



Translation and Semantic Shift of Islamic Vocabulary in English Abstracts: A Corpus-Based Study at an Indonesian Islamic University

Misbahuddin¹, Srifani Simbuka^{2*}

¹Institut Agama Islam Negeri Manado, Manado, Indonesia. Email: misbahuddin.jamal@iain-manado.ac.id

²Institut Agama Islam Negeri Manado, Manado, Indonesia. Email: srifani.simbuka@iain-manado.ac.id

*Corresponding Author

ARTICLE INFO

Cite this article in APA style as:

Misbahuddin, M., Srifani, S. (2024). Translation and Semantic Shift of Islamic Vocabulary in English Abstracts: A Corpus-Based Study at an Indonesian Islamic University. *Langkawi Journal of The Association for Arabic and English*, 10(1), 128-147.

DOI:

<http://dx.doi.org/10.31332/lkw.v0i0.8326>

History:

Submitted: 2023-12-23

Revised: 2024-05-13

Accepted: 2024-06-28

Published online: 2024-06-30

Copyright Holder:

The Authors.

Published by:

UPT Pengembangan Bahasa IAIN Kendari.

ABSTRACT

Incorporating Nation's vocabulary classification, this study investigates the translation and semantic shifts of Islam-related vocabulary in the abstracts of undergraduate and postgraduate theses. A corpus of 810 English and Bahasa Indonesia abstracts, sampled from a total of 1,928, was analyzed using Antwordprofiler and Antconc software. Findings indicate that translation strategies employed include phonetic alteration, pure borrowing, and one-to-one translation. Specifically, loanwords of Arabic origin were predominantly translated using pure borrowing, with some phonetic and/or orthographic modification in both Bahasa Indonesia and English. The translated vocabulary primarily consisted of Arabic-origin words with Bahasa Indonesia orthographic adaptations rather than their original Arabic forms. Additionally, semantic shifts were observed in the adoption of Islam-related Arabic vocabulary into Bahasa Indonesia. This is attributed to 1) the introduction of Quranic meanings, often differing from the common understanding of Arabic words at the time of the Quran's revelation, and 2) modifications to accommodate the traditions of Indonesian local Muslim communities. The implications of this study highlight the need for translators to be aware of the cultural and religious contexts that influence semantic shifts. Moreover, this research underscores the importance of understanding the dynamic nature of language in the context of religious texts, which can lead to more accurate and culturally sensitive translations.

Keywords: Corpus method; Islamic vocabulary; Vocabulary profile; Semantic shift; Translation

1. Introduction

Islam-related vocabulary, largely derived from Arabic, prompts discourse on its representation in English texts. Studies have demonstrated diverse approaches to conveying Islamic concepts in English, a language not traditionally associated with Islam. Brown (1996) identifies Arabic loanwords, like "ayatollah," that have been assimilated into English lexicons and are recognized as "English words" in dictionaries such as Longman, Oxford Advanced Learners, Collins COBUILD, and Cambridge International. Hassan (2016) outlines three distinct representation types for Islam-related terminology in English texts: direct translation from Arabic, transliteration to accommodate Arabic phonemes, and a combination of transliteration with English description. Online resources like the Cambridge Dictionary's religion section (<https://dictionary.cambridge.org/topics/religion/islam/>) further contribute to this evolving landscape of Islamic representation in the English language.

In the Indonesian context, academic works on Islamic topics frequently employ Arabic words, such as “*surah*,” as direct loanwords without accompanying definitions. Erlina et al. (2016) observe that such practices are prevalent due to the perceived familiarity of these terms among the target audience of Muslim readers and scholars. The assumption is that these words are widely understood and best conveyed in their original Arabic or transliterated forms, rendering further explanation or translation unnecessary within this specific academic discourse.

Complications arise as the Indonesian higher education landscape increasingly demands Islamic scholars to disseminate their research findings in international, interdisciplinary academic journals. This necessitates the publication of Islamic studies research in a language accessible to the wider academic community, predominantly English. However, this poses challenges in translating Islamic vocabulary into academically sound English without compromising its original meaning or diluting its inherent Islamic values (Dakake, 2019; Anis, Nababan, Santosa & Masrukhi, 2022). The abstracts of undergraduate and postgraduate theses in Islamic studies exemplify this issue, with varying approaches to representing Islamic terminology. Some adhere to Hassan's three types of writing Islam-related words, while others follow the practice noted by Erlina et al. (2016), utilizing the original Arabic forms.

This study investigates the translation of Islam-related vocabulary from Bahasa Indonesia into English within the abstracts of undergraduate and master's theses. Employing a mixed-method approach, the research combines quantitative-descriptive corpus linguistics with qualitative content analysis to examine the translation and semantic features of the data. Accordingly, the research questions guiding this study are: 1) What is the most frequent vocabulary (single-word and multi-word) related to Islamic studies found in undergraduate and master's theses at IAIN Manado?; 2) How are these vocabulary items translated into English? Are there differences in the strategies used to translate Islamic studies terminology from Arabic to Bahasa Indonesia and from Bahasa Indonesia to English?; and 3) How accurate is the translation of Islamic studies vocabulary into English from a semantic perspective? Is there a semantic shift from the original Arabic meaning when these terms undergo two layers of translation?

In contrast to previous studies on Islam-related words and terminologies, which predominantly employed qualitative inquiry for manual identification, this study contributes a novel approach. By utilizing corpus linguistics and a quantitative-frequency count technique, we offer an alternative method for identifying Islam-related vocabulary. This distinguishes our research from prior work, such as Hassan, Azmi & Abubakar (2017), who focused on reporting Islam in Nigerian and Malaysian media, and Lien (2022), who examined responses to Islam-related news on social media in Scandinavia. Additionally, our study differs from Embabi (2021), who conducted a qualitative analysis of Quranic and Prophetic translations within the discourse of “Islamic feminists.” Then, by employing a quantitative-frequency count technique within a corpus linguistics framework, this study provides a unique and valuable contribution to the existing body of knowledge on Islamic terminology.

While existing literature extensively discusses the translation of Islamic terms, focusing on debates surrounding translation versus transliteration and challenges in rendering Quranic text into Bahasa Indonesia and local languages (Rizal, E.A & Sobirin, M, 2021; Hussin, M., & Kamal, M. H. 2021; Al Farisi, M. Z. 2023), this study seeks to extend empirical research. By examining specific Islam-related vocabulary

within academic discourse, we aim to critically investigate the emergence of localized Islamic terminology whose meaning is closely tied to Indonesian culture. This approach goes beyond relying solely on original Arabic references, offering a nuanced understanding of the evolving nature of Islamic vocabulary in the Indonesian context.

2. Method

This study employed a mixed-method explanatory research design (Creswell & Plano Clark, 2007), combining quantitative-descriptive corpus linguistics with qualitative content analysis of Islam-related vocabulary. This design was chosen due to its prevalence in corpus-based discourse studies (Dakakke, 2022; Giampieri, 2020; Nugraha, Sujatna, & Mahdi, 2021; Marín Pérez & Sánchez-Lafuente, 2022).

The study utilized two parallel corpora, comprising 186,837 words compiled from 405 paired Bahasa Indonesia and English abstracts (810 total) of undergraduate and postgraduate theses at an Islamic state university. The source population consisted of 1,128 English abstracts and approximately 800 corresponding Bahasa Indonesia abstracts. These abstracts originated from students across 19 undergraduate and 4 postgraduate programs within the university's four schools: Islamic Law, Islamic Education and Teacher Training, Islamic Business and Economy, and Ushuluddin. The selection of abstracts was based on several considerations: 1) utilizing shorter texts mitigated word count bias often associated with longer texts (Akbarian, Ghanbarzadeh, & Shahri, 2017; Coxhead, 2000); 2) shorter texts allowed for greater topic diversity, yielding a wider range of word types compared to longer texts focused on a single topic (Akbarian, Ghanbarzadeh, & Shahri, 2017; Coxhead, 2000); and 3) the involvement of multiple authors across the source texts ensured a more balanced and unbiased representation of word types than longer texts written by a single author.

