

A corpus-based study of sermons to determine the stage of religious development

Raymund T. Palayon, Richard Watson Todd and Sompatu Vungthong
King Mongkut's University of Technology Thonburi

Abstract

Determining the stage of development of a religious group from a corpus linguistics perspective is a research goal that is relatively unexplored. This paper discusses an innovative corpus-based methodology focusing on keyness which serves as a preliminary study to set up the appropriacy of the methodology for further research on determining the stage of development of a cultlike religious group. The paper focuses on the methodological principles including corpora selection based on the stages of religious development (namely mainstream religion, sect, cult, and destructive cult), units of linguistic features for data interpretation, the use of keyness methods and statistic, thresholds, and benchmark corpora preparation for characterizing and categorizing the sermons of a cultlike religious group. These procedures are then used to examine the sermons of a target religious group to determine its stage of development. The criteria we established to address the methodological issues provided us with a systematic process to identify the key linguistic items in the sermons, which allowed us to describe the social characteristics of the stages of religious development as part of the procedures to see the stage of development of a cultlike religious group.

1. Introduction

Religious groups frequently shift and change in their beliefs, philosophies and practices over time (see Blanchard, 2009; Turner, 2013, for the development of Islam; see Conze, 2012, for the development of Buddhism). In Christianity, some religious groups started from being beneficial and transformed into detrimental in the long run (e.g. Palayon, Watson Todd, & Vungthong, in press). These groups are destructive cults whose leaders generally start from mainstream religion. As they develop, they separate themselves and establish their own groups becoming sects. At this stage of development, they remain connected to mainstream religion when it comes to religious philosophy. Later, they develop into cults often viewing the leaders as godlike figures and promoting sets of beliefs which are not associated with the philosophy of mainstream religious groups and sects. In the end, they develop into destructive cults by viewing damaging activities as helpful ways to achieve their goals (see Beaman, 1990; Palayon, Watson Todd, & Vungthong, in press, for further details on mainstream religion, sect, and cult; see Bohm & Alison, 2001; Palayon, Watson Todd, & Vungthong, 2020, for further details on destructive cult). Examples of such destructive cults are Peoples Temple led by Jim Jones (The Jonestown Institute, 2018) and Heaven's Gate led by Marshall Applewhite (Zeller, 2014), whose leaders led their followers to death through committing mass suicide.

Conducting research into religious discourse from a corpus linguistics perspective is a research practice that is underexplored; and, determining the stage of development of a religious group from a corpus linguistics perspective is a research goal that has rarely been attempted. Some religious discourse studies used sermons as corpus characterizing a few elements in the target corpus (e.g. Acheoah & Hamzah, 2015; Agersnap et al., 2020; Esimaje, 2012) and giving supports for critical analyses through patterns of linguistic items (e.g. Mansouri et al., 2017; Salama, 2011); however, none of these studies examined the discourses of cults. Two studies related to cults were conducted and used large data sets and keyness analyses.

These studies focused on the transitions of a cult (Palayon, Watson Todd, & Vungthong, in press) and the linguistic characteristics of destructive cults (Palayon, Watson Todd, & Vungthong, 2020). To the best of our knowledge, the current study is the first research that aims to identify the stage of development of a suspicious religious group from a keyness perspective using large sets of sermons and different keyness methods. This paper discusses the methodological principles that focus on:

- corpora selection based on the stages of religious development namely mainstream religion, sect, cult, and destructive cult;
- units of linguistic features for characterizing keyness of the corpora;
- the use of keyness methods to reveal the key linguistic features;
- the use of keyness statistic to identify the significant linguistic features;
- thresholds for determining the final key linguistic features;
- preparation of the benchmark corpora to be applied for further research on determining the stage of religious development; and,
- characterizing and categorizing procedures to identify the stage of development of a target religious group from a keyness perspective.

Choosing the significant methodological components, identifying the procedures for characterizing and categorizing the data, and knowing their purposes of application enabled us to address the methodological issues which guided us to achieve the goal of the entire study.

2. Stages of religious development

As a starting point in determining the development of a religious world, we identified the stages of religious development based on their social characteristics and we focused on Christianity. We used Holbrook's (2015) 'Extremist Media Index' which comprises three levels allowing us to classify the selected religious groups representing the stages of religious development. These three levels are: (1) moderate level (illustrating conventional ideologies or discourses without endorsement of violence), (2) fringe level (illustrating isolationist ideologies which characterize the transitional point from being moderate to being extreme), and (3) extreme level (illustrating ideologies or discourses which generally promote acts of violence). We were guided by these steps to select the corpora which serve as the sources of linguistic evidence.

In this study, first, we categorized predominant religious groups (e.g. Baptist, Presbyterian, Methodist) as mainstream religion (or in moderate level), those groups that promote sets of doctrines which anchor to the biblical philosophy and are socially beneficial. Second, we categorized those religious groups separated from mainstream religion as sect (or in moderate level), those groups that aim to develop some parts of mainstream cultures but remain connected to mainstream religion philosophically and are socially beneficial. Third, those religious groups separated from mainstream religion and sect which promote belief systems not connected to the philosophy of mainstream religion and sect and view the leaders as godlike figures are categorized as cult (or in fringe level); and fourth, those cults which promote acts of violence with their guiding beliefs are categorized as destructive cult or in extreme level (see Table 1 for more details on the social characteristics of mainstream religion, sect, cult, and destructive cult together with the sample religious groups representing the stages of religious development).

Table 1. Stages of religious development in Christianity from a social-psychological perspective

	<i>1. Mainstream religion</i>	<i>2. Sect</i>	<i>3. Cult</i>	<i>4. Destructive cult</i>
<i>Stages of religious development</i>	*uses biblical concepts as the foundation of its philosophy *promotes biblical philosophy for life development *is socially involved by conducting community services *does not promote the notion of violence	*is an offshoot of mainstream religion *is philosophically connected to mainstream religion *revitalizes some aspects of beliefs and practices of mainstream religion *does not promote the notion of violence	*owns personalized belief systems not traditionally associated with mainstream religion and sect *identifies the leaders as godlike figures *does not promote acts of violence	*is an extreme version of cult *promotes acts of violence (e.g. committing mass suicide and conducting murder)
<i>Religious groups that belong in each stage</i>	*Southern Baptist Convention led by Billy Graham *Saddleback Church led by Rick Warren	*Peoples Temple (1960s to early 1970s) led by Jim Jones *Lakewood Church led by Joel Osteen	*Peoples Temple (1973 to 1976) led by Jim Jones *Divine Truth led by AJ Miller	*Peoples Temple (1977 to 1978) led by Jim Jones *Heaven’s Gate led by Marshall Applewhite in the period leading to mass suicide
<i>References</i>	*Benitez (2015) *Cronshaw (2019) *Holbrook (2015) *Martin (2018) *Tangenberg (2008) *Wallace (1985)	*Abbott (2015) *Barnhart (2013) *Beaman (1990) *Holbrook (2015) *Palayon, Watson Todd, & Vungthong (in press) *Wallace (1985)	*Abbott (2015) *Beaman (1990) *Holbrook (2015) *Palayon, Watson Todd, & Vungthong (in press) *Real Stories (2018) *The Cult Education Institute (2016)	*Abbott (2015) *Bohm & Alison (2001) *Holbrook (2015) *Palayon, Watson Todd, & Vungthong (2020) *Zeller (2014)

