

Messages from the Quran and the Bible in Mandarin through Factor Analysis with Syntactic and Semantic Tags

Kuanlin Liu

National Taipei University of Business
No. 100, Fulong Rd., Pingzhen Dist., Taoyuan City, Taiwan
kennyliu@ntub.edu.tw

Abstract

This paper tries to decipher messages from the Quran and the Bible’s Mandarin translation using the multidimensional factor analysis (MDA) approach. Part-of-speech and word-meaning annotations were employed for data tagging. Seven syntactic and six semantic factors derived from the tagging systems demonstrated how the two scriptures are interpreted on the factor score scales. The analyses indicated that both holy books uphold a “persuade” and “preach” style with higher frequencies of *imperative, advocative, and explanatory* expressions. In addition, both favor the “interpersonal, non-numeric, and indicative” strategies to *impress* followers and practitioners alike with more *elaborative* wordings. The factor analysis approach also revealed that the Bible differs from the Quran by adopting more “motion, direction, and transportation” information, reflecting the deviation in their historical and religious backgrounds.

1 Introduction

Rendering messages of the Gods¹ from scriptures might be the most intriguing yet challenging task for the faithful and spiritual shareholders. Religious texts are passed down from long ago, and practitioners are believed to rely conventionally on a literal interpretation of the scriptures. The messages rendered from the scriptures come through the endeavors of reverends or believers. However, the divine messages might have been altered or diverted

along with the time or varied interpretations. Are there other ways to hear from the Gods based on a deeper examination of the words of doctrine?

As one of the most adopted techniques for investigating linguistic data, Factor Analysis can identify clustered language features in a dataset comprised of various texts. If the factor score schemes (see 3.1) are applied, individual texts (presumably including the Quran and the Bible) will show particular preferences regarding their *linguistic factor* tendencies.

Factor Analysis relies on language tags to annotate each token used in the texts. Traditionally, multidimensional investigations were adopted to conduct genre analysis or stylistic categorization. Most tagging projects employ syntactic Part-of-speech (POS) annotations, and Stylometry has become one of the most prominent fields in linguistic studies. However, the Mandarin versions of the Quran and the Bible did not seem to have been viewed from a multidimensional perspective yet.

This paper aims to hear from the holy books using two factor-score systems: syntactic POS tagging and semantic word-meaning annotation. Some hidden messages are expected to emerge from the syntax-semantic approach to interpreting the religious scriptures, as holy texts like the Quran and the Bible can be distinctively regarded as a specific genre type. Some renderings or interpretations not brought up by early reverends or practitioners might be found from a multidimensional perspective. This paper tries to answer one fundamental question: how would the

¹ This paper withholds the polytheism/monotheism dichotomy; please see the author’s disclaimer in 5.3.

Quran and the Bible be measured by the Mandarin version of the factor score scales? Moreover, how would these scores reveal the untold/unknown messages from the scriptures?

2 Features of religious texts and linguistic multidimensional analysis

This section first reports on some common features found in the literature on religious texts, followed by the two sets of tagging systems: the syntactic POS tags and the semantic annotations.

2.1 Previous linguistic investigations of religious texts

Religious texts have always been an esteemed research direction for many linguists. Islamic and Christian scriptures received significant attention regarding their formality, word choice, or stylistics. The review in this section focuses on the commonly shared features found in religious texts.

Razzaq (2023) identified some linguistic features in Islamic holy scripture, which tend to maintain spirituality, morality, and religious practices using parallelism, repetition, and antithesis expressions. One essential character is the “persuasive strategies” of the Islamic scripture. Emotional, logical, authoritative, story-telling, and analogical expressions are used to achieve the goals.

Otabek et al. (2023) reported that Islamic scriptures are perceived as having a higher status. Religious expressions are not ordinary daily conversations with comparatively more prestigious wordings and lexicons involved. Religious texts adopt more figurative words to describe and more metaphorical expressions to “denote and preach”.

The “persuade” and “preach” purposes are also found in the Bible. Adetuyi (2020) stated that religious texts use graphological elements to *persuade* and win trust and belief. Religious texts achieve this goal by using word choices especially to inform, convince, persuade, or impress. The religious style was distinguished as an individual type of itself according to Crystal’s (1987) classification.

Kapranov et al. (2024) classified religious texts into informative and agitational types by

analyzing the core parameters of objectives, fundamental concepts, genre diversity, communicative purpose, and linguistic features. These core parameters are presumably related to the purpose of *preaching*.

Other functions and features of religious texts were also mentioned. Al-Ebadi (2012), studying specifically the Epistle of James with the system-functional approach, concluded that MP (material phrases), RP (relational phrases), and MnP (mental phrases) constitute more than 96% of religious words. The MPs were used to activate believers, the RPs to describe and identify religious relationships, and MnPs arose desire and belief as a warning to unfaithful people.

Uhunmwangho and Oghiator (2022) studied the syntactic structures in Chapter 9 of the Book of Proverbs and found that substantial repetitions achieve aesthetic, rhythmic, and cohesive values. The religious style is similar to the “contemporary living” one in Crystal and Davy (1969).

Acheoah and Abdulraheem (2015) compared the Gospel of Matthew in the Bible with the Evergreen Islamic Sermon. They found that the former uses more acceptable fragments and the latter more causal (e.g., *so*, *therefore*) and conditional (e.g., *if*) words. The fragments were meant to *impress* followers, while the causal/condition words were to *persuade*.

Several common factors emerged based on the short review of the selected literature reported above. Religious texts tend to *persuade* and *preach* with several distinctive functions, such as repetition, figurative words, and fragments to *impress* or *agitate* believers.

However, it is believed that some other factors also play a part in religious texts, which could be found by the multidimensional technique designed to identify distinctive factors in selected texts. It is assumed that the MDA approach can single out the hidden, untold, or unknown factors and the constituting features contributing to the linguistic factors.

