“Is Selangor in Deep Water?”: A Corpus-driven Account of Air/water in the Malaysian Hansard Corpus (MHC)

Norsimah Mat Awal  
norsimah@ukm.edu.my  
Center for Literacy and Socio Cultural Transformation,  
Faculty of Social Sciences and Humanities,  
Universiti Kebangsaan Malaysia

Azhar Jaludin  
azharj@ukm.edu.my  
Center for Literacy and Socio Cultural Transformation,  
Faculty of Social Sciences and Humanities,  
Universiti Kebangsaan Malaysia

Anis Nadiah Che Abdul Rahman  
p87706@siswa.ukm.edu.my  
Center for Literacy and Socio Cultural Transformation,  
Faculty of Social Sciences and Humanities,  
Universiti Kebangsaan Malaysia

Imran Ho Abdullah  
imranho@ukm.edu.my  
Center for Literacy and Socio Cultural Transformation,  
Faculty of Social Sciences and Humanities,  
Universiti Kebangsaan Malaysia

ABSTRACT

Air or its English equivalent ‘water’ is very important in our everyday life so much so that when the tap runs dry, it even made it as one of the topics debated by politicians. This paper looks at the issues that surround air/water in Malaysian Parliamentary debates by specifically focusing its relation to the state of Selangor. The air/water related issues were examined based on the collocates of air and Selangor in the Malaysian Hansard Corpus (MHC) from Parliament 1 (P1) to Parliament 13 (P13). The findings show that air is consistently present as one of the collocates of Selangor from Parliament 4 (P4) to Parliament 13 (P13). However, air started to show an upward trend starting in Parliament 7 (P7) and continued to Parliament 13 (P13). The recurring issues during those periods are the never-ending water-related problems and the steps taken by the government to overcome the problems. In P7 and P8, the focus is on the source of water as it collocates with pembersihan logi air (water treatment plant) and kawasan tadahan air (water catchment area). In Parliament 10, Parliament 11, Parliament 12 and Parliament 13 the recurring issue with air and Selangor is penyaluran air mentah (the transfer of raw water) from the neighbouring state Pahang to Selangor. Another issue observed is penstrukturuan air (restructuring) of water supply and services which was first observed in Parliament 12 and continues to Parliament 13. Thus, by focusing on the collocates of air this corpus-driven account has managed to show the trend of the parliamentary debates in relation to air and Selangor. Therefore, parliamentary debates where various issues of national interest are often raised offer opportunities for more critical analysis of issues that are important to the public.

Keywords: air (water); Selangor; Malaysian Hansard Corpus; parliamentary debates & collocates
INTRODUCTION

Water is one of the most precious resources, livelihoods and prosperity to human and other life forms. Without water, all forms of life will not survive and would cease to exist. It is an essential element in maintaining life and needed for the survival of all organisms. It is also a critical ingredient in all social and economic endeavours, such as agriculture, industry, energy, transport. Water plays an equally powerful role through its wide impact on the factors such as food production, hygiene, sanitation and health and the environment (Chamhuri & Ferdoushi, 2014; Wen, 2004; Malek et al., 2013). Hence, it can be said that should any country be hit by water crisis, the country would also be facing crisis in humanity, economy and health. Malaysia has had its share of water crisis in 2014, where Selangor, Federal Territory of Kuala Lumpur (FTKL) and Putrajaya were badly effected and water rationing had to be implemented (Syed Muhammad Shahabudin, 2015; Rasyidah Md Khalid, 2018). Since water plays a very important and powerful role in the economic growth and sustainable development of a country, it would be interesting to examine issues pertaining to water are discussed and debated by Malaysian law makers in the parliament.

PARLIAMENT

Parliament is the legislative body of a government. Power (2012) described parliament as:

Parliaments are the indispensable institutions of representative democracies around the world. Whatever their country-specific rules, their role remains the same: to represent the people and ensure that public policy is informed by the citizens on whose lives they impact...Parliament is therefore the central institution of democracy and constitutes an expression of the very sovereignty of each nation. Parliament is a political institution. It is a place for political, and often confrontational, debate. (Power, 2012, pp. 1-2)

In Malaysia, Parliament is the highest legislative body of the country and it is the place where laws are being enacted. Dewan Rakyat or the House of the Representatives is one of the three components in the Parliament of Malaysia, the premier legislative body in Malaysia after the Dewan Negara (Senate) and His Majesty The King. The House of Representatives has 222 members that are elected from the General Election of Malaysia that is held in every five years. Each Member of Parliament (MP) represents a Parliamentary Constituency (The Official Portal of Parliament of Malaysia, 2018). The role of the House of Representative stands as the body to pass, emend and repeal the laws of the country. The parliamentary debates of House of Representative and Senates are all recorded in an official record called Hansard.

HANSARD

Hansard is an official parliament report that contains accurate and verbatim debates during the Dewan Rakyat (House of Representatives) and the Dewan Negara (Senate) sittings. The publication of this report gives the public an opportunity to follow and understand all the debates involving members of both houses. The public will also be able to understand and made aware of the processes and proceedings that occur in parliament without having to attend the sessions. The term ‘Hansard’ is actually taken after the name of Thomas C. Hansard, the publisher of the debates in British Parliament in the early 19th century (Vice & Farrell, 2017). The Malaysian Parliament Hansard contains the Yang di-Pertuan Agong’s speech, the attendance of the Members of Parliament (MPs), the questions and answers of the parliament sittings, bills and acts tabled, MPs’ votes and others. It is important to note that
the Members of Parliament are free to voice out any views during the session without fear of being subjected to legal action as any statements made in Parliament cannot be prosecuted in any court.

Malaysian hansard in its current state (in isolated PDF forms) cannot be fully utilized for in-depth research. One of the ways to harness the data in the hansard is to turn the PDF form of the Hansard to a digital corpus form. The construction of hansard corpus has already taken place in other parts of the world, such as researchers in Britain, Canada and the European countries. They have been building corpus of Hansard Parliament of their respective countries such as the British Hansard corpus (or collection of texts) which contains nearly every speech given in the British Parliament from 1803-2005 (https://www.hansardcorpus.org/); The CLARIN ERIC infrastructure, an internet service organization, has listed and provided access to 18 European parliamentary corpora, covering almost all of the languages spoken in European countries (CLARIN, 2019). CLARIN ERIC also highlighted the importance of Parliamentary corpora:

Parliamentary corpora are a very important multidisciplinary language resource that can be approached from many research perspectives, including not only political science, but also sociology, history, psychology, and applicative approaches to linguistics, for instance, critical discourse analysis. (CLARIN, 2019)

Another hansard corpus is the Canadian Hansard Corpus and it is well known for its parallel corpus. This corpus contains debates from the Canadian Parliament in both English and French (the country’s official languages), and has been utilized for translation research and to build bilingual dictionary (Canadian Hansard Corpus, 1996). Australia, too, has its own hansard corpus called the Australian Parliament Question Time Corpus, consisting of the Question Time records of the Australian Federal Parliament House of Representatives from 1998 to July 2014 (Australian Parliament Question Time Corpus, 2015).