3. Corpora selection

After identifying the social characteristics of the stages of religious development, we determined the discourse data coming from mainstream religion, sect, cult, and destructive cult to identify the linguistic features that describe the social characteristics of each stage of religious development, which can be used as benchmark corpora to see the stage of development of a target religious group (for further research). We chose the sermons of the leaders to be the corpora of this study since these texts are likely to illustrate the beliefs and practices of the religious groups as a whole. We compiled: (1) the sermons of Billy Graham (a minister of Southern Baptist Convention) and the sermons of Rick Warren (the leader of Saddleback Church) and combined them into a single corpus representing the stage of mainstream religion, (2) the 1960s to early 1970s sermons or early sermons of Jim Jones (the leader of Peoples Temple) and the sermons of Joel Osteen (the leader of Lakewood Church) and combined them into a single corpus representing the stage of sect, (3) the 1973 to 1976 sermons or the sermons of Jim Jones (the leader of Peoples Temple) from the middle period and the sermons of AJ Miller (the leader of Divine Truth) and combined them into a single corpus representing the stage of cult, and (4) the 1977 to 1978 sermons or late sermons of Jim Jones (the leader of Peoples Temple) and the sermons of Marshall Applewhite (the leader of Heaven’s Gate) in the period leading to mass suicide and combined them into a single corpus representing the stage of destructive cult. These sermons were taken from <https://www.youtube.com/> and transcribed except for the transcribed sermons of Jim Jones which are available at <http://jonestown.sdsu.edu>, and they were converted into text files for comparison.

The current study aims to provide a basis for conducting further research in the future. This future research will examine a religious group that displays cultlike characteristics (e.g. the leader was identified as a godlike figure) or that has raised social concerns. Being able to identify the stage of religious development of such a religious group could have real-world implications. However, to conduct such an identification, benchmark data relating to each of the stages of religious development is needed. Therefore, the four sets of sermons prepared in this paper serve as the benchmark corpora to be used for further study on characterizing and categorizing the sermons of a target religious group to see its stage of development.

Taking a corpus approach specifically keyness approach may allow us to find evidence such as key linguistic features in the sermons characterizing the aboutness and communication style features in the discourse which serve as significant indicators to identify the stage of religious development. Also, this approach is automated and suggests clear procedures which enable us to view and interpret objectively the important features in the large data sets. In this study, since the log-likelihood mean value of each data comparison is critically observed and the log-likelihood values (that show keyness of the data) are corpus-size dependent, the sizes of the five data sets (which are the four sets of sermons in this paper and the sermons of a target cultlike religious group for future study) should be comparable (Pojanapunya, 2017). According to Rayson et al. (2004) cited in O’Halloran (2011) and Pojanapunya (2017), the sizes of the data should not differ by more than a factor of 10. This principle was applied in this study to see that the log-likelihood values across key linguistic feature lists are statistically comparable (see Table 2).

Table 2. Sizes of the corpora

<i>Corpora</i>		<i>Text files</i>	<i>Word tokens each data set</i>	<i>Overall sizes of the data</i>		
				<i>Word tokens</i>	<i>Semantic tag tokens</i>	<i>Part-of-speech tag tokens</i>
Mainstream sermons	Billy Graham’s sermons	19	91,074	189,544	184,990	211,114
	Rick Warren’s sermons	9	98,470	(1.8 x 10 ⁵)	(1.8 x 10 ⁵)	(2.1 x 10 ⁵)
Sect sermons	Jim Jones’ early sermons	14	104,027	195,987	181,089	209,995
	Joel Osteen’s sermons	20	91,960	(1.9 x 10 ⁵)	(1.8 x 10 ⁵)	(2.0 x 10 ⁵)
Cult sermons	Jim Jones’ sermons (mid period)	8	106,344	197,508	182,579	213,003
	AJ Miller’s sermons	8	91,164	(1.9 x 10 ⁵)	(1.8 x 10 ⁵)	(2.1 x 10 ⁵)
Destructive cult sermons	Jim Jones’ late sermons	10	97,246	190,381	175,557	202,178
	M. Applewhite’s sermons	11	93,135	(1.9 x 10 ⁵)	(1.7 x 10 ⁵)	(2.0 x 10 ⁵)
Sermons of a cultlike religious group		approximately 190,000		(1.9 x 10 ⁵)	---	---

4. Linguistic units for characterizing keyness

Keyness is a characteristic of a corpus or a text-dependent quality (Scott, 2010) and this can be viewed from the perspectives of aboutness or content (Hutchins, 1978; Cheng, 2009) and communication style (De Vries, Bakker-Pieper, & Oostenveld, 2010) through linguistic features with markedly higher relative frequencies in a target corpus than in a benchmark corpus. These key linguistic items provide a characterization of the register in a target corpus in terms of its aboutness and communication styles displaying keyness (Bondi, 2010; Gerbig, 2010; Scharl & Weichselbraun, 2008). From a keyness perspective, having different levels of linguistic features to be analyzed tends to provide greater insights into different aspects of the register of a target corpus.

To view the aboutness and communication styles which display the keyness of sermons, we need to identify the units of linguistic features to be analyzed. Thus, we focused on keywords (or at the lexical level) which provide the aboutness and communication style information, key semantic tags (or at the semantic level) which more likely display the aboutness information, and key part-of-speech tags (or at the grammatical level) which exhibit the communication style information. Keywords are lexical items (Baker, 2004; Pojanapunya & Watson Todd, 2018), key semantic tags are semantic groups of words (Piao, Rayson, Archer, & McEnery, 2005), and key part-of-speech tags are grammatical groups of words with higher relative frequencies in the target corpus when compared to the benchmark corpus (Culpeper, 2009; Palayon, Watson Todd, & Vungthong, 2020). Keywords show the main findings on aboutness and communication styles, while key semantic tags and key part-of-speech tags confirm the keyword findings and highlight other features when words are grouped together based on the semantic and grammatical relationships which are considered key. These linguistic features can be identified using keyness methods which are presented in the next section.

5. Keyness analyses

The keyness methods applied in this study are keyword analysis, key semantic tag analysis, and key part-of-speech analysis to examine the aboutness and communication style features in the sermons. These methods were applied since they can quantify the key linguistic features which are useful for characterization and categorization from the aboutness and communication style perspectives. Keyword analysis reveals the important lexical items in the target corpora using AntConc 3.4.4 (Anthony, 2014), key semantic tag analysis shows the meaningful semantic tags in the target corpora using the UCREL semantic tagger (Archer, Rayson, Piao, & McEnery, 2004) and AntConc 3.4.4, and key part-of-speech analysis displays the significant grammatical tags in the target corpora using the Multidimensional Analysis Tagger (MAT) 1.3 (Nini, 2015) and AntConc 3.4.4 (see Palayon, Watson Todd, & Vungthong, 2020, for more details on these methods).

5.1 Keyness statistic

In corpus linguistics studies, the use of statistic to measure the keyness of linguistic items in a target corpus depends upon the purpose of the study (Gabrielatos, 2018; Pojanapunya & Watson Todd, 2018). For studies focusing on genre and register, log-likelihood (LL) is normally applied which allows the analysts to generally view the issues on similarity and difference between a target corpus and a benchmark corpus; and for critically-oriented studies, odds ratio (OR) is normally used which allows the analysts to conduct in-depth examination of key features (Lien, 2021; Pojanapunya & Watson Todd, 2018). In this study in which the goal is more on register-style characterization and categorization of the data, LL is more applicable than other statistical metrics since it allows us to describe and classify the sermons based on the aboutness and communication styles revealed by the key linguistic features which serve as our ways to identify the stage of development of a target religious group (Palayon, Watson Todd, & Vungthong, in press; Rayson & Garside, 2000). Although LL has been criticized theoretically by Gabrielatos (2018) regarding the sensitivity of p-values to item frequency and the sizes of the corpora, the comprehensive use of LL and its applicability to studies from genre and register perspectives make this statistic suitable for this research.