2.2 MDA and stylistic studies

The MDA of languages has a time-honored tradition (Biber, 1986a, 1986b, 1988, 1992), and it has contributed significantly to genre and stylistic investigations in several languages, including

English, Nukulaelae Tuvaluan, Korean, and Somali (Biber, 1995). In Sardinha and Pinto (2014), linguistic features in Brazilian Portuguese and Spanish were identified by conducting MDA. Some specialized genres have also been further identified and discussed, such as language uses on the Internet and pre-Internet eras (Sardinha, 2014), languages in movies (Pinto, 2014) and pop songs (Bertoli-Dutra, 2014), and linguistic feature differences in *Time* magazine (de Souza, 2014). This paper follows the perspective of Crystal (1987) and Crystal and Davy (1969) that religious texts should also be classified as a specific genre.

The MDA approach has been applied to investigate languages used in function-specific scenarios. For example, Xiao (2009) studied the linguistic types and English dialects in multiple geographic distributions. With MDA, register-diversified corpora (Biber, 1993) became an effective tool for language studies. Using semantic tags, MDA has also been applied to business fields (e.g., Piao et al., 2015). Moreover, Cao and Xiao (2013) used the MDA in English to examine the contrast between native and non-native speakers. Huang and Ren (2019) compared different editorial styles used in *China Daily* and *The New York Times*. Ren and Lu (2021) compared the discussions in Chinese and American corporate annual reports.

Nevertheless, religious texts in Mandarin remain a genre seldom examined by the MDA approach. The analysis of linguistic tags appears to bear promising potential when paired with the appropriate interpretation. Applying the MDA to religious texts might lead to more fruitful results. This paper holds that Factor Analysis is a plausible approach to a deeper understanding of the two holy scriptures.

2.3 The factor components

Multidimensional analysis has been used to investigate stylistic differences or genre-type variation based on linguistic features reflected by co-occurring POS tags (as factors). The constituting features (the annotating tags) are crucial in conducting multidimensional analysis and linguistic stylometric studies. Factor analysis relied on the identified factors to compare how genre types differ in each factor.

This section reports on the factor distributions across languages. In Biber (1988, 1995), English

can be analyzed with six factors: (1) involved vs. *informational*; (2) *narrative* vs. non-narrative; (3) situation-dependent vs. elaborated; (4) overtly *argumentative* vs. not overtly argumentative; (5) non-abstract vs. abstract; (6) online informational vs. edited or not informational. (The 7th factor, academic hedging, was not included in the factor score scheme).

On a dialect in Chinese, Tiu (2000) identified the five factors for Taiwan Southern Min (TSM): (1) interpersonal vs. informational; (2) the personal expression of emotion; (3) *persuasion: logical* vs. temporal linking; (4) *narrative*; (5) involved exposition vs. precise reportage.

In Mandarin Chinese, seven factors were identified: (1) interpersonal vs. *informational*; (2) descriptive vs. vocal; (3) *elaborative* vs. non-elaborative; (4) explanatory vs. *narrative*; (5) locative vs. non-locative; (6) numeric vs. non-numeric; (7) indicative vs. casual.

The three sets of factor components in different languages share certain common factors. For example, delivering *information* is crucial in all three languages mentioned above. How one *narrates* is also a common concern. The factor components show that these three languages try to impress/persuade people from slightly different perspectives: “argumentative” in English, “logical” in TSM, and “elaborative” in Mandarin.

The three sets of factor components indicated that some factors are highly related to the religious texts’ word choices, especially the “persuade” and “preach” expressions (through elaborative, explanatory, or indicative factors), as they are found in both the literature and the factor listings. The typical “narrative” factor would be a vital indicator for analyzing religious texts, as it is reported in all three languages’ factor sets. These initial findings should be examined with prudence. This paper adopted a methodology based on the factor score schemes to see how religious texts are mapped on the scoring systems. The focus is on whether these factors in religious texts stand out when compared to the model/other text types.

3 Methodology

This section presents the two adopted analytical schemes and how to arrive at the factor scores. The factor scores are calculated based on a model corpus with 20 genre types of 28 million tokens (see Appendix (1)). Through MDA, 33 out of 71 POS tags and 24 out of 129 semantic tags were included. Appendices (4) and (5) listed

Code	Features	Factor loadings
Factor 1: Interpersonal vs. informational		
FPP3	Third-person pronoun	0.811
VE	Active verb with a sentential object	0.745
VK	Stative verb with a sentential object	0.558
Nh	Pronoun	0.525
Caa	Conjunctive conjunction	-0.374
A	Non-predicative adjective	-0.479
Na	Common noun	-0.549
VHC	Stative causative verb	-0.594
Factor 2: Descriptive vs. vocal		
VF	Active verb with a verbal object	0.765
Nc	Place noun	0.514
Nd	Time noun	0.470
EMPH	Emphatics	-0.395
DWNT	Downtoners	-0.514
AMP	Amplifier	-0.553
SHI	SHI “copula”	-0.554
Factor 3: Elaborative (vs. non-elaborative)		
Cbb	Correlative conjunction	0.768
CONC	Concessive adverbial	0.709
V_2	you “have/possess”	0.497
VJ	Stative transitive verb	0.478
Ng	Postposition	0.407
Factor 4: Explanatory vs. narrative		
CAUS	“because”	0.641
PRMD	Predictive modal	0.500
PERF	Perfect tense	0.418
DE2	DE “attribute/possessive marker”	-0.606
Factor 5: Locative (vs. non-locative)		
Ncd	Localizer	0.845
VCL	Active verb with a locative object	0.740
Factor 6: Numeric (vs. non-numeric)		
Neu	Numeral determinatives	0.894
Neqb	Post-quantitative determinatives	0.640
Factor 7: Indicative vs. casual		
P	Preposition	0.790
Dfb	Post-verbal adverb of degree	-0.472
I	Interjection	-0.673

Table 1: 7 POS factors and components

the exemplary tokens of the included tags. The factors listed in Tables 1 and 2 are based on the tokens and genre types as the *model*. The figures in Appendices (2) and (3) are the sum of standardized tag frequencies of the identified factor components (grouped sets of tags). The scores in each scale are used as indexes to indicate genre-type differences between the religious texts and the model texts.

