## Lexical Density Results

The results show that there are 10 positive keyword lexicons that appear in the keyword list. The keyword list is a lexical form that was compared using the SMP test (Brezina 2018). Once content filtering was performed, a list of keywords in the form of content was submitted for the discussion of lexical density or lexical richness in classical text. Content words are lexical forms that have a dominant relationship with the text. Figure 1.0 represents the distribution of content key words in the classical KPMR Malay text.

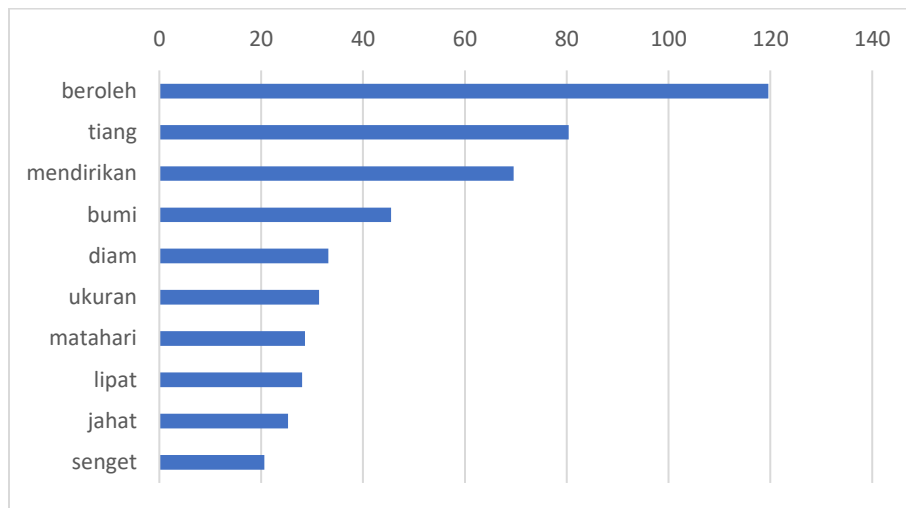


Figure 1.0 Distribution of 10 Lexicons in the Keyword List

Figure 1.0 shows the content words in the list of keywords whereby a significant SMP value is a lexicon with 119.5768032 and the tenth lexical '*senget*' is an adjective which had a significant value of SMP v 20.60332658. All of these lexicons have a positive representational value in KPMR compared to KBH. This is owing to that lexical domination is widely used in the KPMR as a content descriptor for home-building tips. The dominant nouns in these 10 lexical terms are *tiang*, *bumi*, *ukuran*, *matahari*. This is followed by verbs such as the lexicons *beroleh*, *mendirikan*, *diam*, *lipat*. Lastly are the adjectives *jahat* and *senget*.

## Geometric Pattern

As a result of this lexical density result, the lexicon can be plotted, based on the co-occurrence relation among lexicons. This involves a collocation graph to examine the position of lexicons in classical Malay texts. A collocation graph is a lexical form or plot based on the position between lexicons for content words and represents a co-occurring lexical collocation (other than lexical content) (Brezina, McEnery and Wattam 2015; Bruening 2019; Gablasova, Brezina and Mcenery 2017; Seretan 2008; Stuart and Botella 2009; Xiao 2015). Figure 2.0 shows the collocation of 10 content word lexicons that are included in the keyword list.

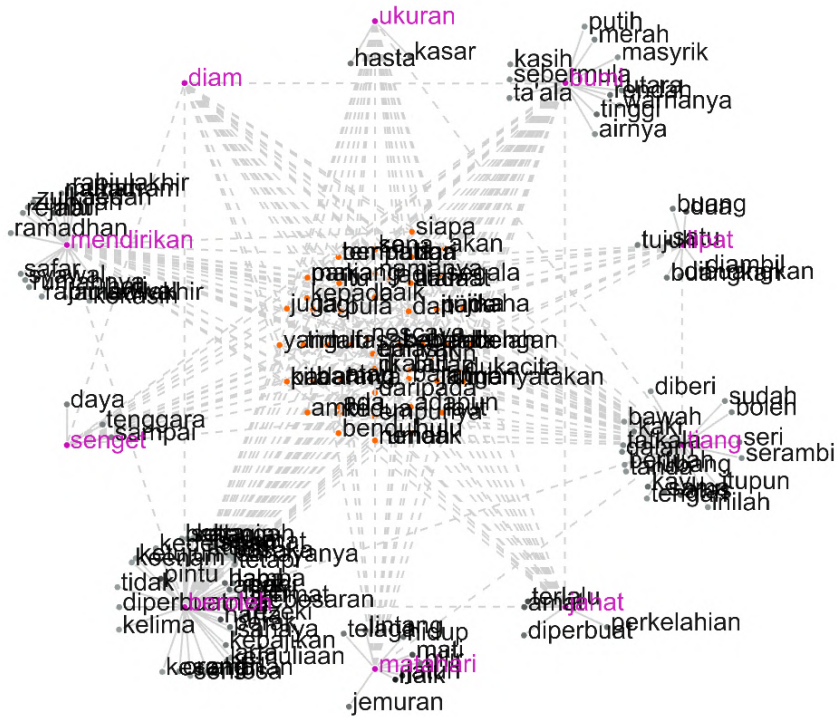


Figure 2.0 Geometric Pattern Representative Visualization of 10 Content Words

Figure 2.0 shows a representative visualization of geometrical patterns that map 10 interconnected content words. The dashed lines indicate the relationship between lexicons as a co-occurrence. The 10 lexical content words are *ukuran*, *bumi*, *lipat*, *beroleh*, *tiang*, *jahat*, *matahari*, *senget*, *mendirikan* and *diam*. Only two lexicons have a co-occurrence value among lexicons by connecting four other lexical content words, namely *tiang* and *beroleh*. The lexicon *tiang* co-occurs with the lexis *ukuran*, *lipat*, *matahari* and *beroleh*. Meanwhile, the lexical *beroleh* co-occurs with the lexis *mendirikan*, *diam*, *tiang* and *jahat*. The co-occurrence relationship suggests the complexity of a classical Malay text that is loaded with knowledge (Amer Hudhaifah 2017; Muhammad Akmal R. Azmi and Faudzinain 2018). This is owing to that the complexity of the past community in developing thought shows significance towards the rating of a specific knowledge (Hassan 2016; Muhamad Fadzllah Zaini and Hashimah 2019).