5.2 Thresholds

After the corpus tools revealed the lists of key linguistic features, we need to identify the final linguistic items in the lists. Thus, establishing criteria for identifying the thresholds should be considered since the quality of choices in selecting the features determines the quantity of features to be used for data interpretation which might also affect the quality of findings (Pojanapunya & Lieungnappar, 2017). A sufficient number of key items need to be identified which cover the important aspects in the target corpus for characterization and categorization.

There are three possible choices in setting up the thresholds, namely, a cutoff LL value or its associated probability value (e.g. Esimaje, 2012), the Top N method (e.g. Palayon, Watson Todd, & Vungthong, 2020), and a proportion of the range of LL values that can be identified by using a percentage or a z-score (e.g. Pojanapunya & Watson Todd, 2021). The main problem with the first choice is that the LL values are heavily influenced by the sizes of the corpora being compared. Therefore, the LL value is problematic when it is used as a cutoff point, and this method is not appropriate in this study (see Gabrielatos, 2018, for further details). The Top N method may be applicable in this study; however, it is unclear how to identify the N value and the way it is identified may involve subjective decisions. Thus, we applied the third choice which is a z-score cutoff point since it does not require subjective decisions in identifying the final items. This method provides us with straightforward and reliable indicators in identifying the thresholds as one of the fundamental steps for the entire analyses (see Palayon, Watson Todd, & Vungthong, in press; Pojanapunya & Watson Todd, 2021, for more details on a z-score cutoff point).

Since the lists of the three analyses show different sizes of the total numbers of items wherein the total numbers of word types ranged from 7,628 to 8,882, the total numbers of semantic tag types ranged from 334 to 347, and the total numbers of part-of-speech tag types ranged from 88 to 89, we applied different z-scores to have a sufficient number of key items to examine the keyness of sermons. For keyword analysis, we used a z-score of 3 that covers a portion of keywords. For key semantic tag analysis, we used a z-score of 2 that covers a portion of key semantic tags. And, for key part-of-speech analysis, we used a z-score of 1 that covers a portion of key part-of-speech tags. Therefore, any word items with a z-score greater than 3 were considered keywords, any semantic tags with a z-score greater than 2 were considered key semantic tags, and any part-of-speech tags with a z-score greater than 1 were considered key part-of-speech tags. Presented in Table 3 are the total numbers of final key linguistic items to be used for characterizing keyness of the benchmark corpora.

Table 3. Total numbers of final key linguistic items from different comparisons

<i>Comparisons</i>	<i>Keyword lists</i>	<i>Key semantic tag lists</i>	<i>Key part-of-speech tag lists</i>
Mainstream sermons vs. Sect sermons	41	7	3
Mainstream sermons vs. Cult sermons	35	3	4
Mainstream sermons vs. Destructive cult sermons	25	5	2
Sect sermons vs. Mainstream sermons	35	9	4
Sect sermons vs. Cult sermons	38	7	4
Sect sermons vs. Destructive cult sermons	27	3	3
Cult sermons vs. Mainstream sermons	27	5	3
Cult sermons vs. Sect sermons	41	7	3
Cult sermons vs. Destructive cult sermons	34	4	3
Destructive cult sermons vs. Mainstream sermons	18	2	1
Destructive cult sermons vs. Sect sermons	29	5	3
Destructive cult sermons vs. Cult sermons	31	6	8

6. Preparation of the benchmark corpora

Having all fundamental methodological components (such as corpora, linguistic units, methods, statistic, and thresholds), we need to characterize linguistically first the four sets of sermons to verify their social characteristics before using them as the benchmark corpora for further research on determining the stage of development of a cultlike religious group. Following the principle of Pojanapunya and Watson Todd (2021) on analyzing features from multiple comparisons for register characterization, we compared the mainstream sermons, sect sermons, cult sermons, and destructive cult sermons with each other in multiple ways to identify the important linguistic features in each data set (see Table 4). This method of comparisons may provide patterns of key linguistic items which indicate the reliability of information to describe the language in the sermons. After the lists of items were revealed, we applied a z-score cutoff point to determine the proportions in the lists to be the final items for data interpretation. Lastly, we calculated the LL mean value of the final items in each list which allows us to see the similarity and difference across multiple comparisons.

Table 4. Multiple comparisons of the data

<i>Target corpora</i>	<i>Benchmark corpora</i>		
1. Mainstream sermons	vs. Sect sermons	vs. Cult sermons	vs. Destructive cult sermons
2. Sect sermons	vs. Mainstream sermons	vs. Cult sermons	vs. Destructive cult sermons
3. Cult sermons	vs. Mainstream sermons	vs. Sect sermons	vs. Destructive cult sermons
4. Destructive cult sermons	vs. Mainstream sermons	vs. Sect sermons	vs. Cult sermons

6.1 Mainstream sermons

The sermons of Billy Graham and the sermons of Rick Warren display common aboutness features that generally anchor to the religious philosophy found in the Bible (e.g. God, Christ, Bible, prayer in Table 5; S9 Religion and the supernatural, Z1 Personal names in Table 6) and associate with life development topics (e.g. life, goodness, heart in Table 5; L1+ Alive, E1 Emotional actions, states and processes general, X7+ Wanted in Table 6). They promote biblical topics with an aim to develop positive attitudes and views to make the members as productive individuals. They also contain common communication styles of personal involvement generally involving the audience, the presence of the discourses, and Jesus or God through the frequent use of personal pronouns (e.g. he and my in Table 5; SPP2 Second person pronouns in Table 7) and elaborating style through the frequent use of phrasal coordination (e.g. and in Table 5; PHC Phrasal coordination in Table 7), verb phrases (e.g. says, ask, loves in Table 5; VPRT Present tense, SUAV Suasive verbs in Table 7), and noun phrases (e.g. NN Total other nouns in Table 7) which allows the religious leaders to present the general goal of their discourses (on strengthening life spiritually and physically) and enables the audience to see the purpose of religious philosophy.

Table 5. Keywords in the mainstream sermons

Rank	<i>Mainstream sermons</i>					
	<i>vs. Sect sermons</i>		<i>vs. Cult sermons</i>		<i>vs. Destructive cult sermons</i>	
	<i>Keywords</i>	<i>LL</i>	<i>Keywords</i>	<i>LL</i>	<i>Keywords</i>	<i>LL</i>
1	and	801.1	your	707.5	god	2329.5
2	says	286.3	he	657.4	you	1105.7
3	gonna	234.8	christ	601.9	your	886.4
4	pray	232.8	god	572.1	christ	660.2
5	bible	225.4	says	547.3	life	466.0
6	prayer	199.0	you	541.4	bible	459.8
7	christ	191.0	pray	413.2	says	429.3
8	wanna	163.8	jesus	406.3	pray	420.7
9	cross	154.4	lord	362.2	lord	388.6
10	ask	142.2	prayer	301.2	he	373.0
11	daniel	137.8	life	291.9	prayer	329.1
12	sin	127.8	and	285.9	jesus	291.9
13	emotions	126.4	gonna	243.0	and	261.4
14	life	125.2	his	235.2	heart	240.9
15	then	104.5	bible	192.4	gonna	205.9
16	christian	99.3	cross	182.4	verse	191.9
17	judgment	96.1	daniel	168.5	day	186.6
18	world	94.5	praying	156.4	sins	165.7
19	verse	92.5	mercy	138.3	daniel	164.1
20	okay	90.1	wanna	124.5	mercy	141.6
21	breakthrough	83.6	heaven	121.5	praying	134.4
22	praying	82.3	the	117.2	cross	128.3
23	birthright	80.9	breakthrough	115.7	church	125.8
24	sins	80.0	scripture	114.6	sin	125.7
25	really	78.2	in	113.9	scripture	121.4
26	emotion	77.7	verse	106.8		
27	or	73.4	come	106.1		
28	worry	68.4	goodness	91.5		
29	number	66.7	ask	87.9		
30	this	63.3	may	87.4		
31	screen	62.5	heart	87.2		
32	rest	61.7	my	84.6		
33	first	60.8	him	81.8		
34	day	60.1	birthright	81.4		
35	loves	59.3	man	79.8		
36	goodness	58.0				
37	because	55.0				
38	esau	54.0				
39	choice	53.3				
40	grace	53.2				
41	repentance	52.5				
	<i>LL mean</i>	<i>100.1</i>	<i>LL mean</i>	<i>188.6</i>	<i>LL mean</i>	<i>293.3</i>

