



All Theses and Dissertations

2017-07-01

Applying Corpus-Assisted Critical Discourse Analysis to an Unrestricted Corpus: A Case Study in Indonesian and Malay Newspapers

Sara LuAnne White
Brigham Young University

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>

 Part of the [Linguistics Commons](#)

BYU ScholarsArchive Citation

White, Sara LuAnne, "Applying Corpus-Assisted Critical Discourse Analysis to an Unrestricted Corpus: A Case Study in Indonesian and Malay Newspapers" (2017). *All Theses and Dissertations*. 6478.
<https://scholarsarchive.byu.edu/etd/6478>

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu.

Applying Corpus-Assisted Critical Discourse Analysis to an Unrestricted Corpus:
A Case Study in Indonesian and Malay Newspapers

Sara LuAnne White

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of
Master of Arts

William Eggington, Chair
Mark Davies
Chad Emmett
Jesse Egbert

Department of Linguistics and English Language
Brigham Young University

Copyright © 2017 Sara LuAnne White

All Rights Reserved

ABSTRACT

Applying Corpus-Assisted Critical Discourse Analysis to an Unrestricted Corpus: A Case Study in Indonesian and Malay Newspapers

Sara LuAnne White

Department of Linguistics and English Language, BYU
Master of Arts

In 2008 Baker et al. proposed a nine-step method that combines quantitative corpus linguistics with qualitative critical discourse analysis. To date this cycle has only been used to analyze a single language with a restricted corpus. Can this method, originally designed for this narrow focus, be applied cross-culturally to an unrestricted corpus? There are two over-arching goals for this paper, one linguistic and one methodological. The first goal is to learn about language ideologies in Indonesian and Malay newspapers; the second goal is to evaluate the efficacy of a mixed-methods corpus-driven approach to discourse analysis using the methods proposed by Baker et al. Our research will be based on the cross-cultural analysis of two 4-million-word corpora of newspaper articles; one Indonesian and one Malay. Malaysia and Indonesia are home to two peoples, living side by side and sharing a common language background, but reacting to the Islamic fundamentalist movement in different ways. Applying Baker et al.'s cycle, we will use keyword analysis, collocation, concordance lines, and qualitative analysis in this study. Whereas Baker employed a corpus restricted to articles about refugees, asylum seekers, immigrants, and migrants, our corpus encompasses articles on any topic; whereas their study focused solely on English, ours will compare Indonesian and Malay. To build a "useful methodological synergy" between qualitative and quantitative analysis (Baker, et al., 2008), this corpus-driven study will consider how Islam and related terms are being represented by government, historical, and religious sources. The results of this study will help us discern how these two countries are reacting to the fundamentalist movement. This study will also help evaluate the applicability of Baker et al.'s proposed methods to other types of sociolinguistic research and bring to light any modifications that could be made.

Keywords: Discourse Analysis, Corpus, Indonesian

ACKNOWLEDGMENTS

I would like to first off thank my committee chair, Dr. Eggington. Thank you for your support through this process and for making me feel at home in this program. Dr. Davies, thank you for your help with my corpus and for considering me a colleague during our drives to and from the Iowa airport. Dr. Emmett, thank you for helping me continue my research with Indonesian from advanced grammar class to studying in Salatiga – not to mention sing-alongs across Java. Dr. Egbert, thank you for your help and support through my different applications and for having so much faith in my work.

To my mom, dad, and family, thank you for your love and support and for giving me the opportunities to have the life that I have and pursue my education. Your sacrifices have made this possible.

To Michael, for all the late nights of homework, reading over drafts of my papers and applications, sitting with me through my tears, and still wanting to marry me – thank you. Thank you for standing by me through the challenges and hard decisions and still having faith the fruits of our labors will come.

Thank you.

TABLE OF CONTENTS

TITLE PAGE	i
ABSTRACT	ii
ACKNOWLEDGMENTS	iii
TABLE OF CONTENTS	iv
LIST OF TABLES	vi
LIST OF FIGURES	viii
1 INTRODUCTION	1
2 REVIEW OF LITERATURE	4
2.1 Linguistic Ecology	4
2.2 Corpus Linguistics.....	6
2.3 Discourse Analysis	6
3 METHODOLOGY	13
3.1 Corpus Design	13
3.2 Corpus Creation.....	15
3.3 Quantitative	18
3.4 Qualitative	21
4 RESULTS	23
4.1 Keyword Lists	23
4.2 Analysis by Keywords.....	30
4.3 Analysis by Category	43

4.4 Additional Possible Influences	49
5 CONCLUSION	51
5.1 Answers to Questions	51
5.2 Limitations	54
5.3 Future Research	55
5.4 Implications	56
APPENDIX	58
WORKS CITED	66

LIST OF TABLES

Table 1. Research Questions/Goals	2
Table 2. Possible stages in corpus-assisted critical discourse analysis (Baker et al., 2008).....	10
Table 3. Research Questions/Goals	13
Table 4. Content of IndoCorp and MalayCorp	16
Table 5. Counts of texts throughout creation process.....	18
Table 6. Steps in corpus creation	18
Table 7A. First 30 keywords from IndoCorp/Table 7B. First 30 keywords from MalayCorp	24
Table 8A. Initial religious keywords from IndoCorp	26
Table 8B. Initial religious keywords from MalayCorp.....	27
Table 9A. Islamic keywords from IndoCorp/Table 9B. Islamic keywords from MalayCorp.....	29
Table 10A. Collocates of ‘Islam’ from MalayCorp.....	31
Table 10B. Collocates of ‘Terrorist’ from IndoCorp.....	32
Table 11. N-grams of ‘Islam’ from MalayCorp.....	34
Table 12. N-grams of ‘Islam’ from IndoCorp	35
Table 13. Current events in Indonesia and Malaysia during the time of corpora creation	37
Table 14. Collocates of ‘Allah’ from MalayCorp.....	40
Table 15. Collocates of ‘Tuhan’ from IndoCorp	42
Table 16A. Location keywords from IndoCorp/Table 16B. Location keywords from MalayCorp	43
Table 17A. Political keywords from IndoCorp/Table 17B. Political keywords from MalayCorp	46

Table 18A. Terrorism keywords from IndoCorp/	Table 18B. Terrorism keywords from
MalayCorp	47
Table 19A. Islamic keywords from IndoCorp/	Table 19B. Islamic keywords from MalayCorp..
Appendix Table 1. IndoCorp List of Online News Sources	58
Appendix Table 2. MalayCorp List of Online News Sources	63

LIST OF FIGURES

Figure 1. Translated Concordance lines of ‘Islam’ from MalayCorp.....	33
Figure 2. Translated Concordance lines of ‘terrorist’ from IndoCorp.....	36
Figure 3A. Translated Concordance lines of ‘Bangladesh’ from MalayCorp.....	44
Figure 3B. Translated Concordance lines of ‘Bangladesh’ with ‘Islami’ from MalayCorp.....	44
Figure 3C. Translated Concordance lines of ‘Brunei’ from MalayCorp.....	45
Figure 4. Concordance lines for ‘wars’ in IndoCorp.....	46

1 INTRODUCTION

Malaysia and Indonesia are home to two peoples, living side by side and sharing a common language background, but reacting to the Islamic fundamentalist movement in different ways. Both countries boast populations which are Muslim majorities and together are home to more Muslims than any of their cousins in the Middle East. However, the religious background and colonization of Malaysia and Indonesia have led to each country following different paths as their respective Islamic movements become more prominent. One way to analyze discourse is to compare the language surrounding Islam in each country through the analysis of associated discourse. Through discourse analysis, we may be able to determine if these different pathways and reactions to Islamic fundamentalism are realized in language use.

However, a possible weakness of established discourse analyses involves their subjective nature and susceptibility to researcher bias. To combat this subjectivity, I have conducted a corpus-assisted discourse analysis combining qualitative and quantitative research following a model presented by Baker (2008) (see Table2), as explained further below. In addition to employing this model, I have chosen to use unrestricted corpora in order to represent all articles and topics published at that time.

For this research, I designed and built two corpora consisting of newspapers articles, one in Bahasa Indonesia (IndoCorp) and one in Bahasa Malay (MalayCorp) through web-crawling Google News and extracting all newspaper articles written in these languages during a 6-month period. By doing this, the data drives and directs the analysis thus increasing the objective reliability of the results.

After completion of the historical analysis stage of Baker et al.'s (2008) methods, the research questions were determined; IndoCorp and MalayCorp were designed specifically to answer these questions. In sum, this research project is a corpus-based synchronic examination of the discursive presentation of Islamic themes in Indonesian and Malaysian online newspapers during a 6-month period that contrasts the discourses used by Malaysian versus Indonesian online newspapers.

These research questions (see Table 1) were modeled after Baker et al.'s own questions in his 2008 RASIM (refugees, asylum seekers, and immigrants) study, but have been adapted to help better discover the unique relationship and history between Indonesia and Malaysia.

Table 1. Research Questions/Goals

Subject Matter/Linguistic
1. How are Islamic themes portrayed in the newspapers? How do they differ by country?
2. How can this inform us on each country's reactions to the Islamic fundamentalist movement?
3. What are the frequent topics and issues discussed in articles relating to Islam?

Baker et al.'s cycle was applied through keyword analysis, collocation, concordance lines, and qualitative analysis in this study. Whereas Baker employed a corpus restricted to articles about refugees, asylum seekers, immigrants, and migrants, our corpus encompasses articles on any topic. Whereas their study focused solely on English, ours will compare Indonesian and Malay. To build a "useful methodological synergy" between qualitative and quantitative analysis (Baker et al., 2008), this corpus-driven study considers how Islam and related terms are being represented by government, historical, and religious sources.

The results of this study may help us discern how these two countries are reacting to the fundamentalist movement. This study will also help evaluate the applicability of Baker et al.'s

proposed methods to other types of sociolinguistic research and may bring to light any possible modifications. A key aspect of the Baker et al. cyclical approach is a balanced, representative corpus.

2 REVIEW OF LITERATURE

It is important, before looking at the results from this research to first conduct a review of previous studies and literature surrounding this topic, in order to set the stage for this study. I will first give an overview of the linguistic ecology, or linguistic landscape of Indonesian and Malay, to help the reader better understand the significance of unique relationship between the two languages. I will first discuss the linguistic landscape of Indonesia and Malaysia to illustrate the historical religious and linguistic ties in this region, to show why an analysis of discourse surrounding Islam in these two countries is relevant. After, I will address the field of corpus linguistics, more specifically, I will discuss what a corpus is and how it has been used in discourse analyses. I will then introduce a more in-depth look at Baker et al.'s research and how it is used in this research, and where this research adds to it. Next, I will talk about the corpora used in this study, including their design and development.

2.1 Linguistic Ecology

Linguistic ecology or language ecology is “the study of interactions between any given language and its environment” (Haugen, 1972). The unique linguistic ecology in Indonesia and Malaysia can be connected to the spread of Islam through these two nations. Unlike many other Islamic nations which where Islam was purposefully spread to the inhabitants, Islam was brought to the Malay/Indonesian Archipelago peacefully by traders from Arabia and India. “The spread of Islam was not through the use of force –from Arabia, for example – but rather by way of slow penetration over centuries and allowing for the accommodation of local beliefs and cultures” (Azra, 2006). This slow implementation lead to a distinctive form of Islam being practiced in the island kingdoms.

The oldest recorded Islamic texts in the Malay/Indonesian Archipelago are found on two inscribed stones dating back to the 14th century. They are from Ache and Terengganu (formerly Trengganu). The stone found in Terengganu is inscribed with a proclamation that all civil leaders and rulers must strictly obey the Islamic faith (Sneddon, 2003). This inscription begins a timeline for the introduction of the Islamic faith into the Malay Archipelago. While Islam was brought to these sister nations in the same manner and at the same time, they have reacted to the Islamic fundamentalist movements in different ways. For example, recently Malaysia solidified a law prohibiting the use of the word *Allah* by non-Muslims. The Malaysian government ruled that, “‘*Allah*’ is a sacred word that symbolizes the Muslims and they have to defend the word so that it will not deviate from its true meaning. It is the obligation of all Muslims to protect the divinity of Islam” (Bin Mas’od, n.d.). Indonesia has yet to make a ruling like this and still supports the use of *Allah* by both Christians and Muslims to refer to *God*.

While there has been previous research into the influence of Islam on Indonesian and Malay language change (Campbell, n.d.; Muryantina, n.d.; Sneddon, 2003; Teeuw, 1961; Versteegh, 2014) and perception of how specific Arabic words should be used (Bin Mas’od, n.d.), these studies looked more closely at the history of the languages and word borrowing from Arabic. Because of the large Muslim population in both countries, much of the lexicon of the language, specifically the religious and academic words have been borrowed from Arabic. However, these studies have not conducted discourse analyses on the portrayal of Islam in media. Therefore, there is a need for research on how the topic of Islam is being discussed in these countries. Specifically, a corpus-based discourse analysis can be used to conduct a thorough and balanced analysis of language surrounding Islam in these two nations.

