



INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES



Recognising Metaphor through the Lens of Corpus-based Approach

Norasyikin Abdul Malik, Mohamad Syafiq Ya Shak, Faizah Mohamad

To Link this Article: <http://dx.doi.org/10.6007/IJARBSS/v12-i11/15337> DOI:10.6007/IJARBSS/v12-i11/15337

Received: 06 September 2022, **Revised:** 08 October 2022, **Accepted:** 29 October 2022

Published Online: 13 November 2022

In-Text Citation: (Malik et al., 2022)

To Cite this Article: Malik, N. A., Shak, M. S. Y., & Mohamad, F. (2022). Recognising Metaphor through the Lens of Corpus-based Approach. *International Journal of Academic Research in Business and Social Sciences*, 12(11), 1741 – 1755.

Copyright: © 2022 The Author(s)

Published by Human Resource Management Academic Research Society (www.hrmars.com)

This article is published under the Creative Commons Attribution (CC BY 4.0) license. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this license may be seen at: <http://creativecommons.org/licences/by/4.0/legalcode>

Vol. 12, No. 11, 2022, Pg. 1741 – 1755

<http://hrmars.com/index.php/pages/detail/IJARBSS>

JOURNAL HOMEPAGE

Full Terms & Conditions of access and use can be found at
<http://hrmars.com/index.php/pages/detail/publication-ethics>



INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES



www.hrmars.com

ISSN: 2222-6990

Recognising Metaphor through the Lens of Corpus-based Approach

¹Norasyikin Abdul Malik, ²Mohamad Syafiq Ya Shak, ³Faizah Mohamad

¹Akademi Pengajian Bahasa, Universiti Teknologi MARA Cawangan Perak, 32610 Seri Iskandar Perak, Malaysia, ²Akademi Pengajian Bahasa, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

Abstract

Metaphor analysis is a qualitative method that allows researchers to examine the use of metaphors in a text and help them to attain understanding of the metaphors used. Generally, the method involves a researcher's intuition as a way of identifying the metaphors. To date, there has been an increasing body of research that adopts corpus linguistics in analysing metaphors which has contributed a lot to the development of metaphor theory and analysis in various ways. This paper intends to demonstrate corpus-based approach in metaphor analysis. The application of corpus approach in metaphor analysis is believed to be more reliable and valid as it produces both quantitative and qualitative results. Hence, relevant manifestations of metaphorical patterns in a text can be established.

Keywords: Metaphors, Metaphor Analysis, Corpus-Based Approach

Introduction

Corpus-based approaches have been recognised as one of the major empirical approaches in linguistics analysis and have been employed to research issues relating to linguistic structure (e.g., lexis and grammar) and to numerous facets of language use (Stefanowitsch, 2006). Corpus linguistics is a method in language study which is not connected with any general or definite theory (Semino, 2017); thus, it is preferred by many linguistic researchers due to its flexibility. The quantitative approach can be integrated with any qualitative-based linguistic research and permit a large collection of authentic texts to be compiled and analysed systematically as they are supported by empirical evidence. In the field of metaphor study, implementing corpus approach has been an advantage. Metaphor researchers can use keyword list, collocation, and concordance lines in processing large amount of data to identify metaphor. Hence, contemporary researchers of metaphors are experimenting more systematic and efficient techniques to investigate metaphors to reduce the researcher's objectivity in metaphor identification (Abdul Malik et al., 2022).

The birth of the Conceptual Metaphor Theory (CMT) has greatly impacted the course of metaphor study (Lakoff & Johnson, 1980; Lakoff, 1987; Johnson, 1987; Lakoff & Turner, 1989; Lakoff, 1993). Nevertheless, the field of metaphor and metonymy research is still lagging

slightly behind with regards to its methodological foundations that has a strong emphasis researcher's intuition in identifying metaphor. Since CMT only involves manual identification of metaphors, many researchers have come out with more systematic approaches in identifying traces of metaphorical expressions used in a discourse (e.g., Adams, 2017; Ahrens & Jiang, 2020; Charteris-Black, 2004; Partington, 2007; Pragglejazz Group, 2007), but there are some issues with some of the developed approaches. Charteris-Black (2004) for instance, coined Critical Metaphor Analysis (henceforth CMA) that combines corpus linguistics with cognitive linguistics and Critical Discourse Analysis (CDA) in metaphor identification. Yet, the initial phase of examining metaphors in the method is still conducted manually. In other words, the identification of salient keywords as clues to identify metaphors is done via close-reading and inspecting the text manually. If the keywords are initially identified based on the researcher's intuition, an issue relating to reliability and validity of the findings can be questioned.