Table 6. Key semantic tags in the mainstream sermons

Rank	Mainstream sermons					
	vs. Sect sermons		vs. Cult sermons		vs. Destructive cult sermons	
	Key semantic tags	LL	Key semantic tags	LL	Key semantic tags	LL
1	E1 Emotional actions, states and processes general (e.g. self-control, compassion, emotion)	205.6	S9 Religion and the supernatural (e.g. soul, heaven, spirit)	843.3	S9 Religion and the supernatural (e.g. soul, heaven, spirit)	1257.4
2	G2.2- Unethical (e.g. sinful, cheating, defile)	181.1	Z1 Personal names (e.g. Jesus, Daniel, John)	272.6	L1+ Alive (e.g. life, alive, live)	369.0
3	L1+ Alive (e.g. life, alive, live)	145.2	L1+ Alive (e.g. life, alive, live)	257.3	G2.2+ Lawful (e.g. moral, mercy, forgiveness)	202.0
4	X7+ Wanted (e.g. plans, purpose, choice)	143.7			G2.2- Unethical (e.g. sinful, cheating, defile)	181.7
5	N1 Numbers (e.g. three, millions, thousands)	139.0			E1 Emotional actions, states and processes general (e.g. self-control, compassion, emotion)	171.7
6	S9 Religion and the supernatural (e.g. soul, heaven, spirit)	118.1				
7	G2.2+ Lawful (e.g. moral, mercy, forgiveness)	111.7				
	<i>LL mean</i>	<i>146.1</i>	<i>LL mean</i>	<i>389.6</i>	<i>LL mean</i>	<i>311.1</i>

Table 7. Key part-of-speech tags in the mainstream sermons

Rank	Mainstream sermons					
	vs. Sect sermons		vs. Cult sermons		vs. Destructive cult sermons	
	Key part-of-speech tags	LL	Key part-of-speech tags	LL	Key part-of-speech tags	LL
1	VPRT Present tense (e.g. is, are, say)	237.2	SPP2 Second person pronouns (e.g. you, your, thou)	953.5	SPP2 Second person pronouns (e.g. you, your, thou)	1694.3
2	SUAV Suasive verbs (e.g. ask, pray, requires)	155.3	NN Total other nouns (e.g. Bible, Jesus, Christians)	412.3	VPRT Present tense (e.g. is, are, say)	253.1
3	PHC Phrasal coordination (e.g. and)	143.9	SUAV Suasive verbs (e.g. ask, pray, requires)	249.0		
4			PUBV Public verbs (e.g. say, confess, acknowledge)	198.1		
	<i>LL mean</i>	<i>174.4</i>	<i>LL mean</i>	<i>373.2</i>	<i>LL mean</i>	<i>654.8</i>

6.2. Sect sermons

The early sermons of Jim Jones and the sermons of Joel Osteen exhibit aboutness features such as religious philosophy anchored to the biblical concepts (e.g. Almighty, Holy Ghost, Jesus, Abraham, scripture in Table 8; S9 Religion and the supernatural, S7.1+ In power, Z3 Other proper names in Table 9), life and well-being concepts (e.g. healed, healing, yes indicating positivity, joy, thought in Table 8; B2- Disease, B1 Anatomy and physiology, B3 Medicines and medical treatment in Table 9), and communication styles of personal involvement through the frequent use of personal pronouns (e.g. she, he, you, you in Table 8; SPP2 Second person pronouns and TPP3 Third person pronouns in Table 10) and elaborating style through the frequent use of phrases such as verb phrases (e.g. VBD Past tense in Table 10) and noun phrases (e.g. NN Total other nouns in Table 10), which associate with the mainstream sermons. This may happen since sect is philosophically connected to mainstream religion (Beaman, 1990). However, the sect sermons frequently employ negating style (e.g. t, didn't, not, never in Table 8) which generally restricts individuals to avoid committing sins. This implies that sect is generally more restrictive than mainstream religion when it comes to behavior.

Table 8. Keywords in the sect sermons

Rank	Sect sermons					
	vs. Mainstream sermons		vs. Cult sermons		vs. Destructive cult sermons	
	Keywords	LL	Keywords	LL	Keywords	LL
1	she	468.5	he	538.4	god	1931.4
2	her	322.5	your	447.6	you	786.8
3	yes	265.1	god	362.2	your	598.4
4	sister	194.8	you	320.8	he	283.5
5	uh	170.9	his	268.4	yes	262.0
6	didn	130.5	jesus	220.3	faith	216.5
7	almighty	128.0	lord	189.8	lord	213.4
8	ghost	120.4	name	180.1	spirit	212.5
9	favor	100.1	him	152.9	shall	176.8
10	healing	94.4	christ	135.8	christ	175.2
11	thought	93.2	may	134.4	heart	147.8
12	out	88.3	david	133.7	sister	143.2
13	scars	88.3	destiny	126.9	bles	138.5
14	tongues	86.6	was	118.9	jesus	137.5
15	t	79.9	yes	118.6	blessed	132.6
16	hallelujah	76.4	blessed	117.6	she	121.9
17	keep	75.1	come	116.5	life	119.8
18	would	74.5	scripture	106.1	said	118.5
19	clasped	73.1	almighty	104.5	almighty	117.8
20	joel	70.2	scars	100.4	name	115.3
21	move	67.2	favor	97.8	scripture	112.9
22	able	66.0	abraham	93.8	abraham	112.7
23	healed	64.2	thought	93.4	my	112.4
24	blessed	62.9	keep	90.9	favor	107.2
25	cancer	62.4	mind	90.3	destiny	107.0
26	was	60.8	joel	87.8	holy	101.0
27	see	59.6	blessing	84.7	ghost	98.6
28	scar	58.2	ghost	83.6		
29	upset	58.2	joy	81.1		
30	got	58.0	my	80.9		
31	lady	56.3	didn	80.0		
32	joseph	54.8	never	79.9		

33	addiction	54.0	promise	76.7		
34	back	53.9	tongues	74.1		
35	pentecost	52.7	father	73.3		
36			thank	72.7		
37			dreams	72.5		
38			discouraged	66.9		
	<i>LL mean</i>	87.8	<i>LL mean</i>	121.3	<i>LL mean</i>	175.9

