Research Journal of Applied Sciences 5 (6): 429-433, 2010
ISSN: 1815-932X

Lexical Associations of Malayness in Hikayat Abdullah: A Collocational Analysis

[1]Imran Ho-Abdullah, [1]Ruzy Suliza Hashim and [2]Norhafizah Mohamed Husin
[1]Pusat Pengajian Bahasa and Linguistik, Faculty of Sains Sosial and Kemanusiaan,
University of Kebangsaan, Malaysia
[2]Program Teknologi Bahasa, Bhg. Peristilahan and Leksikologi,
Jabatan Pembinaan Bahasa dan Sastera

Abstract: Lexical collocation is a universal phenomenon in any language. Repetitions of any lexical item with other words form specific patterns. From the point of view of natural language processing, collocations can be divided into two categories. First, collocation which is formed by chance because of the high usage of any particular word in a text. The other category are words which collocate with other words for some significant reason. The patterns of collocation are different between texts and users. This study explores lexical collocations of the word Melayu in Hikayat Abdullah, the nineteenth century text renown as the first Malay autobiography by Munsyi Abdullah. This quantitative study investigates the strength of association of the collocates using the mutual information scores and the semantic fields of the collocates to discern the representation of Malayness in the Hikayat. The results indicate that the collocates with strong association with Melayu to the semantic fields of bahasa (language); bangsa (race); institusi (institutions) and nilai (values).

Key words: Hikayat Abdullah, corpus linguistics, collocation analysis, Mutual Information (MI) score, strength of association, Malaysia

INTRODUCTION

Abdulah bin Abdul Kadir better known as Munsyi Abdullah is often described as the father of modern Malay literature. Munsyi Abdullah was born in 1797 (a contemporary of Jane Austen) in Melaka and lived during the colonial British era in Malaya. Hikayat Abdullah which was completed in 1843 is an autobiography and represents his later work in contrast to his earlier research Kesah Pelayaran Abdullah which provides an account of his travels to east coast of the Peninsular Malay in 1837. Alatas (1977) has argued that Abdullah while critical of the Malays, he does not uncritically absorb colonial attitudes but transforms them in a quest for modernity. While reproducing British colonial stereotypes depicting Malays as lazy for instance, Abdullah nonetheless finds a different reason for such behaviour.

This study seeks to examine the representation of Malayness in Munsyi Abdullah's Hikayat using a corpus linguistics approach. Specifically, the study uses collocational analysis (Baker and McEnery, 2005; McEnery *et al*., 2006) in the analysis of the Hikayat. The analysis concentrates on the study of the distribution of the strength of association using the mutual information scores of the collocates with the word Melayu. In addition, the study also look for recurrent semantic fields strongly associated with Melayu among the collocates. As Sinclair (1991) argues, collocates can contribute to the semantic analysis of a word and the examination of collocation patterns can reveal systematic semantic associations.

MATERIALS AND METHODS

There are several versions of Hikayat Abdullah. They include 1914 edition published by Methodist Publishing House, Singapore; the 1965 version by published by Malaya Publishing House Limited, Singapore and the 1997 edition published by Pustaka Antara, Singapore. The collocation analysis is based in the digitised version of Hikayat Abdullah published by Pustaka Antara in 1997 which is held in the Dewan Bahasa and Pustaka Corpus Database (B00679). The text contains 106,065 words in total. The basic statistics of the corpus is shown in Table 1. The collocates for the word Malay were generated using Oxford Wordsmith (version 5) with a collocate horizons spanning 5 to the left and 5 to the right of the search word. The strength of the collocational relationships of each collocate with the search word (also called the association measures) were calculated using the

Corresponding Author: Imran Ho-Abdullah, Pusat Pengajian Bahasa and Linguistik, Faculty of Sains Sosial and Kemanusiaan, University of Kebangsaan, Malaysia

Table 1: Basic statistics of Hikayat

Parameters	Values
Number of words (tokens)	106,065
Number of types	6,682
Type:Token ratio	6,11

Mutual Information (MI) score. The Mutual Information (MI) score in Wordsmith is derived using the formula of Gaussier, Lange and Meunier described in Oakes (1998). The formula is a probabilistic logarithm of two events, x and y appearing together divided by the probablity of both occuring independently. A score of 0 means there is not relationship between the two words while a score greater than 0 indicates a propensity for the two words to collocate is more than a matter of chance. The higher the score, the stronger the association between the two words (McEnery *et al*., 2006). Hunston (2002) suggests that a MI score >3 can be accepted as evidence of a valid collocate.