This type of research will both better help researchers understand the scope and abilities of discourse analysis research and increase the understanding of the relationship that exists between public discourse and ideologies. Corpora offer a way to fill this gap in research with naturally occurring language data from a variety of sources. While research has been done with corpora of Indonesian, the corpora, such as the SEAlang Library Corpus, have been more limited in the overall number of texts. This research aims to fill this gap through utilizing a larger, unrestricted corpus of newspapers to conduct a corpus-based discourse analysis on newspapers.

2.2 Corpus Linguistics

A corpus is a large collection of texts (written, spoken or both), which is collected to be studied as a representative sample of language (Johnson, 1999) and which are gathered or “built according to explicit design criteria for a specific purpose” (Atkins et al, 1992). Corpora are designed to be samples of natural language and can represent any language – from English to Arabic – and any genre – from Yelp reviews to English student essays. Corpora can be used for numerous types of research. Due to the nature of corpora, more specifically, their content of naturally occurring data, they have become starting points for many studies on the connection between society and language (Baker, 2010). These studies can be called a variety of names such as: discourse analysis, critical discourse analysis, and corpus-assisted discourse analysis. We will address these briefly in this section.

2.3 Discourse Analysis

While there has been much research done using corpus data in critical discourse analysis (CDA) (this will be explained in further detail in the following sections) (Baker, 2006; Baker et al., 2008; Baker, 2010; Baker, Gabrielatos & McEnery, 2013; Partington, 2010), and even corpus

research on the discussion of Islam in the British press (Baker, 2013; Bednarek, 2015), there have not been any studies using corpus data to compare the discussion of Islam in Indonesian and Malaysian press despite their large Muslim populations.

Traditionally, this relationship between society and language has been researched using a method usually referred to as discourse analysis. For the scope of this paper, the term discourse will be defined using Baker's (2010) definition which is a combination of definitions given by previous discourse researchers, Burr (1995), Parker (1992), and Foucault (1972). In short, in this paper discourse is defined as "a set of meanings, metaphors, representations, images, stories, statements and so on that in some way together produce a particular version of events ... Surrounding any one object, event, person etc., there may be a variety of different discourses, each with a different story to tell about the world, a different way of representing it to the world" (Burr, 1995) and are "practices which systematically form the objects of which they speak" (Foucault, 1972) or a "system of statements which constructs an object" (Parker, 1992).

Commonly, dictionaries are used to determine the meaning of any given word, however, discourse analysis linguistics argues that the meaning of a word is not that which is prescribed in a dictionary, but rather determined by its actual use in speech. As Firth (1957) explained, "You shall know a word by the company it keeps." What Firth means is that the true meaning of a word is determined by its context. The purpose of a discourse analysis is to use context to determine what a certain set of words means. For example, take the word *break*. If we look at the Oxford English Dictionary, there are more than 25 unique definitions for 'break,' and that is only for the uses of *break* as a verb, not to mention all the possibilities of its use as a noun. How, without context, is a reader or speaker supposed to know which of these is the intended

definition? That is where Firth's words (1957) apply – “You shall know a word but the company it keeps.” Discourse analysis looks at the use of words in speech (written or spoken) to determine what the words mean from context and aims to shed light on the relationship between what a person or group of people say and the values of societies they come from.

Critical discourse analysis falls under the umbrella of discourse analysis; however, critical discourse analysis is generally characterized by the researcher being “motivated by the desire to inspire or cause some sort of social change; for example, by highlighting inequalities of power” (Baker, 2010) and usually gives attention to topics such as gender roles or social inequalities. While these are topics of merit, the desire to incite change often leads to researcher bias, whether assumed or actual. For this reason, critical discourse analysis has often been criticized as being too focused on affecting social change, and less focused on allowing the data and discourses to emerge naturally from the corpus.

In CDA, researchers often tend to rely on the use of “qualitative techniques, as well as considering analysis of the social, political, historical and intertextual contexts which go beyond analysis of the language within texts (Baker et al., 2008). This is in slight contrast to corpus-based analysis which “utilizes a collection of different methods which are related by the fact that they are performed on large collections of electronically stored, naturally occurring texts (Baker et al., 2008).

The threat of researcher bias, one of the common criticisms of CDA research, could lead some to support an entirely quantitative approach to critical discourse analysis; however, this also has its weaknesses and can lead to misinterpretation of data and statistics. For this reason, linguists have developed a number of methodological ideologies to increase the objectivity of

these types of studies (Fairclough 1989, 1995; van Dijk 1991, 1993; Reisigl and Wodak 2001; Blommaert 2005, Baker et al. 2008). These methods all emphasize the need for multi-level analysis of texts and balanced, representative corpora. This need for multi-level analysis research has led to the combination of qualitative and quantitative analysis of texts. Baker (2010) explained that “a corpus-based analysis of discourse or ideology can be extremely powerful, but it is intended to enhance rather than replace small- scale qualitative analyses.” The quantitative prong can support the reduction of researcher bias and the qualitative side ensures that data is not misinterpreted by a computer.

For example, when researching collocates of *abortion* in the Corpus of Contemporary American History, the 7th collocate is *supreme*. If no qualitative research is conducted, it could be misinterpreted to mean that Americans speak of abortion as a supreme act or a supreme right. However, when qualitative analysis is added, it is discovered that *supreme* refers to the Supreme Court and the cases of abortion rights which have been tried there. While this may seem like a trivial example, this paper will show how misinterpretations such as these can be avoided through following Baker’s cyclical methodology.

One of the leaders in combining CDA with CL is Baker (2008), who suggests that using a corpus in discourse analyses creates a more objective CDA. This type of research is usually referred to as corpus-assisted discourse analysis (CADS). In his research with critical discourse analysis, Baker et al. (2008) undertook a nine-step cyclical method which combined quantitative corpus linguistics with qualitative critical discourse analysis to show “the extent to which methods normally associated with corpus linguistics can be effectively used by critical discourse analysts” (Baker et al., 2008). To date, however, this cycle has only been used to analyze a single

language with a restricted corpus. RASIM, the corpus constructed for Baker et al.'s (2008) study was compiled by searching for articles containing specific keywords, such as refugees, asylum seekers, immigrants and migrants. Baker et al.'s (2008) study looked at the discourses surrounding refugees and asylum seekers in the UK press from a diachronic perspective, 1996-2005. The current study however will apply this same methodology to look at discourse surrounding Islam in Indonesia and Malaysia from a synchronic perspective. The complete cycle can be found in Table 2 below.

Table 2. Possible stages in corpus-assisted critical discourse analysis (Baker et al., 2008)

1. Context-based analysis of topic via history/politics/culture/etymology. Identify existing topoi/discourses/strategies via wider reading, reference to other CDA studies.
2. Establish research questions/corpus building procedures.
3. Corpus analysis of frequencies, clusters, keywords, dispersion, etc. – identify potential sites of interest in the corpus along with possible discourses/topoi/strategies, relate to those existing in the literature.
4. Qualitative or CDA analysis of a smaller, representative set of data (e.g., concordances of certain lexical items or of a particular text or set of texts within the corpus) – identify discourses/topoi/strategies (DH approach).
5. Formulation of new hypotheses or research questions.
6. Further corpus analysis based on new hypotheses, identify further discourses/topoi/strategies, etc.
7. Analysis of intertextuality or interdiscursivity based on findings from corpus analysis.
8. New hypotheses.
9. Further corpus analysis, identify additional discourses/topoi/strategies, etc.

As explained in Table 2, it is important to conduct background research on the topic of choice. Then the researcher develops questions and designs a corpus able to help answer the research questions. The cycle then goes through a series of qualitative and quantitative steps that each work to refine the research questions and the corpus design. Baker et al. (2008) explains that a researcher can enter the cycle at any point; however, it is important to go through all of the steps so that the end result is a balanced corpus that will produce representative results.

Through implementing the cycle proposed by Baker and his colleagues (2008), a CADS can provide better insight into discourses, with less chance of interference from the researcher. This cycle works similar to a system of checks and balances, each step refining the research questions and results, and erasing possible bias. However, these steps have not yet been applied cross-linguistically with unrestricted corpora – this paper seeks to do that.

Applying Baker et al.'s (2008) cycle, keyword analysis, collocation, concordance lines, and qualitative analysis were employed for this study. Whereas Baker employed a corpus restricted to articles about refugees, asylum seekers, immigrants, and migrants, our corpus encompasses articles on any topic. Whereas their study focused solely on English, ours will compare Indonesian and Malay. To build a “useful methodological synergy” between qualitative and quantitative analysis (Baker, et al., 2008), this corpus-driven study considers how Islam and related terms are being represented by government, historical, and religious sources. The results of this study may help us discern how these two countries are reacting to the fundamentalist movement. This study will also help evaluate the applicability of Baker et al.'s (2008) proposed methods to other types of sociolinguistic research and may bring to light any possible modifications. A key aspect of the Baker et al. (2008) cyclical approach is a balanced, representative corpus.

In 2015, a similar study was conducted to look at the language ideologies, public discourses and ethnonationalism in the Balkans through analysis of SERBCORP (research corpus) and SERBCOMP (reference corpus) (Ajsic, 2015). These corpora consist of articles from relevant publications. The SERBCORP corpus was built by sampling articles that had 1+ hits for the keyword *jezi** and then sampling again for articles with 5+ hits for the keyword *jezi**. The

reference corpus, SERBCOMP, was compiled of all articles without any hits for *jezi**. While similar to the current study, Ajsic's had a much larger scope than the current study and was concerned with the discourse surrounding ethnonationalism. Ajsic's study looked at keyword analysis, collocation analysis, exploratory factor analysis, synchronic and diachronic variation, cluster analysis, and a CDA from a discourse-historical approach. The current proposed study will use similar methodology, but only for the keyword analysis and collocation analysis, and will be looking at the discourse of two nations instead of an entire region.

Pasha (2015) conducted a corpus-assisted CDA to look at the representation of the Muslim Brotherhood in Egyptian newspapers. The purpose of this study was to answer, "what would the Egyptian government do to halt the Brothers' political growth and potential threat?" (Pasha, 2015). The corpus used by Pasha was created from the front-page news articles published in the online version of the Egyptian newspaper *al-Ahram* from the years 2000-2005. The search was further narrowed by only retrieving texts that included the phrases *the Brothers*, *the banned*, or *The Muslim Brotherhood* in both the headlines and texts.

This literature review points to an additional methodological research question with respect to how we use corpora in linguistic research. This prompts the question, can Baker et al.'s (2008) methodology, originally designed for a narrow focus, be applied cross-culturally to an unrestricted corpus? Consequently, there are now two overarching goals for this paper, one linguistic and one methodological. The first goal is to learn about language ideologies in Indonesian and Malay newspapers; the second goal is to evaluate the efficacy of a mixed-methods corpus-driven approach to discourse analysis using the methods proposed by Baker et al. (2008). The amended research questions can be found in Table 3. The methodological

question was added later as the background research for this project progressed, as in keeping with Baker et al.'s model to amend and readdress research goals throughout the process.

Table 3. Research Questions/Goals

Subject Matter/Linguistic
1. How are Islamic themes portrayed in the newspapers? How do they differ by country?
2. How can this inform us on each country's reactions to the Islamic fundamentalist movement?
3. What are the frequent topics and issues discussed in articles relating to Islam?
Methodological
4. Evaluate the efficacy of a mixed-methods corpus-driven approach to discourse analysis using the methods proposed by Baker et al. (2008).

The above research questions have been constructed to mirror those used by Baker et al. (2008) as closely as possible. In Baker et al. (2008) however, only texts that contained keywords referring to refugees were used in the corpus. This study will allow the corpus to determine the statistically relevant keywords through a keyword analyses. I will first run the keyword analysis to find words relating to Islam that are unique to each language, and then use these keywords to start building possible discourses.

3 METHODOLOGY

3.1 Corpus Design

When designing a corpus, representativeness is basic to most designs. This means that a corpus needs to be representative of the population which is being researched. "A corpus is thought to be representative of the language variety it is supposed to represent if the findings based on its contents can be generalized to the said language variety" (Leech, 1991). For example, if a researcher would like to represent American English, they cannot only have samples from California. Or if the researcher would like to represent Californian English, they cannot have samples from Los Angeles alone.

For the scope of this study, a specialized corpus was created to be representative of newspapers in Indonesia during a 6-month period. However, unlike previous research, this corpus is an unrestricted corpus, meaning that all available newspaper articles were extracted from Google News for that period and added to the corpus. The use of an unrestricted corpus will decrease research influence or bias.

3.2 Corpus Creation

Texts

The scope of this research required the creation of two corpora. These corpora are called IndoCorp and MalayCorp. Each corpus consists of approximately 4 million words of newspaper texts. These corpora are designed specifically for a corpus-assisted discourse analysis of Indonesian and Malay newspapers on the topic of Islam as it was portrayed in the media at the time of text gathering. For a complete list of online newspaper sources, see Appendix Tables 1 and 2.

Often, corpora constructed for CADS research are referred to as restricted corpora because they only contain texts from certain newspapers or on certain topics. The MalayCorp and IndoCorp are unrestricted corpora, meaning, they contain all newspaper articles from that day as available on Google News. The target domain is discourse about Islam as portrayed in the news media found on Google News from 15 Aug 2015 – 25 Feb 2016. The basic counts for each corpus as they now stand can be seen below in Table 4.