Table 9. Key semantic tags in the sect sermons

Rank	<i>Sect sermons</i>					
	<i>vs. Mainstream sermons</i>		<i>vs. Cult sermons</i>		<i>vs. Destructive cult sermons</i>	
	<i>Key semantic tags</i>	<i>LL</i>	<i>Key semantic tags</i>	<i>LL</i>	<i>Key semantic tags</i>	<i>LL</i>
1	Z3 Other proper names (e.g. Ephesians, Ghost, Christ)	254.9	Z1 Personal names (e.g. Luke, John, Matthew)	432.7	S9 Religion and the supernatural (e.g. heaven, pray, salvation)	611.1
2	B2- Disease (e.g. sick, addiction, illness)	235.6	S9 Religion and the supernatural (e.g. heaven, pray, salvation)	329.0	E6+ Confident (e.g. confidence, faith, peace)	180.6
3	B3 Medicines and medical treatment (e.g. heal, healed, healing)	211.0	M1 Moving, coming and going (e.g. go, pass, getting)	269.5	B1 Anatomy and physiology (e.g. physical, heart, arm)	131.1
4	M1 Moving, coming and going (e.g. go, pass, getting)	161.9	Z3 Other proper names (e.g. Ephesians, Ghost, Christ)	261.6		
5	A9+ Getting and possession (e.g. belong, keep, receive)	150.5	S4 Kin (e.g. brother, sister, family)	165.7		
6	B5 Clothes and personal belongings (e.g. wear, yoke, robe)	138.9	B1 Anatomy and physiology (e.g. physical, heart, arm)	159.7		
7	M2 Putting, pulling, pushing, transporting (e.g. bring, send, raise)	123.3	S7.1+ In power (e.g. Almighty, King, Lord)	139.4		
8	B1 Anatomy and physiology (e.g. physical, heart, arm)	121.9				
9	X4.1 Mental object: Conceptual object (e.g. principle, vision, dream)	108.5				
	<i>LL mean</i>	160.3	<i>LL mean</i>	232.8	<i>LL mean</i>	243.7

Table 10. Key part-of-speech tags in the sect sermons

Rank	<i>Sect sermons</i>					
	<i>vs. Mainstream sermons</i>		<i>vs. Cult sermons</i>		<i>vs. Destructive cult sermons</i>	
	<i>Key part-of-speech tags</i>	<i>LL</i>	<i>Key part-of-speech tags</i>	<i>LL</i>	<i>Key part-of-speech tags</i>	<i>LL</i>
1	VBD Past tense (e.g. heard, said, came)	309.0	NN Total other nouns (e.g. God, Bible, life)	780.6	SPP2 Second person pronouns (e.g. she, he, them)	1278.8
2	TPP3 Third person pronouns (e.g. she, he, them)	231.9	SPP2 Second person pronouns (e.g. you, your, yourself)	643.5	NN Total other nouns (e.g. God, Bible, life)	410.6
3	PRMD Predictive modals (e.g. will, would, shall)	139.5	VBD Past tense (e.g. heard, said, came)	550.0	VBD Past tense (e.g. heard, said, came)	252.2
4	GER Gerunds (e.g. healing, praying, saying)	134.4	TPP3 Third person pronouns (e.g. she, he, them)	265.1		
	<i>LL mean</i>	<i>191.5</i>	<i>LL mean</i>	<i>520.2</i>	<i>LL mean</i>	<i>509.7</i>

6.3 Cult sermons

The sermons of Jim Jones from the middle period and the sermons of AJ Miller contain dual aboutness features since they contain non-religious and religious features. The cult sermons promote belief systems which are not traditionally associated with the philosophy of mainstream religion and sect. Jim Jones’ sermons promote sociopolitical concepts (e.g. socialist, laws, organization in Table 11; G1.2 Politics, G2.1 Law and order in Table 12) with the integration of some religious concepts (e.g. God, faith, church, spirit, Bible in Table 11), while AJ Miller’s sermons generally uphold psychological concepts (e.g. emotions, feeling, conscience in Table 11; E1 Emotional actions, states and processes general, X2.1 Thought, belief in Table 12) with the integration of some religious concepts (e.g. sins, sinning, God in Table 11). Their discourses contain communication styles which are not frequent in the mainstream sermons, for instance, othering style (e.g. the frequent use of we and they in Table 11) that tends to view non-members as outsiders, and intensifying style (e.g. the frequent use of so, really, actually, obviously, and probably in Table 11; RB Total adverbs and EMPH Emphatics in Table 13) that tends to magnify the content elements in their sermons. They also contain elaboration style through the frequent use of phrases such as verb phrases (e.g. VPRT Present tense in Table 13), pronoun it (e.g. PIT Pronoun it in Table 13), and causative adverbial subordinators (in Table 13) that tends to expound generally the content features or communication goals of their sermons, and negating style (e.g. t and isn’t in Table 11) through the frequent use of negation guiding the members to see the purposes of their beliefs.

Table 11. Keywords in the cult sermons

Rank	<i>Cult sermons</i>					
	<i>vs. Mainstream sermons</i>		<i>vs. Sect sermons</i>		<i>vs. Destructive cult sermons</i>	
	<i>Keywords</i>	<i>LL</i>	<i>Keywords</i>	<i>LL</i>	<i>Keywords</i>	<i>LL</i>
1	we	504.0	sin	872.3	sin	855.5
2	sin	380.5	so	537.6	god	707.9
3	so	320.6	we	536.2	yeah	392.5
4	yeah	318.0	conscience	345.5	conscience	358.3
5	her	284.8	yeah	268.6	love	198.7
6	she	248.3	that	234.0	right	187.9
7	conscience	237.2	really	214.8	faith	156.9
8	truth	215.5	truth	189.6	so	147.7
9	they	187.9	because	173.9	s	121.3
10	right	167.8	or	165.8	you	108.2
11	that	151.9	s	139.2	desire	107.7
12	got	128.4	and	134.5	truth	105.1
13	desire	112.7	actually	130.6	harmony	99.8
14	sister	111.7	know	129.0	laws	95.7
15	cause	111.5	desire	123.2	law	94.7
16	d	98.2	then	122.6	church	90.5
17	uh	97.0	lot	115.9	spirit	88.4
18	process	96.5	obviously	115.8	problem	87.9
19	see	95.0	organization	95.6	actually	85.2
20	condition	94.1	about	94.7	why	83.8
21	harmony	88.7	laws	94.4	attitudes	83.7
22	t	88.5	action	92.8	good	83.0
23	it	84.5	harmony	91.4	loving	82.6
24	laws	84.1	process	85.0	bible	80.3
25	like	83.5	sinning	84.1	feeling	77.1
26	organization	78.0	isn	82.9	emotion	74.8
27	socialist	75.4	emotionally	82.7	about	74.6
28			quite	80.5	people	72.4
29			pretty	78.8	vision	71.5
30			attitudes	76.6	sister	70.0
31			emotion	76.6	sinning	67.5
32			also	75.7	sins	67.4
33			question	75.4	her	66.4
34			it	73.4	person	66.3
35			law	72.8	emotionally	66.2
36			moral	71.2		
37			probably	70.8		
38			feeling	70.3		
39			condition	65.6		
40			thing	65.1		
41			questions	64.6		
	<i>LL mean</i>	<i>142.3</i>	<i>LL mean</i>	<i>119.8</i>	<i>LL mean</i>	<i>111.7</i>