RESULTS AND DISCUSSION

Based on the span of 5 to the left and 5 to the right a total of 92 collocates for the word Melayu was generated. Of these only 82 words have a MI score ≥3. In addition, only collocates that have a frequency five times or more is included in the analysis (Scott, 1996) (Appendix A). However, the definition of collocation advocated by Choueka (1988) reminds us that collocation does not only refer to two or more words which are adjacent to each other (technical collocates) but also have syntactical and semantical characteristics and the meaning or the connotative meaning cannot be expressed directly from each of the collocational component independently.

Hence, the reliance on the MI score to establish the measure of association between the words that have been identified as true collocates and eliminate collocates that exist by chance. The highest MI score was for the word umpamaan (MI = 7.34). The word umpamaan has the highest significant association with the word Melayu in Hikayat Abdullah.

The high score means that we can confident that umpamaan appears in the same context as Melayu, i.e., in the phrase umpamaan Melayu. In 13 of the 14 occassions when umpamaan occurs, it collocates with Melayu (the other instance being umpamaan Cina). In contrast, the word bahasa with a frequency of 467 times occupies second position in terms of the MI score. Bahasa collocates with Melayu 289 times out of the 467 instances of bahasa. Other words with a MI score of 6 includes cara, nahu, bertutur, dipakai, karangan, canggung, tanya, Keling and belajar. Interestingly, only two of these collocates, namely bahasa and belajar occurs in high frequencies in Hikayat (f>100). In the text, the concept bahasa is highly associated with the concept Melayu in the phrase bahasa Melayu. An examination of the concordance for bahasa reveals that bahasa is also used in the phrases bahasa Hindu, bahasa Keling, bahasa Portugis, bahasa Cina and bahasa Inggeris. In the case of the verb belajar (to learn) the association with Melayu is seen in phrases such as belajar Melayu, bangsa Melayu yang harus belajar menjadi berilmu, Melayu harus belajar bahasanya dan semangat belajar orang Melayu. As with the word bahasa, the MI score >6 suggests that the collocation of belajar and Melayu is highly associative and significant. Other words that are significantly associated and collocate with Melayu MI >6 are cara, nahu, karangan and Keling.

Although, most of these words occurs infrequently; cara (method, ways) occurs 15 times; nahu (grammar) 5 times, karangan (composition) 6 times; it does not affect the strong propensity to collocate with Melayu. In other words, these items are found exclusively with Melayu in the Hikayat. In contrast, Keling which occurs frequently in the Hikayat (f = 65) only collocates 23 times with the word Melayu.

Other instances of the word Keling in the Hikayat relates to bangsa, budaya, cara, negara, saudagar and the name of a place in Melaka-Tanjung Keling. The strong association of Keling in the context of Melayu is interesting and revealing, considering the fact that the word Inggeris which has a higher frequency (f = 395) has lower affinity with the word Melayu (MI = X). The collocation analysis also show that the word Melayu has strong affinities with the verb bertutur (to converse) in phrases such as kepentingan bertutur bahasa Melayu and bertutur bahasa-bahasa lain.

Another verb relating to speech act that is found to strongly collocate with Melayu is the word tanya (to ask) in the context relating to kitab, pelajaran, perkataan and usaha untuk mempelajari sesuatu. The most significant adjectival collocate for Melayu in the Hikayat is canggung (awkward). Almost all instances of canggung refers to awkward noises, words and language of a person. There is only a single instance of use of the word canggung in relation to female behaviour.

The word dipakai (used) as a colloate of Melayu mainly relates to perkakasan, perkataan and kias ibarat Melayu. A second group of collocates of Melayu with a MI score of 5 includes undang-undang, menjadikan, huruf, mengerti, membaca, senang, jalan, kitab, hikayat, raja, asal, betul, mengajar, surat, cap and diketahui.

In terms of frequencies, these group of collocates can be divided into two groups: those that occur with relatively high frequency (f>20) and those that occur in

lower frequencies. Collocates with high frequencies of occurrence are jalan 35 times, kitab 48 times, raja 52 times, surat 60 times. Collocates with lower frequency of occurences include undang-undang occurs 8 times in the text; huruf 12 times; hikayat 16 times, asal 9 times and cap 6 times. The higher frequency of the words jalan, kitab, raja and surat also reveals to us the construal of Malayness in the Hikayat.

In the case of the words with lower frequencies, the context for asal also appears with usul, asli and asalkan. Huruf and aksara in relation to Melayu falls within the domain of discussion of the Malay language. Hikayat relates to kisah atau cerita while the word cap relates to mohor, jenis hukuman dan kaedah percetakan. Four verbal collocates of Melayu are found with MI scores of 5- menjadikan, membaca, mengajar and diketahui. The four words occur between 20-80 times in the texts and collocate with Melayu around 20 times each. Most collocates with a MI score of 4 are nouns and adjective. Noun collocates include Cina, bunyi, Inggeris, saudagar, rahsia, adat, perkataan, kata, hukum, bangsa, nama, guru, perintah, tuan, anak dan orang.