IndoCorp and MalayCorp aim to represent the population of newspapers in each country. Ideally, this research would be able to represent the entire population of Indonesian and Malaysian residents, however that is not in the scope of this research. This research will look at how Islam is portrayed by the population of newspaper media in these two countries in hopes that it will help increase understanding of the relationship between how Islam is portrayed in each country and how each country is reacting to the Islamic fundamentalist movements.

Table 4. Content of IndoCorp and MalayCorp

	Text Count	Word Count	Dates Represented	Source	Sources Count	Ave. Text Length
IndoCorp	15,065	3,915,940	15 Aug 15 – 25 Feb 16	Google News	180	259 words
MalayCorp	11,895	3,766,350	15 Aug 15 – 25 Feb 16	Google News	113	315 words

Text Gathering

These corpora were built with the help of Mark Davies from Brigham Young University. The process to build these corpora follow the same basic steps which were used by Davies to build the NOW Corpus (News on the Web), a monitor corpus which automatically updates every 24 hours; the largest difference being that the NOW corpus updates every 24 hours.

The first step in building IndoCorp and MalayCorp was collecting the metadata from Google News for Indonesian and Malay newspaper articles to retrieve the links for each article on Google News. No text from the articles themselves was taken. This was done by copying the metadata for each day from August 2015 to February 2016, and then saving it in .txt files. This means that the texts in the corpora are all from the articles which appeared on Google News for each of those days. The metadata was then sorted to retrieve each URL. We used Google News because it is efficient and contains a large amount of newspaper articles which are easily accessible. Using Google News as the text source means that we are relying on Google to find the news articles that will be used in the corpus. However, Google News is representative of the news that people see if they search online. Google News is not only a valuable tool in corpus creation because of its accessibility and ease of use, but also because it is representative of what

someone living in Malaysia or Indonesia would find if they conducted a web search for a news topic.

The article text was then extracted with HTTrack, a web crawler which downloads web pages onto one's computer. In the case of IndoCorp and MalayCorp, this means that it not only extracted the article's text, but also text from advertisements, navigation links, headers, footers, etc. Usually 20%-40% of the URLs are not accessible and no text is extracted from them. For IndoCorp we ran 18,131 URLs and ended up with 15,119 text files; for the MalayCorp we ran 17,829 URLs and ended up with 12,281 text files. This means 83% of the Indonesian URLs were accessible and 69% of the Malaysian URLs.

Text Cleaning

After data was collected from each URL, the texts were run through JusText to eliminate all irrelevant text, including navigation links, headers, etc. Each remaining text was then saved in .txt files with a unique identifying number for the file name. The unique number corresponds to a row in an excel file that contains all relevant extra-linguistic identifiers; specifically: language, date, country, URL, title, snippet, source, filename, word count.

The corpora were then cleaned for duplicates of texts. For example, text duplication can occur when two newspapers reprint the same article from a newswire. Duplicated deletion was done by using the option in Excel to highlight duplicate fields in the Excel file containing titles of all the articles and then deleting the corresponding text files. The counts of texts throughout the creation process can be seen below in Table 5. The corpus has also been tested for balance by splitting the corpora in half at arbitrary points and then running the same test on each half to

check for balance in the results. This will help ensure that the data is representative and not skewed.

Table 5. Counts of texts throughout creation process

	URLs found from Google News	URLs after HTTrack	Texts after duplicates erased
IndoCorp	18,131	15,119	15,100
MalayCorp	17,829	12,281	12,252

Each step in the corpus creation, from determining the research questions to testing for balance in the corpus are important in creating a corpus that will give reliable results. If a corpus is created on an unstable foundation, or without strong research questions to guide it, it will not be reliable or representative. The steps are outlined in Table 6.

Table 6. Steps in corpus creation

1. Gather metadata from Google News
2. Extract URLs from metadata and save in excel file
3. Run URLs through HTTrack
4. Run texts through Jus Text
5. Check for duplicates/Balance
6. Delete English texts
7. Correct country of origin

For this research, in accordance with Baker et al.'s (2008) model, both quantitative and qualitative research was conducted. I will first discuss the quantitative research and then move on to the qualitative research.

3.3 Quantitative

AntConc was used for quantitative analysis the data. In keeping with generally accepted CADS research methodology and Baker et al.'s 2008 research (Ajsic, 2015; Baker, 2008; Gabrielatos & Baker, 2008; Pasha, 2015), the quantitative analysis consists of keyword, collocates, n-grams, and concordance line analysis.

To create keyword lists, AntConc first takes each corpus and makes a list of every word that appears in it and how many times it appears. It then uses a statistical analysis, in this case log-likelihood, to evaluate the significance of each word to one corpus compared to the other. When a word is statistically more relevant to one of the corpora, it is added to that corpus' keyword list. If a word is not statistically relevant to either corpus, it is left off both lists. This results in keyword lists of the words that are statistically characteristic of the corpus they represent.

Collocate analysis is done by first picking a specific word that the researcher wishes to analyze more closely. AntConc then finds every occurrence of that word in the corpus and the words that appear next to it, called collocates. AntConc then makes a list of the most relevant collocates for the researcher to analyze.

N-grams, like keywords, are also found using statistics. Again, the researcher specifies a target word to be analyzed more closely and AntConc finds the most common 3, 4, 5, etc. strings of words that appear with it. The researcher specifies if they would like to search for 3-word strings or combinations, 4-word combinations, or more. Then AntConc makes a list of the most statistically relevant combinations found in the corpus. For example, in English, if we did a search of the word *please*, we might find common strings such as, *please and thank you*, *please may I go*, etc.

Concordance lines, or key words in context, are like collocates in that they help the researcher see the word as it appears in the text files. They are longer than collocates however and show a snippet of each sentence in which the target word appears.

Representativeness in Keyword Analysis

Mirroring Baker et al.'s research (2008), keywords were calculated and then sorted by topic, after which the religious words were analyzed in further detail and sorted into more specific categories. Each religious word was then analyzed in context through a qualitative look at collocates and concordance lines. Collocates were calculated through combining MI and Log (MI + Log ($p > .05$)) (Gabrielatos & Baker, 2008).

It is possible during keyword analysis for the results to be skewed or weighted if the word appears at a high frequency, but all from one author or text. This would make it appear that the word is characteristic to the corpus, but actually is only characteristic of one text in the corpus. To try to avoid this from occurring, AntConc's concordance plot tool was used which shows the dispersion of the word throughout the corpus. This helps the researcher see if the word is well-dispersed in the corpus and therefore more representative of the target population. This tool was used in this research to increase representativeness. If a word appeared in a small number of text files, it was left out of the analysis. It is common to set a limit to the minimum number of times a word must appear in other types of analysis, like collocates. However, this type of check on a keyword analysis is not common. It is hoped that by adding in this extra step or check, the final wordlist to be analyzed will be more representative of the target population.

Unrestricted Corpora

For this research, open or unrestricted corpora were built to decrease the possibility of research influence on the results. As mentioned in Baker et al, "Subjective researcher input is, of course, normally involved at almost every stage of the analysis" (2008). While subjectivity is, as said, "normally involved in almost every stage of the analysis," beginning with open corpora

helps to lower the researcher subjectivity. Allowing the trends and patterns in the data to appear first permits the researcher to narrow the research based on the data-driven patterns. For example, when deciding to conduct an Islamic discourse analysis, a researcher, may choose to extract articles containing words such as *Islam, Muslim, terrorism, fundamentalism, radical, etc.*

While these are words almost guaranteed to produce a corpus of articles relating to Islam, the approach would miss any terms the researcher did not consider. For examples, in my initial research I considered the word *fundamentalisme* ‘fundamentalism.’ I thought that this would be a very common term used to discuss Islamic Fundamentalism, as it is used in English very frequently. However, it did not appear on either language’s keyword list. Also, when presented in focus groups conducted in Malaysia and Indonesia with native speakers, none of the groups were familiar with the term. Allowing the data to produce the search terms can help prevent similar issues.

3.4 Qualitative

This leads into the topic of qualitative analysis. As Baker (2010) explains, “a corpus-assisted analysis of discourse is therefore reliant on qualitative methods of analysis as well as those that are more traditionally quantitative.” The researchers are responsible for examining the smaller sets of data and investigating outside events, such as current events, which could have influenced the patterns found in the data.

This paper will mirror the methodology of Baker as used in Baker et al. (2008) and Gabrielatos and Baker (2008) where “The *RASIM* words and keywords were then qualitatively examined via detailed line-by-line concordance analyses, in order to identify semantic/discourse prosodies” and “The *RASIM* words and keywords were then qualitatively examined via detailed

line-by-line concordance analyses, in order to identify semantic/discourse prosodies”
(Gabrielatos & Baker, 2008). This important step helps to alleviate any outside influence from the researcher and ensure that each significant keyword is given its due research and attention. By employing a combination of qualitative and quantitative analysis, this research hopes to discover more generalizable, unbiased discourses.

4 RESULTS

For the paper, research was conducted following Baker et al.'s (2008) model of qualitative and quantitative analysis. This section will address the different steps of the discourse analysis, starting with a creating a keyword list for each corpus, then narrowing those lists to lists of religious words, then further sorting the religious keyword lists into subcategories, followed by an in-depth qualitative analysis of the words in these lists. The categories which appeared in the data were: function words, time, locations, government/military, sports, verbs, business/finance, technology/transportation, media/entertainment, people, religion, current events, and numbers. Religious words made up a relatively small percent of the words found in each list, .5% of the first 3,000 words searched in the corpus. This is a smaller number and shows that while religion is discussed, it is discussed much less compared to the other topics.

By following the afore mentioned steps, this research is replicating the methodology used by Baker et al., rotating back and forth between qualitative and quantitative analysis to allow the natural patterns in the data to drive the research as opposed to the researcher's own bias or motivations.

4.1 Keyword Lists

The first step of analysis on the unrestricted corpora was to run a keyword list for each corpus. Each keyword list consists of the words that are statistically characteristic of the corpus is represents. As can be seen in Tables 7A and 7B, at first glance the results of the keyword analysis appear to produce or reflect primarily basic, topic-related and lexical differences; however, upon further inspection, patterns begin to appear. These patterns will be discussed

further in the following section and used to illustrate the importance of rotating between qualitative and quantitative research in a discourse analysis.

Table 7A. First 30 keywords from IndoCorp *Table 7B. First 30 keywords from MalayCorp*

IndoCorp	Gloss	MalayCorp	Gloss
<i>bisa</i>	can	<i>berkata</i>	say
<i>saat</i>	moment	<i>malaysia</i>	Malaysia
<i>karena</i>	because	<i>pandangan</i>	view
<i>jakarta</i>	Jakarta	<i>datuk</i>	progenitor
<i>dari</i>	from	<i>kuala</i>	estuary
<i>indonesia</i>	Indonesia	<i>lumpur</i>	mud
<i>bahwa</i>	that	<i>daripada</i>	from
<i>tersebut</i>	mentioned	<i>kerana</i>	because
<i>tak</i>	so	<i>beliau</i>	he/she
<i>pemerintah</i>	government	<i>selepas</i>	after
<i>tim</i>	team	<i>katanya</i>	he/she said
<i>mengatakan</i>	say	<i>kerajaan</i>	kingdom
<i>harus</i>	must	<i>rm</i>	Malaysian Ringgit
<i>para</i>	the (pl. people)	<i>polis</i>	police
<i>hal</i>	thing	<i>sini</i>	here
<i>rp</i>	Indonesian Rupia	<i>syarikat</i>	company
<i>terjadi</i>	happen	<i>berkenaan</i>	pertain
<i>setelah</i>	after	<i>peribadi</i>	personal
<i>kasus</i>	case	<i>akaun</i>	account
<i>tapi</i>	but/however	<i>peribadi</i>	personal
<i>lainnya</i>	more	<i>bertanggungjawab</i>	responsible
<i>persen</i>	percent	<i>seri</i>	series
<i>ujar</i>	say	<i>sinarharian</i>	Sinar Harian ¹
<i>saja</i>	just	<i>pendirian</i>	stand
<i>film</i>	film	<i>diutarakan</i>	be explained
<i>korban</i>	sacrifice/victim	<i>semestinya</i>	necessarily
<i>wib</i>	Waktu Indonesia Barat ²	<i>padi</i>	rice paddy
<i>Partai</i>	party ³	<i>bahar</i>	Bahar ⁴

¹ Malaysian daily newspaper published in Selangor, Malaysia.

² Time zone: Western Indonesian Time

³ Used only for political parties.

⁴ Khai Bahar, Malaysian singer/artist

After generating the keyword lists from each corpus in AntConc, the keywords were then sorted into naturally occurring topics or themes, such as religion, locations, economy, etc. This process is slow, but ensures that no words are neglected and that the data is truly allowed to produce natural patterns. Each list was sorted through the first 3,000 words. At 3,000 words, the number of times the words appeared in the corpus was minimal and so they were not included in the analysis.