In Malaysia, the Malaysian Hansard Corpus (henceforth MHC) has been developed by a team of researchers from Universiti Kebangsaan Malaysia (UKM) to make available the hansard documents in electronic form which can be accessed by the public, upon request. The texts and information contained in MHC can be utilized as a set of data in cross-discipline researchs such as corpus linguistics, sociolinguistics, discourse analysis, political science and other disciplines, including the study of culturomics that can be carried out in the quantitative analysis of digital texts.

**PREVIOUS STUDIES ON HANSARD**

Various studies have used data taken from parliamentary debates. Baker (2006) examined keywords in the House of Commons in 2002 and 2003 debates on fox hunting in the UK. Using concordance analyses, the study identified different discourses that speakers utilized in order to persuade others of their standpoints. The study further explored ways in which keyness can be used to find salient language differences. The study postulated that a keyword list is useful for identifying lexical differences between texts. Bachmann (2011) also utilized the British Hansard to examine the language used in civil partnership debates in both Houses of the UK Parliament. The study employed a corpus-driven approach. By grouping keywords thematically and analysing them in context, scrutinising collocations and concordance lines, he was able “to see how (recurrent) uses of language construct gay and lesbian relationships” (Bachmann, 2011, p. 77).

Employing a corpus-based approach, Bayley et al. (2004) investigate lexical choices used by English, German and Italian MPs to communicate fear and reaction to fear in parliamentary debates. Linguistically, Italian MPs are said to use low-intensity words to
evoke danger and fear. Bayley and San Vicente (2004) also employed a corpus approach to investigate how UK and Spanish MPs talk about ‘work’. The researchers employed concordance and collocation tools to study how collocation patterns characterise ‘work’ and to see whether there are any similarities or differences between English and Spanish. They concluded that concordancing software allows for identifying certain linguistic details that cannot be observed by simple perusal.

Valvason (2017) employed a corpus-based study to explore Hungarian politicians’ attitudes towards the EU. This study utilized the corpus of Hungarian parliamentary debates (the HUNPOL corpus). Semantic preference and semantic prosody were spotted in the corpus through collocation analysis and concordance lines. The study showed that, in the years preceding Hungary’s joining the European Union, Hungarian politicians referred to the EU with mixed voices, alternating positive stances by claiming the need to join the EU to develop further and faster, and negative standpoints by signaling warning that EU-membership could cause a loss of national strength.

A corpus-based study by Bijeikienė and Utka (2006) focused on gender-specific features in parliamentary discourse from a combined sociolinguistic and corpus-linguistic methodological perspectives. The source data for the spoken language corpus consists of stenographs of the Lithuanian Parliament. It is shown that gender-related language variation in political communications does not follow a dichotomous pattern.

In Malaysia, several studies have been conducted utilizing the Malaysian Hansard. Ngeow et al. (2010) examined the use of sexist language and gender stereotypes in the parliamentary debates in the Malaysian parliament. The main data elicitation source of this study is the hansard in the Parliament official website from 2004 to 2008. This study did not adopt the corpus approach in its methodology, instead it employed qualitative content analysis to identify instances of gender stereotyping and sexist language use in Malaysian parliament. This study identified the emergence of three categories of frequent episodes involving gender stereotypes and/or sexist language, namely, beautiful women are sexual objects for lusting; female divorcees are sexually promiscuous and menstruation is demonised.

Tan et al. (2017) studied the trend of vague language, in the form of adjunctive and disjunctive phrases, used by politicians in the Malaysian Parliament. The study utilized the Malaysian Parliament Hansard corpus specifically, from 1970 to 1979 and comparing it to the corpus from 2000 to 2009. Utilizing corpus tools viz collocation, concordance and keywords analysis the study concluded that, the trend of usage from the 70’s to 80’s showed a high preference for the use of adjunctives over disjunctives based on frequency per million. The findings showed that in the year 2000s the function of adjunctive phrase such as dan sebagainya has been expended to include both issues and solutions while the disjunctive phrase such as atau lebih has been extensively to quote legal acts.

Azianura et al. (2017) also utilized the Malaysian Hansard Corpus to look at how issues pertaining to crime were discussed in Malaysian Parliament. They found that there was a dramatic increase in the mention of the lexical jenayah or crime in Malaysian Parliament from the year 1995 to 2016. She concluded that this increase correlated with the increase of crime statistics in Malaysia from 1980 to 2004.

Past studies utilizing hansard corpus have delved into various issues brought up in parliamentary debates such as fox hunting or civil partnership in the UK parliamentary debates. There are also comparative studies on expressions of and reaction to fear by English, German and Italian MPs and another that looked at how British and Spanish MPs characterized ‘work’. On the local front, there are various studies that utilized the Malaysia hansard data such as Ngeow et al. (2010) which looked at the use of sexist language and
gender stereotypes. Tan et al. (2017) and Azianura et al. are two studies to date that utilized the MHC.

There are obviously various issues that can be explored using the corpus on parliamentary debates. Kwabena (2018) through UK and Ghanian parliamentary data has shown that a corpus-driven method can help in determining one’s research focus in the study of parliamentary debates. Therefore, this study aims to examine the issues that are frequently associated with one of the states in Malaysia, specifically, Selangor. Selangor was chosen as the focus of this study because it has the highest Gross Domestic Product (GDP) in Malaysia (Department of Statistics Malaysia, 2017). Hence, it would be interesting to look into the issues related to Selangor discussed in parliament. The lexical ‘Selangor’ was also chosen because of its consistently high frequency in the MHC (refer to Figure 2) compared to other lexicals which represent states in Malaysia such as Johor, Kuala Lumpur and Pulau Pinang and frequency of co-occurrence of linguistics variable is the most commonly used statistic in corpus linguistics (Gries, 2010).