In another view, Pragglejaz Group (2007) who proposed Metaphor Identification Procedure (MIP) lists some comprehensive measures to detect metaphors. Yet, the proposed method requires exhaustive and intricate analysis on the researcher's behalf. The procedure is more suitable for group research instead of individual work as dealing with thousands of words corpus requires a high level of concentration (Adnan, 2014). Meanwhile, Ahrens and Jiang (2020) proposed a replicable method to identify source domain by using collocational patterns in Sketch Engine to study the building metaphor in the Hong Kong Political Speeches Corpus. Apart from Sketch Engine, they also used other corpus-based linguistic software such as SUMO (Suggested Upper Merged Ontology), WordNet, and an online dictionary. Even though the use of collocational patterns help researcher to identify metaphor signals based on statistical evidence, there are too many tools involved in the metaphor analysis making it to be a meticulous procedure. Meanwhile, Adnan (2014) integrated a corpus-based approach and CMT to provide statistical evidence for the identification of time metaphor in literary texts. She used Wmatrix, USAS tagger, and POS tagger CLAWS to compare the use of metaphors in the original and simplified version of a literary work. The use of the taggers helps researchers to classify the keywords easily and help to find clues of metaphors systematically. Nevertheless, not all corpus tools possess feature for word taggers. Most of the available corpus tools usually provide the basic features such as wordlist, keyword list, collocation, and concordance; thus, making it a troublesome to some researchers.

Partington (2006) proposed Corpus-Assisted Discourse Studies (CADS), an integration of corpus approach and discourse analysis, to study metaphor. The approach includes a step-by-step procedure that involves analyses of frequency list, keyword list, and concordance. The procedure can be easily replicated and followed by other researchers. Nonetheless, collocational strength of the keywords is not considered in the approach. If collocation analysis is added to the procedure before the concordance analysis, the evidence of metaphor in a discourse can be more credible. Collocation permits a researcher to make a strong claim of how the signals of metaphorical expression can be identified. Even though collocation can be done instinctively, it is naturally a poor guide to perform collocation (Xiao, 2015). Collocation analysis needs to be quantified statistically since it is more trustworthy than depending on human's intuition (Hunston, 2002); thus, it should be included in the process of metaphor identification.

The empirical data provided by the corpus approach helps the metaphor analysis to be done more efficiently and systematically. By merging corpus linguistics method, namely to the qualitative-based CMT, a large collection of authentic texts can be compiled and analysed systematically supported by empirical evidence. These past literatures exemplify how corpus approach offers the ease in detecting metaphor through the analysis of frequency lists, collocational analysis, and in-depth analysis of the concordance lines. It also shows the increasing interest to explore the full use of the empirical values provided by this approach in determining metaphorical expressions in a text. Corpus methods have indeed contributed massively to the development of metaphor theory and analysis in various ways, but there are still rooms for improvement. Therefore, the present study aims to fill in the loopholes in terms of methodology in studying metaphor.

Literature Review

Definition

Corpus can be defined as a relatively large collection of naturally occurring texts, which are stored in machine-readable form (Meyer, 2002, cited by Deignan, 2005). Concerning the machine-readable form, Abdul Malik et al (2022) explained that a corpus must be stored electronically, where “sophisticated corpus tools [are used] to analyse language in a collection of naturally occurring (either written or spoken) texts.” (p.513). Perhaps Biber et al (1998, p.4) summarized the characteristics that are essential to the corpus-based approaches best: They are empirical; they analyse the actual patterns of language use in natural texts; they utilize an extensive collection of natural texts or corpus, as the basis for their analysis; they employ the use of computers or electronic tools extensively for analysis purposes; they depend on quantitative and qualitative analytical techniques.

Meanwhile, some scholars proposed that studies using corpus-based approaches “are those which make use of the corpus predominantly to propose, investigate or demonstrate a particular theory of language” (Tognini-Bonelli, 2001, p. 65). However, some others, like Semino (2017), believed that such approaches do not have to be necessarily linked to any specific language theory.

Weaknesses of other approaches to metaphor research

There are other approaches to metaphor analysis, such as introspective methods. However, scholars such as Orts Llopis and Lopez (2009); Kispal (2010) highlighted concerns about these methods. Orts Llopis and Lopez (2009), for instance, observed that “the listing of metaphorical expressions is mainly based on the [analysts’] own intuitions and knowledge of the language and the previous descriptive results of other [analysts] on the same topic” (p.183).

At the same time, Benboudriou and Hamlaoui (2021) cautioned that with the absence of sophisticated electronic corpus tools, analysts would dedicate a vast amount of time to reading the whole corpora, as they need to refer to a dictionary for the meaning of every word before deducing its contextual and basic meaning. Once again, this process involves analysts' intuitions while being impractical and almost impossible to be completed.

Strengths of corpus-based approaches?

One of the significant reasons corpus-based approaches are employed in analysing metaphors is their systematicity (Caruso, 2011; Stefanowitsch, 2006). Caruso (2011), for example,

observed that Conceptual Metaphor Theory (CMT) and Conceptual Metaphor Analysis (CMA) “[do] not provide a systematic methodology that can be applied to the study of metaphor when considering expressions in a corpus of naturally occurring data” (p.2). Not only that but regarding the CMT, for instance, Koller et al (2008) claimed that it provides less empirical basis despite having a solid theoretical foundation.

Furthermore, corpus-based approaches enable research on metaphors to be more authentic and extensive. According to Orts Llopis & Lopez (2009), these linguistics methods have provided “a way to base metaphor research on authentic data, covering a wider range of metaphors, from primary to high-level ones” (p.183). With authentic data, Wikberg (2008) believed that metaphors could be studied in their context since “immediate context is crucial for deciding whether a word or phrase is used figuratively [or literally]” (p.37). Wikberg also concluded that such authentic data could provide explicit information about the collocational tendencies of a word or phrase.