Table 8A and 8B show the results of the Religious category of the IndoCorp keyword list and the MalayCorp keyword list. All words which could potentially have a religious association, such as *korban* ‘sacrifice’ were sorted into this category at the first phase. The decision of which words were religious or not was based on the researcher’s own knowledge; however, each word in this category was checked for religious ties and taken off the list if there seemed to be none. For this first run of the lists, as many words with potentially religious ties were included with the knowledge that they would be checked and narrowed down in later steps. While this does present the possibility of over-looked words, this methodology enables the inclusion of more words than only focusing on 4 or 5 researcher-determined words. In line with Baker et al.’s (2008) methodology, it is better to start broad and slowly narrow in than to start narrow and overlook potential keywords. This methodology of starting broad and slowly narrowing the research is done to reduce researcher bias and influence. It is important to note here that additionally words that were not on the keyword list were still researched if they appeared relevant to the research, such as *Sharia*, which appeared in MalayCorp, but not IndoCorp.

Table 8A. Initial religious keywords from IndoCorp

IndoCorp	Gloss	Frequency	Keyness
<i>jumat</i>	Friday ⁵	2106	2015.101
<i>suriah</i>	Syria	1200	1305.427
<i>aceh</i>	Aceh ⁶	1718	1298.814
<i>isis</i>	ISIS	1215	1011.332
<i>natal</i>	Christmas	769	964.6
<i>teror</i>	terror	590	723.681
<i>terorisme</i>	terrorism	619	669.724
<i>menikah</i>	marry	773	605.482
<i>pasal</i>	article/verse ⁷	649	479.308
<i>pasalnya</i>	the article/verse	695	592.288
<i>teroris</i>	terrorist	481	559.453
<i>irak</i>	Iraq	419	371.183
<i>gereja</i>	church	411	324.772
<i>sarinah</i>	Sarinah	237	228.992
<i>prostitusi</i>	prostitution	184	202.807
<i>iran</i>	Iran	523	148.351
<i>radikalisme</i>	radicalism	167	136.759
<i>jamaah</i>	pilgrims	332	131.948
<i>tuhan</i>	Lord/God	367	121.715
<i>alquran</i>	Al-Qur'an	132	119.35
<i>PKB</i>	The National Awakening Party ⁸	135	115.769
<i>Sumatra</i>	Sumatra	146	115.529
<i>Shalat</i>	solat	113	100.026
<i>Nahdlatul</i>	Nahdlatul ⁹	76	26.417
<i>Nahdatul</i>	Nahdlatul	9	4.741
<i>Nadhatul</i>	Nahdlatul	3	4.126
<i>korban</i>	sacrifice/victim	3000	2796.469

⁵Important day for worship in Islam where all members are encouraged to gather in the Mosque for prayers and listen to teachings from the priest.

⁶ Only province in Indonesia where Sariah law is practiced.

⁷ Commonly refers to religious verses of scripture or an article in a legal paper. Can be used in both Muslim and Christian discourse.

⁸ The National Awakening Party is a moderate Islamic and a conservative political party in Indonesia.

⁹ Largest Muslim organization in Indonesia; appx. 30 million members; created PKB; conservative Sunni Islamic group; advocate against radical Islam

<i>salat</i>	solat ¹⁰	152	161.418
--------------	---------------------	-----	---------

Table 8B. Initial religious keywords from MalayCorp

MalayCorp Keyword	Gloss	Frequency	Keyness
<i>mohd</i>	Mohd	3575	4882.159
<i>umno</i>	UMNO ¹¹ (United Malays National Organization)	3232	4384.299
<i>terengganu</i>	Terengganu ¹²	1875	2579.525
<i>shah</i>	shah	2365	2546.335
<i>persembahan</i>	religious offering ¹³	1547	1989.913
<i>mahamad</i>	Mahammad	1124	1293.686
<i>fenomena</i>	phenomenon	1560	1165.023
<i>islam</i>	Islam	3981	1010.766
<i>mohamed</i>	Mohamed	936	958.619
<i>akhbar</i>	news	522	729.416
<i>haram</i>	illegitimate ¹⁴	758	411.196
<i>jemaah</i>	congregation ¹⁵	854	230.978
<i>hussein</i>	Hussein	260	208.533
<i>bangladesh</i>	Bangladesh	288	192.853
<i>kelayakan</i>	worthiness	273	179.076
<i>hussin</i>	Hussein	124	173.271
<i>quran</i>	Al-Qur'an	349	164.282
<i>syariah</i>	sharia	440	144.131
<i>kahwin</i>	marry	100	139.735
<i>ibadat</i>	worship	119	136.51
<i>rohani</i>	spiritual	152	135.625
<i>muhammad</i>	Muhammad	1004	133.419
<i>ummah</i>	community	99	128.513
<i>mengharamkan</i>	forbid, prohibit, ban	107	120.578
<i>syurga</i>	heaven	77	107.596
<i>Brunei</i>	Brunei	201	106.657

¹⁰ Muslim ritual prayer.

¹¹ One of the first words associated with Islam that appears in the corpus is UNMO. This is the United Malays National Party. One of the focuses of this party is to "uphold and expand Islam across Malaysia" (Wikipedia and their website).

¹² Cite of fist Islamic Arabic texts.

¹³ Usually refers to a religious tittle or offering. Can be used in both Muslim and Christian discourse.

¹⁴ Used only in Islam.

¹⁵ Can refer to both Muslim and Christian congregations.

<i>Allah</i>	God	1053	105.241
<i>agama</i>	religion	1584	105.045
<i>Muhamad</i>	Muhammad	185	105.018
<i>Haji</i>	Title of Muslim who has completed Hajj	1188	103.695
<i>Halal</i>	permitted, clean, legal	353	96.353
<i>Pengharaman</i>	ban	68	95.02
<i>Allahyarham</i>	God keep her ¹⁶	340	475.099
<i>Solat</i>	solat	378	481.754

Following Baker et al.'s (2008) cycle, the words were then looked at qualitatively to determine if they were indeed religiously used in context. Meaning, each word was looked at in context briefly to determine of the usage was indeed religious in the majority of uses. For example, the word *Jumat* 'Friday' was included initially on the 'Religious' list because in Islam Friday is the day of worship, similar to Sunday in Christianity. However, when looked at more closely in the qualitative step, it was determined that when Friday was discussed, it was not in relation to religion at all.

After this step, the lists were cleaned and narrowed according to the qualitative analysis. I then repeated the same steps on the MalayCorp (see Table 7B), resulting in two lists of religious words, one from each corpus. Following Baker's (2008) methodology, the religious words in these lists were then sorted into more descriptive sub-categories which can be found in Table 8A and 8B. The sub-categories found in the lists were: Worship, Teachings, Locations, Terrorism, and Politics. It is important to note that each of these categories emerged naturally from the word lists as they were sorted and were not created prior to the sorting. During this process, the words were also looked at more closely to determine if there was a connection with the topic of Islam

¹⁶ Arabic-based phrase used to speak of the deceased.

or not. If the words were decided to not have a connection with Islam, there were deleted from the lists. For example, the word *akhbar* ‘great’ was included in the initial list, however, after further investigation, it was taken off the religious list due to a lack of association with the topic of Islam.

An initial look at the Islamic keywords from each corpus may seem alarming (see Table 9A and 9B), especially when looking at the Terrorism category which is blank in MalayCorp and full in IndoCorp. Without a deeper qualitative analysis, it is tempting to draw conclusions that as the largest Muslim country in the world it has a media that is supportive of terrorism, or that terrorism is more prevalent in Indonesia when compared to Malaysia. For this reason, Baker et al.’s (2008) cycle is critical to produce accurate and reliable results which truly attempt to reflect an honest depiction of the discourse of a nation. In the next step, concordance lines and collocates were analyzed for the words in this list.

Table 9A. Islamic keywords from IndoCorp

Table 9B. Islamic keywords from MalayCorp

Category	Keyword	Gloss	Category	Keyword	Gloss
Worship & Teachings	<i>salat</i>	salat/prayer	Worship & Teachings	<i>Islam</i>	Islam
	<i>jamaah</i>	pilgrims		<i>solat</i>	salat/prayer
	<i>tuhan</i>	lord		<i>haram</i>	illicit
	<i>alquran</i>	Al Qur'an		<i>jemaah</i>	congregation
				<i>syariah</i>	Sharia'
				<i>persembahan</i>	religious offering
				<i>ibadat</i>	worship
				<i>rohani</i>	spiritual
				<i>allah</i>	God
				<i>agama</i>	religion
				<i>haji</i>	title of Muslim who has completed Hajj
				<i>halal</i>	permitted/clean /legal
				<i>syurga</i>	Heaven

Locations	<i>aceh</i> <i>irak</i> <i>iran</i> <i>suriah</i> <i>sarinah</i>	Aceh Iraq Iran Syria Sarinah	Locations	<i>terengganu</i> <i>bangladesh</i> <i>brunei</i>	Terengganu Bangladesh Brunei
Terrorism	<i>isis</i> <i>teror</i> <i>terorisme</i> <i>teroris</i> <i>radikalisme</i>	ISIS terror terrorism terroris radicalism	Terrorism		
Politics	<i>wars</i> <i>nahdlatul</i> <i>pkb</i>	wars Nahdlatul PKB	Politics	<i>unmo</i>	UNMO

The following sections will address different keywords from each corpus which were significant in uncovering the discourses surrounding Islam in each corpus. Each section will highlight a different word or set of words where applicable and the steps which were taken to analyze them in concordance with Baker et al.'s (2008) model.

4.2 Analysis by Keywords

While keyword lists give a quick picture of the unique characteristics of a corpus, they are not significant enough on their own to produce reliable and generalizable discourses. Keyword lists are generated through a statistical analysis and focus primarily on quantitative data. In order to produce generalizable discourses, multiple types of analysis are needed, as Baker et al. (2008) illustrated. The next step will focus on step 4 of Baker's (2010) model, by conducting an "analysis of a smaller, representative set of data (e.g., concordances of certain lexical items or of a particular text or set of texts within the corpus) – identify discourses/topoi/strategies." Concordance lines, or Key Words in Context (KWIC) give a wider context for each keyword in the list, which will better help us determine how the words are being used to discuss Islam in each corpus. It also helps reduce researcher bias. The category-level and word-

level analysis offer a better illustration of this principle. We will first look at the significant words, then move to categories, and then to the entire religious keyword list.

Terrorist and Islam

We will begin with a qualitative examination of the word *Islam* from the MalayCorp and the word *teroris* ‘terrorist’ from IndoCorp as these are words which could be susceptible to misrepresentation without a detailed, corpus-assisted qualitative analysis. A close look at these words will help address the methodological goal of this paper to ‘evaluate the efficacy of a mixed-methods corpus-driven approach to discourse analysis using the methods proposed by Baker et al. (2008).’ By closely examining these words, it will be possible to show the effectiveness of a combined qualitative and quantitative research model. If we only look at the keywords, we may think that terrorism is more prevalent in Indonesia and Islam is more popular in Malaysia. However, this is based only on frequency and needs to be substantiated through further, deeper, qualitative research before it can be claimed. By following Baker’s (2010) model, this paper aims to overcome researcher bias and misinterpretation of corpus data.

For both words, collocates were first analyzed to determine in which context the keywords were being used. Collocates for each keyword can be seen below in Table 10A and 10B.

Table 10A. Collocates of ‘Islam’ from MalayCorp

Rank	Freq	Freq (R)	Freq (L)	Stat	Word	Gloss
1	555	518	37	9.70665	<i>umat</i>	people
2	19	6	13	9.68486	<i>berkemajuan</i>	progress
3	26	0	26	9.64261	<i>ppim</i>	PPIM
4	139	133	6	9.64027	<i>beragama</i>	religious
5	20	20	0	9.57444	<i>syiar</i>	symbols
6	12	0	12	9.57444	<i>solok</i>	gift

7	62	61	1	9.54678	<i>memeluk</i>	hug
8	14	0	14	9.5338	<i>maidam</i>	time (immemorial)
9	24	1	23	9.52713	<i>umatnya</i>	his/the people
10	42	0	42	9.4558	<i>melor</i>	Jasmine
11	80	76	4	9.42388	<i>syariat</i>	Sharia
12	106	106	0	9.20644	<i>perkuburan</i>	cemetery

Table 10B. Collocates of 'Terrorist' from IndoCorp

Rank	Freq	Freq (R)	Freq (L)	Stat	Word	Gloss
1	46	46	0	11.38134	<i>terduga</i>	assured
2	12	8	4	8.81349	<i>radikal</i>	radical
3	86	82	4	8.67082	<i>kelompok</i>	group
4	56	52	4	8.54209	<i>serangan</i>	attack
5	13	12	1	8.18932	<i>menangkap</i>	catch
6	17	4	13	7.84744	<i>paris</i>	Paris
7	19	18	1	7.73681	<i>jaringan</i>	network
8	16	11	5	7.7076	<i>bom</i>	bomb
9	10	5	5	7.38876	<i>teroris</i>	terrorist
10	10	10	0	7.06743	<i>ancaman</i>	threat
11	18	4	14	6.89991	<i>isis</i>	Isis
12	11	11	0	6.7905	<i>organisasi</i>	organization

Islam

When we look at the keywords if Islam in the table, we see words like *Sharia*, which will be discussed more in its own section and PPIM, a non-profit for Muslim consumers. From the first look at the collocates of Islam, there is no strong connection between Islam and the Government, aside from these previously mentioned words. However, there is a connection here between Islam and the way it is practiced, such as *cemetery*, *gifts*, and *symbols*. These words show that while other evidence supports a strong connection with Government, there is still some association with the way the religion is practiced and its beliefs. Now we need to look deeper to see what else can be uncovered.