METHODOLOGY

The current research adopts a corpus-driven approach utilizing both quantitative and qualitative methods. According to McEnery (2013:6) corpus-driven approach is “an inductive process where corpora are investigated from the bottom up and patterns found therein are used to explain linguistic regularities and exceptions of the language variety/genre exemplified by those corpora”. Hence, by utilizing the MHC we will explore the patterns of usage of the lexical ‘Selangor’.

THE CORPUS

The corpora used in the present study is the Malaysian Hansard Corpus (MHC). MHC was developed from 2016 to 2018 from the reports of Malaysian Parliamentary Debates in the House of Representative (Dewan Rakyat). To date, the corpus contains texts from 3,511 files of reports from 13 parliamentary proceedings (Parliaments 1 – 13). The first parliamentary proceeding was convened from September 1959 to January 1964 while the last entry of the corpus was taken from the 13th parliamentary proceeding which was convened from June 2013 to April 2018 (the 13th Parliament was dissolved on 7 April 2018). This diachronic corpus of parliamentary discourse contains over 157 million words.

THE COMPOSITION OF THE MHC

The existing MHC is divided into smaller subcorpora which are sorted according to the Parliaments (Parlimen), Sessions (Penggal) and Meetings (Mesyuarat). Each meeting has a number of reports because it is held for several days. Each day of the meeting is reported in one file. Therefore, a Meeting in a Session could comprise of at least 2 report files, for example in Parliament 1, Session 1, Meeting 2 could have as many as 95 files as is the case in Parliament 8, Session 1, Meeting 2. One report denotes one sitting day.

To date, there are a total number of 3,511 Hansard reports of the House of Representatives collected in MHC with the total of over 157 million words (token).
TABLE 1. The Divisions of House of Representative according to Parliaments and Sessions: Last Measured in April 2018 (Imran et al., 2018)

<table>
<thead>
<tr>
<th>Parliament Session</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>50</td>
<td>43</td>
<td>63</td>
<td>108</td>
<td>109</td>
<td>103</td>
<td>109</td>
<td>70</td>
<td>4</td>
<td>68</td>
<td>80</td>
<td>54</td>
<td>182</td>
</tr>
<tr>
<td>2</td>
<td>46</td>
<td>60</td>
<td>50</td>
<td>72</td>
<td>59</td>
<td>66</td>
<td>65</td>
<td>96</td>
<td>84</td>
<td>77</td>
<td>77</td>
<td>70</td>
<td>57</td>
<td>241</td>
</tr>
<tr>
<td>3</td>
<td>44</td>
<td>56</td>
<td>45</td>
<td>61</td>
<td>64</td>
<td>40</td>
<td>46</td>
<td>72</td>
<td>86</td>
<td>77</td>
<td>91</td>
<td>83</td>
<td>93</td>
<td>44</td>
</tr>
<tr>
<td>4</td>
<td>38</td>
<td>31</td>
<td>71</td>
<td>21</td>
<td>24</td>
<td>67</td>
<td>72</td>
<td>61</td>
<td>85</td>
<td>63</td>
<td>53</td>
<td></td>
<td></td>
<td>138</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>46</td>
<td>67</td>
<td>70</td>
<td>57</td>
<td></td>
<td>182</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>46</td>
<td>67</td>
<td>70</td>
<td>57</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>182</td>
<td>241</td>
<td>138</td>
<td>267</td>
<td>231</td>
<td>256</td>
<td>255</td>
<td>276</td>
<td>385</td>
<td>286</td>
<td>321</td>
<td>366</td>
<td>334</td>
<td>3,511</td>
</tr>
</tbody>
</table>

Based on Table 1, the total number of 3,511 reports are compiled from Parliament 1 to Parliament 13. Parliament 1 has the least number of reports with 182 reports while Parliament 9 has the most number of reports (385). Customarily, each Parliament has 5 sessions except for Parliament 13 which has 6 sessions. The number of reports denotes the number of days of the parliamentary sittings.

THE SIZE OF MHC

Figure 1 shows the token (running words) in texts of each parliamentary sessions.

![Figure 1: Division of Basic Statistical Information according to Parliaments (Generated using WordSmith 5.0)](Imran et al., 2018)

Based on figures illustrated in Figure 1, there is an upward trend in the number of token over the years. Parliament 3 has the least number of word token while Parliament 12 has the most number of word token. The increasing number of word token may indicate that the number of debates have also increased in the parliament.

THE CORPUS TOOLS

The corpus tool utilized to process and analyze the target corpus for the current research is WordSmith Tools (henceforth, WST), version 5 (Scott, 2008). There are three main tools within WST, namely; Concord, KeyWord and Wordlist and 12 utilities. A brief explanation of each of the main tools are as follows.
CONCORD

The primary purpose of a concordance is to see many lexical items in context (Key Word in Context, KWIC format). Scott (2008, p. 106) states that “you get a much better idea of the use of a word by seeing lots of examples of it, and it’s by seeing or hearing new words in context lots of times that you come to grasp the meaning of most words in your native language”. The concordance allows a researcher to analyse a word, part of a word or lexical chunk in its linguistic context, and thereby notice recurring lexical features or idiosyncratic usages of a lexical item or chunk (Guillén Nieto et al., 2008).

KEYWORDS

The purpose of the KeyWords is to compare two word lists: a reference corpus and a study corpus. Scott (2008, p. 156) states that “key-words provide a useful way to characterise a text or a genre”. The comparison results in a list of keywords where the frequencies are significantly different between the reference corpus and the study corpus. For example, if a word has a frequency of 6% in the reference corpus, and 5% in the study corpus it will not be a keyword, even though it may be the most frequently occurring word (Scott, 2008). The KeyWord tool is useful for observing lexical similarities or differences between two texts and lastly, noticing the writer’s stylistic preferences (Guillén Nieto et al., 2008).

WORDLIST

Scott (2008) describes the purpose of the WordList tool as (1) to analyse the vocabulary used, (2) to identify common word clusters, (3) to compare the frequency of a word in different text files or across genres, (4) to compare the frequencies of cognate words or translation equivalents between different languages and, (5) to get a concordance of one or more words in a list. The current research, however, uses only two out of the three WST main tools, namely Concord and Wordlist tools.

THE PROCEDURE

The following 4 steps were followed in order to meet the aim of the research:

Step 1: Producing the frequency count of the lexical ‘Selangor’ in each parliamentary sitting i.e. from the first parliament through to the thirteenth parliament. The frequency count (raw frequencies) for ‘Selangor’ in each parliament was determined using WST and the frequency count was then normalized. Normalisation is a process to standardise word count which is done to accurately compare corpora (or sub-corpora) of different sizes (Evison, 2010). The line graph is then plotted and the plot indicates the trend of occurrence for the word ‘Selangor’ spanning from the first parliament to the thirteenth parliament.