It is crucial to note that this study employed a specific criterion for data collection: only abstracts with both Bahasa Indonesia and English versions were included. This was essential as the study required paired versions of each abstract to construct two parallel corpora, the Bahasa Indonesia Abstract Corpus and the English Abstract Corpus. Consequently, out of the 1,128 English abstracts in the population, only 800 were utilized because they had corresponding Bahasa Indonesia versions. The remaining 328 abstracts were excluded due to file conversion or formatting issues, rendering either the Bahasa Indonesia or English version corrupted and incompatible with the corpus software.

2.1. Research Ethics

The data for this research, consisting of undergraduate and postgraduate thesis abstracts, was obtained from the open-access directory of the research site's library. As such, no specific patent requirements were necessary for accessing and utilizing these abstracts as research data. Moreover, to ensure privacy, student personal information was removed during the conversion of the raw data into plain text format. Additionally, the translated versions of the abstracts did not contain any information about the translator's identity, further safeguarding against potential privacy breaches. The data collection process was conducted with the explicit permission of the Head of the Central Library at the research site.

2.2. Data Collection Procedures and Instrumentation

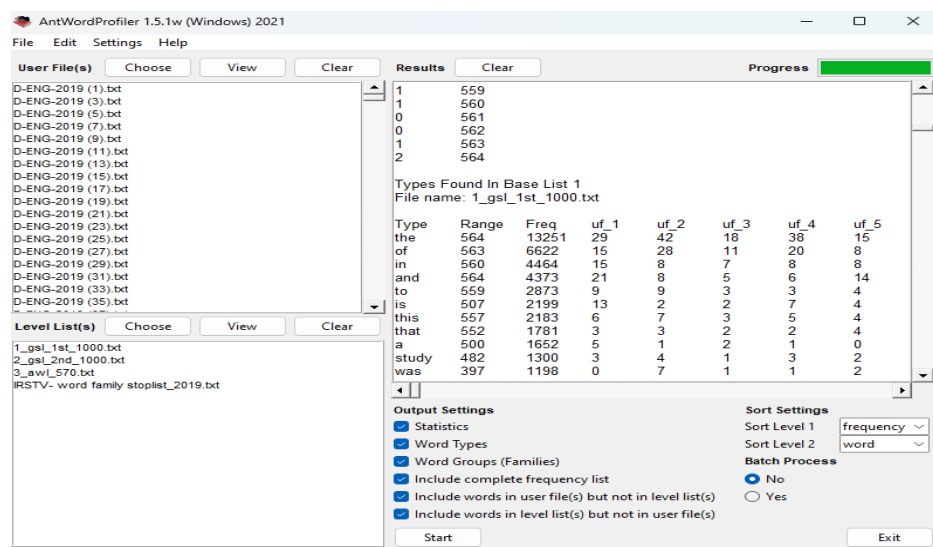
To address the research questions, data collection entailed the compilation of abstracts from undergraduate and postgraduate theses, with a focus on both single-word and multi-word lexical items. Initially, digital versions of the abstracts, predominantly in Microsoft Word format, were gathered. These files were

subsequently converted into plain text (txt) format to ensure compatibility with the designated analysis tools. The resultant text files, constituting the raw data, were then uploaded to the AntConc corpus software (version 4.0.1). Developed by Laurence Anthony, AntConc is a sophisticated corpus analysis tool that facilitated the comprehensive profiling of vocabulary content within the abstract corpus and the systematic identification of Islam-related terminology.

2.3. Data Analysis

To adhere to the study's mixed-method design, a quantitative frequency count analysis of Islam-related terminology was conducted. The results were presented in tables containing descriptive statistics on general, academic, and, most importantly, Islam-related vocabulary found within the abstracts. Islam-related terms were subsequently analyzed qualitatively through meaning and usage verification. Initially saved as txt files, the analysis results were converted into Excel format for enhanced data visualization. Figure 1 illustrates the Antwordprofiler interface displaying the uploaded raw data files, along with the stop list files used to filter vocabulary categories: General Service Lists (GSL 1 and GSL 2) compiled by Michael West (1953), Academic Word List (AWL) by Avril Coxhead (2000), and Islamic Religious Studies Textbook Vocabulary (IRSTV) List by Simbuka (2019). The interface also showcases partial analysis results, indicating the types of words belonging to the GSL 1st-1000 list. The complete results encompass frequency counts for each vocabulary category, providing comprehensive statistics on their coverage within the abstract corpus.

Figure 1. The display of Antwordprofiler analysis



The second phase of the study involved a qualitative analysis to uncover the translation strategies employed by the translators. This aimed to determine whether differences existed in the translation of Islamic terminology from Arabic to Bahasa Indonesia, and subsequently from Bahasa Indonesia to English. Additionally, qualitative analysis was utilized to examine potential semantic shifts arising from these translation processes. Major reference works, such as the Oxford Encyclopedia of the Islamic World and "Kosakata Keagamaan" by renowned Indonesian Islamic scholar M. Quraish Shihab (2020), were consulted to verify the meaning and semantic features of the translated terms. Then, the data analysis technique employed to address the second research question was content analysis. AntConc version 3.4.3 facilitated the search for Islam-related words within their respective contexts (the sentences in

which they occurred). Although corpus software was used, the analysis remained qualitative as it focused on the contextual meaning of the investigated words rather than their frequency data.

3. Findings and Discussion

3.1. General vocabulary profile of the corpus data

This section commences with the presentation of quantitative analysis results, providing descriptive statistics on the frequency of each vocabulary type within the dataset. Each vocabulary category is accompanied by its respective occurrence count, derived from the word profiling analysis conducted using the AntWordProfiler program. This procedure aims to address the first research question: "What are the most frequent vocabulary items (single-word and multi-word) related to Islamic studies that occur in the undergraduate and master's theses at IAIN Manado?"

The displayed data, imported from the program's output, additionally included the total word count (tokens) for each vocabulary category and the number of unique "word types" within each category. The general, academic, and other vocabulary types were tabulated alongside their corresponding tokens, total word types, and the frequency of each word, ranked from highest to lowest. The final output of this initial data analysis step was a vocabulary profile, including the most frequent Islam-related terms, which constitute the primary focus of this study, as presented in Table 1.

Table 1. Vocabulary profile of the English abstract corpus

Level	File	Token	Token%	Type	Type%
1	1_gsl_1st_1000.txt	66208	67.96	1809	29.08
2	2_gsl_2nd_1000.txt	4932	5.06	668	10.74
3	3awl_570.txt	12394	12.72	1014	16.3
	IRSTV- word family stoplist				
4	Simbuka_2019.txt	2186	2.24	225	3.62
5	Kosakata keagamaan-QS-stoplist.txt	190	0.2	29	0.47
0	-	11519	11.82	2475	39.79
Total:		97429		6220	

The analysis of the corpus, as depicted in Table 1, reveals that the undergraduate and postgraduate theses abstracts contained a total of 97,429 words. Most of these words, approximately 73.02%, were classified as high-frequency words, filtered by the General Service List (GSL) 1st and 2nd. Academic vocabulary, filtered by the Academic Word List (AWL), comprised 12.72% of the corpus. Notably, Islam-related vocabulary, identified using the Islamic Religious Studies Textbook Vocabulary (IRSTV) list, constituted 2.24% of the total words. A small fraction, 0.2%, belonged to the "*Kosakata Keagamaan Islam*" by Quraish Shihab (2022), a list of Islamic religious terms. The remaining 11.82% of words did not fall into any of the categories, indicating a diverse range of vocabulary beyond general, academic, and specifically Islamic terminology.