Likewise, corpus-based approaches allow more substantial claims about language to be made. As mentioned by Charteris-Black (2004), since these approaches deal with an extensive set of data and include more variety of texts, it “has a greater potential for making claims about language” (p.31). Deignan (2005) concurred as he highlighted that a metaphor analysis involving a large set of data “regularly finds uses of words that [researchers] would not have predicted [earlier]” (p.85). Shimizu (2010) agreed, noting that the wide-ranging data gathered through corpus-based approaches “expands our intuitions about metaphors which then broadens the horizon of finding and analyzing metaphors” (p.204). Adnan@Anang (2014) later added that the large number of texts analysed through these approaches would contribute to the representativeness of the findings of metaphor research. This representativeness could then increase the generalization of these findings.

Step-by-step Procedure of Corpus-based Metaphor Analysis

This segment explains the procedure involved in identifying metaphor. The procedure encompasses both quantitative and qualitative approaches. Most researchers of metaphors make use wordlist, keyword lists, and concordance list to analyse language patterns in a discourse in unearthing metaphorical expression. Meanwhile, the proposed procedure was designed to make the process in metaphor identification to be more systematic and replicable for other metaphor researchers who are keen to implement corpus approach in their research. To explain this clearly, Table 1 presents the outline of the proposed procedure in discovering metaphors in a discourse

Table 1

Step-by-step procedure in discovering metaphor

No.	Step-by-step procedure
1.	Develop research questions
2.	Select, assemble, create, and edit the intended corpus
3.	Generate wordlist
5.	Generate keyword list
6.	Analyse collocates of the 'key' items
7.	Concordance interesting 'key' items (investigate the context)
8.	Identify semantic tension of the selected 'key' items with the collocates
9.	Categorise and list the identified metaphorical expressions
10.	Perform conceptual mapping for conceptual metaphors.
11.	Administer Inter-Rater Reliability (IRR) Test

Table 1 presents the step-by-step procedure. The procedure begins with designing research question(s), followed by compiling the corpus that want to be examined. Then the data analysis stage begins with quantitative analysis stage comprising analysis of wordlist, keyword list, and collocates. This is followed by qualitative analysis stage which comprises analysis of concordance list, identifying semantic tension, and performing conceptual mapping. Once the metaphors are identified, researchers need to conduct Inter-Rater Reliability (IRR) Test so the identified metaphors can be certified by experts of the field.

Analysing Wordlist

Wordlists are lists of lexical items that are assembled by frequency of occurrence in a corpus. The words are arranged by frequency and/or alphabetically. Other information such as the percentage of each word is used in a text, occurrences of words across a number of texts, or 'range', and plotting distribution of the words in the text is also tabulated and can be examined. However, analysing wordlists alone are not ample to examine the significance of language patterns in a corpus. A wordlist needs to be compared with either a 'reference' corpus or another wordlist. 'Reference corpus' is large set of text and represents some idea of a range of text in a given language like the 100-million-word British National Corpus (Scott, 2020). After the corpus has been built and compiled, wordlists are generated with the help of WordSmith Tools 8.0. Figure 1 displays a sample of a wordlist.

Word list (C:\Users\User\OneDrive\Documents\YM Transcript 1wl.lst)

N	Word	Freq.	%	Texts	%	Dispersion	Lemmas	Set
1	THE	442	5.38%	1	100.00%	0.88		
2	YOU	369	4.49%	1	100.00%	0.89		
3	TO	282	3.43%	1	100.00%	0.88		
4	AND	227	2.76%	1	100.00%	0.93		
5	THAT	223	2.71%	1	100.00%	0.95		
6	IS	189	2.30%	1	100.00%	0.91		
7	OF	148	1.80%	1	100.00%	0.89		
8	IT	146	1.78%	1	100.00%	0.91		
9	S	144	1.75%	1	100.00%	0.86		
10	A	136	1.65%	1	100.00%	0.91		
11	WE	116	1.41%	1	100.00%	0.85		
12	I	105	1.28%	1	100.00%	0.90		
13	ALLAH	103	1.25%	1	100.00%	0.78		
14	IN	101	1.23%	1	100.00%	0.81		
15	RE	98	1.19%	1	100.00%	0.79		
16	WHAT	97	1.18%	1	100.00%	0.91		
17	HAVE	97	1.18%	1	100.00%	0.88		
18	SO	76	0.92%	1	100.00%	0.86		
19	BE	70	0.85%	1	100.00%	0.84		
20	IF	68	0.83%	1	100.00%	0.86		
21	THEY	67	0.82%	1	100.00%	0.67		
22	NOT	66	0.80%	1	100.00%	0.85		
23	HEART	60	0.73%	1	100.00%	0.57		
24	DO	59	0.72%	1	100.00%	0.80		
25	YOUR	58	0.71%	1	100.00%	0.72		
26	WHEN	58	0.71%	1	100.00%	0.77		
27	THIS	57	0.69%	1	100.00%	0.86		
28	ARE	57	0.69%	1	100.00%	0.76		
29	T	52	0.62%	1	100.00%	0.80		

frequency alphabetical statistics filenames notes

1,143 entries Row 1 T S < > Help THE

Figure 1: A sample a wordlist.

