

## A Comparative Multidimensional Study of the English Translation of *Lunyu (The Analects)*: A Corpus-based Analysis

Mengyu He  
[mengyuhe1989@gmail.com](mailto:mengyuhe1989@gmail.com)  
School of Humanities,  
Universiti Sains Malaysia

### ABSTRACT

Although *Lunyu (The Analects)* is well-known and often mentioned in Confucian scholarship, there have been no focused examinations of the comprehensive linguistic features of its English translations. This study seeks to report a comparative multidimensional investigation into the similarities and differences in the lexico-grammatical features of *Lunyu (The Analects)* translated by James Legge and Ku Hungming. The comparison is made along five functional dimensions (involved versus informational production, narrative versus non-narrative concerns, explicit versus situation-dependent reference, overt expression of persuasion, and abstract versus non-abstract information), and the prominent lexico-grammatical features (based on a 67-item feature set) in the two texts are singled out. It is found that there are more private verbs, present tense verbs, *be* as main verb, past tense verbs, third-person pronouns, and public verbs in Legge's *The Analects of Confucius*, whereas Ku's *The Discourses and Sayings of Confucius* uses nouns, adjectives, long words, nominalisations, and time adverbials more often. The identified differences in lexico-grammatical patterns are related to the distinct goals of the two translators. The results demonstrate that the multidimensional (MD) approach is effective in differentiating the linguistic features of the two translation versions and motivating a micro-analysis of the texts, seeking to discern the translators' underlying assumptions about the relations of Confucius and his followers. It is considered that these findings may have implications for the understanding of the translations of *The Analects*.

**Keywords:** multidimensional (MD) approach; *Lunyu (The Analects)*; lexico-grammatical features; James Legge; Ku Hungming

### INTRODUCTION

*The Analects*, as one of the most well-known classics in China, serves as an important means by which ancient Chinese culture and overall morals are transmitted. By presenting the seminal wisdom and classic Chinese thought, *The Analects* exerts a significant influence on Chinese and East Asian thought and values. An instantiation is its influence on traditional Chinese rhetoric which values "collectivism" (Bloch & Chi, 1995). Moreover, the sayings in *The Analects* have proven highly effective for providing psychological support to individuals to work toward a decent manner (Daniel, 2001). Due to its importance, the study of *The Analects* has attracted considerable scholarly attention.

Over the past decades, scholars have carried out the study of *The Analects* in various fields, such as philosophy (e.g., Fung, 2014), education (e.g., Rui, 2010), ethics (e.g., Daniel, 2001) and translation (e.g., Roger & Henry, 1998). Studies that investigate the translation of *The Analects* are especially voluminous (e.g., Huang, 2012; Liu, 2014). The rich variety of these studies on the translation of *The Analects* has revealed the utility of various theories, such as hermeneutics and Chinese translation theory (e.g., 'faithfulness', 'expressiveness' and 'elegance' proposed by Yan Fu in 1898) in the analysis of *The Analects*. These studies

have enhanced our understanding of translated versions of *The Analects* by adopting different approaches. For example, Chen (2009) uses a functional approach to analyse the translation of *The Analects* based on the “Systemic Functional Linguistics” (Halliday, 1994). Chen’s study reveals the applicability of the functional discourse analysis to translation studies.

Despite the considerable insights gained on the translation of *The Analects*, remarkably little is known of the linguistic features of its translation. The multidimensional approach (henceforth MD approach), also known as a computerised tagging program, could be used to discover significant differences in terms of a number of features among different contexts (Grant & Ginther, 2000). However, no study to date has examined the translated versions of *The Analects* by using the MD approach. Multidimensional studies on translation are still embryonic. By selecting two representative translations, i.e., *The Analects of Confucius* translated by James Legge and *The Discourses and Sayings of Confucius* translated by Ku Hungming, this study performs an analysis of the lexico-grammatical features in the two translations. It then presents an investigation that compares the similarities and differences of the two texts. The MD approach can help us greatly grasp the substantial component of language in the two translations. The aim of this paper is to gain insight into the practicability of using MD approach in comparing different translated texts. It is hoped that the results of this analysis will contribute to a thorough and better understanding of *The Analects*.

#### DESCRIPTION OF MULTIDIMENSIONAL (MD) APPROACH

The MD approach was introduced by Biber (1984). It was later developed into a more comprehensive one by Biber (1988) in the studies of register variation. The theoretical underpinnings of the MD method are based on the work by linguists such as Hymes (1974) and Halliday (1988), in which the analyses of registers, genres and text types are emphasised. The use of MD approach enables observation of nearly a hundred linguistic features at the same time. This approach has proven to be useful in discourse analyses. Many researchers have drawn on MD approach to investigate the underlying linguistic questions of various corpora, such as the spoken corpora for stance and dialects (Precht, 2000), business corpora (Jiang & Xu, 2015), written texts (Crosthwaite, 2016). However, studies on translated texts using MD approach are still limited.