Factors	1	2	3
Tags	Exposition	Events	Affection
demonstrative (M)	0.972		
number (M)	0.937		
factual (static v.)	0.713		
degree (M)	0.612		
evaluation (M)	0.533		
certainty (M)	0.395		
time (M)	-0.368		
sequence (M)	0.325		
animal (N)	0.438		
event (N)		0.745	
activity (N)		0.390	
idea (N)		0.353	
joy (N)			0.534
alive (M)			0.499
psych (static v.)			0.366
Factors	4	5	6
Tags	Motion	Places	Items
moving (acts/v.)	0.687		
direction (M)	0.493		
transportation (N)	0.453		
place (N)		0.578	
organization (N)		0.543	
fear (N)		0.315	
item (N)			0.542
academics (N)			0.377
sensation(static v.)			0.331

Table 2: 6 semantic factors and components (M=modifiers, N=nominals, v.=verbs)

3.1 Seven POS factors in Mandarin and calculating factor scores

As reported in 2.3, Mandarin corpus data are represented by seven factors based on POS tags. Table 1 lists the components (clustered features) for each factor. Each feature might have positive or negative factor loadings (ranging from 1 to minus 1). Features with more significant loadings have higher frequencies, whereas features with negative or lesser loadings have lower frequencies. For example, a text high on Factor 1 is interpreted as more “interpersonal”, for having higher “third-person pronouns, active/stative verbs with sentential objects, and pronouns” but fewer “conjunctives, non-predicate adjectives, common nouns, and stative causative verbs” tags. A text with an opposite formation of these tags is regarded as “informational”.

The feature components can be used to calculate factor scores with their normalized and standardized frequencies compared with the averaged frequencies and SDs (standard deviation) of the original model that reported the factor loadings. To illustrate, if one would like to calculate the factor score of a newly collected text. The factor

scores are arrived at by first normalizing the text's tag frequencies to per 1000 tokens. The normalization allows the intended text to be placed on the same frequency scale as the model texts. The normalized tag frequencies are then standardized using each tag's model frequency averages and SDs. For example, a text's Factor 1 score is the sum of the standardized frequencies of the FPP3, VE, VK, and Nh (positive loading features) tags, subtracted by that of Caa, A, Na, and VHC (negative loading features). The POS tag features of Factor 1 to Factor 7 are listed in Table 1. Appendix (2) contains the factor score scales for Factors 1 to 7, in which the score distributions of the four representative genres from the model and the two religious texts are illustrated. In section 4, the POS tag analysis compares the 7-factor scores of the Quran and the Bible.

In addition to the POS tagset, another factor analysis of Mandarin used semantic tags. The semantic multidimensional explorations identified six semantic factors in Mandarin. They are: (1) Exposition; (2) Events; (3) Affection; (4) Motion; (5) Places; and (6) Items. The semantic factor components are listed in Table 2. With the feature components based on semantic tagging, the semantic factor scores can also be calculated using the normalized and standardized semantic tag frequencies. The arithmetic procedures are the same as those reported in 3.1, with only the difference in semantic tag component features and their frequencies. Appendix (3) reports on the semantic factor scores of the representative text types on each factor.

4 Messages rendered by the factor scores

This section discusses the inferring from the religious texts through the POS and semantic factor scores. The two systems reported some common factors the Quran and the Bible shared. A few scripture-specific factors are also identified.

4.1 Hearing from the Gods with POS factor scores

How the two scriptures differ from the original model genre types can be illustrated by the score scales in Appendix (2). Each factor scale lists six representative type scores (two from each of the highest, middle-ranging, and lowest ones, sorted by scores and chosen from the 20 original model

genre types). The relative scale positionings of the religious texts illuminate their characteristics.

For Factor 1, the Quran and the Bible are highly “interpersonal” by the definition of Factor 1 features. However, the literature indicates that religious texts are supposed to contain substantial *information* (e.g., Razzaq, 2023, to “persuade” with story-telling, and Otabek, 2023, to “preach” with prestigious lexicons). These factor-score results do not seem to comply with the findings in the literature by having a high “interpersonal” score. Nevertheless, it is argued that there arises no contradiction here. It should be noted that the content of the two religious texts is *informative*, and their delivery style remains *interpersonal*. To account for this dual-directional division, the purpose of the scriptures should be considered when communicating with the believers and followers, and a friendly preaching approach helps to achieve the goal of offering information in an interpersonal style. The higher frequencies of the interpersonal features are believed to achieve the *imperative* and *advocative* purposes in persuading and preaching.

For Factor 2, both the Quran and the Bible are moderately “descriptive” compared to model text types. This again differs slightly from the literature that religious texts should be of “higher status” (Otabek et al., 2023). The analysis shows that neither holy scripture stood out in *descriptiveness* based on the POS tags. Their Factor 2 scores rank in the middle, similar to W4-newspaper reports and S2-documentary narratives, and maintain a moderately descriptive style. This tendency is believed to be a result of translating the holy books. When Islam and Christianity were introduced to Mandarin-speaking areas, the general public was the target of the religious missions. The Mandarin translations of the classic scriptures were made fairly descriptive (and colloquialized) to be understood by the masses of the society, and the descriptive features were used to *advocate* for certain causes.