Drawing upon established research on word lists (Sutarsyah, Nation & Kennedy, 1994; Hirsh, 1992 cited in Nation & Waring, 1997) and Nation's (2001) concept of ideal vocabulary content, the analysis of the thesis abstracts corpus revealed an approximate alignment with expected coverage for each vocabulary category. High-frequency words, as predicted, constituted around 73.02% of the total word count, falling within the expected range of 78-90%. Academic vocabulary, accounting

for 12.72%, also met the suggested range of 8-12%. However, the coverage of Islam-related vocabulary, categorized as “technical words” by Nation (2001), was lower than the typical 5-7% expectation, comprising only around 2% of the corpus. This deviation is considered acceptable given that abstracts, while academic in nature, are inherently shorter than full-length texts within a specialized field, thus potentially containing a lesser proportion of technical vocabulary. Therefore, the lower coverage of Islam-related terms aligns with the nature of abstracts as concise summaries rather than comprehensive treatises on Islamic studies.

Table 2 illustrates the top 25 most frequent English words within the dataset. These findings align with previous research on typical English vocabulary content, confirming the prevalence of function words such as articles (the, a, an), prepositions (in, of, on), and conjunctions in the English abstracts under investigation.

Table 2. Top 25 Most Frequent Word Types in the English Abstract Corpus

Type	Range	Freq
the	394	9294
of	393	4531
in	392	3114
and	394	3079
to	391	2030
is	355	1571
this	387	1489
that	387	1229
a	348	1196
study	329	886
students	164	804
was	277	785
by	306	746
at	266	734
on	272	661
learning	121	649
are	260	622
with	285	619
as	243	557
for	246	542
were	226	456
it	241	443
out	245	423
from	225	420
used	240	380

Analyses of the Bahasa Indonesia Corpus

The Table 3 presents the vocabulary profile of the Bahasa Indonesia corpus, categorized based on word frequency and type. This analysis utilized two stop lists: the Indonesian most frequent words list by Doyle (no year) and the “*Kosakata Keagamaan Islam*” list based on Quraish Shihab’s (2020) compilation of essential Islamic religious vocabulary in Indonesia.

Table 3. Vocabulary profile of the Bahasa Indonesia Abstract Corpus

Level/ Stop list	File	Token	Token%	Cumtoken%	Type	Type %
1	INDONESIAN stoplist- Doyle.txt	25798	28.85	28.85	229	3.37
2	Kosakata keagamaan- QS-stoplist.txt	665	0.74	29.59	49	0.72
0	-	62945	70.4	99.99	6520	95.91
TOTAL:		89408			6798	

As shown in Table 3, the Bahasa Indonesia abstract corpus comprised 28.85% of the most frequent words in Bahasa Indonesia (Doyle, n.d.), and 0.74% of vocabulary pertaining to Islamic religious concepts "*kosakata keagamaan*". The remaining 70.4% of the corpus consisted of words that were either not highly frequent in general usage or were specific to the field of Islamic studies.

Table 3 presents the 25 most frequent words in the dataset. Consistent with the English abstract corpus, the most frequent words in the Bahasa Indonesia corpus were function words, including the conjunction "*dan*" and prepositions such as "*dalam*," "*di*," and "*dari*."

Table 4. Top 25 Most Frequent Words in the Bahasa Indonesia Abstracts Corpus

Type	Range	Frequency
<i>yang</i>	391	2884
<i>dan</i>	393	2730
<i>dalam</i>	374	1851
<i>ini</i>	390	1688
<i>di</i>	350	1547
<i>dengan</i>	382	1424
<i>dari</i>	316	888
<i>adalah</i>	323	789
<i>pada</i>	280	738
<i>yaitu</i>	270	577
<i>bahwa</i>	307	454
<i>terhadap</i>	183	452
<i>tidak</i>	196	360
<i>oleh</i>	205	341
<i>dapat</i>	194	337
<i>serta</i>	179	301
<i>atau</i>	179	285
<i>bagaimana</i>	150	239
<i>tentang</i>	163	235
<i>tersebut</i>	141	222
<i>hal</i>	142	218
<i>ada</i>	145	216
<i>karena</i>	138	204
<i>sebagai</i>	130	193

A secondary word list was employed to filter and calculate the most frequent words that pertain to Islamic religious vocabulary (in Bahasa Indonesia), the results of which are presented in the following section on Islam-related vocabulary.

The Islamic Vocabulary in the English and Bahasa Indonesia corpora

3.1.1. Islamic vocabulary in the English Corpus

The subsequent section details the findings derived from a qualitative analysis of potential Islam-related terminology within the dataset. This analysis is situated within the theoretical framework of Nation's vocabulary classification system.

The current study employed theIRSTV list (Simbuka et al., 2019; Simbuka & Nagauleng, 2021) as a filtering mechanism ("*stop list*") within the Antwordprofiler corpus tool. This facilitated the categorization of lexical items within the data as "technical vocabulary," as defined by Nation (2001). As previously mentioned, Nation's model encompasses additional categories, including high-frequency words (filtered using the General Service List, GSL 1 & 2; West, 1954) and academic words (filtered using the Academic Word List, AWL; Coxhead, 2001). Results indicate that 3235 tokens (word occurrences), corresponding to 2.32% of the total tokens in the abstract corpus, were classified as belonging to the Islam-related vocabulary list. This list comprised 263 unique word types (distinct words). Table 2, presented below, enumerates these word types, along with their respective frequencies of occurrence and range (the number of texts in which they appear).

Table 5. Top 25 most Frequent Islam-related words in the English corpus

Type	Range	Frequency	Type	Range	Frequency
islamic	137	401	rules	22	28
one	93	136	guidance	18	27
so	107	134	know	26	27
well	64	83	hadith	9	25
muslim	38	59	province	11	23
man	23	50	competence	8	21
terms	36	44	rules	22	28
mosque	10	42	said	18	20
curriculum	13	41	prophet	12	19
divorce	11	40	tribe	3	19
guardian	7	38	goods	6	17
known	29	37			
entitled	35	36			
concept	17	34			
madrassa	10	33			

Table 5 reveals that among the top 25 words of Islam-related vocabulary, "Islamic," "Muslim," and "hadith," all of Arabic origin, have been seamlessly integrated into the discourse of academic Islamic studies. These terms have gained acceptance as English words (Brown, 1996) within this field-specific context. This adoption reflects a broader strategy within academic discourse in Islamic studies to engage a wider readership, including those whose expertise lies outside the field (Hassan, 2016).

When examining field-specific vocabulary, relying solely on word frequency counts derived from corpus methods (i.e., frequency counts of individual words) may

prove insufficient. Terms such as ‘so,’ ‘one,’ ‘concept,’ or ‘curriculum’ might occur frequently but possess limited relevance, or even be irrelevant, to the domain of Islamic studies. Therefore, a more nuanced understanding of how these words contribute to the lexicon of Islamic studies can be achieved through qualitative analysis, examining the specific contexts or sentences in which they appear.

3.1.2. Other representation of Islamic vocabulary data outside the IRTV List

The Antwordprofiler analysis revealed that, in addition to Islam-related words identified through the IRSTV list, numerous terms demonstrated significant relevance to Islam due to their nature as “Arabic loanwords” (Hassan, 2016). The 25 most frequent loanwords are presented in Table 6.