After the keywords were looked at, the concordance lines of Islam were analyzed for content. To fully understand the words, we must look and understand at their surroundings. The concordance lines can be seen in Image 1. The concordance lines have been translated to English for the ease of the reader.

Figure 1. Translated Concordance lines of 'Islam' from MalayCorp

names she. But struggling for Malay and Islamic rights is in the country's constitution. Things Umno and BN AKP is not an Islamic party. There is another other party more also culture and products. "Now the Islamic world has big opportunities in the halal economy, can not be punished, even if Islam is the principle of sin and reward, but it very important," he said. He said that in Islam, certain manners to overcome differences as interference in Islamic affairs. Is becomes a crime, sermons yan awesome smokers. For smokers who are Muslims, are they willing to accept the fatwa scrape around, he remained in his nature. Islam is a religion of nature which Allah SWT resistant and reinforced," he said. He said Islam is a religion of peace and does not teach damaging a place of worship religion Islam is a religion of prosperity and prohibit people public opinion) believes Article 3 (1), namely Islam is the religion of the Federation, has become peru lam the provisions of the Constitution dealing with Islam is the religion of the Federation, which was rejected by now is subject to the fundamental provisions of Islam is the religion of the Federation - quoting the judgment can demonstrate to the global community that Islam is a peace-loving religion. "Denga , "Stigma in the majority Muslim country is very significant, most patients do not constructive and healthy for the good of Islamic institutions are good but make criticisms without was held in the spirit of 'all Muslims are brothers', which should always tamp

As we look closer at the concordance lines, we can see that the context in which the word *Islam* is used shows the way in which Malaysians view Islam. It is used with such words as *rights, party, economy, affairs, and country*. These all show a strong connection to government. We also see the phrase, *Islam is the religion of the Federation* multiple times.

Following the cycle, I then researched the Malaysian government structure in order to better understand the connection between religion and state. It is possible that this strong connection of religion and government can be tied to the underlying government policies of Malaysia, namely that of Sharia Law. Each of Malaysia's 13 states have state branches of

government which practice Sharia Law. “The Sharia court has in recent years become a prominent subject in public discussion, not least of all with the constitutional provision that ‘Islam is the religion of the Federation’” (Dahlan & Faudzi , 2015) (Article 3, Federal Constitution). This contrasts with Indonesia where the constitution states that, “The Syariah Court of the Province of Nanggroe Aceh Darussalam is the judicature body which is free of the influence of parties anywhere in the territory of the Province of the Special Region of Aceh effective for Moslems” (Law No. 18/2001). This also sheds light on why this phrase is found often in the corpus. The word *Islam* appeared significantly more times in the MalayCorp not necessarily because Malaysia is more Islamic in practice, but because Islam is a part of the law in all 13 states as opposed to one province like Indonesia.

This connection is even further supported by the N-grams found in MalayCorp. N-grams were set to a cluster size of 3 minimum and 3 maximum. An excerpt from this can be seen below in Table 10.

Table 11. N-grams of ‘Islam’ from MalayCorp

Rank	Freq.	Range	Cluster	Gloss
1	84	69	<i>islam malaysia (jakim</i>	Islam Malaysia (JAKIM
2	71	59	<i>islam di negara</i>	Islam in the country
3	43	43	<i>islam kampung tasek</i>	islam lake village
4	38	33	<i>islam dan adat</i>	Islam and indigenous
5	36	30	<i>islam wilayah persekutuan</i>	Islamic federal territory
6	29	25	<i>islam di malaysia</i>	islam in malaysia
7	27	20	<i>islam di negeri</i>	Islam in the country
8	27	22	<i>islam malaysia (usim</i>	Islam Malaysia (USIM
9	25	25	<i>islam, tindakan beberapa</i>	Islam, the actions of some
10	23	23	<i>islam malaysia (ppim</i>	Islam Malaysia (PPIM
11	21	20	<i>islam antarabangsa malaysia</i>	international Islamic Malaysia
12	16	14	<i>islam dan bukan</i>	Islam and not
13	14	2	<i>islam di aceh</i>	Islam in Aceh
14	14	14	<i>islam di seluruh</i>	Islam throughout
15	14	14	<i>islam se-malaysia</i>	Islam One-Malaysia
16	14	13	<i>islam sebagai agama</i>	Islam as a religion
17	14	14	<i>islam yang sebenar</i>	the real Islam

18	13	11	<i>islam malaysia (yapeim</i>	Islam Malaysia (YAPEIM
19	13	13	<i>islam selangor (jais</i>	Islam Selangor (Jais
20	10	9	<i>islam yang lain</i>	other Islamic

I then compared these N-gram results to the N-grams for *Islam* in IndoCorp. The results are below in Table 12. Here we see that in contrast to the MalayCorp results, Islam is less frequently used in association with the State. There are cases, such as *Islam in Indonesia*, *Islam and the people*, and *Islam of the country*. However, these cases do not associate Islam with the State as strongly as does MalayCorp. In this list, Islam is more commonly associated with other countries. It is also notable that in this list *Islam and Christianity* appears as it shows that in Indonesia these two religions carry a strong association. The N-grams of *Islam* in the IndoCorp results are more concerned with the characteristics of the religion of Islam, which can be seen in such examples as *moderate Islam*, *merciful Islam*, *hard-line Islam*, and *the kind of Islam*. On this list however, we also find the N-grams of *Islam and Terrorism*, and *Islam or ISIS* the relationship of which will be discussed in the following section.

Table 12. N-grams of 'Islam' from IndoCorp

Rank	Freq.	Range	Cluster	Gloss
1	55	54	<i>islam irak dan</i>	Iraqi Islam and
2	27	19	<i>islam di indonesia</i>	Islam in Indonesia
3	17	5	<i>islam di aceh</i>	Islam in Aceh
4	16	13	<i>islam di irak</i>	Islam in Iraq
5	11	11	<i>islam atau isis</i>	Islam or ISIS
6	10	2	<i>islam dan umat</i>	Islam and the people
7	9	9	<i>islam terbesar di</i>	Islam's largest
8	8	8	<i>islam negeri (uin</i>	Islam of the country (UIN
9	7	1	<i>islam saya' adalah</i>	My Islam is
10	7	6	<i>islam yang moderat</i>	Moderate Islam
11	6	6	<i>islam adalah agama</i>	Islam is a religion
12	6	4	<i>islam rahmatan lil</i>	Islam mercy
13	5	4	<i>islam garis keras</i>	Hard-line Islam
14	5	1	<i>islam jenis itu</i>	That kind of Islam
15	5	5	<i>islam yang rahmatan</i>	The merciful Islam
16	4	1	<i>islam aceh singkil</i>	Islam Aceh Singkil
17	4	4	<i>islam dan kristen</i>	Islam and Christianity

18	4	2	<i>islam dan terorisme</i>	Islam and terrorism
19	4	3	<i>islam di kota</i>	Islam in the city
20	4	2	<i>islam di nigeria</i>	Islam in Nigeria

Extremism

For the Extremism category, before even looking at the collocates, we can see a significant difference in the use of words relating to the terrorism. There are no directly terror-related words in the Malaysian list. This leads to an initial possible discourse that Indonesian discourse sees Islam as closely related to terrorism. However, we need follow Baker et al.'s cycle to see if the initial possible discourse is supported by the deeper analysis. By doing the same analysis as used for the Islam section, and looking at collocates of *terrorist* in IndoCorp we see that terrorism is spoken of in a non-supportive and negative manner. Figure 2 below shows a snapshot of the translated results.

Figure 2. Translated Concordance lines of 'terrorist' from IndoCorp

Indonesia who has ever experienced a terrorist attack but also in many other countries enforcing the discussion is motivated terrorist attacks that just happened in Jakarta let baseball has radical thinking like terrorists again. The trick pass approach, we invite and Turkey is to save the terrorists who continue to experience defeat ISIS good results broadly in combating terrorist group ISIS. If we discount the strategy, 4 prohibits all forms of support for terrorist organizations. Unknown, Syrian government forces tel is stems from feelings of concern by terrorists who killed a lot of innocent people as well, which puts to Stop Terrorism and Terrorists. Stop Terrorism application allows peng Jakarta, Wednesday (01/27/2016). While the game is an arcade game Terrorists Tumpas light, which is me users. This game which invites gamers to crush terrorists is troubling, by way of knock out t terrorists were troubling, by way of crippling terrorist with slingshots and stones. "Applications game

Here we see that *terrorist* is used in association with such words and phrases as *terrorist attack, radical, combating, and killed innocent people*. Terrorism is also referenced when speaking of a popular video game where users are encouraged to *crush terrorists*, and [*cripple*] *terrorists with slingshots*. We can see here that *terrorist* is not necessarily key to Indonesia's

media because they are supporting it, but possibly because they are condemning it and supporting its combating. There was also a terrorist attack in Jakarta, the capital of Indonesia, during the time IndoCorp was created, which could lead to the higher frequency of the word. This illustrates the need to conduct outside qualitative research, as explained in Baker et al. (2008).

The next step, moving through the cycle (Baker, 2010), is to look into the outside influences or factors that could affect the corpus data. This means moving back to step 1 for a “context-based analysis of topic via history/politics/culture/etymology. Identify existing topoi/discourses/strategies via wider reading, reference to other CDA studies.” Due to the nature of this paper and the synchronic design of the corpus, this step includes looking at current events of each country at the time in which these events occurred. Part of the qualitative analysis as proposed by Baker et al. (2008), this step will help bring understanding to the keyword list as it pertains to each country at the time the newspapers were gathered. To do this I looked at the events that occurred during this time according to the different news sources in the corpus. The relevant events can be seen below in Table 13.

Table 13. Current events in Indonesia and Malaysia during the time of corpora creation

Country	Event
Indonesia	Sarinah Plaza bombing (Jakarta, Jan. 2016) Terrorism video game Government program against terrorism Indonesia’s Aceh province enacts enforcement of Islamic sharia criminal code
Malaysia	Syariah Courts debated Islamic government organizations

These first two keywords, *terrorism* and *Islam*, highlight both the effectiveness of Baker et al.’s (2008) model and the key discourse surrounding Islam in each of these countries,

summarized here. For Malaysia, Islam has close Government associations and is viewed as the religion of the State. For Indonesia, Islam has close ties to the culture and is viewed as the religion of the majority of its citizens. For the remainder of this paper, this initial observation of the relationship between Islam and these two countries will be referred to as a “running hypothesis.” This hypothesis emerged naturally from the initial investigations of the keyword lists from both corpora. We will continue to analyze the religious keywords to further test this running hypothesis, and then move on to the category and entire list analysis.

Sharia

Another significant word which emerged from the data is *Sharia* referring to the Islamic law that is practiced in many Muslim societies and countries. The word *Sharia* appeared on the MalayCorp keyword list, meaning it was more statistically characteristic to Malay newspaper discourse when compared to Indonesian newspaper discourse. To research this further, I looked at both the context in which the word was used and looked at other sources to uncover why *Sharia* would be characteristic of Malaysian discourse. It is also relevant with the rise of the fundamentalist movements in many Islamic nations. These rises often are associated with Sharia law, and so a deeper look at this word as it is used in Malaysia is relevant to a discourse analysis surrounding Islam in these two countries.

After conducting qualitative research, I discovered that in Malaysia, Sharia law or Sharia courts are present in all 13 states. It is possible that the word *Syariah* is more common in MalayCorp because the “Syariah court has in recent years become a prominent subject in public discussion, not least of all with the constitutional provision that ‘Islam is the religion of the Federation’” (Dahlan & Faudzi , 2015) (Article 3, Federal Constitution).

On the other hand, in Indonesia Sharia' law is only officially recognized in one province, the province of Aceh. "The Syariah Court of the Province of Nanggroe Aceh Darussalam is the judicature body which is free of the influence of parties anywhere in the territory of the Province of the Special Region of Aceh effective for Moslems." (Law No. 18/2001)

This qualitative research supports the quantitative data from the keyword analysis. It is possible that this is the reason the word *Sharia* appears more statistically more frequently in MalayCorp when compared to IndoCorp. From these results and the previous ones found when researching *terrorism* and *Islam* a pattern starts to emerge that Islam is thought about differently in these two countries. It seems that thus far in the research, the discourse surrounding Islam in Malaysia tends to associate it more with the Government and the State, while the discourse surrounding Islam in Indonesia seems to equate it more with the culture of the peoples living there. According to Baker et al.'s (2008) model, this theme should continue to be tested in this research process and should also guide the next steps of the research.

Allah

The keyword *Allah* has a keyness of 105.241 and a frequency of 1053 in MalayCorp. When comparing the word *Allah* in each corpus, it was found that Malaysian newspapers use *Allah* 8.6 times per 100 articles, while Indonesian newspapers use *Allah* 4.2 times per 100 articles. However, as Baker et al.'s (2008) model reminds us, we cannot make assumptions regarding discourse based on statistics and quantitative data alone.