For Factor 3, both religious texts are moderately “elaborative”. However, the Quran is slightly more “elaborative” than the Bible. Compared to the model genre types, the Quran has a factor score similar to that of W1-fictional works regarding elaborateness. It is not as detailed as

W6-commentaries (newspaper editorials), yet it maintains a touch of seriousness. The Bible, on the other hand, has a score similar to that of S8-TV variety shows. It adopts a more conversational approach when conveying messages. Again, the endeavor of the religious texts to *preach* in a tone similar to a daily-life conversation was made explicit from the factor score perspective. The *elaborative* tendency indicated the explanatory purpose of the religious texts.

For Factor 4, both scriptures are powerfully “explanatory” with relatively high factor scores. It is assumed that religious texts need to explain things in more detail. Both scriptures adopt a style similar to that of S3-TV news magazines. In this way, the religious texts can report and explain issues in a fashion that is as detailed as possible.

For Factor 5, the Bible is significantly more “locative” than the Quran. This situation might result from the comparatively higher frequencies of *instructions* and *calls for action* in the Bible. This does not mean the Quran did not use directional or action words at all. The contrast only indicated that the Quran uses *localizers* and *location-related verbs* more conservatively.

For Factor 6, both holy books are comparatively “non-numeric”. This lower use of numerical information can be expected as number terms are not the main subject of religious texts. Numbers are relatively refrained in the religious texts as more focus would be placed on the spiritual mentality.

For Factor 7, both scriptures are highly indicative, which reflects their “preaching” purpose. The *indicativeness* is realized by using the “interjection” and “degree adverbs”. It is assumed that the purpose of *impressing* followers also resulted in the higher use of these features.

The analysis based on POS tags demonstrated several features of the two religious scriptures under discussion. The common ones reflect that the religious texts are more interpersonal (F1), descriptive (F2), explanatory (F4), non-numeric (F6), and indicative (F7). The reasons behind these same features are believed to be that the religious texts must communicate with believers, explain ideas, give guidance, or convince/impress followers without relying on numbers. A couple of

text-specific features are also identified. The Quran is comparatively more non-elaborative (F3) and non-locative (F5) than the Bible. The Quran might resort to a style that uses fewer colloquial explanations to communicate with believers. As mentioned, both scriptures use fewer direction/action words, and the Quran cuts down even more on location information.

4.2 Interpreting the religious texts with semantic factor scores

The semantic factor score scheme for the Quran and the Bible also tells something exceptional that had not been revealed before. The score scales in Appendix (3) list the six sets of semantic factor scores of the model text types and that of the two holy scriptures. Some common factors are found among them; a few scripture-specific factors also emerged.

According to Factor 1 scores, both holy texts are low on the “exposition” factor, in which the exposition refers to a public-talking and spoken style. That is to say, both scriptures utilized comparatively *fewer* “factual, degree, evaluation, and certainty” features used in expressions. This strategy is believed to be due to the need for the holy texts to keep a *reserved, diplomatic*, and relatively *formal* style, to win the trust of followers and practitioners alike. Therefore, the *frank* and *direct* “exposition” factor is restrained for the religious texts, and a style similar to W10-captions (in a fashion analogous to story-telling) and W02-announcements (for raising awareness) was detected. It could be summarized that the way the religious texts convey to followers is *interpersonal* and *tactful*, using a “story-telling” style to *persuade* and “agitational” words to *preach*. This echoes and explains the perceivable POS contradiction of the bidirectional use of both informational and interpersonal features simultaneously, as discussed for the syntactic POS Factor 1.

For Factor 2, neither of the holy texts is high on the “events” factor. The expressions of “event and activity (see Table 2 for factor components)” seem cut down in the religious texts. This situation is probably because the expressions of *ideology, mentality, and spirituality* are more stressed in the holy texts.

For Factor 3, both are low on the “affection” factor. The “joy, alive, and psych (Table 2)” expressions constitute the affection factor. Re-examining the original texts with annotation tags, the “joy” tag is mainly tagged to *love* words, the “alive” to *life* words, and the “psych” to *willing/intend* verbs. This could prove that the Quran and the Bible do not employ a pretentious approach to communicating with followers. The holy texts involve more expressions to showcase *care, commitment, guidance for life, precaution, or advice* (partly illustrated by the excerpt in Appendix (6), in which a focus on “psych” words was used).

For Factor 4, the Bible is higher, while the Quran is moderate on the “motion” scale. Both Factor Analyses with POS and semantic tags identified the “localizer/moving” related factor as one of the key components in Mandarin. The Bible is believed to contain more descriptions of “moving, direction, and transportation (Table 2)” expressions. The POS Factor 5 discussed in 4.1 has already pointed out this unique characteristic of the Bible. As Appendix (7) demonstrates, wordings such as those in the Book of Exodus reported higher frequencies of semantic “motion” tags (e.g., “direction” and “moving” tags).

However, if the scope is limited to “place names” only (Factor 5), the Quran and the Bible are relatively low on the scale, indicating that place names are relatively rarely used in both the religious texts.

Finally, both holy texts are low on Factor 6 “items”, as sacred texts would abstain from materialistic expressions. The religious texts would focus more on the faithfulness and the spiritual aspects of human life.

The semantic factor score scheme identified five common factors shared by the Quran and the Bible (both are low on “exposition (F1), events (F2), affection (F3), places (F5), and items (F6)” factors). The semantic scales managed to notice the non-materialism in the holy texts. One distinctive factor of the Bible has been noted: there are comparatively more “motion (F4)” expressions (with the “moving, direction, and transportation” tags) than that in the Quran, and this is the only factor in the Quran that differs drastically from the Bible.