Compared to a frequency word list analysis of Hikayat where the words orang (f = 2,253) and Inggeris (f = 395) appear as two of the most frequent words, the collocation analysis reveals that these words (despite their high frequencies) are not as strongly associated with the word Melayu in the Hikayat as some other words.

The word Cina which has and occurrence of only 185 times in the text is found to have similar affinities to collocate with Melayu. The ratio of frequency of occurrence in the text to frequency of collocations with Melayu for the respective collocates in this group (MI>4) are bunyi (64:9), saudagar (72:10), rahsia (36:5), adat (132:18), perkataan (234:31), kata (230:16), hukum (45:6), bangsa (120:14), nama (114:12), guru (63:6), perintah (67:6), tuan (1452:25), anak (307:25) dan orang (2254:183). An adjective with the MI score of >4 is pandai which is used 135 times in the text but collocates 10 times with the word Melayu.

Other adjectives which are found to collocate with Melayu within the same MI range are putih, sedikit, habis and tahu. Of the remaining 82 collocates, 28 falls into the range of MI score >3. These words can still be regarded as strong collocates of Melayu. The distributions of collocates according to their strength of association with Malays are shown in Fig. 1.

In terms of the distributions of collocates, 90% of the 91 collocates (i.e., 82 collocates) have a MI score >3. Only 10% (9 collocates) have strength of association <3 on the MI score. All these words have very high frequency of occurrence in the text (between 200-1000) such as aku, hari, mereka and besar.

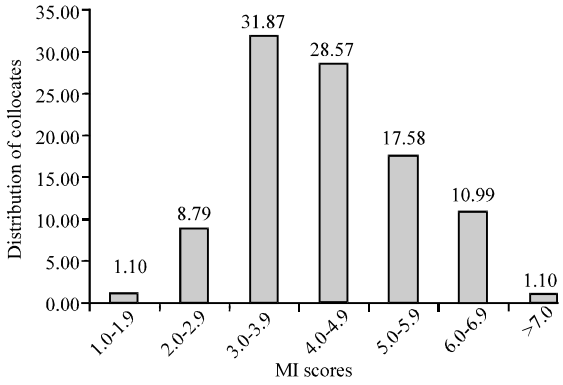

Fig. 1: Distribution of collocates and MI scores

CONCLUSION

This study has utilized the corpus methodology of collocation analysis to observe the strength of association between collocates using the mutual information scores. The method has allowed us to differentiate and group the collocates according to their strength of collocation. Based on their MI scores, 82 words were found to have a strong collocation with the word Melayu in the Hikayat. The relationship between the words can reveal both syntactic (phrases) and semantic motivations (lexical propensities and lexical restrictions) of Munsyi Abdullah's representation of Malayness in the Hikayat through the lexical choices that he makes.

Based on the collocations, an image of the concepts in relation to the search word Melayu is formed. The collocates and their semantic prosodies can produce a pattern that reveals the collocative meanings of Malayness in the text. At least four groups of collocative meanings can be identified: bahasa, bangsa, institusi and nilai. Munsyi Abdullah's observation of the social changes and his exposure to various knowledge and cultures have extended his concerns of Malayness which encompass ilmu, sastera, wilayah and seni budaya. The collocative meaning of Malayness in bahasa is reflected and amplified in numerous collocates such as kitab, kata, hikayat, belajar, bunyi, menulis, membaca, mengajar, karangan, nahu, surat, jalan, guru, cara, bertutur, perkara, hal dan bunyi. This semantic field of bahasa also extends to a wider and more comprehensive notion of the Malay language such as dialek, persuratan, peribahasa, pengarang, manuskrip, pantun, perkataan, tatabahasa, naskhah, and perkamusan. The changes in the choice of collocates from nahu to tatabahasa, bahasa to linguistik and umpamaan to peribahasa reveals the enlargement and progression of Munsyi Abdullah's vocabulary in relation to language studies over time. Introduction of new concepts such as dialek, versi, perkamusan, pustaka dan naskhah also reveals the development in the field of Malay language studies. At the same time the collocates

also reveals the influences from other languages. Dialek and versi are derived from English while naskhah shows the influence of Arabic in the Hikayat. A second group of collocates revolves around the collocative meaning of bangsa. Colocations such as Inggeris, Cina, Holanda and Keling reveals that the Malays have established an active social (and commercial-saudagar) relationships with these group of people and have included them in their discourse. The collocates manusia, umat, kaum, tamadun, rumpun, bumiputera and kelompok show Abdullah's concern with bangsa in a wider context.