Table 6. Islam-related vocabulary filtered by the Islamic Religion Vocabulary by Quraish Shihab

Type	Range	Frequency	Type	Range	Frequency
<i>syariah</i>	14	37	<i>Bid'ah</i>	1	1
<i>akhlak</i>	8	31	<i>Haji</i>	1	1
<i>sakinah</i>	4	22	<i>Hikmah</i>	1	1
<i>allah</i>	15	21	<i>Ilmu</i>	1	1
<i>ayat</i>	1	14	<i>Karimah</i>	1	1
<i>fatwa</i>	7	9	<i>Muhasabah</i>	1	1
<i>sunnah</i>	6	9	<i>Mungkar</i>	1	1
<i>iman</i>	1	8	<i>Nikah</i>	1	1
<i>akidah</i>	3	7	<i>Shalawat</i>	1	1
<i>halal</i>	2	7	<i>Syafaat</i>	1	1
<i>mahram</i>	3	3			
<i>ulama</i>	3	3			
<i>hidayah</i>	1	2			
<i>mut'ah</i>	1	2			
<i>aurat</i>	1	1			

Table 6 reveals that most Islam-related terms identified outside the IRSTV list are of Arabic origin and function as loanwords within the Indonesian language. Consequently, terms such as “*sharia*” (or “*syari'ah*”), “*madrassa*,” “*hajj*,” “*zakat*,” “*aqidah*,” and “*aliyah*” occur with relatively high frequency. Notably, the spelling of these Arabic-origin words reflects an Indonesian adaptation of their original orthography.

3.1.3. Islamic Vocabulary in the Bahasa Indonesia Corpus

To identify Islam-related terms within the Bahasa Indonesia abstracts, the AntWordProfiler tool was employed, utilizing distinct stop lists compared to the English corpus analysis. These stop lists included a list of the most frequent Indonesian words compiled by Damian Doyle (no year) and the “*Kosakata Keagamaan*” list developed by Quraish Shihab (2020). While the latter is not a corpus-linguistically derived list, it represents a qualitatively curated collection of prominent Islamic terminology by a renowned Indonesian Islamic scholar. The resulting top 25 Islam-related terms identified within the Bahasa Indonesia abstract corpus are presented in Table 7.

Table 7. Islamic Vocabulary by Quraish Shihab ‘List’

Type	Range	Frequency	Type	Range	Frequency
<i>syariah</i>	33	118	<i>hadis</i>	4	13
<i>akhlak</i>	28	102	<i>doa</i>	6	12
<i>ayat</i>	19	61	<i>ikhlas</i>	5	11
<i>nikah</i>	12	45	<i>fatwa</i>	7	10
<i>haji</i>	5	35	<i>halal</i>	4	10
<i>shalat</i>	5	30	<i>iman</i>	5	10
<i>akidah</i>	10	24	<i>sunnah</i>	7	10
<i>allah</i>	17	24	<i>tasawuf</i>	3	10
<i>ilmu</i>	15	24	<i>adil</i>	7	7
<i>sakinah</i>	3	21	<i>darurat</i>	4	4
<i>hikmah</i>	3	18	<i>talak</i>	3	4
<i>ulama</i>	10	16	<i>amal</i>	2	3

3.2. Translation of Islam-Related Vocabulary into English

To address the second research question regarding the translation of Islamic vocabulary into English, a mixed-methods approach was employed. This involved quantitative analysis utilizing AntConc corpus software to ascertain word frequency, complemented by qualitative manual searches to identify instances of Islamic concepts and their corresponding translations (or lack thereof) within the corpora.

Utilizing comparative corpora comprising abstracts in Bahasa Indonesia and their corresponding English translations (English-abstract corpus), we traced the trajectory of Arabic-origin terms, examining their adaptation into Bahasa Indonesia and subsequent rendition in English.

Table 8. Comparative profile of the Islam-related words in the Bahasa Indonesia and English Corpora

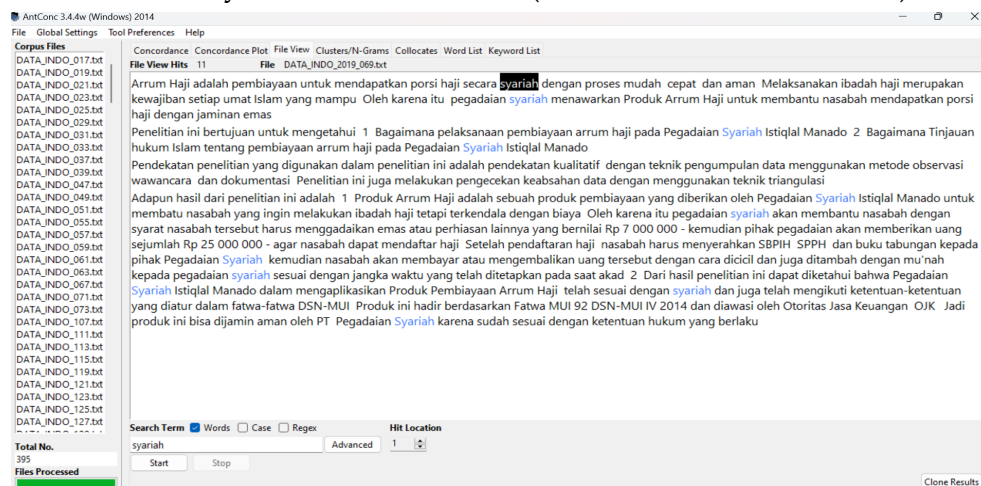
Bahasa Indonesia Corpus			English Corpus		
Type	Range	Frequency	Type	Range	Frequency
<i>syariah</i>	33	118	<i>syariah</i>	14	37
<i>akhlak</i>	28	102	<i>akhlak</i>	8	31
<i>ayat</i>	19	61	<i>sakinah</i>	4	22
<i>nikah</i>	12	45	<i>allah</i>	15	21
<i>haji</i>	5	35	<i>ayat</i>	1	14
<i>shalat</i>	5	30	<i>fatwa</i>	7	9
<i>akidah</i>	10	24	<i>sunnah</i>	6	9
<i>allah</i>	17	24	<i>iman</i>	1	8
<i>ilmu</i>	15	24	<i>akidah</i>	3	7
<i>sakinah</i>	3	21	<i>halal</i>	2	7
<i>hikmah</i>	3	18	<i>mahram</i>	3	3
<i>ulama</i>	10	16	<i>ulama</i>	3	3
<i>hadis</i>	4	13	<i>hidayah</i>	1	2
<i>doa</i>	6	12	<i>Mut'ah</i>	1	2
<i>ikhlas</i>	5	11	<i>aurat</i>	1	1
<i>fatwa</i>	7	10	<i>Bid'ah</i>	1	1
<i>halal</i>	4	10	<i>haji</i>	1	1
<i>iman</i>	5	10	<i>hikmah</i>	1	1
<i>sunnah</i>	7	10	<i>ilmu</i>	1	1

tasawuf	3	10	karimah	1	1
adil	7	7	muhasabah	1	1
darurat	4	4	mungkar	1	1
talak	3	4	nikah	1	1
amal	2	3	shalawat	1	1
aurat	3	3	syafaat	1	1
ihsan	1	3	syubhat	1	1
karimah	3	3	tadarus	1	1
mahram	3	3	tahlil	1	1
rezeki	2	3	zikir	1	1

Table 8 demonstrates that certain Islam-related terms present in the Bahasa Indonesia corpus were not consistently translated into English within the corresponding abstracts. For instance, the term “*syariah*” appeared 118 times across 33 texts in the Bahasa Indonesia corpus, but only 37 times in 14 texts within the English corpus. This discrepancy suggests that the term was often retained as a loanword in the English abstracts, albeit with Indonesian-adapted spelling rather than its original Arabic orthography. Conversely, approximately 81 instances of “*syariah*” were translated into its English equivalent, “Islamic law” (Baalbaki, 1998:669).