In Figure 1, the wordlists display not just the frequency of occurrence (first column), but also the frequency percentage (second column), texts (third column), text percentage, as well as 'dispersion' (sixth column) of the words in the respective corpus. The 'dispersion' of words exemplifies where and how words are distributed in a text. From the 28 words displayed, the most frequent word in the wordlist is *the* that occurs 442 times. Meanwhile, the least frequent word is *are* with only 57 occurrences. In relation to the metaphor study, the information of the overused and underused words can be helpful in detecting signals of metaphorical expressions. This is because the overused or underused words of functional and content words can indicate a strong dominance of metaphor use in a discourse. Nevertheless, since this simple and basic frequency count of a corpus just displays the basic descriptive statistics (McEnery & Hardie, 2012), thus, the generated wordlists need to be compared to generate a keywords list where statistical significance tests such as chi-square test, t-test, and log-likelihood test are involved. The result of these tests permits the researcher to strongly determine whether the result of an analysis is significant or not (McEnery & Hardie, 2012).

Analysing Keyword List

Keyword list can be generated by using most of available corpus tools. The procedure compares all the words in two different corpora and reports the keywords that seem significantly more frequent in one than the other (with the help of statistical significance test). This also includes the keywords that appear more than a minimum number of times in one even if they do not appear at all in the other (Scott, 2020). Figure 2 illustrates a sample of a keyword list.

N	Key word	freq. in YM Transcript 1wl	%	Texts	freq. in NAK Transcript 1wl	Rc. %	BIC	Log_L	Log_R	P	Lemmas	Set
1	HEART	60	0.73%	1	5	0.08%	45.99	55.71	3.61	0.000000000		
2	SWT	33	0.40%	1	0		36.54	46.26	141.04	0.000000000		
3	FOUNDATION	28	0.34%	1	0		29.54	39.25	140.80	0.000000000		
4	SOLID	25	0.30%	1	0		25.33	35.05	140.64	0.000000000		
5	BUILDING	19	0.23%	1	0		16.92	26.63	140.24	0.000002429		
6	FEAR	19	0.23%	1	0		16.92	26.63	140.24	0.000002429		
7	HEALTHY	19	0.23%	1	0		16.92	26.63	140.24	0.000002429		
8	ISTIQAHAH	18	0.22%	1	0		15.52	25.23	140.17	0.000005053		
9	TREE	15	0.18%	1	0		11.31	21.03	139.90	0.000045254		
10	BODY	15	0.18%	1	0		11.31	21.03	139.90	0.000045254		
11	BUILD	14	0.17%	1	0		9.91	19.63	139.80	0.0000094185		
12	WILL	45	0.55%	1	13	0.16%	9.46	19.18	1.81	0.0000118930		
13	TAQWA	13	0.16%	1	0		8.51	18.22	139.70	0.0000196413		
14	HARAM	20	0.24%	1	2	0.02%	7.66	17.37	3.34	0.0000306978		
15	ROOT	12	0.15%	1	0		7.11	16.82	139.58	0.0000410588		
16	TAUGHT	12	0.15%	1	0		7.11	16.82	139.58	0.0000410588		
17	BECAUSE	31	0.38%	1	46	0.55%	-7.00	-2.71	-0.55	0.0994921550		
18	OVER	7	0.09%	1	15	0.18%	-6.86	-2.85	-1.08	0.0911154374		
19	INTO	5	0.06%	1	12	0.14%	-6.85	-2.86	-1.24	0.0906528756		
20	NEED	5	0.06%	1	12	0.14%	-6.85	-2.86	-1.24	0.0906528756		
21	THINGS	11	0.13%	1	21	0.25%	-6.69	-3.03	-0.91	0.0819725022		
22	STARTED	3	0.04%	1	9	0.11%	-6.67	-3.05	-1.56	0.0808617994		
23	EVEN	13	0.16%	1	24	0.29%	-6.56	-3.15	-0.86	0.0758120418		
24	SAID	13	0.16%	1	24	0.29%	-6.56	-3.15	-0.86	0.0758120418		
25	THEY	67	0.82%	1	91	1.09%	-6.42	-3.30	-0.42	0.0693512857		
26	T	52	0.63%	1	75	0.90%	-5.87	-3.84	-0.51	0.0500251018		
27	WORLD	3	0.04%	1	10	0.12%	-5.85	-3.87	-1.71	0.0491819344		
28	USE	3	0.04%	1	10	0.12%	-5.85	-3.87	-1.71	0.0491819344		
29	POLICE	11	0.13%	1	0		5.70	15.42	139.45	0.0000860758		

Figure 2: A sample of a keywords list.