Table 12. Key semantic tags in the cult sermons

Rank	<i>Cult sermons</i>					
	<i>vs. Mainstream sermons</i>		<i>vs. Sect sermons</i>		<i>vs. Destructive cult sermons</i>	
	<i>Key semantic tags</i>	<i>LL</i>	<i>Key semantic tags</i>	<i>LL</i>	<i>Key semantic tags</i>	<i>LL</i>
1	A1.1.1 General actions/ making (e.g. close, create, make)	195.3	G2.2- Unethical (e.g. sin, sinful, wrong)	689.8	G2.2- Unethical (e.g. sin, sinful, wrong)	683.6
2	G1.2 Politics (e.g. communists, fascist, revolution)	186.9	A2.2 Cause & Effect/ Connection (e.g. effect, get, connected)	292.5	E1 Emotional actions, states and processes general (e.g. emotions, emotional, emotionally)	145.5
3	G2.2- Unethical (e.g. sin, sinful, wrong)	178.7	A5.2+ Evaluation: True (e.g. truly, fact, truthful)	257.2	X2.1 Thought, belief (e.g. attitudes, feelings, conscience)	125.1
4	A2.2 Cause & Effect/ Connection (e.g. effect, get, connected)	173.6	G1.2 Politics (e.g. communists, fascist, revolution)	216.6	S3.2 Relationship: Intimacy and sex (e.g. love, sexually, homosexual)	121.6
5	G2.1 Law and order (e.g. penalty, laws, punished)	165.7	E1 Emotional actions, states and processes general (e.g. emotions, emotional, emotionally)	176.9		
6			G2.1 Law and order (e.g. penalty, laws, punished)	168.3		
7			X7+ Wanted (e.g. choice, desire, will)	142.7		
	<i>LL mean</i>	<i>179.8</i>	<i>LL mean</i>	<i>241.4</i>	<i>LL mean</i>	<i>197.2</i>

Table 13. Key part-of-speech tags in the cult sermons

Rank	<i>Cult sermons</i>					
	<i>vs. Mainstream sermons</i>		<i>vs. Sect sermons</i>		<i>vs. Destructive cult sermons</i>	
	<i>Key part-of-speech tags</i>	<i>LL</i>	<i>Key part-of-speech tags</i>	<i>LL</i>	<i>Key part-of-speech tags</i>	<i>LL</i>
1	RB Total adverbs (e.g. so, constantly, exactly)	217.6	RB Total adverbs (e.g. so, constantly, exactly)	386.1	VPRT Present tense (e.g. is, are, see)	136.8
2	PIT Pronoun it	126.4	EMPH Emphatics (e.g. more, really, most)	175.6	SPP2 Second person pronouns (e.g. you, your, yourself)	118.6
3	PRMD Predictive modals (e.g. will, would, shall)	125.1	CAUS Causative adverbial subordinators (e.g. because)	171.0	RB Total adverbs (e.g. so, constantly, exactly)	103.7
	<i>LL mean</i>	<i>151.0</i>	<i>LL mean</i>	<i>226.3</i>	<i>LL mean</i>	<i>118.9</i>

6.4 Destructive cult sermons

The late sermons of Jim Jones and the sermons of Marshall Applewhite in the period leading to mass suicide contain features which are not traditionally associated with the mainstream sermons and sect sermons (e.g. soviet, shit, overcoming, information, vehicle in Table 14; G1.2 Politics, G1.1 Government, Z2 Geographical names, N3.7 Measurement: Length & height, P1 Education in general in Table 15), and these findings confirm the previous findings of Palayon, Watson Todd, and Vungthong (2020) on the aspects of language of destructive cults. Jim Jones' sermons heavily focus on sociopolitical concepts with an emphasis on the concept of revolutionary suicide, and Marshall Applewhite's sermons heavily focus on science-fiction stories with an emphasis on entering the heavenly kingdom (or outer space) by committing mass suicide. Their sermons contain unconventional communication styles such as othering style (e.g. through the frequent use of we and they in Table 14; TPP3 Third person pronouns in Table 16), intensifying style through the frequent use of intensifiers (in Table 14), and excessive elaboration through the frequent use of that-clauses (e.g. that in Table 14; THVC That verb complements in Table 16), demonstratives (e.g. DEMO Demonstratives in Table 16), noun phrases (e.g. NN Total other nouns and NOMZ Nominalizations in Table 16), prepositional phrases (e.g. PIN Total prepositional phrases in Table 16), and verb phrases (e.g. VBD Past tense, PEAS Perfect aspect in Table 16) while leading their followers to committing mass suicide (see Palayon, Watson Todd, & Vungthong, 2020, for further information on the language of destructive cults).

Table 14. Keywords in the destructive cult sermons

Rank	<i>Destructive cult sermons</i>					
	<i>vs. Mainstream sermons</i>		<i>vs. Sect sermons</i>		<i>vs. Cult sermons</i>	
	<i>Keywords</i>	<i>LL</i>	<i>Keywords</i>	<i>LL</i>	<i>Keywords</i>	<i>LL</i>
1	kingdom	685.7	kingdom	968.2	kingdom	1053.5
2	that	461.5	that	600.4	human	346.9
3	they	342.8	human	422.6	father	325.5
4	human	316.0	we	297.2	level	178.3
5	we	275.4	or	285.7	overcoming	169.4
6	shit	193.5	shit	198.2	heaven	165.0
7	vehicle	190.7	vehicle	184.8	vehicle	159.8
8	level	178.6	soviet	169.9	ti	145.2
9	information	157.6	then	154.4	soviet	140.9
10	overcoming	146.8	overcoming	150.8	heavenly	115.2
11	ti	141.0	information	149.4	overcome	104.3
12	their	139.0	they	149.4	usa	97.5
13	soviet	130.8	and	145.9	member	96.1
14	our	122.5	might	137.1	certainly	93.8
15	she	121.9	member	134.7	instruction	92.5
16	father	119.3	ti	127.3	union	91.7
17	would	109.3	so	121.9	that	88.3
18	certainly	108.1	because	116.6	students	82.1
19			even	112.7	crew	79.3
20			certainly	109.8	representative	76.9
21			union	106.7	classroom	73.2
22			usa	106.2	lucy	72.6
23			humans	104.7	goddamn	72.4
24			level	104.5	information	72.1
25			session	103.3	his	71.8
26			father	97.2	older	70.1
27			our	96.7	task	68.0
28			really	92.8	shit	67.7

29		as	91.8	as	66.1
30				stoen	64.0
31				age	63.0
	<i>LL mean</i>	<i>186.9</i>	<i>LL mean</i>	<i>156.3</i>	<i>LL mean</i>
				<i>108.7</i>	

Table 15. Key semantic tags in the destructive cult sermons

Rank	<i>Destructive cult sermons</i>					
	<i>vs. Mainstream sermons</i>		<i>vs. Sect sermons</i>		<i>vs. Cult sermons</i>	
	<i>Key semantic tags</i>	<i>LL</i>	<i>Key semantic tags</i>	<i>LL</i>	<i>Key semantic tags</i>	<i>LL</i>
1	G1.2 Politics (e.g. communists, capitalism, election)	416.5	G1.1 Government (e.g. kingdom, minister, nation)	496.8	G1.1 Government (e.g. kingdom, minister, nation)	380.9
2	G1.1 Government (e.g. kingdom, minister, nation)	332.2	G1.2 Politics (e.g. communists, capitalism, election)	455.3	P1 Education in general (e.g. class, lesson, study)	274.0
3			Z2 Geographical names (e.g. Africa, America, China)	331.6	Z1 Personal names (e.g. Lucy, Mao, Stoen)	182.5
4			P1 Education in general (e.g. class, lesson, study)	209.0	Z2 Geographical names (e.g. Africa, America, China)	176.5
5			G3 Warfare, defence and the army; weapons (e.g. war, bombs, weapons)	161.7	N3.7 Measurement: Length & height (e.g. elevation, level, levels)	166.7
6					G3 Warfare, defence and the army; weapons (e.g. war, bombs, weapons)	139.2
	<i>LL mean</i>	<i>371.9</i>	<i>LL mean</i>	<i>302.5</i>	<i>LL mean</i>	<i>206.7</i>