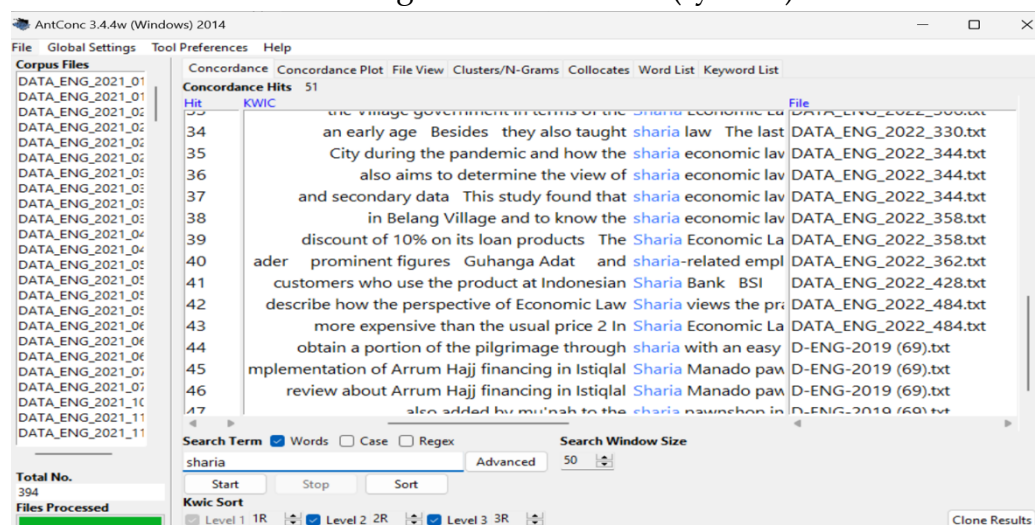
The following excerpts illustrate the usage of the term “*syariah*” within the dataset. Excerpt a): The term “*syariah*” is retained in its original Arabic form, albeit with adapted Indonesian spelling, without translation into English. This usage is exemplified in the following context: “...*pembiayaan untuk mendapatkan porsi haji secara syariah dengan proses mudah cepat dan aman*” (DATA_INDO_2019_069.txt). The AntConc 3.4.4 concordance tool facilitates the display of data containing the word “*syariah*,” as depicted in the subsequent figure.

Figure 2 The Display of the Antconc file view tool showing the data that contain the word ‘syariah’ in an abstract (Bahasa Indonesia version)



Excerpt b) demonstrates a different treatment of the term “*syariah*” within the same abstract but in the translated (English) version: “*Arrum Hajj is a financing to obtain a portion of the pilgrimage through sharia with an easy, fast, and safe process*” (DATA_ENG_2019_069.txt). This excerpt, displayed in the AntConc software as illustrated in Figure 3.

Figure 3. The display of Antconc software showing the data of English abstracts containing the word 'sharia' (syariah)



The data reveals that certain Arabic-origin terms underwent a two-step translation process, first into Bahasa Indonesia and subsequently into English. This phenomenon was particularly evident for common Islamic vocabulary adopted into Bahasa Indonesia, despite having existing English equivalents. Consequently, translation strategies varied. In the Bahasa Indonesia abstracts, many well-known Islamic terms were not translated but rather retained in their original Arabic forms, albeit with adapted Indonesian spelling (Hassan, 2016). For instance, terms such as “syariah,” “nikah,” “haji,” and “akad” were incorporated into the Bahasa Indonesia spelling system without transliteration, a common practice for Arabic loanwords. In contrast, in the English abstracts, “syariah” was rendered with a slightly modified spelling, “sharia” (Baalbaki, 1998:669), and was not translated into equivalents like “Islamic law” (Baalbaki, 1998:669). This translation strategy, known as “Alteration” (Nida, 1964 in Molina & Albir, 2002), is employed when the target language (English) lacks a direct graphemic representation for a specific sound present in the source language. In this case, alteration appears to be a preferred strategy for translating concepts with strong ideological or cultural values, such as Islamic teachings and law (Anis, Nababan, Santosa & Mashruki, 2022). Alongside the strategy of pure borrowing, alteration allows translators to preserve the original ideology and values embedded within the terms while gradually introducing them into the target language. Similar patterns were observed for terms like “sakinah,” “akhlak,” “ayat,” “fatwa,” “sunnah,” “iman,” “akidah,” “halal,” “mahram,” and “ulama.” Conversely, other Islam-related words, including “hidayah,” “mut’ah,” “aurat,” “bid’ah,” “muhasabah,” “shalawat,” “syubhat,” and “karimah,” exhibited consistent frequencies in both corpora, suggesting a uniform approach to their translation or retention as loanwords in both languages.

The second observed strategy for translating Islam-related words from Arabic into Bahasa Indonesia and subsequently into English involved pure borrowing. In this approach, the original Arabic terms were retained in their unaltered form, either through maintaining the original orthography or utilizing transliteration. Exemplary instances of such terms include “tahlil” and “Allah.” Figure 4 illustrates the usage of “tahlil” in both the Bahasa Indonesia and English abstract corpora (Figure 5).

Figure 4. The use of the word '*tahlil*' in the corpus of abstracts in Bahasa Indonesia (DATA_INDO_108.txt)

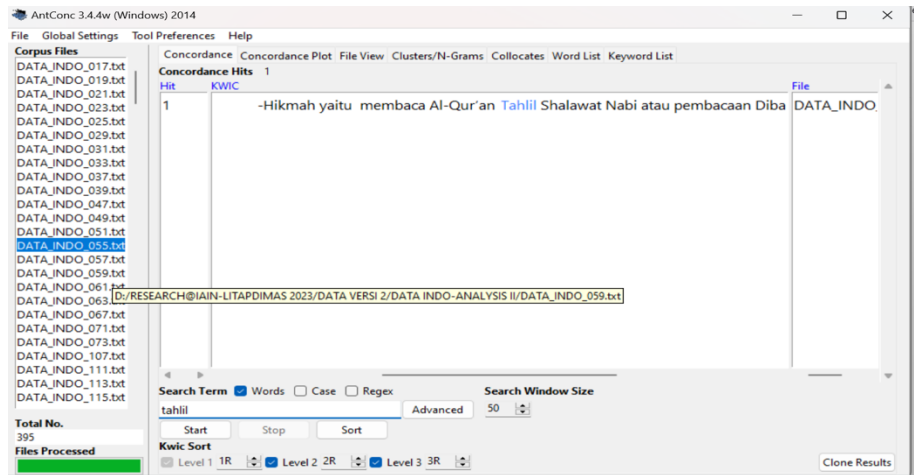
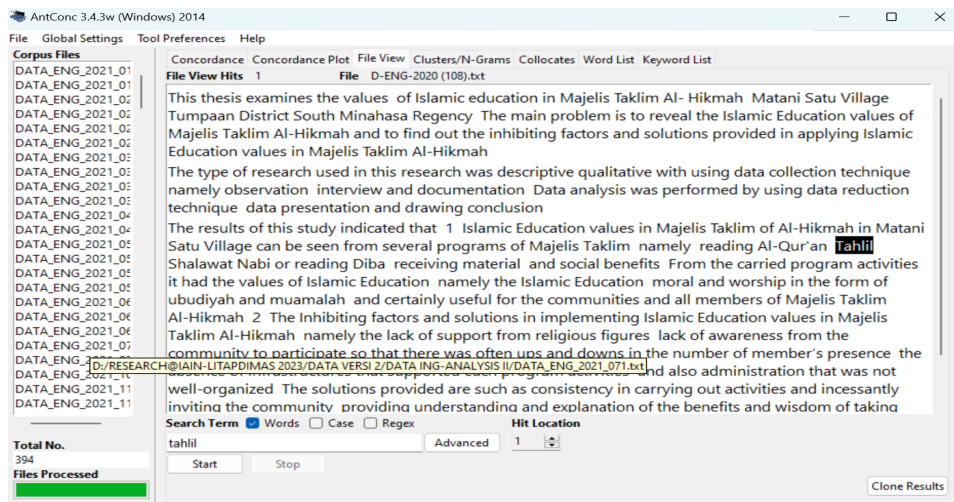


Figure 5. The word '*tahlil*' in the data (DATA_ENG_2020_108.txt)



The data presented in Figure 5 reveals that additional Islam-related words, such as "*shalawat*" and "*diba*," were subject to the same translation strategy as "*tahlil*," wherein the original Arabic forms were borrowed entirely. This observation aligns with the findings of Anis et al. (2022), suggesting that translators opt for pure borrowing due to the specialized nature of these terms, which are primarily understood within the Islamic community. Üstün Külük (2023) further elucidates this phenomenon through the concept of the "intra-ummah paradigm," emphasizing the use of shared vocabulary among Muslims, transcending national boundaries, to facilitate effective communication. Terms like "*jihad*," "*Sunnah*," "*ayah*," and Quranic verses exemplify this shared linguistic currency. The preference for pure borrowing over direct literal translation, which can introduce inaccuracies, is particularly evident when dealing with longer stretches of language, such as Quranic metaphors (Nurbayan, 2019).