Figure 2 allows the researcher to strongly determine whether the result of an analysis is significant or not (McEnery & Hardie, 2012) because instead of just having a list of most to the least frequent words in a corpus, keyword lists tabulate the overused and underused words in the study corpus as compared to the reference corpus. This comparison allows the researcher to learn the significant lexical items of a corpus as compared to another corpus. This keywords list is different than the wordlist as displayed in Figure 1. Keyword lists do not only display merely a simple frequency counts, but rather tabulate the saliency of the words based on statistical significance tests (e.g., chi-square test, t-test, and log-likelihood test). The words in the keywords list are arranged according to their 'keyness' (statistical significance), with the most statistically significant words are mentioned first followed by the least statistically significant words that are mentioned later/last. Thus, the most significant keyword in the keywords list in Figure 2 is *heart* with the LL value of 55.71 (see column Log_L), even though its frequency is only 60 occurrences as compared to the most frequent word in the wordlist (see Figure 1) is *the* that occurs 442 times. With regards to metaphor study, this statistical evidence in the keywords list is paramount in allowing a metaphor analyst to make a strong claim of a dominant use of metaphor in a discourse (Semino, 2017). As the orthodox approach of the early detection of metaphor is through close-reading and manual detection, this statistic result reduces the tendency for the discourse analysts to interfere the result of a study (Partington, 2007). Thus, key items that indicate the possible existence of metaphorical expressions in a text can be performed systematically.

Analysing Collocates

Collocation refers to words that systematically co-occur in a corpus "to create a range of cross-associations that can be visualized as networks of nodes and collocates" (Brezina, 2016, p.90). In other words, the collocation exemplifies significant relationship between words used in a corpus using statistical evidence (Williams, 2001). Collocation allows a discourse analyst to analyse how the keyword is used, thus providing the 'atmosphere' of a word (Baker et al., 2008). In relation to the study of metaphor, apart from concordance, collocation allows the researcher to make a strong claim of how the signals of metaphorical expression can be

identified. Even though collocation can be done intuitively, it is classically a poor guide to perform collocation (Xiao, 2015). Hunston (2002) argues it is vital for collocation analysis to be measured statistically because it is more reliable than through one's intuition. There are several statistical formulae that can be used in calculating the collocational strength such as Mutual Information (MI), t-test, z-score test, and log-likelihood test (Xiao, 2015). Nevertheless, the most common statistical formulae are Mutual Information (MI) value and t-score (Cheng, 2012; Hunston, 2002). Hunston (2002) states that to determine whether two lexical items are considered as significant collocates, the MI score must be 3.0 or above. Whilst the MI test calculates the collocational strength, t-test calculates the confidence of association between lexical items (Church & Hanks, 1990). Thus, both values allow researchers to determine the significant lexical items in a text. Figure 3 exemplifies a sample of collocates.

N	Word	Dice	MI	Texts	Total	Total Left	Total Right	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	HEALTHY	21.00	19.02	1	21	1	1		1				19					1
2	FEAR	10.00	17.95	1	10	1	9	1						9				
3	AND	0.02	7.71	1	7	4	3	1			1	2			2	1		
4	YOU	0.02	7.58	1	6	2	4		1	1				1	1	1		1
5	THE	0.01	6.66	1	5	4	1		1	1		2				1		
6	HAVE	0.05	9.38	1	5	3	2			1	1					1		1
7	TAQWA	0.77	13.36	1	5	0	5				1	1		2		1	2	

Figure 3: A sample of collocates.

The collocation view presented in Figure 3 displays the six surrounding words (left and right) that have the closest proximity with the word *healthy* (centre), and *fear* as the most significant collocate with the highest MI score. In relation to the study of metaphor, this empirical data allows the researcher to make a strong claim of how the signals of metaphorical expression can be identified. This is because the collocation displays the significant co-occurrence (based on MI score) of surrounding words and possible word patterns in a text. In this case, the word *healthy*, that is usually indicating a health state of a physical entity, has the closest proximity with the word *fear* that is an abstract entity indicating a certain feeling or emotion. This friction of meaning may indicate an evidence of metaphor usage.

Analysing Concordance Lines

After computing the relationship between the lexical items, the collocates that have a high confidence level of association will undergo a further qualitative analysis, which is the analysis of the concordance lines. Sinclair (1991: p.32) defines a concordance as “a collection of the occurrences of a word-form, each in its own textual environment. In its simplest form it is an index. Each word-form is indexed, and a reference is given to the place of occurrence in a text”. The concordance feature in WordSmith Tools is useful in sorting out the salient lexical items and looking for the linguistic patterns in a text. In this study, the analysis of the concordance lines allows the researcher to have a basic understanding of the word in the chosen corpus/corpora. To detect metaphorical signs in a text, there is a need to analyse the surrounding words of the salient key items, giving the contextual characteristics of the word. To have a clearer view of how this is done, Figure 4 presents some of the surrounding words at either side of *healthy* in the concordance lines.