5 Conclusion

This paper investigated the two holy scriptures from a multidimensional perspective. The perceived features in the sacred texts were first reviewed in the literature. These features were then re-examined using two factor-score systems: the POS and the semantic tagging schemes. Some common factors and certain distinctive/scripture-specific characteristics were identified and discussed (see 4.1 and 4.2). In essence, both of the holy books use *impressive* and *elaborative* strategies to *persuade* and *preach to* followers. The Bible differs from the Quran in being more elaborative, locative, and motion-oriented. The author assumes these differences reflect the deviation in their religious/historical backgrounds. This section ends the paper by reporting on this study’s limitations, possible bias, and ethical concerns.

5.1 Limitations in building tagsets

This study adopted the POS and semantic tagsets in Mandarin. Compared to the available full POS tagset (all tokens can be POS tagged), the semantic one is still under development. The semantic factor score system has managed to semantically tag 67% of the tokens in the collected scriptures. Therefore, some might wonder why both of the religious texts tend to score low on the five semantic factors. It is contended that both religious texts performed the way reflected by the semantic factor scores, and tests using other genre types supported the accuracy of the semantic factor scales. The relatively insensitive prediction or identification power, if any as shown in 4.2, is probably due to the gap between an entirely constructed semantic tagset and the ongoing semi-completed one used in this study. A more complete semantic tagset will improve its categorizing and identifying abilities.

The annotating and tagging project required a substantial amount of manual work. It awaits further efforts to improve the effectiveness of the semantic Factor Analysis scheme. Future works can alleviate this limitation by perfecting the semantic tagging project. With a more complete semantic tagset, the analysis could be more precise, and the referencing results could be more persuasive.

5.2 Possible bias in the translated religious texts

This study focuses on the Mandarin versions of the Quran and the Bible because the author did not possess the necessary command of Arabic, Greek, or Hebrew (the original languages of the scriptures being written). The author resorted to their first-language translated versions. For the Quran, the included corpus was the traditional Chinese version of Muhammad Ma Jia (1906-1978), who studied in Egypt. The included corpus of the Bible was that of the Chinese Union Version (made by the Union Bible Societies). Therefore, the perspectives reflected Sunni Islam (not Shia Muslims) and Protestantism (not Catholicism or Orthodoxism). The findings through the factor analysis in this paper only partly reflect these two sects.

5.3 Ethical considerations and disclaimer

This study is simply an academic exploration based on the multidimensional approach with POS and semantic tags. The statements made by the author are of no intentional disrespect. On investigating the two holy books of Islam and Christianity, it is of no blasphemy intended to assume polytheism. The author himself is neither an Islamic nor a Christian, and the discussions were made with reasonings as objective as possible. Hopefully, this paper's findings, results, and claims could answer the research questions, including what untold messages could be rendered from the multidimensional perspective.

References

- John Emikr Acheoah and Hamzah Abdulraheem. 2015. Style in Christian and Islamic Sermons: A Linguistic Analysis. *American Research Journal of English and Literature*. 1(2), P23-31.
- Chris Adetuyi and Charles Patrick Alex. 2020. A stylistic analysis of selected Christian religion print advertisement in Ibadan metropolis, Oyo state. *Romanian Journal of English Studies*. (17), P81-90.
- Hani Al-Ebadi. 2012. A Systemic-Functional Analysis of Religious Texts concerning the Epistle of 'James'. *Journal of Thi Qar Arts*. 2. P1-20.
- Patrica Bertoli-Dutra. 2014. Chapter 2.2 Multidimensional analysis of pop songs. In Tony Berber Sardinha and Marcia Veirano Pinto. (Eds.) *Multidimensional Analysis, 25 years on*, P149-176.
- Douglas Biber. 1986a. On the investigation of spoken/written differences. *Studia Linguistica*, 40(1), 1-21.
- Douglas Biber. 1986b. Spoken and Written Textual Dimensions in English: Resolving the Contradictory Findings. *Language*, 62(2), 384-414.
- Douglas Biber. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Douglas Biber. 1992. The Multidimensional Approach to Linguistic Analyses of Genre Variation: An Overview of Methodology and Findings. *Computers and the Humanities*, 26(5), 331-345.
- Douglas Biber. 1993. Using Register-Diversified Corpora for General Language Studies. *Computational Linguistics*, 19(2), 219-241.
- Douglas Biber. 1995. *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge: Cambridge University Press
- Yan Cao and Richard Xiao. 2013. A multidimensional contrastive study of English abstracts by native and non-native writers. *Corpora*. 8(2), 209-234.
- David Crystal. 1987. *The Cambridge encyclopedia of language*. Cambridge University Press.
- David Crystal and Derek Davy. 1969. *Investigating English Style*. London and New York, NY: Routledge.
- Ying Huang and Wei Ren. 2019. A novel multidimensional analysis of writing styles of editorials from China Daily and The New York Times. *Lingua*, 235. Doi: 10.1016/j.lingua.2019.102781.
- Yan Kapranov, Olesya Cherkhava, and Anna Wierzchowska. 2024. Parameters of religious popular discourse within theolinguistic frameworks. *Journal of Modern Science*. DOI: doi.org/10.13166/jms/187200. P239-262.
- Bektoshev Otabek, Shodmonov Sherzod Hursanaliyevich, and Tirkashev Dilshod Sharobiddin ugli. 2023. The Specificity Of Religious Language In Modern Linguistics. *Journal of Positive School Psychology*. 7(1), P914-920.
- Scott Piao, Xiaopeng Hu, and Paul Rayson. 2015. Towards a Semantic Tagger for Analyzing Contents of Chinese Corporate Reports. *Proceedings of ISCC 2015*.
- Marcia Veirano Pinto. 2014. Dimensions of Variation in North American Movies. In Tony Berber Sardinha and Marcia Veirano

Pinto (Eds.) *Multidimensional Analysis, 25 years on-A tribute to Douglas Biber*. Amsterdam: John Benjamins Publishing.