The third identified translation strategy involved rendering Islam-related vocabulary into their English equivalents. For instance, the term "*nikah*," appearing 45 times across 12 abstracts, and it's derived from "*pernikahan*" (occurring 44 times in 16 abstracts), were consistently translated as "marriage" (121 instances in 26 abstracts) or "marriages" (20 instances in 10 abstracts), as depicted in Figures 6 and 7.

Figure 6: The word 'nikah' in the DATA_INDO_(2021)_055

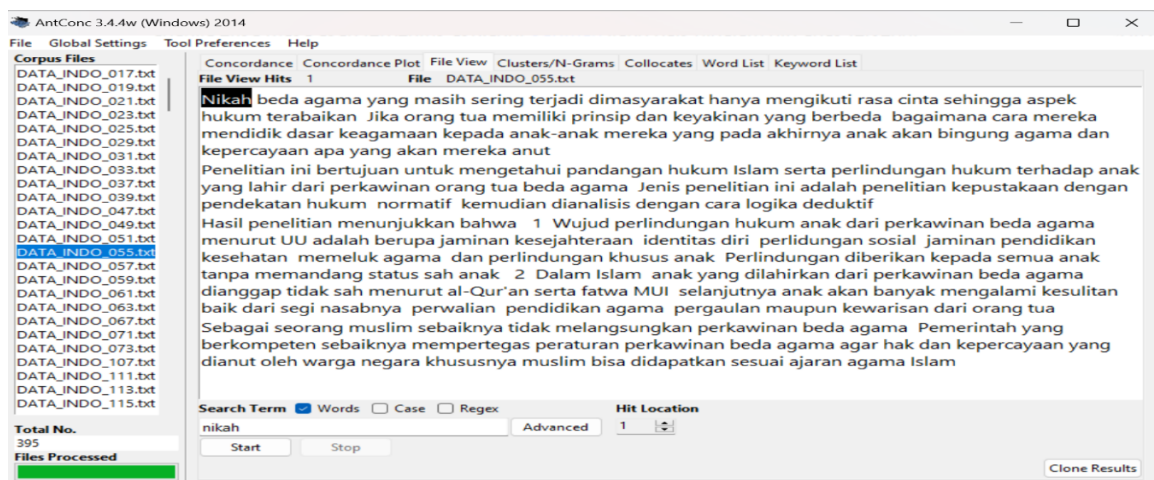


Figure 7. The word 'marriage(s)' in the DATA_ENG_2021_055

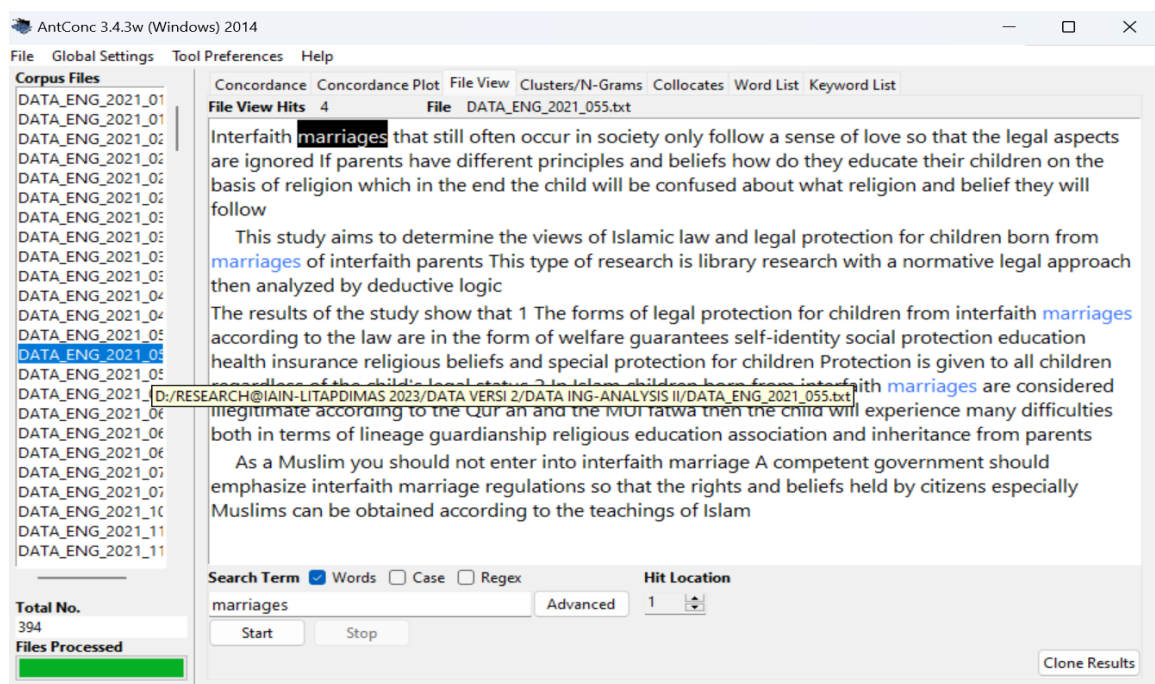


Figure 6 illustrates the occurrence of the term “*nikah*” within the Bahasa Indonesia abstracts. Upon comparing this data with the corresponding English abstracts, as depicted in Figure 7, it becomes evident that “*nikah*” is consistently translated into its English equivalent, “*marriage(s)*.” This translation strategy, termed “*Globalization*” (Farkhan, 2018), involves the direct conversion of words from the source language (Bahasa Indonesia) into English, under the premise that the Islamic vocabulary in question is universally understood and somewhat familiar to adherents of other Semitic religions due to its presence in their respective traditions.

3.3. Semantic Shift in the Translation of Islamic Vocabulary

To address the third research question concerning potential semantic shifts in the translation of Islam-related vocabulary into English, a qualitative analysis was undertaken. This analysis involved comparing entries in Arabic-Indonesian dictionaries, Arabic-English dictionaries, and scholarly references on Islamic terminology. Table 9 presents a sample of Islam-related words, displaying their forms

in Arabic, Bahasa Indonesia, and English, along with their meanings as indicated in reference materials and observed within the data.