A brief look at the collocates of *Allah* in Malay newspapers compared to Indonesian newspapers shows the Muslim initialism *SWT*, meaning *Subhanahu wa Ta'ala* or 'Allah the Exalted and Most High' to be common in both languages. However, *SWT* is used with *Allah*

almost three times as often in Malay newspapers (3.8 per 100) than Indonesian newspapers (1 per 100) and is the most frequent collocate for *Allah* in MalayCorp (frequency of 330 when looking at 3R and 3L collocates). The use of this abbreviation shows that *Allah* is used more frequently in specifically Muslim contexts in Malay newspapers compared to Indonesian.

As the most populated Muslim country in the world, Indonesia would have been more logical as the country to reference *Allah* more in its news. However, the appearance of *Allah* in the MalayCorp keyword list and its association with Islam, not other religions such as Christianity, supports the running hypothesis of this paper that in Malaysia there is a stronger tie between Islam and the state than in Indonesia. It is also significant that *Allah* appeared in the top 3,000 keywords at all since news generally does not talk about God specifically, while it may talk about religion in general.

To gain a better understanding of the contexts in which *Allah* is used and the discourse surrounding it, concordance lines and collocates were analyzed. A list of the top 20 collocates can be found below in Table 14.

Table 14. Collocates of ‘Allah’ from MalayCorp

Rank	Freq.	Freq. L	Freq. R	Stat	Word	Gloss
1	2	1	1	12.80856	<i>kami beramal</i>	we do
2	4	1	3	12.22360	<i>taala</i>	Most High
3	330	12	318	11.87110	<i>SWT</i>	God almighty ¹⁷
4	128	125	3	11.84278	<i>insya</i>	willing
5	1	1	0	11.80856	<i>yaseen</i>	Yasin ¹⁸
6	2	0	2	11.80856	<i>watallahi</i>	and God
7	2	0	2	11.80856	<i>wallahi</i>	by God
8	1	1	0	11.80856	<i>wahablumminannas</i>	relationship with people ¹⁹

¹⁷ From the Arabic-rooted phrase used to speak about God.

¹⁸ Person’s name and Quranic chapters ‘Surah Yaseen’

¹⁹ From the Arabic-rooted phrase used in Malay, ‘Hablum min Allah Wahablum min annas,’ roughly

9	2	0	2	11.80856	<i>wabillahi</i>	and with God
10	1	0	1	11.80856	<i>tetapkanlah</i>	inscribe
11	1	1	0	11.80856	<i>tertamu</i>	be a guest
12	1	1	0	11.80856	<i>talaq</i>	revocable
13	1	1	0	11.80856	<i>takwanya</i>	righteous
14	1	1	0	11.80856	<i>takutlah</i>	fear
15	2	0	2	11.80856	<i>takdirkan</i>	destined
16	1	1	0	11.80856	<i>sumpahnya</i>	affirmation
17	1	1	0	11.80856	<i>sembahlah</i>	worship
18	1	0	1	11.80856	<i>sasarnya</i>	goal
19	2	2	0	11.80856	<i>renunglah</i>	meditate
20	1	0	1	11.80856	<i>redai</i>	a wire

The second most common collocate to appear with *Allah*, with a frequency of 128 (when looking at 3R and 3L collocates), is *insya* ‘if want,’ meaning together ‘If God wills/wants it.’ This phrase is often used in a similar context to the word ‘hopefully’ in English. However, it is still significant as it is a Muslim phrase, and associated only with members of that faith. This is even more certain in Malaysia where at the time MalayCorp was gathered, the use of *Allah* was prohibited by all non-Muslims. In Table 11 we can see that aside from *SWT* and *insya*, all other collocates appear a minimal number of times with *Allah* and are less significant. The significance of ‘Allah’ in the MalayCorp keyword list is that *Allah* occurs at all on the list of newspaper keywords and supports the running hypothesis of this paper.

Tuhan

We can learn even more about the discourse surrounding Islam in these two nations when we look at the word *Allah* in comparison to the word *Tuhan* ‘Lord,’ which appeared on the IndoCorp keyword list with a keyness of 121.715 and frequency of 367. The appearance of these two words on their respective lists supports the running hypothesis of this paper. In Indonesian,

translated to ‘relationship with God and relationship with mankind.’

Tuhan ‘Lord’ is used to refer to a general God that is not specific to one religion or another. This contrasts with Malaysia use of *Allah*, which has been restricted by the Malaysian government to use by Muslim only. It is significant that when Indonesian’s talk about this deity they use the word what includes all religions which are recognized there, while in Malaysia they use the word that excludes all religions except Islam.

Table 15. Collocates of ‘Tuhan’ from IndoCorp

Rank	Freq.	Freq. L	Freq. R	Stat	Word	Gloss
1	2	1	1	14.36256	<i>commandements</i>	commandments
2	3	2	1	13.94752	<i>mahaesa</i>	almighty
3	1	1	0	13.36256	<i>semulya</i>	as noble
4	1	0	1	13.36256	<i>semedo</i>	Semedo ²⁰
5	1	0	1	13.36256	<i>sebaris</i>	In line
6	1	0	1	13.36256	<i>samakah</i>	Is it the same
7	1	1	0	13.36256	<i>sabur</i>	patient
8	1	1	0	13.36256	<i>qayyum</i>	Al-Qayyum ²¹
9	1	1	0	13.36256	<i>qawiyuu</i>	Al-Qawiyuu ²²
10	1	1	0	13.36256	<i>pasrahkan</i>	resigned
11	1	0	1	13.36256	<i>nyiptain</i>	to create
12	1	0	1	13.36256	<i>nksa</i>	nksa
13	1	1	0	13.36256	<i>ngabdi</i>	subserve
14	1	1	0	13.36256	<i>mulyanya</i>	noble/his nobility
15	1	0	1	13.36256	<i>mengasihi</i>	love
16	1	1	0	13.36256	<i>membuatkannya</i>	create one
17	1	1	0	13.36256	<i>malikul</i>	reverse
18	1	0	1	13.36256	<i>mahatinggi</i>	exalted
19	1	0	1	13.36256	<i>imani</i>	pious
20	2	1	1	13.36256	<i>hebohnya</i>	excitement

When we conduct a similar analysis to *Allah* and look at collocates (3R and 3L collocates), we see that only a few of the of the collocates of *Tuhan* are related to Islam, *qayyum* and *qawiyuu*, two of the 99 Muslim names for God. There are other religious words, such as “almighty” and “to create” that could be Islamic references, but they are not exclusive to Islam in

²⁰ City in Central Java.

²¹ One of the 99 Muslim names for God.

²² One of the 99 Muslim names for God.

the way that many of the collocates of *Allah* were exclusive to Islam. This shows that both countries' newspapers discuss religion and God at a high frequency, however when God is discussed in Malaysian newspapers, it is the God of Islam. On the other hand, when God is discussed in Indonesian newspapers, it is a non-denominational God that could be claimed by any of the religions in Indonesia.

4.3 Analysis by Category

Locations

We will now move to the category of Locations. The locations which were found in the keyword lists are countries which have strong ties to Islam in Indonesia, Malaysia, or outside countries. Looking at how these are talked about and in what contexts will possibly help us better understand the discourses surrounding Islam in Indonesia and Malaysia. We will first look at MalayCorp, and then move on to IndoCorp. See Tables 16 A and B for words in this category.

Table 16A. Location keywords from IndoCorp Table 16B. Location keywords from MalayCorp

Category	Keyword	Gloss	Category	Keyword	Gloss
Locations	<i>Aceh</i>	Aceh	Locations	<i>Terengganu</i>	Terengganu
	<i>Irak</i>	Iraq		<i>Bangladesh</i>	Bangladesh
	<i>Iran</i>	Iran		<i>Brunei</i>	Brunei
	<i>Suriah</i>	Syria			
	<i>Sarinah</i>	Sarinah			

The MalayCorp keyword locations were *Bangladesh* and *Brunei*. To determine the significance of these words in relation to the MalayCorp discourses surrounding Islam, collocates and concordance lines were investigated. From the concordance line results, it appears that *Bangladesh* is discussed mostly in the context of workers from Bangladesh who have come to Malaysia seeking employment. Examples from the results can be found below in Figure 3A.

Additionally, when looking at collocates for Bangladesh, the word *Islami* ‘Islamic’ is only used twice, and both occurrences are from the same text file. It is notable however that to single reference connecting Islam and Bangladesh is in a text file is discussing leaders of and Islamic group from Bangladesh who were sentenced with the death penalty (Text File 8_17918), see Figure 3B. This further supports the running hypothesis of this paper that in Malaysia, Islam has a stronger connection to the Government and organization of the State than it does to the culture of the people. As we move on to look at the keyword *Brunei* and its concordances and collocates, there is no clear connection made with Islam; most contexts are references to the different ASEAN countries, see Figure 3C.

Figure 3A. Translated Concordance lines of ‘Bangladesh’ from MalayCorp

entry of foreign workers, including those from	Bangladesh, there are the angry and menyinga. word
the government's intention to bring in 1.5 million workers	Bangladesh is motivated 'profit businesses
(MoU) signed yesterday by the government	Bangladesh is a common agreement as in
signed a total of 1.5 million migrant workers from	Bangladesh, is to meet the demands of magic
United take strict action against the Government	Bangladesh in order to comply with the principles of human rights
wrap the balloon at a factory in	Bangladesh. - Child labor AGENCY forced m
Riot denies 1.5 million foreign workers	Bangladesh will be brought to this country, but
was reported as saying by 1.5 million workers from	Bangladesh will be brought to Malaysia for a role
! Close! Finished! No further 1.5 million workers	Bangladesh will go into the country. "Www.
comedian from Bangladesh. Certainly, the character guaranteed	Bangladesh will continue to be held by local people

Figure 3B. Translated Concordance lines of ‘Bangladesh’ with ‘Islami’ from MalayCorp

political opposition parties, particularly Jamaat	Islami Bangladesh, it is clearly seen as a
now some of the leaders of Jamaat	Islami Bangladesh were recently sentenced to death

Figure 3C. Translated Concordance lines of 'Brunei' from MalayCorp

there are 26 November 2015 in Malaysia, Singapore and	Brunei.
26 November simultaneously in Malaysia, Singapore and	Brunei.
two members of the Armed Forces	Brunei (ABDB). Team Shipping MALBATT 850-3
and Brunei. Vietnam, Philippines, Malaysia da	Brunei is a member of ASEAN and they always
Asean members, Malaysia, Vietnam, Singapore and	Brunei is among 12 countries that try
United Airlines, owned by Brunei Darussalam, Royal	Brunei Airlines (RB) will start flying b
of all states, including contingent invitation,	Brunei will participate in the Games this time
has been in the Singapore market and	Brunei. Shukri said Aliff, Malaysia persain
Harimau Muda coach as well	Brunei. He also regularly appears in televisy
Music, Elizabeth Tan and singer of	Brunei, Aziz Harun.

The locations found in IndoCorp are *Syria, Aceh, Iraq, Sarinah, Iran, and Sumatra*. As with the MalayCorp locations, these are all locations found in the keyword list, either located in the country of interest or not, that could potentially have a tie to Islam. The location *Sarinah* will not be addressed here as it is addressed in the section on current events. *Aceh* will also not be addressed in detail here as it is address in the section on Sharia' law. The other locations, *Syria, Iraq, Iran, and Sumatra* will be discussed here briefly. First, we will look at *Sumatra* in the data. The only evident connection with *Sumatra* and Islam is the when the Islamic province of Aceh is referenced. Next, *Syria, Iraq, and Iran* are all Muslim majority countries. These countries appear most commonly in IndoCorp because of the current events with terrorist attacks, wars, and immigrants.

Politics

By investigating the words in each list which have political associations, and not just the words associated with worship or practices, we are able to keep our researched more balanced. We can see how Islam is portrayed in different realms, and help keep our results more generalizable to the larger population. We will first look at IndoCorp, and then move on to MalayCorp. See Tables 17 A and B for words in this category.

Table 17A. Political keywords from IndoCorp Table 17B. Political keywords from MalayCorp

Category	Keyword	Gloss	Category	Keyword	Gloss
Politics	<i>wars</i>	wars	Politics	<i>UNMO</i>	UNMO
	<i>Nahdlatul</i>	Nahdlatul			
	<i>PKB</i>	PKB			

In the Political category of IndoCorp, we find the words *wars*, *Nahdlatul*, and *PKB* (The National Awakening Party is a moderate Islamic and a conservative political party in Indonesia). The first word of significance is *wars*. First, it is notable that the English word is used here. At first glance, it would seem that ‘*wars*’ are very significant in IndoCorp, however, thanks to Baker et al.’s (2008) cycle, when we look further at the concordance lines and collocates, we see that ‘*wars*’ is referring to the film ‘Star Wars’ and not actual wars occurring in the world at that time. See Figure 4.