Step 2: Generating the collocates of ‘Selangor’ in each Parliamentary sitting. According to McEnery (2013), collocation is “a co-occurrence relationship between words or phrases. Words are said to collocate with one another if one is more likely to occur in the presence of the other than elsewhere ‘. Even though there are a number of statistical tests that can be used to measure collocational strength, for example the MI (mutual information), z, t, log-likelihood, log-log, and MI3 scores (Xiao and Mcenery 2006), the current research opted to use MI (mutual information) as a measure of collocation strength. A higher MI score means a stronger association and thus a more coherent relationship between words (Salazar 2014). MI score works well with content words rather than function words. MI score of 3 or
higher indicates that the two lexical items frequently co-occur (Xiao & Mcenery, 2006). However, this current research sets the minimum MI score to 6, and items must have a minimum co-occurrence frequency of 10 (within a 4–4 window span) to be considered as a collocate of a given node word. These settings allow for greater statistical significant differences of the collocations.

**Step 3:** Identifying the most consistent collocate/s of ‘Selangor’.
This step was done manually. The MI value and the normalized co-occurrence frequencies of the identified collocate/s were recorded and then plotted on a line graph.

**Step 4:** Identifying the context of use by concordance analysis (Keyword in context) of the collocate identified in step 3. WST is used here to determine the context of use of the collocate in relation to the lexical term ‘Selangor’.

**FINDINGS AND DISCUSSION**

The research findings will be presented in three categories:
The frequency of ‘Selangor’ in the MHC.
The collocates of ‘Selangor’ in each parliamentary session in the MHC.
The most consistent collocates of ‘Selangor’ throughout the 13 Parliamentary sessions – ‘air’/water.

**‘SELANGOR’ IN MALAYSIAN HANSARD CORPUS (MHC)**

This study focuses on the profile or trend of one of the states, namely ‘Selangor’. As the most developed state in Malaysia (The Sultan of Selangor Opening Address in the 3rd Session of the 13th State Assembly, Hansard Selangor, 2015) ‘Selangor’ or issues pertaining or related to Selangor would often be brought up in the Malaysian Parliamentary debates. Utilizing WST, ‘Selangor’ was set as the node to determine its frequency in 13 parliamentary sessions to plot the trend of its occurrences in those parliamentary debates. Figure 2 indicates the trend of ‘Selangor’ in MHC.

![Figure 2. The Trend of “Selangor” in MHC](image)

Figure 2 shows that ‘Selangor’, indeed, is widely mentioned in Parliament 1 to Parliament 13 particularly in P3, P7, P11 and 12. Hence, our next aim is to look at the issues that surround ‘Selangor’. In order to get an insight on the issues discussed or debated
pertaining to ‘Selangor’, the collocates of ‘Selangor’ from P1 to P13 were generated. The collocates of ‘Selangor’ was generated for all 13 parliamentary periods. The aim is to look at the pattern of the top 15 collocates of the node ‘Selangor’ with Mutual Information (MI) score and frequency set as > 6 and > 10 respectively (see Appendix A). The list is grouped according to the Parliament and sorted according to MI score. Based on the word list generated, most of the top collocates with ‘Selangor’ are nouns, specifically proper noun such as names of states Perak, Pahang, Kedah or names of persons such as Chen, Lee, Daud or Ibrahim. Another trend observed from the list is the consistent presence of the lexical air or water as one of the collocates of ‘Selangor’ from P4 to P13, although air/water did not occupy the upper rank of the collocates list. Hence, we will next look at the context of occurrence of air/water in MHC.

‘AIR’/WATER AS A CONSISTENT COLLOCATE OF SELANGOR

As one of the consistent collocates for Selangor, a detailed information on air/water needs to be examined. Figure 3 shows the trend of the collocate air/water with the node Selangor based on its MI score for each Parliament. The two dotted lines signify the parameters of the MI score for the study. The study sets the collocate results of MI score at 6.0 and above.

![Figure 3: The MI Score of ‘air’/water as a collocate for ‘Selangor’ from P1-P13](image)

Figure 3 shows that air/water is one of the consistent collocates of the node Selangor from P4 to P13. This, however, does not mean that air/water did not appear in P1, P2 or P3 (1959-1973), on the contrary, air/water did appear but not as a collocate of Selangor. Additionally, air/water was spelled as ayer during those periods until the formation of the Majlis Bahasa Indonesia – Malaysia (MBIM) – Indonesia-Malaysia Language Council in 1972, which also launched the standardization of the romanised spelling system of the Malay language in Malaysia and Indonesia (Awang Sariyan, 2002). air/water was also referred to by its English equivalent ‘water’ during those three parliamentary periods.

In terms of frequency, air/water also shows an interesting trend as demonstrated in Figure 4.
Figure 4 shows the trend of air/water as the collocate of Selangor. The figures are generated from the frequencies of the collocate air/water with the node Selangor. To compare between the parliaments, each of the frequencies were normalized. air/water started to show an upward trend in P7 but went down in P8. However, from P8 air/water seems to regain momentum as it continuously became one of the issues frequently brought up in parliamentary debates and reached its peak in P11. air/water, however, took a downturn in P12 but still continues to be one of the frequently discussed topics in P13 as shown in Figure 4. In light of this interesting trend, it is felt that the issues that surround air/water should be closely examined by looking at its contextual usage. This paper will focus on P7 to P13 because beginning with P7 the MI score for air/water is 6 which is the MI score set for this study (see Figure 3 for details). We will next look at the concordances of air/water and its contextual usage.

THE CASE OF AIR/WATER

The issue of water and Selangor has a long history. In 1998, the water crisis in Selangor reached its peak when residents in the Klang Valley were subjected 150 days of water rationing and it has been dubbed as ‘the country’s worst water shortage’ (The Star Online, 26 February 2014a). The dry spell was caused by drought-inducing El-Nino phenomenon with no rainfall in the catchment area between January and March 1998 (Rasyikah Md Khalid, 2018). As a consequence of the water shortage, 3.2 million residents in Klang Valley had to be put on water rationing exercise from April 20, 1998 to September 16 during the same year. In 2014 Selangor again experienced a severe water shortage. This time it is blamed on ‘over development and lack of planning (The Star, March 19, 2014b). With 2 major water crisis in Selangor, we will now turn to the Malaysian Hansard Corpus (MHC) to observe the trend of air/water as it is debated in parliament.