The group of collocates associated with nilai revolves around character and attitude such as putih, baik, tahu, canggung, senang, banyak, diketahui, bernama and sedikit. The collocates in the semantic field of nilai also include items valued by the Malay such as sejarah, kebudayaan, tradisional, kebangsaan and nasionalisme. Finally, the collocates connected to the notion of institusi also occur in some abundance. Words such as raja, negeri, adat, hukum, undang-undang, perintah and cap represent the Malay administrative institutions. Similarly, collocates such as masyarakat, persatuan, negeri, kesultanan, parti, pemimpin, kesatuan, keluarga, pertubuhan, istiadat, rumah, pembesar, sekretariat and empayar all contribute to building a rich collocative meaning of the Malay institusi in the Hikayat.

The collocational analysis of Melayu in the Hikayat have constructed for us Abdullah's world view of the Malays that is comprehensive and encompassing. Apart from the language, the institutions, the values and the race, concerns with ilmu, sastera, wilayah and seni budaya also stands out. The concern with ilmu is seen in the collocates pengajian, sekolah, penuntut, buku, aliran, pemikiran, kosmologi and sarjana. The concern with sastera is reflected in the collocates puisi, novel, sasterawan, cerpen, karya, klasik and penyair. The concern with territory and borders is seen in the collocates tanah, dunia, alam, kepulauan, kampung dan perkampungan. While collocates such as filem, budaya, lagu, drama, teater, perfileman, silat, album, muzik, persembahan, seni and sinema emphasise his concerns with Malay arts and culture.

APPENDIX

A. List of collocates, MI score and frequency

Collocate	MI score	Frequency
Umpamaan	7.338	11
Bahasa	6.993	289
Cara	6.734	15
Nahu	6.685	5
Bertutur	6.560	11
Dipakai	6.548	10
Karangan	6.463	6
Canggung	6.307	5
Tanya	6.270	6
Keling	6.187	23
Belajar	6.003	67

Appendix continue

Collocate	MI score	Frequency
Undang-undang	5.931	8
Menjadikan	5.685	5
Huruf	5.627	12
Mengerti	5.570	9
Membaca	5.556	16
Senang	5.533	9
Jalan	5.520	35
Kitab	5.509	48
Hikayat	5.496	16
Raja	5.479	52
Asal	5.463	9
Betul	5.432	13
Mengajar	5.330	17
Surat	5.326	60
Cap	5.183	6
Diketahui	5.100	5
Menulis	4.985	12
Cina	4.961	28
Bahasanya	4.908	7
Bunyi	4.855	9
Inggeris	4.841	55
Saudagar	4.837	10
Rahsia	4.837	5
Adat	4.811	18
Perkataan	4.782	31
Hukum	4.779	6
Memakai	4.685	7
Bangsa	4.586	14
Nama	4.438	12
Kata	4.395	23
Tahu	4.343	14
Habis	4.307	5
Bernama	4.303	7
Guru	4.293	6
Sedikit	4.278	18
Mengetahui	4.226	6
Putih	4.213	10
Perint	4.204	6
Mengatakan	4.194	8
Anak	4.105	25
Pandai	4.077	10
Orang	4.075	183
Encik	3.985	9
Dahulu	3.954	7
Holanda	3.878	6
Ada	3.769	28
Negeri	3.749	26
Sebelah	3.747	6
Sahaya	3.739	24
Dia	3.630	24
Tempat	3.613	17
Dapat	3.584	12
Hal	3.544	14
Tiada	3.523	55
Kita	3.508	10
Pekerjaan	3.487	11
Ia	3.485	68
Lain	3.432	13
Membawa	3.412	6
Baik	3.400	18
Tanah	3.378	5
Masuk	3.376	6
Melihat	3.364	8
Menjadi	3.300	17
Membuat	3.259	8
Tuan	3.210	29
Sekarang	3.174	5
Perkara	3.058	7
Sama	3.049	8
Sampai	3.042	13
Datang	3.027	12

REFERENCES

Alatas, S.H., 1977. The Myth of the Lazy Native. Frank Cass, London, pp: 267.

Baker, P. and A.M. McEnery, 2005. A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts. J. Language Politics, 4: 197-226.

Choueka, Y., 1988. Looking for needles in a haystack. Proceedings of the RIAO Conference, RIAO'88, Cambridge, MA., pp: 609-623.

Hunston, S., 2002. Corpora in Applied Linguistics. Cambridge University Press, Cambridge.

McEnery, T., R. Xiao and Y. Tono, 2006. Corpus-Based Language Studies: An Advanced Resource Book. Routledge, Oxon Hill, MD USA., pp: 408.

Oakes, M.P., 1998. Statistics for Corpus Linguistics. Edinburgh University Press, Edinburgh, UK., pp: 287.

Scott, M., 1996. Wordsmith Tools Manual. Oxford University Press, Oxford.

Sinclair, J., 1991. Corpus Concordance Collocation. 3rd Edn., Oxford University Press, Oxford, pp: 179.