Nasir Razzaq. 2023. Language and Religious Discourse: An Analysis of the Linguistic Features and Persuasive Strategies in Islamic Sermons. *IQAN*(10), 5(2), P17-34.

Chaowang Ren and Xiaofei Lu. 2021. A multidimensional analysis of the management's discussion and analysis narratives in Chinese and American corporate annual reports. *English for Specific Purposes*, 62, 84-99.

Tony Berber Sardinha and Marcia Veirano Pinto. (Eds.) 2014. *Multidimensional Analysis, 25 years on-A tribute to Douglas Biber*. Amsterdam: John Benjamins Publishing.

Tony Berber Sardinha. 2014. 25 years later: comparing Internet and pre-Internet registers. In Sardinha & Pinto (Eds.) *Multidimensional Analysis, 25 years on-A tribute to Douglas Biber*. Amsterdam: John Benjamins Publishing.

Renata Condi de Souza. 2014. Dimensions of variation in TIME magazine. In Sardinha & Pinto (Eds.) *Multidimensional Analysis, 25 years on-A tribute to Douglas Biber*. Amsterdam: John Benjamins Publishing.

Hak-khiam Tiu. 2000. A Multidimensional Analysis of Spoken and Written Taiwanese Register. *Language and Linguistics*, 1(1), 89-117.

Amen Uahunmwangho and Florence Etuwe Oghiator. 2022. The language of religion as discourse: analysis of the syntactic structures in Proverbs Chapter Nine. *International Journal of Culture and Religious Studies*. 3(1), P31-47.

Richard Xiao. 2009. Multidimensional analysis and the study of world Englishes. *World Englishes*, 28(4), 421-450.

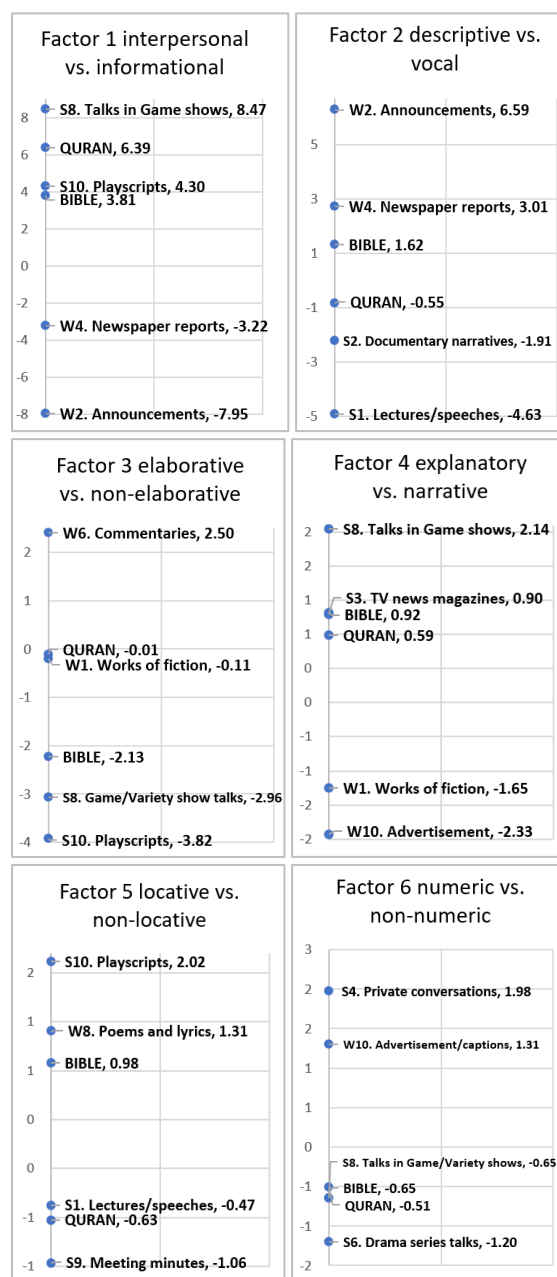
Appendices:

(1) Number of tokens by genre texts

Genres	No. of tokens
S1. Lectures/speeches	2,862,521
S2. Documentary narratives	3,712,698
S3. TV News magazines	2,987,873
S4. Private conversations	125,154
S5. Interviews (public conversations)	4,269,528
S6. Drama series talks	467,073
S7. Group/panel discussions	4,152,415
S8. Talks in game/variety shows	379,204
S9. Meeting minutes	9,767
S10. Play scripts	2,825

W1. Works of fiction	1,994,370
W2. Announcements	37,822
W3. Letters	79,985
W4. Newspaper reports	5,467,737
W5. Prose works	881,290
W6. Commentaries	1,118,735
W7. Biographies and diaries	27,637
W8. Poems and lyrics	35,858
W9. Manuals and handbooks	101,812
W10. Advertisements/pic. captions	23,429
Total (S1~W10 model texts)	28,737,733
The Quran	180,091
The Bible	792,600