AIR/WATER IN PARLIAMENT 7 (1986-1990)

In Parliament 7, air/water collocates mostly with pembersihan which literally means cleaning. However, contextual usage from concordances show that pembersihan referred to a different kind of cleaning i.e. the treatment of water in water treatment plant a shown in examples below:
Concordance 1:
The treatment of water

2 December 1986
a. “Mengenai penswastaan bekalan air Selangor itu, saya telah difahamkan bahawa Kerajaan Selangor telah menimbang cadangan untuk menswastaan operasi loji pembersihan air Sungai Semenyih, Bentuk penswastaan yang sedang dipertimbangkan ialah secara kontrak operasi dan penyelenggaraan loji pembersihan air kepada sektor swasta. Mengikut cara penswastaan ini, Kerajaan Negeri Selangor akan membayar kontraktor bekalan jumlah air yang dibekalkan daripada loji pembersihan air.”

“I have been informed that the Selangor Government has considered the proposal to privatize the water treatment plant in Sungai Semenyih. The privatization being considered is privatizing the operation and maintenance of water treatment to the private sector. According to the terms of the privatization, the State Government of Selangor will pay the contractor for the amount of water supplied from the water treatment plant.”

The issue at hand is the proposal by the Selangor state government to privatize the water treatment plant in Sungai Semenyih. This plan is signaled by the word cadangan or proposal. This debate took place on 2 December 1986 which is before the big water crisis in 1998.

AIR/WATER IN PARLIAMENT 8 (1990-1994)

In Parliament 8, kawasan or area is the top collocate for air/water, specifically water catchment area:

Concordance 2:
Water catchment area

4 August 1993
b. “Saya lebih satu dekad di Dewan Undangan Negeri sentiasa membangkitkan hal ini, mengenai perhutanan di Negeri Selangor yang mempunyai hanya 30% humi negeri itu adalah hutan. Itupun 30% adalah hutan itu untuk 'catchment area', kawasan tadahan air.”

“I have raised this issue in the State Assembly for more than a decade and, about forestry in the state of Selangor which covers only 30% of the state. And that 30% is designated as ‘catchment area’.”
25 October 1991

c. “Tuan Yang di-Pertua, saya hendak jelaskan di sini, di Sungai Buluh yang dekat dengan Forest Reserved Institute, kawasan itu memang hutan simpanan yang kekal. Forest Reserve Institute kita tidak boleh ditebang, ini yang di tepi yang dekat dengan Sungai Buluh dan juga di kawasan Tadahan Air.”

“The Honourable Speaker, let me clarify here, in Sungai Buluh which is close to the Forest Reserve Institute, a permanent forest reserved area. Cutting trees in the Forest Reserve Institute is not allowed. This area is adjacent to Sunggai Buluh and it is also in the water catchment area.”

From both examples (b) and (c), the debates were about areas designated as water catchment area. In (b) the concern was about limited area in Selangor (about 30%) that can serve as as water catchment area. In (c) it is about a water catchment area which is adjacent to the Forest Reserved Institute.

AIR/WATER IN PARLIAMENT 9 (1995-1999)

In Parliament 9, there were two collocates which frequently co-occur with air/water and they were Enakmen (Enactment) and penswastaan or privatization.

18 December 1999

d. “Saya mengucapkan banyak terima kasih kepada pihak kementerian dan juga kepada kerajaan seterusnya atas inisiatif ini untuk mengemaskinikan dan meminda undang-undang yang termaktub pada tahun 1974 untuk diselaraskan dengan pindaan pada undang-undang yang sama Enakmen Bekalan Air Selangor.”

“I would like to thank the ministry and the government for this initiative to update and amend the law that was formed in 1974 so that it is in line with the amended law within the Selangor Water Supply Enactment.”

Concordance 3:
Privatisation of Water Supply


19 April 1999
e. “Proyek empangan Sungai Selangor berharga RM2 bilion ialah satu produk penswastaan skim bekalan air sungai Selangor fasa III yang telah diserahkan kepada TSWA Sendirian Berhad iaitu syarikat milik Tan Sri Wan Azmi, Gamuda Berhad dan kumpulan Darul Ehsan Berhad yang ditubuhkan untuk mengatasi masalah kekurangan air yang dijangka melanda sekitar Lembah Klang tahun 2003.”

“The Sungai Selangor dam project valued at RM2 billion is the product of the Phase III in the privatization of Sungai Selangor water supply that was given to TSWA Sendirian Berhad. It is a company owned by Tan Sri Wan Azmi, Gamuda Berhad and the Darul Ehsan Berhad group that was established with the purpose to overcome water shortage problem that is expected to hit the Klang Valley area in 2003.”
In Parliament 9, the top collocates for air/water were enakmen or enactment and penswastaan or privatization. Both examples, (d) and (e), were taken from meetings in 1999 – a year after the 1998 water crisis which has been labelled as the worst water shortage to have happened in the country. Hence, the Selangor Water Supply Enactment was amended and measures were taken to overcome water shortage by privatizing the water supply.

**AIR/WATER IN PARLIAMENT**

For Parliament 10, the two collocates closely associated with air/water were dikorporatkan or corporatization and penyaluran or to transfer (water).

**Concordance 4:**
Corporatization of JBS

11 December 2001


“Members of the House – for your information, the Selangor State Assembly had passed the Company Replacement for Water Supply Enactment on 1 November 2001. This is to enable JBAS to be corporatized to become Perbadanan Urus Air Selangor or PUAS Berhad. It is expected that the corporatization will take place next year.”

**Concordance 5:**
Water Transfer Project from Pahang to Selangor

26 November 2000

g. “Antara projek bekalan air sambungan yang besar adalah projek penyaluran air mentah antara negeri iaitu dari Pahang ke Selangor yang memerlukan peruntukan sebanyak RM2.6 bilion dalam Rancangan Malaysia Kelapan.”

“Among the big water supply connection is the project to transfer raw water between states which is from Pahang to Selangor that needs an allocation of RM2.6 billions in the Eighth Malaysian Plan.”
In Parliament 10 the debates on air/water centered on efforts taken to improve water supply services when the Selangor State Government took step to corporatize the Selangor Water Supply Department. Efforts were also taken to avoid another major water crisis in Selangor. To this end, the government has proposed to transfer raw water from neighbouring state Pahang to Selangor.