Table 9. Sample of Translation of Islam-related Vocabulary from Arabic into Bahasa Indonesia and English

Arabic language / Original	Bahasa Indonesia (in the data)	Meaning in Bahasa Indonesia		Meaning in English	
		Original meaning	Specific meaning by Islamic concepts/terminology	In references (dictionary and or encyclopedia)	In the data (by the translator)
شريعة	syariah	<i>Bekas aliran air/ sumber air yg selalu diambil orang</i> (Nafis,2011:17)	<i>jalan yang telah digariskan oleh Allah</i> (Nafis,2011:17)	the <u>holy laws</u> of <u>Islam</u> that <u>cover all parts</u> of a Muslim's <u>life</u> (Cambridge online Dictionary), Sharia, Islamic Law, Law of Islam (Baalbaki, 1998:669)	Sharia (DATA_ENG_2 022_086.txt DATA_ENG_20 22_090.txt DATA_ENG_20 22_094.txt DATA_ENG_20 22_104.txt, and 47 other) Islamic Law (DATA_ENG_20 21_039.txt DATA_ENG_20 21_039.txt DATA_ENG_20 21_039.txt DATA_ENG_20 21_055.txt, and 77 others)
نكاح	nikah	<i>Persetubuhan</i> (Badawi,200 8: 962)	<i>perjanjian pernikahan</i>	Marriage, matrimony, wedlock (Baalbaki, 1995:1190)	Marriage (s) (DATA_ENG_2 021_055.txt DATA_ENG_20 21_201.txt DATA_ENG_20 22_072.txt, and 14 others)
تهليل تحليل	tahlil	1) <i>Dzikir</i>	Laa ilaaha illaa Allah -dzikir membaca kalimat tauhid (Shihab, 2020:62)	Acclamation, applause, cheer,hail (Baalbaki, 1995:384)	Tahlil -no decription/translation/explanation (DATA_ENG_2 020_118.txt)

The Table 9 illustrates instances of semantic shifts observed in the data. The most notable shifts occurred in the translation of Arabic words related to Islamic concepts into Bahasa Indonesia. As exemplified, the original meanings of the Arabic words "شريعة" (*sharia*), "نكاح" (*nikah*), and others were significantly altered to align with Islamic conceptual frameworks. This adaptation was driven using these words within Islamic discourse, specifically in the Quran, where their meanings often diverged from

those commonly understood by contemporary Arabic speakers. Shihab (2020) attributes these shifts to the Quran's unique usage of these words, which often differed from their prevailing meanings in the Arabic language of the time.

A distinct phenomenon emerged regarding the semantic shifts observed in the translation of the Arabic term "تهليل" or "تحليل". Originally, "تحليل" signified acclamation, applause, cheer, or hail (Baalbaki, 1995:384). However, the derived term "تهليل" acquired a new meaning: the utterance of "*Lā Ilāha illa Allāh*" (Sirait, 2016). This semantic shift stemmed from the practice of reciting this phrase by followers of Nahdlatul Ulama, a prominent Indonesian Islamic organization, during specific religious gatherings. The alteration of meanings in Arabic words to align with specific Islamic societal needs reflects a dynamic aspect of language evolution. Misunderstandings or partial comprehension of these terms can trigger such shifts, sparking debates among scholars regarding the appropriate response. Shihab (2020) advocates for a gradual reintroduction of the "true" meanings of Islam-related words based on their Quranic context, rather than relying on widely accepted interpretations among Muslims. Conversely, Dhona (2024) suggests employing Islamic discourse as a critique of all forms of dominance, including knowledge perceived as Islamic, and re-examining the history of Islamic communication within specific local contexts. This approach embraces Islamic terminologies imbued with local meanings as legitimate forms of Indonesian Islamic communication. The importance of considering local context and semantic sense in translation is echoed in Choi & Kim's (2021) study on translating Islamic core concepts and terms into Korean. They emphasize the need to account for South Korean social and political contexts, as well as the translator's ideological perspective (Anis, Santosa & Masrukhi, 2022; Totoli, 2023). Furthermore, Gunawan (2022) highlights the significance of considering the target readers' socio-political and cultural background in translation choices.

This study aligns with Shihab's (2020) position by providing empirical evidence of semantic shifts in Islamic vocabulary and their subsequent translation into English within academic literature. Through quantitative and qualitative analyses of abstracts produced at an Islamic university in Manado, we demonstrate how the modified translation of Arabic-origin terms in Bahasa Indonesia has led to similarly nuanced renditions in English. Our findings contribute to both translation studies and Islamic studies by extending the concept of incorporating the translator's subjective stance and the readers' socio-political and cultural context, as proposed by Choi & Kim (2021), Gunawan (2022), and Totoli (2023), to the translation of specific Quranic vocabulary within Islamic studies literature. This approach recognizes that readers may encounter these terms with diverse, non-religious purposes, necessitating a nuanced translation approach that considers both linguistic accuracy and cultural sensitivity.

However, this study does not intend to evaluate the translational quality of the analyzed abstracts. Rather, based on our findings, we posit that the retention of Islam-related vocabulary in its original Arabic form within English-language theses abstracts is generally permissible when the intended audience possesses prior knowledge of Islamic studies or operates within the specific context of Islamic scholarship. Nevertheless, should authors aspire to reach a wider readership, it is recommended that translators incorporate explanatory notes alongside Arabic terms or Indonesian transliterations to facilitate comprehension for those unfamiliar with Islamic terminology.

In conclusion, this study demonstrates the utility of corpus tools in facilitating the identification and frequency-based ranking of specific vocabulary within a given dataset, aligning with the advantages highlighted by Anthony (2013), Masrai & Milton (2021), and Yuliawati & Ekawati (2023). This method offers a time-efficient alternative to manual qualitative searches, particularly in analyzing texts with extensive word counts. Pedagogically, the resultant frequency rankings of words specific to a particular discourse can be leveraged for teaching English for Specific Purposes (ESP). However, a limitation of this method is the potential inclusion of terms not directly related to Islamic teachings, but rather associated with broader fields contributing to Islamic studies, such as philosophy and law.

4. Conclusion

This study concludes with three primary findings that address the research questions. First, the vocabulary profile of the English abstract corpus aligns with Nation's (2001) theory of vocabulary classification, demonstrating the expected distribution of high-frequency, academic, and technical vocabulary, including Islam-related terminology. Second, the predominant translation strategies employed for Islam-related vocabulary involve phonetic alteration and direct borrowing from the Bahasa Indonesia version into the English version, thus preserving the already shifted meanings (and occasionally, spellings) within the translated abstracts. Lastly, regarding the semantic shift of the studied Islam-related vocabulary, the findings reveal that changes in the meaning of Arabic-origin words occurred within the Bahasa Indonesia abstracts, where most of these words were adopted/borrowed/incorporated with altered meanings. This semantic shift is attributed to two factors: 1) changes in meaning due to their specialized usage in the Quran, often differing from the meanings understood by contemporary Arabic speakers, and 2) semantic shift due to the traditions of Indonesian Muslim groups. This shift subsequently influenced the translation of these words into English. The implication of this study is that translators of field-specific vocabulary should consider the intended readership, considering the specific context and targeted publication. This approach ensures that the translated texts are both accurate and accessible to the intended audience.

Certain limitations of this study highlight the potential for future research utilizing mixed methods, combining the quantitative approach of corpus linguistics with qualitative discourse or critical discourse analysis. This would enable a more nuanced understanding of the most frequent Islam-related words identified in the data. Furthermore, additional studies are warranted to enhance the validity of theIRSTV list, thereby improving its utility in identifying Islamic vocabulary.

Acknowledgments

This research was generously funded by the IAIN Manado-Litapdimas 2023 research grant. We extend our sincere gratitude to the Rector, the Head of the Research and Publication Center of IAIN Manado, and our research assistant, Raihan P. Tahir, for their invaluable support and contributions.

References

- Akbarian, I. H., Ghanbarzadeh, Z., & Afzali Shahri, M. (2017). Profiling vocabulary in psychology journal abstracts: A comparison between Iranian and Anglo-American journals. *Iranian Journal of Language Teaching Research*, 5(1), 51-69.