. When you're using false fear or false change on **healthy** shame to erm.. to control someone's
 What's wrong with fear? I'm gonna tell you there's a **healthy** fear and there's an unhealthy fear. Okay,
 fear and there's an unhealthy fear. Okay, there's a **healthy** fear, and there's an unhealthy fear, let me
 themselves from the person. Right? Is that a **healthy** way of to deal with it? Is that solved the
 taubah. That's what we're supposed to do. That's **healthy** fear That's a healthy response when you
 we're supposed to do. That's healthy fear. That's a **healthy** response when you make mistake. Is that
 face Allah. That's an unhealthy type of shame. But the **healthy** fear is called taqwa. The healthy fear is
 of shame. But the healthy fear is called taqwa. The **healthy** fear is called taqwa. And taqwa will make you
 Allah and the religion completely and the deen. So, **healthy** fear will motivate us to get closer to Allah and
 firm in our deen, to have istiqamah, we have to have **healthy** fear of Allah SWT called taqwa. Healthy fear
 have to have healthy fear of Allah SWT called taqwa. **Healthy** fear is taqwa. Now how do we build healthy
 taqwa. Healthy fear is taqwa. Now how do we build **healthy** fear? You know what the, you know.. you
 know what the, you know.. you know what the seed of **healthy** taqwa is? This is interesting. Let me put it
 your culture. Now let's go to the root or the seed of a **healthy** taqwa. And and healthy istighfar, repentance,
 go to the root or the seed of a healthy taqwa. And and **healthy** istighfar, repentance, and healthy taubah,
 taqwa. And and healthy istighfar, repentance, and **healthy** taubah, went back to Allah. Do you know
 like you think about the body, alright? In order to have a **healthy** life, you have to have a body that is fed, a
 that is fed. a body that has oxvaen. a body that has **healthy** food. right? Water? But what if you're eating

Figure 4: A sample concordance lines

Figure 4 displays how the word *healthy* is used in the corpus. As presented in the figure, it can be seen how *fear*, the significant collocate for *healthy*, is used repeatedly with the word *healthy* in the corpus. It is also evident that *fear* is linked to 'Allah' (God) and the phrase 'healthy fear' is used repeatedly to describe the idea of *taqwa*, an Arabic word which means "God-consciousness or God-fearing piety; also rendered as god-fearing, right conduct, virtue, wariness" (Oxford Islamic Studies Online, 2020) and being closer to God. Apart from that, it is also evident that the word *unhealthy* is also used to describe the notion of *fear*. Interestingly, the words *root* and *seed* are also used to describe the phrase 'healthy fear'. This textual evidence illustrate how concordance analysis helps the researcher to examine what the salient keywords are associated with in a corpus and make sense of the significant relationship established in the collocation analysis. Nonetheless, to determine whether the words are considered metaphorical or literal, there is a need to examine the dictionary definition and the context meaning of the target lexical items.

Identifying 'Semantic Tension'

As explained previously, to determine whether a word is metaphorically or literally used in a text, a 'semantic tension' results from the use of the term in the context that contradicts or differs the literal meaning is required (Charteris-Black, 2004). To identify this 'semantic tension', any legit dictionary can be used to determine the actual meaning/dictionary definition of the target item apart from the context of the items in concordance analysis. As an illustration, words such as *fear*, *taqwa*, *healthy*, *unhealthy*, *root*, and *seed* in Table 2 indicate semantic tension that may lead to potential figurative language, in this case metaphor.

Table 2

Dictionary Definition of the Target Items.

No	Target Items	Dictionary Definition
1.	<i>fear</i>	(n) the bad feeling that you have when you are in danger or when a particular thing frightens you
2.	<i>healthy</i>	(adj) having good health and not likely to become ill
3.	<i>root</i>	(n) the part of a plant that grows under the ground and takes in water and minerals that it sends to the rest of the plant
4.	<i>seed</i>	(n) the small hard part produced by a plant, from which a new plant can grow

From the table, it can be seen the definition of the word *healthy* indicates an adjective to describe a physical entity such as a human, animal, or plant. Meanwhile, the word *fear* that is used together with *healthy* means an abstract entity of an emotion or feeling. Since *healthy* is supposed to describe a health condition of a human, animal, or plant, there is evidence of semantic tension between the two domains/concepts that are being compared. Apart from that, the definition of *root* and *seed* relate to parts of a plant, indicating another semantic tension as they are used to describe the phrase 'healthy fear'. These findings further support the existence of metaphorical expression (henceforth linguistic metaphor) in the corpus. Nevertheless, a further qualitative analysis of conceptual mapping is required to seek conceptual metaphor.

Performing Conceptual Mapping

After empirical analysis is completed, the identified recurring patterns of metaphorical expressions will be mapped according to the Conceptual Metaphor Theory framework (see 2.1.2) (Lakoff & Johnson, 1980; Kovecses, 2004; Holyoak & Stamenković, 2018). This technique follows from many studies that have incorporated it in analysing metaphors (Holyoak & Stamenković, 2018). To reinstate, conceptual mapping is carried out when a Source Domain (henceforth, SD) is mapped onto or connected to a Target Domain (TD). As a reminder, SD is generally more concrete while TD is more abstract. A sample of conceptual mapping of the identified linguistic metaphors is displayed in Figure 5.

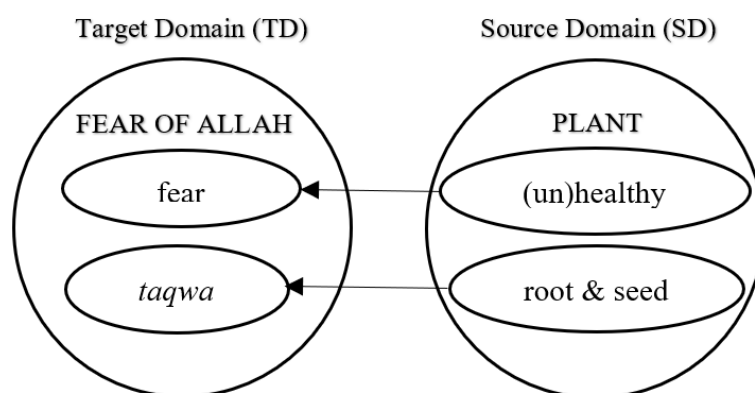


Figure 5: Conceptual Mapping of FEAR OF ALLAH IS A PLANT.