AIR/WATER IN PARLIAMENT 11 (2004-2007)

The top collocate for air in Parliament 11 is *penyaluran* or to transfer which can be observed in the examples below:

Concordance 6:
Water Transfer Project from Pahang to Selangor

23 November 2005

h. “Saya hendak tanya apakah punca perkara ini berlaku dan seterusnya secara ringkasnya saya hendak tanya, mengenai bekalan air, penyaluran air daripada Pahang ke negeri-negeri lain terutamanya negeri Selangor.”

“I would like to ask what is the reason for this to happen and next, briefly I would like to ask about water supply, water transfer from Pahang to other states especially Selangor.”

27 November 2007

i. “Saya ingin merujuk kepada butiran 050200 - Pasukan Projek Penyaluran Air Mentah Pahang Ke Selangor. Sebagai mana yang telah disebut oleh timbalan menteri dan rakan-rakan yang lain bahawa perjanjian penyaluran air mentah, Pahang ke negeri Selangor telah ditandatangani pada 22 November yang lalu.”

“I would like to refer to article 050200 – the project on Raw Water Distribution from Pahang to Selangor. The Deputy Minister and friends have mentioned that the agreement on the transfer of raw water from Pahang to Selangor has been signed on 22 November.”

The issue of transferring raw water from Pahang to Selangor continues to dominate debates on air/water. In 2005 the issues of water supply and water transfer from Pahang was the focus in debates on air/water and Selangor. The same issue was again mentioned in 2007 debates but this time it seemed to be reporting on its progress. This is signaled by the lexical *perjanjian* or ‘agreement’ and *ditandatangani* or ‘signed’ in (i). The Selangor State Government and the Pahang State Government signed a sale and purchase agreement of raw water on 22 November 2007 (Memorandum Persefahaman Antara Kerajaan Malaysia Dan Kerajaan Negeri Selangor Berhubung Penstrukturan Semula Industri Bekalan Air Selangor, 2014).
In Parliament 12 and Parliament 13 there were two recurring issues brought up in relation to water, namely penyaluran air or the transfer of water and second is penstrukturan air or the restructuring of water services. Contextual information in the form of concordances are shown below:

Concordance 7:
Water Transfer Project from Pahang to Selangor in Parliament 12

Concordance 8:
Water Transfer Project from Pahang to Selangor in Parliament 13

The first component is the Pahang-Selangor raw water transfer project which includes other components such as the Kelau Dam in Pahang, tunnel to transfer raw water, and the Semantan Pump Station and pipeline and other related works.

The inter-state raw water transfer project from Pahang to Selangor continued to be one of the issues debated in Parliament 12 and Parliament 13. The issue was debated over two parliamentary periods and this could be attributed to the fact that in the 2008 General Election (P12) when Selangor fell to the opposition, hence the project that was agreed upon by the previous government needed to be re-examined. It could also because of the recurring
incidents of water crisis as brought up by one of the MPs in (j). Water crisis caused by severe drought hit Selangor in 1998 and 2014 (Rasyikah Md Khalid, 2018) and therefore efforts taken to stem it need to be debated.

We will next look at penstrukturan air or the restructuring of water services.

**Concordance 9:**
Restructuring of Water Services in Parliament

4 March 2009

1. “Oleh yang demikian, dasar dan pendekatan yang digunakan untuk merundingkan penstrukturan semula perkhidmatan bekalan air bagi negeri Selangor tiada banyak bezanya dengan apa yang telah dilaksanakan oleh Kerajaan Persekutuan bagi negeri-negeri lain”

Therefore, the policy and approach used to negotiate the restructuring of water supply services for the state of Selangor is not much different from what has been implemented by the Federal Government in other states.

**Concordance 10:**
Restructuring of Water Services in Parliament

9 July 2013

m. “Sehubungan dengan itu, Kerajaan Persekutuan telah pun memulakan perbincangan dengan Kerajaan Negeri Selangor bagi mengatasi masalah ini dan sekali gus menyelesaikan penstrukturan semula industri perkhidmatan bekalan air negeri Selangor, Wilayah Persekutuan Kuala Lumpur dan Putrajaya.”

In line with that, the Federal Government has started talks with the Selangor State Government to overcome this problem and at the same time complete the restructuring exercise on water supply services in Selangor, Federal Territory of Kuala Lumpur and Putrajaya.
The restructuring exercise on water supply and services in Selangor continues to make its way in the parliamentary debates. In (I), the MPs were informed that the policy and approaches used in restructuring water services is similar to what has been practiced in other states despite the fact that Selangor is now controlled by the opposition. In Parliament 13 the water restructuring exercise is again given attention in the debates. However, the focus of the debate was on the problems encountered in the water restructuring exercises as shown by the lexical ‘problem’ in (m) and (n). The continuous problems faced in the restructuring exercise has prompted one of the MPs to remark in (n) that there seemed to be no end to the water problem.

The analysis of the collocates for air/water in the MHC has shown an interesting trend in the discussions and debates on air/water in Parliament. Starting with Parliament 7, the top collocate for air/water is pembersihan and its literal translation is cleaning. However, the cleaning that is referred to is the treatment of water, specifically water treatment plant. In Parliament 8, the top collocate for air/water is kawasan or its English translation is area. Specifically, the area referred to in Parliament 8 is water catchment area.

In Parliament 9 there were two top collocates for air/water, namely enakmen or enactment and penswastaan or privatization. Enakmen became one of the top collocates in Parliament 9 is probably driven by the country’s worst water crisis in 1998. As a result of the water crisis, residents in the Klang Valley (which includes part of Selangor) were forced to endure 150 days of water rationing exercise (Rasyikah Md Khalid, 2018). New enactment on water were discussed to perhaps be better prepared in case severe droughts were to occur in the future. Another recurring theme is penswastaan or privatization. Privatization of water services was discussed as one of the ways to provide better services to the public.