Figure 4. Concordance lines for ‘*wars*’ in IndoCorp

IV, V and VI, namely Star	Wars: A New Hope; The Empire Strikes
glowing. As a rabid fan of Star	Wars, Abrams seems to fulfill the thirst
about 1.4 billion US dollars. after Star	Wars, there was the comedy "Daddy's Home"
encouraged me to become a fanboy Star	Wars. There is a scratch in the liver. Certain
Star Trek and Star now	Wars are processed J.J Abrams
those who follow the movie series Star	Wars, is a character that is not foreign
a critique of the movie series Star	Wars is because the film was too "boys
"Star Wars" or star wars. Star	Wars is the designation for the country berlomba-
certainly remind fans loyal Star	Wars Luke Skywalker scene pulled the sword
faithful waited with pounding. Is Star	Wars will be sweet like cotton candy? Or
Vader was in fact ... argh! (Listening Star	Wars will know what I mean).
directly on the style of the original trilogy Star	Wars. However, The Force Awakens also

Additionally, this explains the use of the English word as opposed to the Indonesian word. Without Baker et al.’s (2008) methodology, it may have been possible to say that in addition to Islam having strong ties to extremism or terrorism, it also has strong ties to war and violence. However, when looked at more qualitatively, we are able to better understand each

keyword found. The other two words *Nahdlatul* and *PKB*, are both political parties associated with Islam, but are not used significantly on their own.

MalayCorp references the political organization, *UMNO* or United Malays National Organization. MalayCorp does not have any other significant political parties or references aside from *Sharia*, which has already been discussed. One of the main goals of UMNO is to “defend and spread Islam, the official religion of the country and respect the principle of religious freedom” (UMNO's Constitution). The keywords and concordance lines for *UMNO* in MalayCorp are not significant enough to expand the discourses found in the corpus.

Terrorism

We have already addressed the Terrorism category at the beginning of the Analysis by Keywords section, so we will only briefly discuss it again here. As a reminder, the words found in this category are below in Tables 18 A and B.

Table 18A. Terrorism keywords from IndoCorp Table 18B. Terrorism keywords from MalayCorp

Category	Keyword	Gloss	Category	Keyword	Gloss
Terrorism	<i>ISIS</i>	ISIS	Terrorism		
	<i>teror</i>	terror			
	<i>terorisme</i>	terrorism			
	<i>teroris</i>	terrorist			
	<i>radikalisme</i>	radicalism			

In the previously mentioned section, terrorism was discussed in terms of how it is significant to IndoCorp, however, we did not discuss how it is significant to MalayCorp. More accurately, we should say, how the lack of terrorism words is significant to MalayCorp. It is significant that there were no terrorism words in the MalayCorp keyword list. It is also significant to note that the words on this list were only included if they related to Islam. Any

other terrorism words which were not connected to Islam were not included on the list. This was to ensure that other terror or terrorism words did not affect the analysis of this category.

Worship & Teachings

Lastly in this section, we will look at the category of Worship & Teachings. See Tables 19 A and B for words in this category. Looking at the categorized table of keywords, it is notable that there are more words in this category from MalayCorp than from IndoCorp. This is significant firstly because as the largest Islamic population, it would seem more fitting that Indonesia would have more words in this category because this category is applicable to more of its inhabitants.

Table 19A. Islamic keywords from IndoCorp Table 19B. Islamic keywords from MalayCorp

Category	Keyword	Gloss	Category	Keyword	Gloss
Worship & Teachings	<i>salat</i>	salat/prayer	Worship & Teachings	<i>Islam</i>	Islam
	<i>jamaah</i>	pilgrims		<i>solat</i>	salat/prayer
	<i>tuhan</i>	lord		<i>haram</i>	illicit
	<i>alquran</i>	Al Qur'an		<i>jemaah</i>	congregation
				<i>syariah</i>	Sharia'
				<i>persembahan</i>	religious offering
				<i>ibadat</i>	worship
				<i>rohani</i>	spiritual
				<i>Allah</i>	God
				<i>Agama</i>	religion
				<i>Haji</i>	title of Muslim who has completed Hajj
				<i>Halal</i>	permitted/clean /legal
				<i>Syurga</i>	Heaven

However, as we have mentioned above, our running hypothesis is now that Islam in Malaysia has stronger ties with Government and the State than Islam in Indonesia. This is supported by the content in the Worship & Teachings category. The lack of Worship & Teachings words on the IndoCorp list also supports the hypothesis that Islam is not as relevant in

the media world in Indonesia as it is in Malaysia. The log-likelihood analysis found that words relating to the worship and teachings of Islam were not statistically characteristic of how Islam is discussed or portrayed in Indonesian newspapers. If Islam was more relevant in Indonesian media than Malaysian media, more of these words would appear on the keyword list.

4.4 Additional Possible Influences

We have already looked at the linguistic ecology, and the current events in Indonesia and Malaysia and how these factors could possibly influence the discourses found in the respective corpora. In addition to looking at these factors, it is important to exercise due diligence and investigate additional possible influences into the trends found in the corpora. This will help researchers determine if trends are from the data or more strongly influenced by unrelated outside events. This section will highlight and analyze the effects of freedom of the press and religious populations in each country.

Freedom of the Press

The first possible influence to be addressed is the press laws in each country. When analyzing discourse found in newspapers, it is important to also examine the laws that govern these newspapers as this may influence what each country can and cannot print. Malaysian and Indonesian newspapers are both heavily regulated by their governments. According to the Freedom House ratings, Malaysian press is considered “Not Free” with a score of 65/100 (0 = best, 100 = worst), while Indonesian press is considered “Partly Free” with a score of 49/100 at the time of corpus creation (Freedom, 2015). This is significant as a discourse analysis based on newspaper articles is only as accurate as the articles it uses. However, without further research it is not possible to know which articles are more influenced by government regulation and which are

not. Because of this, for the purpose of this study, all articles have been treated as equally significant, but the researcher notes the possibility of censorship interference.

Religious Populations

We will not look at the population of Muslims in both countries. This is relevant to our study because if a religious population is the majority in a country, it is likely that the news will report differently on it than if it is the minority in a country. According to the International Religious Freedom Report for 2015 from the Bureau of Democracy, Human Rights and Labor, 61.3% of the population of Malaysia practices Islam, compared to 87% of the population of Indonesia (United, 2015). So, Indonesia has the higher percentage of Muslims between these two countries. This could have influenced the data; however, the scope of this paper does not cover further research into the religious bias of each newspaper in the corpora.

5 CONCLUSION

5.1 Answers to Questions

In this section I will briefly review and summarize the research goals and questions, starting with the methodological goal and moving to the discourse analysis questions. I will then address the weaknesses of the research and suggest possible solutions and future related research topics.

RQ 4. Evaluate the efficacy of a mixed-methods corpus-driven approach to discourse analysis using the methods proposed by Baker et al. (2008).

While there is still much research to be done on the discourse surrounding Islam in Indonesia and Malaysia, the initial results presented in this paper illustrate the differences in Islamic discourses in these two neighboring, historically connected nations. Baker et al.'s (2008) model insures that while conducting discourse analyses, researcher bias and influence, whether intentional or not, can be minimized to allow the data to lead the researcher to trends in the corpora.

The methods of Baker et al.'s model (2008) are successful in producing a more objective cross-cultural discourse analysis on unrestricted corpora. The cycle, as it moves back and forth from qualitative to quantitative analysis, is valuable in its role to increase objectivity; there is also an added effectiveness when combined with unrestricted corpora as this process ensures that the researcher begins with a more open set of data. By enlisting a cycle of qualitative and quantitative analysis on an unrestricted corpus, the researcher can provide more reliable and representative results.

The act of using unrestricted corpora allowed the data to produce the search terms used in this research as opposed to the researcher determining the search terms prior to beginning the analysis. After the keyword list was produced using AntConc, each word on the list was categorized into categories as they naturally formed from the list. Then, as a result of Baker et al.'s cyclical model, the list was re-evaluated multiple times and words were shifted into new categories when KWIC, collocates, and/or n-grams showed the originally category to be incorrect. This ensured that as Baker et al.'s model was followed; the keyword list became more and more accurate.

Unrestricted corpora also led to a more complete list of keywords to be searched by using unrestricted corpora, more words appeared in the lists of words to be searched than if I had thought of them on my own.

By following Baker et al.'s (2008) model, this research was able to successfully investigate the discourses surrounding Islam in Malaysian and Indonesian newspapers from August 2015 to February 2016. The main discourses found have been referred to throughout this paper as a "running hypothesis," meaning, it was a hypothesis that emerged naturally from the data as the Baker et al.'s (2008) model was followed. This method allowed the hypothesis to be formed independent of researcher background or bias.

However, it is important to note that Baker et al.'s model is not the only way to conduct a discourse analysis. There are other methods that have been used and can be used. The benefit of Baker et al.'s model is that it attempts to combine the strengths of the two main schools of thought in discourse analysis – qualitative and quantitative methods. This being said, it is not perfect and could use further testing and evaluation. The goal of this study was not to prove that

Baker et al.'s study was the only method for conducting an unbiased corpus-assisted critical discourse analysis, but to assess how the method worked on a unrestricted, cross-cultural set of corpora.

RQ 1. How are Islamic themes portrayed in the newspapers? How do they differ by country?

The main discourse or “meanings, metaphors, representations, images, stories, statements” (Burr, 1995) surrounding Islam found in MalayCorp is that Islam in Malaysia is closely tied to the Government and Government organizations. This discourse emerged as the keywords, collocates, and concordance lines were investigated following Baker et al.'s (2008) model. Islam is also viewed as the religion of the State. This was evidenced in the keywords found in MalayCorp and the concordance lines of several of the more significant keywords such as ‘Islam.’ Islam is viewed less as one of the various religions found in the country, and more as the identity of the country.

On the other hand, for Indonesia, Islam has close ties to the culture and is viewed as the religion of the majority of its citizens. This was also evidenced in the way Islam was discussed and the keywords found in association with it. The words had less association with Government, laws, and regulations, and more to do with the culture of the people. In Indonesia, Islam is viewed as one of the religions practiced by its inhabitants. It could be concluded that Malaysia is an Islamic nation, while Indonesia is a Muslim-majority nation.

RQ 2. How can this inform us on each country's reactions to the Islamic fundamentalist movement?

The results of this study can help us better understand the reactions of each country to the Islamic fundamentalist movement in that the results help us better understand the relationship

between Islam and the governments of these two nations. These two countries have Muslim-majority populations, however, how they identify themselves in regard to religion is more significant to understanding their actions and reactions at this time.

Because Malaysia's discourse is that it is an Islamic nation, it is understandable that as the fundamentalist movement grows, Malaysia would lean towards government-supported decisions and regulations to protect Islam as the Government sees fit. This is supported by the case referenced at the beginning of the paper in which the Malaysian government prohibited the use of the word *Allah* by all non-Muslims. This paper is neither agreeing nor disagreeing with this decision, it is merely stating that there is a connection between this discourse and the actions of Government. It is possible that this law was put in place to prevent arguments or disagreements between the two religions in the country.

Indonesia's discourse of it being a Muslim-majority country is also significant in better understanding their actions during this time. Indonesia's government is less active in making law regarding Islam, and more focused on recognizing the diversity of religions in the country. Islam is viewed more as a cultural identity than a political distinction.

RQ 3. What are the frequent topics and issues discussed in articles relating to Islam?

According to the data found in the discourse analysis, the frequent topics and issues discussed in the articles which relate to Islam are: terrorism, Government, rights of the people versus rights of the Government, religious traditions, and laws relating to religion.

5.2 Limitations

This research was limited in a few key areas. First, we will discuss the limitations of the corpora sizes. The corpora used in this study were synchronic and only sampled newspaper

articles spanning a 6-month time period. The research conducted here could have been more substantial with a diachronic corpus to show that certain trends were indeed linked to events occurring in the country at the time of text collection. For example, while this research suggested that it is possible that terrorism is referenced so frequently in IndoCorp because of the bombings which occurred in Sarinah plaza, a more diachronically-designed corpus could show the frequency of 'terrorism' before, after, and during the bombing. Doing this would also be more in-line with what Baker et al.'s (2008) model suggests – re-working the corpus throughout the cycle to support more accurate data and analysis, and a more representative conclusion.

This research was also limited in the amount of time available to conduct the research. Since this is a master's thesis and not a dissertation, this research acts more as starting point for future research.

5.3 Future Research

Baker et al.'s methodology (2008) is also applicable to additional fields outside the focus of linguistics and should be tested in these fields, such as constitutional law and literary analysis. In future research, as mentioned above, it would be beneficial to expand this corpus to become a diachronic corpus, spanning at least one year.

It would also be beneficial to expand the corpus to include other genres in addition to newspapers. This would give researchers to opportunity to compare discourses as found in a variety of genres. Or to do a more general discourse analysis as opposed to one only discussing the discourses found in newspapers.

There is also much potential for future research on fundamentalism in Malaysia and Indonesia with an expanded corpus. While this corpus is useful for research on the portrayal or

Islam in newspaper media, and that relationship with fundamentalist movements, an expanded corpus would better represent the entire population of Indonesia. I would like to expand the corpus to incorporate social media, such as Twitter, Facebook, and blogs. This would enable research to look at the voice of the media and the voice of the citizens of each country and their discourse. This type of more expansive web corpus of Indonesian and Malay would also make the corpus more useful for government use and analysis, and will be further addressed in the following Implications section.

5.4 Implications

In conducting this research, it is tempting to look at these results and draw quick conclusions, if this research reveals everything needed to understand the responses of each of these countries to the Islamic fundamentalist movements, however, this study is not meant for that. While it does bring some greater understanding of why each country may be reacting in the way it is, this study is more useful in what it means to the bigger picture of how this research could be used outside of the linguistic milieu.