Figure 5 exemplifies how the conceptual metaphor of FEAR OF ALLAH (GOD) IS A PLANT is constructed by using the conceptual mapping framework. The linguistic metaphors like *healthy*, *unhealthy*, *root*, and *seed* in describing the fear of Allah (God) or *taqwa* construct the underlying notion of PLANT metaphor. The same process is performed to all linguistic metaphors that have been identified based on corpus techniques that involved the identification of keywords and their statistically significant collocates to discover the underlying conceptual metaphor.

Administering Inter-Rater Reliability (IRR) Test

A common analysis practice in qualitative research is coding, a reiterative process seeks to identify “a word or short phrase that captures and signals what is going on in a piece of data in a way that links it to some more general analysis issue” (Rossman & Rallis, 2011, p. 282). Since the later part of the metaphor identification procedure is executed qualitatively, there is a need to conduct an Inter-Rater Reliability test to certify the identified metaphors are not solely based on the researcher’s judgement. The existence of codebook can help researchers to describe the identified themes or findings with a concrete definition and example quote from the data (Creswell, 2013). Table 3 presents the codebook developed for the purpose of metaphor study

Table 3

Codebook on Metaphor Identification

Conceptual Metaphor	Definition	Conceptual Mapping	Semantic Tension
Oriental	It is based on spatial relations like up and down.	Source Domain = Target Domain	Contrast between contextual meaning and basic meaning
Structural	It uses a highly abstract concept in structuring a concrete concept.		
Ontological	It is divided into three categories: a) <i>Entity</i> - identifies abstract experiences as entities or substances. b) <i>Container</i> - refers to abstract experience as containers. <i>Personification</i> - represents an abstract experience as an entity or a person.		

This codebook can then be used by multiple researchers within the project or future researchers conducting similar studies. Guided by the developed codebook, the coders will check the identified conceptual metaphors and state their agreements and disagreements in a coding table as presented in Table 4.

Table 4

Sample of Coding Table

Conceptual Metaphor	Conceptual Mapping	Semantic Tension	Agreement?	Coder	Date
e.g. HEART IS A CONTAINER	e.g. yes/no	e.g. yes/no	e.g. yes/no	e.g. Coder 1 / Coder 2	e.g. 17-Mar-2021

Agreements and disagreements between the coders will be calculated using the IRR formula (Miles & Huberman, 1994) and tallied. Walther, Sochacka, and Kellam (2013) suggested IRR to “mitigate interpretative bias” and ensure a “continuous dialogue between researchers to maintain consistency of the coding” (p. 650). To calculate the coders’ responses, a formula described in Miles and Huberman (1994) is suggested to be used

$$\text{reliability} = \frac{\text{number of agreements}}{\text{number of agreements} + \text{disagreements}}$$

Miles and Huberman (1994) suggest that an IRR of 80% agreement between coders on 95% of the codes is sufficient agreement among multiple coders. This calculation is a method that can be used to measure consistency between coders which then validates the reliability of the results drawn from the qualitative analysis specifically in metaphor study.

Conclusion

Metaphor is universal in human lives and this ubiquity influences the way people think and communicate (Steen, 2010). It also has an immense role in how people use language in a society and culture which includes disseminating ideology, creating gender identity or gender dominance, explaining complex ideas, and advertising products or services. Due to its massive role, metaphor research is deemed to be essential. Nevertheless, the metaphor researchers are still experimenting the best approach and procedure to analyse metaphor since it still requires human interpretation to detect the evidence of metaphor in a discourse. The present study exemplifies how corpus approach can be advantageous to metaphor research in investigating salient keywords and the words associated with them to unearth the underlying metaphorical expression in a discourse. In addition, the present study also reflects how the use of corpus approach reduces validity and reliability issues in metaphor identification process. The study also demonstrates the amalgamation of corpus linguistics method with the qualitative-based CMT allows a large collection of authentic texts to be compiled and analysed systematically, supported by empirical evidence. Thus, it can be concurred that corpus linguistics approach and tools offer a lot of opportunities not just to linguistics studies but also to metaphor research.

Acknowledgement

This study was carried out under the funding of the GKIPP Research Grant Scheme (600-TNCPI 5/3/DDN (08) (008/2020) awarded by Universiti Teknologi MARA, Malaysia.