Next, in Parliament 10 the two top collocates for air/water were dikorporatkan or corporatized and penyaluran or to transfer. The House was informed that the Selangor Department of Water Supply has undergone corporatization exercise and is now known as PUAS. The other collocate, penyaluran or to transfer was also one of the top collocates in Parliament 11, Parliament 12 and Parliament 13. In all four parliamentary periods (P10, P11, P12 and P13) penyaluran was discussed in context of the transfer of raw water from Pahang to Selangor. In Parliament 10, penyaluran was introduced to inform members of the House of the suggestion and plan to transfer raw water from Pahang to Selangor. This was put forth after the 1998 severe drought that led to water crisis in the Klang Valley. The debate on the transfer of raw water from Pahang to Selangor continued in Parliament 11, Parliament 12 and Parliament 13. Another collocate that dominated the debates on water in Selangor is penstrukturkan or restructuring. In both Parliament 12 and Parliament 13, penstrukturkan was debated in terms of the restructuring of the water supply and services in Selangor. The trend of the debate on water in Selangor has not only shown that it is interesting but it has also shown that water is linked to various issues.
CONCLUSION

This corpus-driven study has led us to ‘Selangor’. Through statistical analysis we were able to determine the issue or issues closely related to Selangor and air or water has been identified as one of the issues closely related to Selangor. By utilizing a corpus data, we were able to chart the trend of air/water throughout the thirteen parliaments. For example, the major water crisis experienced in 1998 which led to initiative and plan by the government to transfer raw water from Pahang to Selangor as discussed in P10 or the issue of transferring raw water from Pahang to Selangor which has spanned over four parliamentary periods.

Utilizing large data such as the Malaysian Hansard Corpus (MHC) has opened up a lot of possibilities for further research. Parliamentary corpora are multidisciplinary language resource (CLARIN 2019) and as such such research from multi discipline is made possible. This is shown by Azianura et al. (2017) who conducted a lexical analysis of the word jenayah and observed its dramatic increase in occurrence in the period 1995 to 2016 which correlates with the increase in crime statistics in Malaysia. As a multidisciplinary language resource, the MHC not only offers data for research in language and linguistics but also research in other fields such as political science, history and sociology.

Parliament represents the people and as such various topics of national interest are raised during the parliamentary debates. This paper has managed to present the trend one of the issues raised frequently raised i.e. ‘water’ and its relation with Selangor. There are, however, more issues of national interest that can be explored utilizing the Malaysian Hansard Corpus (MHC).

ACKNOWLEDGMENT

This study was supported by Universiti Kebangsaan Malaysia research grant KRA-2018-005.

REFERENCES


McEnery, T. (2013). Corpus: Some key terms. The ESRC Centre for Corpus Approaches to Social Science (CASS). Lancaster University, UK.


## Appendix A

### TOP 10 AND 11 COLLOCATE LIST FOR “SELANGOR” IN MHC

<table>
<thead>
<tr>
<th>No</th>
<th>Word</th>
<th>Relation</th>
<th>f</th>
<th>No</th>
<th>Word</th>
<th>Relation</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>SELANGOR</td>
<td>12.62954521</td>
<td>996</td>
<td>1</td>
<td>NEGERI</td>
<td>6.009408951</td>
<td>481</td>
</tr>
<tr>
<td></td>
<td>PERAK</td>
<td>8.827691078</td>
<td>87</td>
<td>2</td>
<td>WING</td>
<td>10.84006596</td>
<td>167</td>
</tr>
<tr>
<td></td>
<td>COMMERCE</td>
<td>8.784480995</td>
<td>73</td>
<td>3</td>
<td>SUM</td>
<td>10.02421474</td>
<td>165</td>
</tr>
<tr>
<td></td>
<td>STATE</td>
<td>6.346147537</td>
<td>58</td>
<td>4</td>
<td>CHEN</td>
<td>9.20233345</td>
<td>165</td>
</tr>
<tr>
<td></td>
<td>LABOUR</td>
<td>7.171729565</td>
<td>44</td>
<td>5</td>
<td>MICHAEL</td>
<td>10.76914406</td>
<td>164</td>
</tr>
<tr>
<td></td>
<td>SEMBILAN</td>
<td>9.354406357</td>
<td>37</td>
<td>6</td>
<td>RAJA</td>
<td>7.31487131</td>
<td>133</td>
</tr>
<tr>
<td></td>
<td>PAHANG</td>
<td>8.403063774</td>
<td>32</td>
<td>7</td>
<td>ISKAN</td>
<td>8.568361282</td>
<td>131</td>
</tr>
<tr>
<td></td>
<td>TRENGGANU</td>
<td>6.694219112</td>
<td>29</td>
<td>8</td>
<td>NASRON</td>
<td>8.872180939</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>JOHORE</td>
<td>7.566857338</td>
<td>28</td>
<td>9</td>
<td>PERAK</td>
<td>7.625770569</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>SULTAN</td>
<td>8.256325722</td>
<td>26</td>
<td>10</td>
<td>WILAYAH</td>
<td>6.68702507</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>AIR</td>
<td>3</td>
<td>6</td>
<td></td>
<td></td>
<td>3.5966</td>
<td>12</td>
</tr>
<tr>
<td>P2</td>
<td>STATE</td>
<td>6.686161995</td>
<td>153</td>
<td>1</td>
<td>NEGERI</td>
<td>6.19451751</td>
<td>443</td>
</tr>
<tr>
<td></td>
<td>PERAK</td>
<td>8.321522713</td>
<td>148</td>
<td>2</td>
<td>PERSEKUTUAN</td>
<td>6.991550446</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td>SEMBILAN</td>
<td>9.549764633</td>
<td>95</td>
<td>3</td>
<td>WILAYAH</td>
<td>6.92737886</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td>AFFAIRS</td>
<td>7.499453068</td>
<td>93</td>
<td>4</td>
<td>UMNO</td>
<td>8.59337616</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>PAHANG</td>
<td>8.112201515</td>
<td>62</td>
<td>5</td>
<td>PERAK</td>
<td>7.54276671</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>NEGERI</td>
<td>8.785064697</td>
<td>51</td>
<td>6</td>
<td>SEMBILAN</td>
<td>8.576512337</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>LABOUR</td>
<td>6.12132597</td>
<td>40</td>
<td>7</td>
<td>PINANG</td>
<td>6.566469193</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>KEDAH</td>
<td>6.978283707</td>
<td>39</td>
<td>8</td>
<td>JOHOR</td>
<td>6.2121259</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>MELAKA</td>
<td>6.463280678</td>
<td>38</td>
<td>9</td>
<td>BARAT</td>
<td>7.597874641</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>TRENGGANU</td>
<td>6.207083225</td>
<td>36</td>
<td>10</td>
<td>LAUT</td>
<td>6.418653011</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>AIR</td>
<td>3.5966</td>
<td>12</td>
<td></td>
<td></td>
<td>3.5966</td>
<td>12</td>
</tr>
<tr>
<td>P3</td>
<td>WING</td>
<td>10.68507195</td>
<td>89</td>
<td>1</td>
<td>NEGERI</td>
<td>6.02914903</td>
<td>285</td>
</tr>
<tr>
<td></td>
<td>SUM</td>
<td>9.404016495</td>
<td>89</td>
<td>2</td>
<td>WILAYAH</td>
<td>6.817350388</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>MICHAEL</td>
<td>10.66673322</td>
<td>88</td>
<td>3</td>
<td>PERAK</td>
<td>8.337896347</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>CHEN</td>
<td>8.033283234</td>
<td>86</td>
<td>4</td>
<td>PERSEKUTUAN</td>
<td>6.423883438</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>RAJA</td>
<td>6.802962303</td>
<td>84</td>
<td>5</td>
<td>PERAK</td>
<td>7.916993332</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>ISKAN</td>
<td>8.073349953</td>
<td>75</td>
<td>6</td>
<td>KIM</td>
<td>9.310194969</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>STATE</td>
<td>6.45844841</td>
<td>75</td>
<td>7</td>
<td>KIM</td>
<td>9.38002634</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>CHIK</td>
<td>7.553602695</td>
<td>72</td>
<td>8</td>
<td>JOHOR</td>
<td>7.707167625</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>PERAK</td>
<td>6.523229122</td>
<td>53</td>
<td>9</td>
<td>PULAU</td>
<td>6.037938595</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>SEMBILAN</td>
<td>7.935597424</td>
<td>40</td>
<td>10</td>
<td>PAHANG</td>
<td>8.67684269</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>AIR</td>
<td>4.37</td>
<td>8</td>
<td></td>
<td></td>
<td>4.37</td>
<td>8</td>
</tr>
<tr>
<td>P4</td>
<td>WILAYAH</td>
<td>8.866760352</td>
<td>74</td>
<td>6</td>
<td>WILAYAH</td>
<td>8.556471653</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>PERSEKUTUAN</td>
<td>6.376464844</td>
<td>74</td>
<td>7</td>
<td>PERAK</td>
<td>9.345826149</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>PULAU</td>
<td>6.591117859</td>
<td>68</td>
<td>8</td>
<td>LUMPUR</td>
<td>6.76533556</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>SUBRAMANIAM</td>
<td>8.061209679</td>
<td>63</td>
<td>9</td>
<td>JOHOR</td>
<td>7.15233326</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>PINANG</td>
<td>7.310675621</td>
<td>55</td>
<td>10</td>
<td>SUNGAI</td>
<td>6.495744705</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>AIR</td>
<td>4.540642586</td>
<td>9</td>
<td></td>
<td></td>
<td>4.540642586</td>
<td>9</td>
</tr>
<tr>
<td>P5</td>
<td>PERAK</td>
<td>8.972867979</td>
<td>675</td>
<td>1</td>
<td>NEGERI</td>
<td>7.017210484</td>
<td>212</td>
</tr>
<tr>
<td></td>
<td>DAN</td>
<td>18.05793953</td>
<td>537</td>
<td>2</td>
<td>TARIHET</td>
<td>11.90260887</td>
<td>321</td>
</tr>
<tr>
<td></td>
<td>YANG</td>
<td>20.12871361</td>
<td>376</td>
<td>3</td>
<td>DAUD</td>
<td>11.84880543</td>
<td>321</td>
</tr>
<tr>
<td></td>
<td>PALANIVEL</td>
<td>11.78103352</td>
<td>363</td>
<td>4</td>
<td>HULU</td>
<td>9.220503807</td>
<td>320</td>
</tr>
</tbody>
</table>