Table 16. Key part-of-speech tags in the destructive cult sermons

Rank	<i>Destructive cult sermons</i>					
	<i>vs. Mainstream sermons</i>		<i>vs. Sect sermons</i>		<i>vs. Cult sermons</i>	
	<i>Key part-of-speech tag</i>	<i>LL</i>	<i>Key part-of-speech tag</i>	<i>LL</i>	<i>Key part-of-speech tag</i>	<i>LL</i>
1	DEMO Demonstratives (e.g. that, this, those)	177.6	DEMO Demonstratives (e.g. that, this, those)	260.9	PIN Total prepositional phrases (e.g. of, for, in)	77.9
2			THVC That verb complements (e.g. that)	183.3	PEAS Perfect aspect (e.g. has, have, had)	62.7
3			NOMZ Nominalizations (e.g. agreement, revolution, action)	159.0	NN Total other nouns (e.g. communism, suicide, space)	53.9
4					VBD Past tense (e.g. saw, sent, were)	53.4
5					TPP3 Third person pronouns (e.g. they, them, their)	53.1
6					POMD Possibility modals (e.g. can, could, may)	52.7
7					THVC That verb complements (e.g. that)	46.7
8					DEMO Demonstratives (e.g. that, this, those)	43.8
	<i>LL mean</i>	<i>177.6</i>	<i>LL mean</i>	<i>196.6</i>	<i>LL mean</i>	<i>54.7</i>

Viewing the LL mean values in all comparisons allows us to see the similarities and differences between the target corpora and the benchmark corpora (see Table 17). Lower LL mean values display similarity, while higher LL mean values show difference. From the results, the values in the *mainstream sermons vs. sect sermons* and the values in the *sect sermons vs. mainstream sermons* are lower when compared to the values in the *mainstream sermons vs. cult sermons*, *mainstream sermons vs. destructive cult sermons*, *sect sermons vs. cult sermons*, and *sect sermons vs. destructive cult sermons* which suggest that the mainstream sermons and sect sermons are related to each other. The values in the *mainstream sermons vs. cult sermons*, *mainstream sermons vs. destructive cult sermons*, *sect sermons vs. cult sermons*, and *sect sermons vs. destructive cult sermons* are higher when compared to the values in the *mainstream sermons vs. sect sermons* and *sect sermons vs. mainstream sermons* which suggest that the mainstream sermons and sect sermons are different when compared to the cult sermons and destructive cult sermons.

The values in the *destructive cult sermons vs. cult sermons* are lower when compared to the values in the *destructive cult sermons vs. mainstream sermons* and *destructive cult sermons vs. sect sermons*, which suggest that the destructive cult sermons may be similar with the cult sermons in some way (e.g. non-religious elements, othering style, and intensifying style). The values in the *destructive cult sermons vs. mainstream sermons* and *destructive cult sermons vs. sect sermons* are higher when compared to the values in the *destructive cult sermons vs. cult sermons* which suggest that the destructive cult sermons are different when compared to the mainstream sermons and sect sermons. The values in the cult sermons comparisons show an unpredictable pattern which suggests that cult is a transitional stage (Palayon, Watson Todd, & Vungthong, in press).

Table 17. LL mean values of all data comparisons

<i>Data comparisons</i>	<i>Keywords</i>	<i>Key semantic tags</i>	<i>Key part-of-speech tags</i>
Mainstream sermons vs. Sect sermons	100.1	146.1	174.4
Mainstream sermons vs. Cult sermons	188.6	389.6	373.2
Mainstream sermons vs. Destructive cult sermons	293.3	311.3	654.8
Sect sermons vs. Mainstream sermons	87.8	160.3	191.5
Sect sermons vs. Cult sermons	121.3	232.8	520.2
Sect sermons vs. Destructive cult sermons	175.9	243.7	509.7
Cult sermons vs. Mainstream sermons	142.3	179.8	151.0
Cult sermons vs. Sect sermons	119.8	241.4	226.3
Cult sermons vs. Destructive cult sermons	111.7	197.2	118.9
Destructive cult sermons vs. Mainstream sermons	186.9	371.9	177.6
Destructive cult sermons vs. Sect sermons	156.3	302.5	196.6
Destructive cult sermons vs. Cult sermons	108.7	206.7	54.7

7. Determining the stage of development of a religious group

After displaying the keyness of the mainstream sermons, sect sermons, cult sermons, and destructive cult sermons, these four data sets will be applied in future research aiming to determine the stage of development of a cultlike religious group. To see the stage of development of a target religious group, we need to characterize the sermons of its leader linguistically to identify the aboutness and communication style features which allow us to categorize the target sermons based on the benchmark corpora. Thus, the sermons of a target religious group will be compared to the mainstream sermons, sect sermons, cult sermons, and destructive cult sermons using keyword analysis, key semantic tag analysis, and key part-of-speech analysis. The analyses may reveal key linguistic features characterizing the language in the sermons of a target religious group as indicators to describe its social characteristics, and there might be three possible general results and implications from the comparisons.

First, key linguistic features that associate with the features of the mainstream sermons and sect sermons which imply that a target religious group does not contain any destructive characteristics. Second, key linguistic features that associate with the features of the cult sermons which imply that a target religious group is in the transitional point and a close monitoring or regular examination should be conducted into the discourse of the leader to view its development. Third, key linguistic features that associate with the features of the destructive cult sermons which imply that a target religious group may possess destructive characteristics, and this finding may serve as a warning system before this group can do harmful activities.

To verify the results, the LL mean value of each comparison will be calculated to see which benchmark corpus the sermons of a target religious group will be associated from a keyness statistic perspective, as one way to confirm the stage of development of a target religious group. If the LL mean value of one comparison is lower when compared to the values of other comparisons, this means that the target corpus is likely associated or similar with the benchmark corpus. If the LL mean value of one comparison is higher when compared to the values of other comparisons, this implies that the target corpus is likely unrelated or different when compared to the benchmark corpus.

8. Methodological contributions

This paper focused on the methodological principles that provide us with a systematic process to determine the stage of development of a target religious group from a keyness perspective. From the findings, there is a clear pattern of progression of key linguistic features in the stages of religious development which matches the social constructs of the types of religion. Given the predictability of the progression, this allows us to be able to identify the stage of development of a target religious group.

The keyness findings from different levels of linguistic features may allow us to reduce our subjective interpretation of the data on characterizing and categorizing the sermons as our steps to identify the stage of development of a cultlike religious group. They may also serve as a warning system before this group can do dangerous activities if the language in its sermons contains destructive characteristics. We hope that this paper offers a new perspective in conducting a corpus-based research particularly when it is related to religious discourse.