References

- Abdul Malik, N., Ya Shak, M. S., Mohamad, F., & Joharry, S. A. (2022). Corpus-based studies of metaphor: An overview. *Arab World English Journal*, 13(2), 512–528. DOI: <https://dx.doi.org/10.24093/awej/vol13no2.36>
- Adams, A. M. (2017). Linguistic and Conceptual Metaphors Of 'Heart' In Learner Corpora. [Master Thesis]. University of Kentucky.
- Adnan@Anang, N. B. (2014). *A corpus-based study of time metaphors in the text simplification of 'Around the World in 80 days'* [Unpublished Master dissertation]. University of Malaya, Malaysia.
- Ahrens, K., & Jiang, M. (2020). Source Domain Verification Using Corpus-based Tools. *Metaphor and Symbol*, 35(1), 43–55, <https://doi:10.1080/10926488.2020.1712783>
- Benboudriou, R., & Hamlaoui, F. Z. (2021). *A corpus-driven comparative analysis of metaphor use in master dissertations and doctoral theses including the impact of gender* [Unpublished Doctoral dissertation]. Larbi Ben M'hidi University, Algeria
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Exploring language structure and use*. Cambridge University Press
- Brezina, V. (2016). Collocation networks. In P. Baker & J. Egbert (Eds.), *Triangulating Methodological Approaches in Corpus Linguistic Research*, 102–119. New York/London: Routledge.
- Caruso, A. (2012). *A corpus-based metaphor analysis of news reports on the Middle East Road Map peace process* [Paper presentation]. Proceedings of the Corpus Linguistics 2011 conference, University of Birmingham, UK
- Charteris-Black, J. (2004). *Corpus approaches to critical metaphor analysis*. Palgrave MacMillan.
- Cheng, W. (2012). *Exploring corpus linguistics: Language in action*. London/New York: Routledge.
- Church, K., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16, 22–29.
- Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches (Third Edit)*. Los Angeles: Sage Publications.
- Deignan, A. (2005). *Metaphor and corpus linguistics*. John Benjamins Publishing.
- Holyoak, K. J., & Stamenkovi'c, D. (2018). Metaphor Comprehension: A Critical Review of Theories and Evidence. American Psychological Association. *Psychological Bulletin*. 144(6), 641–671. <http://dx.doi.org/10.1037/bul0000145>
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Kispal, T. (2010). Introspective data and corpus data. Combination instead of confrontation in the study of German metaphorical idioms of life. *Sprachtheorie und Germanistische Linguistik*, (20)1, 57-78.
- Lakoff, G. (1993). The contemporary theory of metaphor. In O. Andrew (Ed.). *Metaphor and thought (2nd ed.)*. Cambridge, UK: Cambridge University Press.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago; London: University of Chicago Press.
- Lakoff, G., & Johnson, M. (2003). *Metaphors We Live By*. London: The University Chicago Press.
- Lakoff, G., & Turner, M. (1989). *More than cool reason: A field guide to poetic metaphor*. Chicago: University of Chicago Press
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded*

- Sourcebook (Second Ed.)*. Sage Publications
- Orts Llopis, M. A., & Lopez, R. A. M. (2009). *Metaphor framing in Spanish economic discourse: a corpus-based approach to metaphor analysis in the Global Systemic crisis* [Paper presentation]. First International Conference of AELINCO (Corpus Linguistics Spanish Association). The University of Murcia, 7–9 May 2019.
- Partington, A. (2007). Metaphors, motifs and similes across discourse types: Corpus-assisted discourse studies (CADS) at work. In A. Stefanowitsch & S. T. Gries (Eds.). *Corpus-based Approaches to Metaphor and Metonymy*, 171, 267. Berlin: Walter de Gruyter.
- Pragglejaz Group. (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 18, 125–146. <http://dilbilim.info/yukseklisans/DB%20520/METHODOLOGY/pragglejaz%20copy.pdf>
- Rossman, G. B., & Rallis, S. F. (2011). *Learning in the field: An introduction to qualitative research (Third edit)*. Los Angeles: Sage Publications.
- Scott, M. (2020). WordSmith Tools version 8, Stroud: Lexical Analysis Software.
- Semino, E. (2017). Corpus linguistics and metaphor. In B. Dancygier (Ed.). *The Cambridge handbook of Cognitive Linguistics*, 463–476. Cambridge University Press.
- Shimizu, T. (2010). A metaphor identification procedure for corpus-based 'Mental Distance' analysis. *Osaka Keidai Ronshu*, 61(1), 203–215.
- Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford, UK: Oxford University Press.
- Stefanowitsch, A. (2006). Corpus-based approaches to metaphor and metonymy. In *Corpus-based Approaches to Metaphor and Metonymy*, edited by Anatol Stefanowitsch and Stefan Th. Gries, 1-16. Walter de Gruyter.
- Tognini Bonelli, E. (2001). *Corpus linguistics at work*. John Benjamins Publishing.
- Walther, J., Sochacka, N. W., & Kellam, N. N. (2013). Quality in interpretive engineering education research: Reflections on an example study. *Journal of Engineering Education*, 102(4), 626–659. <http://doi.org/10.1002/jee.20029>
- Wikberg, K. (2008). The role of corpus linguistics in metaphor research. In: Johanneson, N. L. & Minugh, d. (Ed.). *Selected Papers from the 2006 and 2007 Stockholm Metaphor Festivals*. Department of English, Stockholm University.
- Williams, G. (2001). Mediating between lexis and texts: Collocational networks in specialised corpora. *Asp, la revue du GERAS*, 31, 63–76.
- Xiao, R. (2015). Collocation. In D. Biber & R. Reppen (Eds.). *The Cambridge Handbook of English Corpus Linguistics*, (pp. 106–124). Cambridge: Cambridge University Press. <https://doi:10.1017/CBO9781139764377.007>