(2) Scores Comparison in Seven Factors

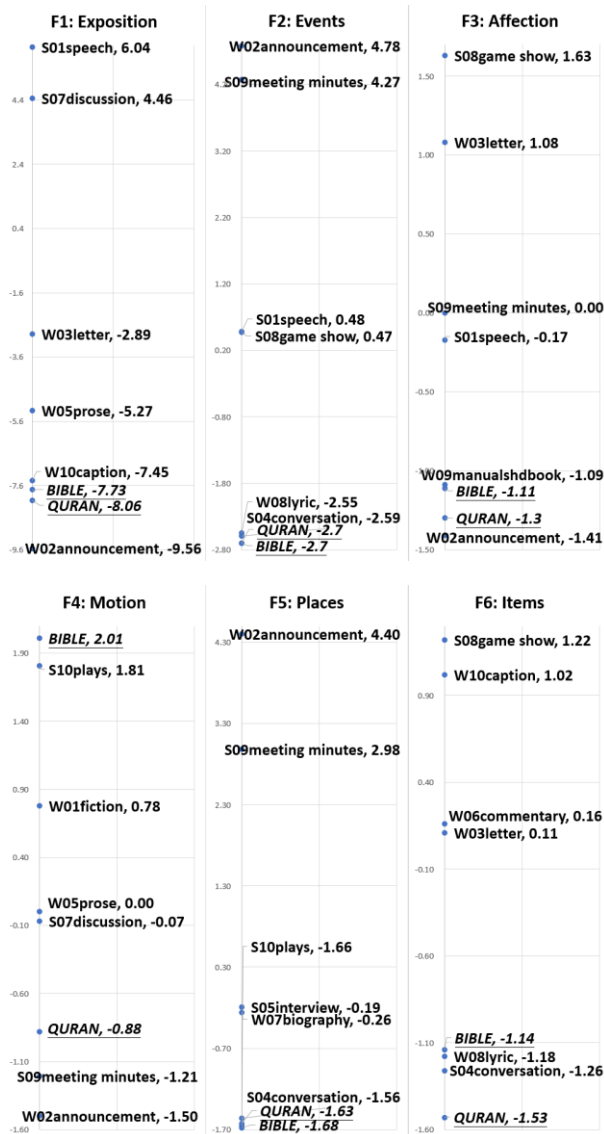




(4) The POS tagset (31 out of 77tags listed)

No	Tag	Full name	Exemplary tokens for each feature (tag)
A7	FPP3	Third person pronoun	<i>ta</i> 'he', <i>da-jia</i> 'everyone', <i>dui-fang</i> 'they'
B37	VE	Active verb with a sentential object	<i>suo</i> 'say', <i>gao-shu</i> 'tell'
B44	VK	Stative verb with a sentential object	<i>jiang-jiu</i> 'be strict about', <i>zhi-dao</i> 'know'
B27	Nh	Pronoun	<i>da-jia</i> 'everyone', <i>dui-fang</i> 'they'
B2	Caa	Conjunctive conjunction	<i>he</i> 'and'
B1	A	Non-predicative adjective	<i>ge-shi-ge-yang</i> 'various'
B15	Na	Common noun	<i>shou</i> 'hand', <i>guan-zhong</i> 'audience'
B41	VHC	Stative causative verb	<i>ping-heng</i> 'balance', <i>chang-sheng</i> 'produce'
B38	VF	Active verb with a verbal object	<i>ji-xu</i> 'continue'
B17	Nc	Place noun	<i>jia</i> 'home', <i>can-ting</i> 'restaurant'
B19	Nd	Time noun	<i>yi-qian</i> 'before', <i>zao-qi</i> 'early times'
A21	EMPH	Emphatics	<i>jiu-shi</i> 'exactly', <i>zhen-de</i> 'truly'
A18	DWNT	Downtoners	<i>ji-hu</i> 'almost', <i>hen-shao</i> 'rarely'
A20	AMP	Amplifiers	<i>jue-dui</i> 'absolutely', <i>que-shi</i> 'indeed'
B29	SHI	SHI	<i>shi</i> 'is' copula
B5	Cbb	Correlative conjunction	<i>ke-shi</i> 'but', <i>dan-shi</i> 'but'
A14	CONC	Concessive adverbial	<i>sui-ran</i> 'however'
B46	V_2	<i>You</i>	<i>you</i> 'have'
B43	VJ	Stative transitive verb	<i>shou-dao</i> 'affected by', <i>cheng-sian</i> 'show'
B26	Ng	Postposition	<i>shou(shang)</i> 'in hand'
A13	CAUS	"because"	<i>yin-wei</i> 'because'
A25	PRMD	Predictive modal	<i>yao</i> 'will', <i>ying</i> 'should', <i>xu</i> 'need to'
A2	PERF	Perfect tense	<i>yi-jing</i> 'already', <i>ceng-jing</i> 'ever'
B8	DE2	<i>De</i> (2)	<i>de</i> marker (<i>de</i> other than <i>de</i> (1))
B18	Ncd	Localizer	<i>shang-fang</i> 'on top', <i>di-bu</i> 'bottom'
B35	VCL	Active verb with a locative object	<i>lai (dao)</i> 'come to', <i>zhun-bei (shang)</i> 'prepare to'
B24	Neu	Numeral determinatives	<i>shi</i> 'ten', <i>si</i> 'four'
B22	Neqb	Post-quantitative determinatives	11 <i>dang (duo)</i> 'minutes past 11'
B28	P	Preposition	<i>zai</i> 'at', <i>dang</i> 'when'
B10	Dfb	Post-verbal adverb of degree	... <i>de duo</i> 'even more ...'
B14	I	Interjection	<i>dui-le</i> 'oh yes', <i>oh</i> 'oh'

(3) Semantic Scores in Six Factors



(5) The two-level semantic tags (24 out of 129 semantic tags listed)