eISSN: 2550-2131  
ISSN: 1675-8021
5 PERAK 18.40701485 171 5 AIR 6.392780781 59
6 WILAYAH 8.206079483 137 6 SUNGAI 6.705155373 61
7 PERSEKUTUAN 6.93635591 133 7 LUMPUR 7.3737607 93
8 PULAU 18.60754776 131 8 PAHANG 9.129398346 67
9 BESAR 14.27156162 120 9 PERSEKUTUAN 6.488616467 77
10 JOHOR 15.75632763 118 10 JOHOR 7.094004631 85
11 AIR 13.89605236 12

P9
1 NEGERI 7.173819542 765 1 AHMAD 8.911207199 363
2 ADNAN 12.36036396 175 2 DZULKEFLY 10.82863045 191
3 JAMALUDDIN 8.540949821 174 3 NEGERI 6.520933151 128
4 ABDOL 10.65914059 164 4 PANCHANATHAN 10.82731819 133
5 MULOK 10.64920807 164 5 PINANG 7.293728828 89
6 PERSEKUTUAN 7.434614182 119 6 MIC 7.677032948 85
7 WILAYAH 8.614062309 114 7 KEDAH 6.179595947 97
8 PERAK 9.564443588 100 8 SIKIT 6.282687187 72
9 AIR 6.619233131 97 9 PAKATAN 6.078561783 89
10 SUNGAI 6.898245811 86 10 PERAK 6.421103954 84
11 AIR 5.987871647 73

P12
1 IRMOHIZAM 10.9547348 724
2 SEKINCHAN 9.797624588 14
3 MAKM 9.337225914 11
4 PKNS 9.167301178 11
5 IJOK 9.070493939 16
6 IBRAHIM 8.979006767 738
7 SPLASH 8.703296661 28
8 KUALA 8.523644447 3032
9 PENYALURAN 8.242791176 17
10 EXCO 7.870017529 37
11 AIR 6.902613163 74

ABOUT THE AUTHORS

Norsimah Mat Awal (PhD) is an Assoc. Professor at the Center for Literacy and Socio Cultural Transformation, Faculty of Social Sciences and Humanities, UKM. Her areas of interest are semantics, pragmatics and corpus linguistics.

Azhar Jaludin is a Senior Lecturer at Center for Literacy and Socio Cultural Transformation, Faculty of Social Sciences and Humanities, UKM and his area of expertise is corpus linguistics.

Anis Nadiah Che Abdul Rahman is a PhD candidate Center for Literacy and Socio Cultural Transformation, Faculty of Social Sciences and Humanities, UKM. She is currently pursuing PhD research in semantics and corpus linguistics.

Imran Ho Abdullah (Ph.D), ia Professor at Center for Literacy and Socio Cultural Transformation, Faculty of Social Sciences and Humanities, UKM. His areas of expertise are corpus linguistics and cognitive semantics and is currently the Deputy Vice-Chancellor (Industry & Community Partnerships), UKM.