References

- Abbott, C. B. (2015). The Reverend Jim Jones and religious, political, and racial radicalism in Peoples Temple. (Doctoral dissertation). University of Wisconsin–Milwaukee, United States.
- Acheoah, J. E., & Hamzah, A. (2015). Style in Christian and Islamic sermons: A linguistic analysis. *American Research Journal of English and Literature*, 1(1), 23-31.
- Agersnap, A., Kristensen-McLachlan, R. D., Johansen, K. H., Schjødt, U., & Nielbo, K. L. (2020). Sermons as data: Introducing a corpus of 11,955 Danish sermons. *Journal of Cultural Analytics*, 12, 1-27. doi: 10.22148/001c.18238
- Anthony, L. (2014) AntConc (Version 3.4.4) [Computer Software]. Waseda University. URL <http://www.laurenceanthony.net/software/antconc>
- Archer, D. E., Rayson, P., Piao, S., & McEnery, A. M. (2004). Comparing the UCREL semantic annotation scheme with lexicographical taxonomies. In *Proceedings of the EURALEX-2004 Conference* (pp. 817-827). European Association for Lexicography.
- Baker, P. (2004). Querying keywords: Questions of difference, frequency, and sense in keywords analysis. *Journal of English linguistics*, 32(4), 346-359.
- Barnhart, M. (2013, September 24). Interview: Joel Osteen on power of prayer, forgiveness, trusting in God's timetable. The Christian Post. Retrieved from <https://www.christianpost.com/news/interview-joel-osteen-on-power-of-prayer-forgiveness-trusting-in-gods-time-table.html>
- Beaman, W. J. (1990). From sect to cult to sect: The Christian Catholic Church in Zion. (Doctoral dissertation). Iowa State University, United States.
- Benitez, J. (2015). Hispanic religious outreach in the Upper U.S. South: Missionary outreach, strategies, and institutional praxis among mainstream denominations. (Doctoral dissertation). University of Kentucky, United States.
- Blanchard, C. M. (2009). Islam: Sunnis and Shiites. In C. Malbouisson (Ed.), *Focus on Islamic issues*, 11-20. Nova Science Publishers, Inc.
- Bohm, J., & Alison, L. (2001). An exploratory study in methods of distinguishing destructive cults. *Psychology, Crime & Law*, 7(2), 133-165.
- Bondi, M. (2010). Perspectives on keywords and keyness. In M. Bondi & M. Scott (Eds.), *Keyness in texts*, 1-20. John Benjamins.
- Cheng, W. (2009). Income/interest/net: Using internal criteria to determine the aboutness of a text. In K. Aijmer (Ed.), *Corpora and language teaching*, 157-177. John Benjamins.
- Conze, E. (2012). *Buddhism: Its essence and development*. Massachusetts: Courier Corporation.
- Cronshaw, D. (2019). "Growing Young" in Australian Baptist Churches: Surveying formation, belonging and mission. *Exchange*, 48(2), 156-181.
- Culpeper, J. (2009). Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's Romeo and Juliet. *International Journal of Corpus Linguistics*, 14(1), 29-59.
- De Vries, R. E., Bakker-Pieper, A., & Oostenveld, W. (2010). Leadership= communication? The relations of leaders' communication styles with leadership styles, knowledge sharing and leadership outcomes. *Journal of Business and Psychology*, 25(3), 367-380.
- Esimaje, A. U. (2012). A corpus-based lexical study of sermons in Nigeria. *English Language Teaching*, 5(9), 24-32.
- Gabrielatos, C. (2018). Keyness analysis: Nature, metrics and techniques. In C. Taylor and A. Marchi (Eds.), *Corpus Approaches to Discourse*, 225-258. Routledge.

- Gerbig, A. (2010). Key words and key phrases in a corpus of travel writing. In M. Bondi and M. Scott (Eds.), *Keyness in Texts*, 147-168. John Benjamins.
- Holbrook, D. (2015). Designing and applying an 'Extremist Media Index'. *Perspectives on Terrorism*, 9(5), 57-68.
- Hutchins, W. J. (1978). The concept of 'aboutness' in subject indexing. In Aslib proceedings. MCB UP Ltd. <https://www.emerald.com/insight/content/doi/10.1108/eb050629/full/html>
- Lien, H. Y. (2021). Revisiting keyword analysis in a specialized corpus: Religious terminology extraction. *Journal of Quantitative Linguistics*, 1-14.
- Mansouri, S., Biria, R., Mohammadi Najafabadi, M., & Sattar Boroujeni, S. (2017). Nomination and argumentation strategies in oratory discourse: The case of an English sermon. *SAGE Open*, 7(2), 2158244017702425. <https://doi.org/10.1177/2158244017702425>
- Martin, W. C. (2018). *A prophet with honor: The Billy Graham story* (Updated edition). Zondervan.
- Nini, A. (2015). Multidimensional analysis tagger (Version 1.3). URL <http://sites.google.com/site/multidimensionaltagger>.
- O'Halloran, K. (2011). Limitations of the logico-rhetorical module: Inconsistency in argument, online discussion forums and electronic deconstruction. *Discourse Studies*, 13(6), 797-806.
- Palayon, R. T., Watson Todd, R., & Vungthong, S. (in press). From the Temple of life to the Temple of death: Keyness analyses of the transitions of a cult. *Corpora Journal*.
- Palayon, R. T., Watson Todd, R., & Vungthong, S. (2020). The language of destructive cults: Keyness analyses of sermons. *Communication & Language at Work*, 7(1), 42-58.
- Piao, S. S., Rayson, P., Archer, D., & McEnery, T. (2005). Comparing and combining a semantic tagger and a statistical tool for MWE extraction. *Computer Speech & Language*, 19(4), 378-397.
- Pojanapunya, P. (2017). A theory of keywords. (Doctoral dissertation). School of Liberal Arts, King Mongkut's University of Technology Thonburi, Bangkok, Thailand.
- Pojanapunya, P., & Lieungnapar, A. (2017). The quality of choices determines the quantity of key words. Doing Research in Applied Linguistics 3 and 19th English in South-East Asia Conference 2017. (pp. 115-124). Retrieved from <http://sola.kmutt.ac.th/dral2017/index.php?q=PROCEEDINGS>
- Pojanapunya, P., & Watson Todd, R. (2018). Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory*, 14(1), 133-167.
- Pojanapunya, P., & Watson Todd, R. (2021). The influence of the benchmark corpus on keyword analysis. *Register Studies*. Advance online publication. <https://doi.org/10.1075/rs.19017.poj>
- Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. Annual Meeting of the ACL: Proceedings of the Workshop on Comparing Corpora 9: 1-6. <http://dx.doi.org/10.3115/1117729.1117730>
- Rayson, P., Berridge, D., & Francis, B. (2004). Extending the Cochran Rule for the comparison of word frequencies between corpora. The 7th International Conference on Statistical of Textual Data (JADT 2004). Vol. II, Purnelle, G., Faircon, C. and Dister, A. (Eds.), March 10-12, Louvain-la-Neuve, Belgium, pp. 926-936.
- Real Stories. (2018, February 18). *For the love of God: The man who thinks he's Jesus (Cult Documentary)* | Real Stories [Video File]. Retrieved from <https://www.youtube.com/watch?v=hXpLVmL2esw&t=284s>

- Salama, A. H. (2011). Ideological collocation and the recontextualization of Wahhabi-Saudi Islam post-9/11: A synergy of corpus linguistics and critical discourse analysis. *Discourse & Society*, 22(3), 315-342.
- Scharl, A., & Weichselbraun, A. (2008). An automated approach to investigating the online media coverage of US presidential elections. *Journal of Information Technology and Politics*, 5(1), 121-132.
- Scott, M. (2010). Problems in investigating keyness, or clearing the undergrowth and marking out trails. In M. Bondi & M. Scott (Eds.), *Keyness in texts*, 43-57. Amsterdam: John Benjamins.
- Tangenberg, K. (2008). Saddleback Church and the PEACE plan: Implications for social work. *Social Work & Christianity*, 35(4).
- The Cult Education Institute. (2016, January 28). *Cult documentary – Evil cult leaders (Investigation discovery)* [Video file]. Retrieved from <https://www.youtube.com/watch?v=ILO4JtBLRUg&t=2246s>
- The Jonestown Institute. (2018). Alternative considerations of Jonestown & Peoples Temple. California, San Diego State University. Retrieved from <http://jonestown.sdsu.edu>
- Turner, C. (2013). *Islam: The basics*. Routledge.
- Wallace, D. D. (1985). Sects, cults and mainstream religion: A cultural interpretation of new religious movements in America. *American Studies*, 26(2), 5-16.
- Zeller, B. E. (2014). *Heaven's Gate: America's UFO religion*. NYU Press.