## Conclusion

The Simple Maths Procedure (SMP) test was performed to represent lexical density or lexical richness in KPMR. This test provides a refined keyword list by focusing on previous content words. This is because the examination of the content words provides the focus of KPMR. The results show that lexical verb *beroleh* can precede and end with the lexical adjective *senget*. The lexical pillar precedes the noun to indicate that this lexical representation is significant and widely used. Next, the lexical verb *mendirikan* became dominant so as to refer to construction work. Therefore, it can be seen that this lexical use reflects the socioculture and architectural knowledge of the past. The adjustments used give a reflection of previous architecture. The relationship between text and people's lives is thus closely linked to the value of nation building.

## Acknowledgement

This research has been carried out under Fundamental Research Grants Scheme (FRGS/1/2018/WAB04/UPSI/02/5) provided by Ministry of Education of Malaysia

## References

- Abainia, K. (2019). DZDC12: a new multipurpose parallel Algerian Arabizi–French code-switched corpus. *Language Resources and Evaluation*. <https://doi.org/10.1007/s10579-019-09454-8>
- Amer Hudhaifah, H. (2017). Manuskrip Melayu : Isu Kontemporari dan Lontaran Idea. *International Journal of West Asian Studies*, 9(1), 25–38. <https://doi.org/10.22583/ijwas.2017.09.01.03>
- Brezina, V. (2018). *Statistics in Corpus Linguistics*. In Cambridge University Press (First).
- Brezina, V., McEnery, T., and Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20(2), 139–173.
- Bruening, B. (2019). Idioms, collocations, and structure. In *Natural Language and Linguistic Theory*.
- Gablasova, D., Brezina, V., and Mcenery, T. (2017). Collocations in Corpus-Based Language Learning Research : Identifying , Comparing , and Interpreting the Evidence. *A Journal of Research in Language Studies*, 67(1), 155–179. <https://doi.org/10.1111/lang.12225>
- Gregori-Signes, C., and Clavel-Arroitia, B. (2015). Analysing Lexical Density and Lexical Diversity in University Students' Written Discourse. *Procedia - Social and Behavioral Sciences*, 198(Cilc), 546–556.
- Hammo, B., Yagi, S., Ismail, O., and AbuShariah, M. (2016). Exploring and exploiting a historical corpus for Arabic. *Language Resources and Evaluation*, 50(4), 839–861. <https://doi.org/10.1007/s10579-015-9304-9>
- Hassan, A. (2016). *Bahasa dan Pemikiran Melayu (First)*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Muhamad Fadzllah Zaini, and Hashimah, J. N. (2019). Struktur Seni Bina Melayu “ Tangga “ Dalam Peribahasa Melayu : Analisis Semantik Inkuisitif. *Jurnal Antarabangsa Dunia Melayu*, 12(1), 1–26.
- Muhammad Akmal R.Azmi, and Faudzinaim, B. (2018). Manuskrip MS 666 Karya Shaykh Daud al-Fatani: Satu Refleksi Awal. *Ational Studies and Human Sciences International Journal of Civilizational Studies and Human Sciences*, 1(4), 107–119.
- Norri, J., Junkkari, M., and Poranen, T. (2019). Digitization of data for a historical medical dictionary. *Language Resources and Evaluation*. <https://doi.org/10.1007/s10579-019-09468-2>
- Rubinstein, A. (2019). Historical corpora meet the digital humanities: the Jerusalem Corpus of Emergent Modern Hebrew. *Language Resources and Evaluation*, 1–29. <https://doi.org/10.1007/s10579-019-09458-4>
- Seretan, V. (2008). Collocation extraction based on syntactic parsing. (July), 265. Retrieved from [internal-pdf://92.0.0.10/Seretan-2008-Collocation extraction based on s.pdfCollocation%0Aextraction%0Abased%0Aon%0As.pdf%0Afile:///Users/aa0aa/Library/Application Support/Mendeley Desktop/Downloaded/Seretan - 2008](internal-pdf://92.0.0.10/Seretan-2008-Collocation%20extraction%20based%20on%20s.pdfCollocation%0Aextraction%0Abased%0Aon%0As.pdf%0Afile:///Users/aa0aa/Library/Application%20Support/Mendeley%20Desktop/Downloaded/Seretan%20-2008)
- Stuart, K., and Botella, A. (2009). *Corpus Linguistics, Network Analysis and Co-occurrence Matrices*. *Internation Journal of English Studies*, (Special Issue), 1–28.
- Timmis, I. (2018). The Pragmatics of Poverty in the Essex Pauper Letters, 1731–1837. *Corpus Pragmatics*, 2(3), 243–263.
- Xiao, R. (2015). Collocation. *The Cambridge Handbook of English Corpus Linguistics*, 106–124. <https://doi.org/10.1007/9781139764377.007>