Category	Sub-Category	Tag	Examples
Nominals (Nouns)	(01) Substantivity	(01) Animal	<i>shan-yang</i> 'goat', <i>da-xiang</i> 'elephant'
		(03) Organization	<i>xian-zheng-fu</i> 'council', <i>gong-si</i> 'company'
		(07) Transportation	<i>fei-ji</i> 'airplanes', <i>qi-che</i> 'cars', <i>zi-xing-che</i> 'bicycles'
		(09) Activity	<i>jiang-hua</i> 'talks', <i>hui-yi</i> 'meetings', <i>pai-dui</i> 'parties'
		(10) Place	<i>gong-che-zhan</i> 'bus station', <i>bo-wu-guan</i> 'museum'
		(11) Item	<i>sha-fa</i> 'sofa', <i>cha-zi</i> 'fork', <i>yao-shi</i> 'keys'
		(12) Academics	<i>shu-ji</i> 'books', <i>wen-zhang</i> , 'papers', <i>biao-ge</i> 'forms'
	(02) Positive-emotions	(20) Joy	<i>gao-xing</i> 'happiness', <i>yu-yue</i> 'amusement', <i>kuai-le</i> 'glee'
	(03) Negative-emotions	(25) Fear	<i>dan-you</i> 'concern', <i>dan-xin</i> 'worry'
	(04) Neutral-emotions	(31) Number	<i>yi-xiao-shi</i> 'one hour', <i>shi-nian</i> 'ten years'
		(32) Idea	<i>tui-li</i> 'reasoning', <i>luo-ji</i> 'logics'
		(34) Event	<i>ji-hui</i> 'convention', <i>shi-wu</i> 'affair(s)'
Acts (Verbs)	(05) Acts	(39) Moving	<i>yi-ju</i> 'migration', <i>chu-kou</i> 'exportation'
	(06) Static	(44) Psych	<i>nan-shou</i> 'mourn', <i>nan-guo</i> 'lament', <i>hou-hui</i> 'regret', <i>shi-huai</i> 'be mean', <i>wu-ru</i> 'belittle'
		(46) Factual	<i>cheng-gong</i> 'succeed', <i>jing-jue</i> 'caution', <i>hui-yi</i> 'recall', <i>fan-bo</i> negate', <i>shi</i> copula
Modifiers (adj/ad v)	(10) Status	(73) Alive	<i>sheng</i> 'alive', <i>si</i> 'dead'
	(11) Spatial	(78) Direction	<i>qian-jin</i> 'forward', <i>hou-tui</i> 'backward'
		(86) Demonstrative	<i>na-xie</i> 'those', <i>na-ge</i> 'that', <i>zhe-ge</i> 'this'
	(12) Temporal	(87) Time	<i>zao</i> 'early', <i>chi</i> 'late'
		(90) Sequence	<i>zhi-hou</i> 'after', <i>zui-hou</i> , 'last'
	(17) Judgment	(120) Evaluation	<i>hao-de</i> 'good', <i>xie-er-de</i> 'evil'
(122) Certainty		<i>que-ding</i> 'certain', <i>bu-queding</i> 'unclear/uncertain'	
(128) Degree		<i>quan-xin-quan-yi</i> 'full-hearted', <i>xie-wei</i> 'slightly'	

(6) Example excerpt (Al-Hujurat: 49-15)

xin-shi 'believers', (Na),
, (COMMACATEGORY),
zhi-shi 'only', (DWNT), judgment, effectiveness
que-xin 'believe', (PUBV),
zhen-zhu 'Allah', (Na),
he 'and', (Caa),
shi-zhe 'messengers', (Na),
, (COMMACATEGORY),
ran-hou 'then', (OSUB),
mei-you 'not', (D), judgment, accuracy
huai-yi 'doubt', (PRIV), static, **psych**
, (COMMACATEGORY),
neng 'able', (D), temporal, modal
yi 'by', (P),
zi-ji 'self', (Nh), substantiality, human
de 'possessive', (DE0),
cai-chan 'possessions', (Na),
he 'and', (Caa),
sheng-ming 'life', (Na), neutral-emo, **alive**
wei 'copula', (VG), static, factual
zhu-dao 'main road', (Na),
er 'then', (Cbb),
fen-dou 'fight', (VA),
de 'possessive', (DE0),
ren 'people', (Na), substantiality, human
; (SEMICOLONCATEGORY),
zhe 'this', (Nep), spatial, demonstrative
deng 'these', (Cab),
ren 'people', (Na), substantiality, human
que 'but', (D),
shi 'copula', (SHI), static, factual
cheng-shi 'honest', (VH),
de 'possessive', (DE0),

'The believers are only the ones who have believed in Allah and His Messenger and then doubt not but strive with their properties and their lives in the cause of Allah. It is those who are the truthful.'

(7) Example excerpt (Book of Exodus 19)

Token 'translation', (Syn Code), Sem code1, code2:
ta-men 'they', (FPP3), substantiality, human
li-kai 'leave', (VC)
le 'PAST-marker', (PAST)
li-fei-din 'Rephidim', (Nc)
lai-dao 'came to', (**VCL**), acts, **moving**
xi-nai 'Sinai', (Nc)
de 'possessive-marker', (DE0)
guang-ye 'wildness', (Na)

、 , (PAUSECATEGORY)
 jiu ‘then’ , (D), temporal, sequence
 cai ‘at’ , (P)
 na-li ‘there’ , (**Ncd**), spatial, **direction**
 de ‘possessive-marker’ , (DE0)
 san ‘mountain’ , (Na), substantiality, place
 xia ‘under’ , (PLA), spatial, location
 an-ying ‘camp’ , (VA)
 mo-xi ‘mosses’ , (Nb)
 dao ‘to’ , (P)
 shen ‘god’ , (FW)
 na-li ‘there’ , (**Ncd**), spatial, **direction**
 、 , (PAUSECATEGORY)
 ye-he-hua ‘Jehovah’ , (Nb)
 cong ‘from’ , (P)
 san ‘mountain’ , (Na), substantiality, place
 shang ‘up’ , (PLA), spatial, **direction**
 hu-huan ‘call’ , (VC)
 ta ‘him’ , (FPP3), substantiality, human
 suo ‘say’ , (VE), acts, motion

‘After they set out from Rephidim, they entered the Desert of Sinai, and Israel camped there in the desert in front of the mountain. Then Moses went up to God, and the Lord called to him from the mountain and said: ... ’