- Al Farisi, M. Z. (2023). Acceptability of the Quran Translation. *Al-Jami'ah: Journal of Islamic Studies*, 61(2), 329-363.
- Anis, M. Y., Nababan, M., Santosa, R. & Mohammad Masrukhi (2022) The ideology behind the translation of an Islamic moral ethic book: A case study of Al-Hikam's aphorisms, *Cogent Arts & Humanities*, 9:1, 2129484, DOI: 10.1080/23311983.2022.2129484
- Anthony, L. (2013). A critical look at software tools in corpus linguistics * 1, 30(2), 141-161.
- Anthony, L. (2014b). AntWordProfiler. Tokyo, Japan: Waseda University. Retrieved from <http://www.antlab.sci.waseda.ac.jp/>
- Hassan, I., Azmi, M.N., & Abubakar, U. (2017). The Use of Terminology in Reporting Islam: A Comparative Analysis. *International Journal of English Linguistics*, 7, 236.
- Baalbaki, R. (1998). A Modern Arabic-English Dictionary. Beirut: *Dar el Ilm Lilmalayin*
- Badawi, El-Said M.; Haleem, M. A. Abdel (2008), *Arabic-English dictionary of Qur'anic usage*, Brill Academic Publishers
- Brown, A. (1996). The Treatment of Religious Terminology in English Dictionaries. In R. M. Asraf (Ed.), *English and Islam: Creative Encounters* 96 (pp. 307-314). Malaysia: Department of English Language and Literature, International Islamic University Malaysia.
- Clark, V. L. P., Creswell, J. W., Green, D. O. N., & Shope, R. J. (2008). Mixing quantitative and qualitative approaches. *Handbook of emergent methods*, 363, 363-387.
- Choi, J., & Kim, K. H. (2021). Translating Islamic Knowledge for the New Soil: Two Korean Translations of the Qur'an in South Korea. *Interventions*, 23(7), 973-996. <https://doi.org/10.1080/1369801X.2020.1863837>
- Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, 34(2), 213-238.
- Dakakke, M.M. (2022) Qur'anic terminology, translation, and the Islamic conception of religion, *Religion*, 49:3, 343-363, DOI: 10.1080/0048721X.2019.1622837
- Dhona, H. R. (2024). Islamic communication as an invention of modern-western knowledge: critical analysis toward Islamic communication in Indonesia. *Asian Journal of Communication*, 34(3), 381-398. <https://doi.org/10.1080/01292986.2024.2320902>
- Doyle, D., no year. Indonesian Stopword. <https://www.ranks.nl/stopwords/indonesian>
- Embabi, D. (2021). Islamic feminism and the renegotiation of new knowledge through translation. *Translation Studies*, 15(1), 21-36. <https://doi.org/10.1080/14781700.2021.1927817>
- Erlina, D., Mayuni, I., & Akhadiah, S. (2016). Whole language-based English reading materials. *International Journal of Applied Linguistics and English Literature*, 5(3), 46-56.
- Gunawan, F. (2022). The ideology of translators in Quranic translation: lessons learned from Indonesia. *Cogent Arts & Humanities*, 9(1). <https://doi.org/10.1080/23311983.2022.2088438>
- Giampieri, P. (2020). *Volcanic experiences*: comparing non-corpus-based translations with corpus-based translations in translation training. *Perspectives*, 29(1), 46-63. <https://doi.org/10.1080/0907676X.2019.1705361>
- Hassan, S. (2016). Islamic religious terms in English-translation vs. transliteration in

- Ezzeddin Ibrahim and Denys Johnson-Davies' translation of An-Nawawī's Forty Ḥadīths. *Translation & Interpreting*, 8(1), 117-132. <https://doi.org/10.12807/ti.108201.2016.a08>
- Hussin, M., & Kamal, M. H. (2021). Translation of al-Quran into Malay Language in the Malay World. *IJISH (International Journal of Islamic Studies and Humanities)*, 4(1), 32-50. <https://doi.org/10.26555/ijish.v4i1.3322>
- Kwary, D. A., & Artha, A. F. (2017). The Academic Article Word List for Social Sciences 1. *Mextesol Journal*, 41(4), 1-11.
- Molina, L. & Hurtado Albir, A. (2002). Translation Techniques Revisited: A Dynamic and Functionalist Approach. *Meta*, 47(4), 498-512. <https://doi.org/10.7202/008033ar>
- Nafis, Ph.D., M. Cholil (2011). *Teori Hukum Ekonomi Syariah*. Penerbit Universitas Indonesia. hlm. 17. ISBN 9789794564561.
- Nation, I. S. P. (2001). *Learning Vocabulary in another Language* (First). Cambridge: Cambridge University Press.
- Nugraha, I. S., Sujatna, E. T., & Mahdi, S. (2021). Corpus linguistic study of tweets using# charliehebdo hashtag. *JALL (Journal of Applied Linguistics and Literacy)*, 5(1), 54-70.
- Nurbayan, Y. (2019). Metaphors in the Quran and its translation accuracy in Indonesian. *Indonesian Journal of Applied Linguistics*, 8(3), 710-715.
- Marín Pérez, M. J., & Almela Sánchez-Lafuente, Á. (2022). The representation of migrants in Spanish judicial decisions: using corpus data to refute hate speech.
- Masrai, A., & Milton, J. (2021). Vocabulary knowledge and academic achievement revisited: General and academic vocabulary as determinant factors. *Southern African Linguistics and Applied Language Studies*, 39(3), 282-294. <https://doi.org/10.2989/16073614.2021.1942097>
- Rizal, Efri Arsyad, and Mohamad Sobirin. "Translating the Transcends of God's World in the Arabic-Quran into Balinese". *ESENSIA: Jurnal Ilmu-Ilmu Ushuluddin* 22, no. 1 (May 29, 2021): 63-75. Accessed June 23, 2024.
- Sutarsyah, C, Nation, P. & Kennedy, G. (1994) How useful is EAP Vocabulary for ESP? A Corpus-Based Study. *RELJ Journal*, 25 (2), 34-50
- Shihab, M. Q. (2020) *Kosakata Keagamaan: Makna dan Penggunaannya*. PT Lentera Hati: Tangerang
- Simbuka, S., Hamied, F. A., Sundayana, W., & Kwary, D. A. (2019). A Corpus-Based Study on the Technical Vocabulary of Islamic Religious Studies. *TEFLIN Journal: A Publication on the Teaching & Learning of English*, 30(1).
- Simbuka, S. (2021). Evaluating the Vocabulary Coverage of EFL Textbooks and Teachers' Generated Materials in an Indonesian Islamic Universities: A Corpus-Informed Materials Evaluation. *AL-ISHLAH: Jurnal Pendidikan*, 13(3), 2489-2500.
- Tottoli, R. (2023). *The Qur'an in Europe, A European Qur'an: A History of Reading, Translation, Polemical Confrontation and Scholarly Appreciation*. *Jurnal Studi Ilmu-Ilmu Al-Qur'an Dan Hadis*, 24(2), 285-336. <https://doi.org/10.14421/qh.v24i2.4738>
- Üstün Külük, S. (2023). Translating Islam(ism) for a Muslim society: a novel 'Intra-Ummah translation paradigm' for Islamist translations in Turkey. *Perspectives*, 1-15. <https://doi.org/10.1080/0907676X.2023.2281969>
- West, M. (1953). *A general service list of English words: with semantic frequencies and a supplementary word-list for the writing of popular science and technology*. London:

Misbahuddin, Srifani Simbuka: Translation and Semantic Shift of Islam-Related Vocabulary in The English Version of Undergraduate and Post-Graduate Abstracts at an Indonesian Islamic University: A Corpus-Based Study

Longman, Green & Co.

Yuliawati, S., & Ekawati, D. (2023) Sketching women: A corpus analysis of woman representation in the Sundanese magazine Manglé (1958-2019). *Indonesian Journal of Applied Linguistics*, 13(1), 173-182.