While this research is interesting from an academic perspective and useful in evaluating the methodology as proposed by Baker et al. (2008), it is useful outside of this realm as a result of its ability to inform agencies, government or otherwise, as to why countries react to things in a certain way. If our governments can better understand why a country is doing something, as opposed to just knowing what is happening, then we can make better informed decisions as to how to find a resolution or prevent that reaction. For example, is Malaysia creating laws to restrict the use of the word *Allah* in attempts to alienate non-Muslims? Or is it because they are making a conscious move towards a more radical Islamic state? Or is it because they view the

relationship of Islam with the government differently than does the West? By better understanding why it is happening, we can better understand how to react to it. Instead of just gathering information from news reports of what is happening, we can use corpus-based critical discourse analysis to better grasp how the actions are tied into the culture. I think that this type of research has great potential to be refined and used in evaluating the political and social state of nations.

APPENDIX

Appendix Table 1. IndoCorp List of Online News Sources

Tribunnews
SINDOnews.com (Siaran Pers
Okezone
KOMPAS.com
Detikcom
CNN Indonesia
Bola.net
Republika Online
Bintang.com
News Liputan6.com
ANTARA
Showbiz Liputan6.com
Suara.com
BBC Indonesia
Bola Liputan6.com
BeritaSatu
Tempo.co News Portal
Bisnis Liputan6.com
VIVA.co.id
Serambi Indonesia
Tempo
KapanLagi.com
Tekno Liputan6.com
Tempo Kolomamp; Opini
Bisnis.com
Health Liputan6.com
Panditfootball Indonesia
Tempo Nasional
Metro TV News
Bangka Pos
Wow Keren
Tempo.co
Lifestyle Liputan6.com
RMOL.CO
Banjarmasin Post
Info Komputer
Tribun Jabar

Warta Kota
Global Liputan6.com
Tribun Timur
Hidayatullah.com mengabarkan kebenaran
JPNN.com
Harian Terbit
Tabloidbintang.com
Tribun Batam
Otomotif Liputan6.com
National Geographic
Mongabay.co.id
Tribun Kaltim
Dream
Sriwijaya Post
Tren Tekno
BeritagarID
Make Mac
Tribun Medan
Citizen6 Liputan6.com
TeknoUp
FourFourTwo
Solopos
Tribunjogja.com
Tribun Pekanbaru
Rancah Post
Selular.ID
Ramadan Liputan6.com
Sinar harapan
BallBall
Tribun Pontiaknak
Super Soccer
Tribun Jambi
SIDOMI
Tempo Sport
Waspada Online
Merah Putih (Siaran Pers)
Surya
Pos Kupang
Kontan
ITB Berita
merdeka.com

Oketekno.com - Inspirasi Berita Teknologi Terbaru
News HargaTop
Jawa Pos
Tribun Manado
swadeka
Suara Pembaruan
Tempo World News
Rimanews
Muvila
Tribun Lampung
Poskotanews
LensaIndonesia.com
Kpop Chart
Berita Sore
Tempo Travel
Warta DPR Liputan6.com
Tabloid Pulsa
Indogamers (Sindiran) (Siaran Pers) (Pendaftaran)
Pontianak Post
Harian Analisa
Harian Jogja
Inilah.com
Bhataramedia.com
Chip Online Portal
Tabloid Nova
Cinemags (Pendaftaran)
Otomotifnet.com
jadiBerita.com
Fajarnews.com
m.beritasatu.com
Sooperboy
BerbagiTeknologi.com
Sportsatu
OtoVaria
Inilah Koran
PCplus
BeritaBulukumba.com
SuaraMerdeka
1Health
Gila Bola
Fimela

RADAR INDO
DNA Berita
Cahaya Baru
RiauOnline.co.id
GadgetGan
Autonet Magz
Riau Pos (Siaran Pers)
Indopos
Beritagar (Pendaftaran)
Popmagz.com
Manado Line (Pendaftaran)
Cosmopolitan
JATIMTECH.com
Otosia
Seruu.com
Pojok Satu
BabatPost.com - Portal Berita Terkini
Nakita
Beritagar (Siaran Pers) (Berlangganan)
Merah Putih (Siaran Pers)
Portal Berita Online Indonesia
Aktudas
Lampost
otorider.com
Majalah Otomotif Indonesia
Metroterkini
Go Bekasi
Jitunews
klikdokter.com
halloriau (Siaran Pers)
Otomotif (Siaran Pers) (Pendaftaran)
Esquire
Ongisnade
WartaJazz.com
deliknews (Siaran Pers)
Mother&Baby
cirebonsatu (Sindiran) (Siaran Pers)
Radar Sampit
Otomotif Community
BINTANG OTOMOTIF (Sindiran) (Siaran Pers) (Berlangganan)
Tempo Video

Suster Gadget
joss (Siaran Pers)
Radar Banjarmasin
Tekno (Siaran Pers)
Jaring News
Pojok Jabar
kriminalitas (Siaran Pers)
Fajarnews.com (Siaran Pers)
Today Idol
acehkita.com
riau (Siaran Pers)
Mesinbalap.Com
Parenting
otodriver.com
TangerangNews
Harper Bazaar
koreanwaveindo
Suara Purwokerto
acehterkini
Katadata News (Siaran Pers)

Appendix Table 2. MalayCorp List of Online News Sources

AgendaDaily
ANTARA
Astro Awani
Badminton World Federation
Badzine
BallBall
Bangka Pos
Banjarmasin Post
BBC Indonesia
BerbagiTeknologi.com
Berita Harian
BeritaBulukumba.com
BeritaSatu
Bernama
Bintang.com
Bisnis.com
Bola Liputan6.com
Bola.net
Chip Online Portal
Citizen6 Liputan6.com
CNN Indonesia
Cosmopolitan
Daily Express
Detikcom
DNA Berita
Dream
FourFourTwo
fourthofficial.com
Free Malaysia Today
GadgetGan
Gila Bola
Global Liputan6.com
Harakah Daily
Harian Analisa
Harian Analisa
Harian Metro
Harian Terbit
Health Liputan6.com
Hidayatullah.com mengabarkan kebenaran
Indogamers (Sindiran) (Siaran Pers) (Pendaftaran)
Inilah.com
Islam NU
ITB Berita
jadiBerita.com
Jitunews

KapanLagi.com
KOMPAS.com
Kpop Chart
LensaIndonesia.com
Lifestyle Liputan6.com
m.beritasatu.com
Majalah Otomotif Indonesia
Make Mac
malysiandigest.com
Merah Putih (Siaran Pers)
merdeka.com
Mesinbalap.Com
Metro TV News
Mother&Baby
Muvila
New Straits Times Online
News HargaTop
News Liputan6.com
Oketekno.com - Inspirasi Berita Teknologi Terbaru
Okezone
Ongisnade
otodriver.com
Otomotif Liputan6.com
OtoVaria
Panditfootball Indonesia
PCplus
Pikiran Rakyat
Pos Kupang
Rancah Post
Republika Online
Riau Pos (Siaran Pers)
Rimanews
RMOL.CO
Selular.ID
Serambi Indonesia
Showbiz Liputan6.com
SIDOMI
Sinar harapan
Sinarharian
SINDOnews.com (Siaran Pers)
Suara Pembaruan
Suara.com
swadeka
Tabloid Pulsa
Tabloidbintang.com
TangerangNews
Tekno Liputan6.com

TeknoUp
Tempo.co
Tempo.co News Portal
The Rakyat Post
The Star Online
The Sun Daily
Tren Tekno
Tribun Batam
Tribun Jabar
Tribun Jambi
Tribun Kaltim
Tribun Lampung
Tribun Medan
Tribun Pontiaknak
Tribun Timur
Tribunnews
Utusan Malaysia
VIVA.co.id
Warta Kota
WartaJazz.com
Wow Keren

WORKS CITED

- Ajšić, A. (2015). *Language ideologies, public discourses, and ethnonationalism in the Balkans: A corpus-based study*. Northern Arizona University. Retrieved from https://www.researchgate.net/profile/Adnan_Ajsic2/publication/280156732_Language_ideologies_public_discourses_and_ethnonationalism_in_the_Balkans_A_corpus-based_study/links/55ad25e908aee079921dc41e.pdf
- Atkins, S., Clear, J., & Ostler, N. (1992). Corpus Design Criteria. *Literary & Linguistic Computing*, 7(1), 1–16.
- Azra, A. (2006). *Indonesia, Islam, and democracy: Dynamics in a global context*. Jakarta, Indonesia: Solstice Pub.
- Baker, P. (2006). *Using corpora in discourse analysis*. London; New York: Continuum.
- Baker, P. (2010). *Sociolinguistics and Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Baker, P., Costas, G., & Tony, M. (2013). Sketching Muslims: A Corpus Driven Analysis of Representations around the Word “Muslim” in the British Press 1998-2009. *Applied Linguistics*, 34(3), 255–278.
- Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyzanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, 19(3), 273–306. <https://doi.org/10.1177/09579265080888962>
- Bednarek, M. (2015). Baker, P., Gabrielatos, C., & McEnery, T. (2013). Discourse Analysis and Media Attitudes. The Representation of Islam in the British Press. *International Journal of Corpus Linguistics*, 20(3), 405–412. <https://doi.org/10.1075/ijcl.20.3.06bed>
- Bin Mas'od, M. A. "The Arguments to Reject the Use of ‘Allah’ by the Christians.” Web. http://www.islam.gov.my/sites/default/files/reject_use_of_allah_by_christians.pdf.

- Blommaert, J. (2005). *Discourse*. Cambridge: Cambridge University Press.
- Break. Def. 1. In Oxford English dictionary online, Retrieved from
<http://www.oed.com.erl.lib.byu.edu>
- Burr, V. (1995). *An Introduction to Social Constructionism*. London: Routledge.
- Campbell, S. "The Arabic element in Indonesian: What do students need to know about it?"
 Web. <<http://your.usc.edu.au/wacana/4/campbell.html>>.
- Dahlan, R., & Faudzi, F. S. (2015). The position of the Shariah Court in the Malaysian legal system. *Malay Mail Online*.
- Fairclough, N. (1989). *Language and Power*. London: Longman.
- Firth, J. R. (1957). *Papers in Linguistics 1934-1951*. London: Oxford University Press.
- Foucault, M. (1972). *The Archaeology of Knowledge*. London: Tavistock.
- Freedom House. (April 29, 2015). Freedom of the Press 2015: Harsh Laws and Violence Drive Global Decline. *Freedom at Issue: Insights on the Global Struggle for Democracy, 2015-4*.
- Gabrielatos, C., & Baker, P. (2008). Fleeing, Sneaking, Flooding: A Corpus Analysis of Discursive Constructions of Refugees and Asylum Seekers in the UK Press, 1996-2005. *Journal of English Linguistics*, 36(1), 5–38. <https://doi.org/10.1177/0075424207311247>
- Haugen, E. (1972). *The Ecology of Language*. Stanford, Calif.: Stanford University Press.
- Johnson, K., & Johnson, H. (Eds.). (1998). *Encyclopedic dictionary of applied linguistics: a handbook for language teaching*. Oxford, UK ; Cambridge, Mass: Blackwell Publishers.
- Leech, G. (1991). The state of the art in corpus linguistics. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics*, 8–29. London, England: Longman.
- Muryantina, R. "When to Say Innalillaahi? Explicating the Meaning of an Indonesian Interjection Borrowed from Arabic." *Academia.edu*. N.p., n.d. Web. 5 Dec. 2015.

<https://www.academia.edu/1786120/When_to_say_Innalillaahi_EYplicating_The_Meaning_of_an_Indonesian_Interjection_Borrowed_from_Arabic>.

Nahdlatul Ulama. Berkley Center for Religion, Peace, and World Affairs. Georgetown University. Berkleycenter.georgetown.edu. Retrieved 2017-03-13.

Parker, I. (1992). *Discourse Dynamics: Critical Analysis for Social and Individual Psychology*. London: Routledge.

Pasha, T. (2011) "Islamists in the Headlines: Critical Discourse Analysis of the Representation of the Muslim Brotherhood in Egyptian Newspapers." Dissertation. Order No. 3487287 The University of Utah, 2011. Ann Arbor: *ProQuest*. Web. 9 Apr. 2016.

Reisigl, M., & Wodak, R. (2001). *Discourse and Discrimination: Rhetorics of Racism and Antisemitism*. London: Routledge.

Sneddon, J. N. (2003). *The Indonesian language: its history and role in modern society*. Sydney, NSW: UNSW Press.

Teeuw, A., & Emanuels, H. W. (1961). *A Critical Survey of Studies on Malay and Bahasa Indonesia: Bibliographical Series 5*. Dordrecht: Springer Netherlands.

UMNO Online. UMNO's Constitution: Goal 3.3. From: "*Archived copy*". *Archived from the original on 29 February 2012*. Accessed January, 2017.

United States. (2015). *International Religious Freedom Report for 2015*. Washington, D.C.: U.S. Dept. of State. Accessed January, 2017.

van Dijk, T. A. (1991). *Racism and the Press: Critical Studies in Racism and Migration*. London; New York: Routledge.

van Dijk, T. A. (1993). Principles of critical discourse analysis. *Discourse & Society*, 4(2), 249–283.

Versteegh, K. (2014). *The Arabic Language* (2nd ed.). Edinburgh: Edinburgh University Press.