The MD approach was initially designed to study textual variation between speech and writing, and between different registers (Biber, 1984). Biber (1984) looks at a list of sixty-seven linguistic features by distinguishing spoken and written discourse based on London Lund Corpus and Lancaster Oslo Bergen Corpus. The analytical process in the MD method begins with measuring the frequencies of 67 features. A statistical analysis is then applied to measure the frequencies of these features. On the basis of the linguistic features, five to seven dimensions of register variation are proposed to link together the particular linguistic features.

The overall score for each register on each dimension is different. There are registers with low scores, middle scores or high scores on each dimension. These 67 linguistic features have their own variance with each linguistic feature showing a remarkably different mean score for the particular dimension. It is thus necessary to analyse the pool of shared variance, i.e., sets of linguistic features analysed in every dimension. Each linguistic feature has some relations to each factor and the strength of the relation is represented by factor loadings (Biber, 1988). Factor loadings can range from 0.0 to 1.0. Factor loading of 0.0 shows the absence of any relationship and 1.0 a perfect correlation. When interpreting a factor, features with positive loadings are considered (e.g., factor loading is greater than 0.35). The frequency counts of linguistic features are considered as *factor scores* (or *dimension scores*). In MD

approach, factor scores are standardised to a mean of 0.0 and a standard deviation of 1.0 represents standard deviation units. Once a factor score is calculated for each text, the mean factor score for each register can be computed. This allows a fuller functional interpretation of the underlying dimension to be realised. In the previous study of MD on spoken and written registers, Biber (1988) proposes seven dimensions from 67 linguistic features:

- Dimension 1: 'Involved versus Informational Production'
- Dimension 2: 'Narrative versus Non-Narrative Concerns'
- Dimension 3: 'Explicit versus Situation-Dependent Reference'
- Dimension 4: 'Overt Expression of Persuasion'
- Dimension 5: 'Abstract versus Non-Abstract Information'
- Dimension 6: 'On-line Informational Elaboration Marking Stance'
- Dimension 7: 'Academic Hedging'

Dimensions 6 and 7 have usually not been included in the analysis of register variation since they have few features with important loadings.

The MD analysis has been applied to studies of particular registers and texts. For example, using the MD approach, Conrad (1996) explores variation within academic texts. Other studies seek to explore a particular domain, such as public speaking (McEnery, Xiao, & Tono, 2006), academic written texts (Crosthwaite, 2016) to determine the dimensions of variation within the texts.

Despite the many MD studies, researchers have focused on particular areas of domain. With respect to Chinese studies, there have been few studies that explored Chinese translation. For instance, Hu (2010) compares the original Chinese and translated Chinese. Hu finds that MD analysis can effectively distinguish the differences between literary and non-literary texts, as well as Chinese and translated Chinese texts. He summarises that translated Chinese are more explicitly grammatical while Chinese idioms are usually associated with passive sentences. Based on the previous studies, the current study conducts a multidimensional analysis of the linguistic characterisation on two different translations of *The Analects*. The specific goals of this study are to investigate the linguistic differences across translation sections, to understand the similarities and differences between the two texts and to provide an overall multidimensional characterisation of the two translation versions.

## METHOD

### DATA

The data analysed in this study are versions of *The Analects* translated by Legge (1815-1897) and Ku Hungming (1898). *The Analects of Confucius* was translated by James Legge, typeset by Yi Su. The translation is downloadable from the website of The World of Free Plain Vanilla Electronic Texts. *The Discourses and Sayings of Confucius* was translated by Ku Hungming in 1898. The two texts were collected following the principles of representativeness and comparability. First, Legge's and Ku's versions represent advanced English translations of *The Analects* and are well-recognised by the scholars and the public. Second, the contents of texts are the translation from one text type. Prior to the analysis, illustrative language such as numerical data, footnotes, and examples were excluded in the analysis. This produces two electronic corpora of 28,074 and 37,466 words, respectively (see Table 1).

TABLE1. Description of the Corpus in the Current Study

Category	Tokens	Author
The Analects of Confucius	28,074	James Legge
The Discourses and Sayings of Confucius	37,466	Ku Hungming

### RESEARCH TOOL

The corpus tool, Multidimensional Analysis Tagger 1.3 (MAT) (Nini, 2015) was used in the study for linguistic tagging purposes. MAT corpus tool was developed by Andrea Nini (2015). It can annotate the selected corpus based on their grammatical and linguistic features as well as generate the statistics needed to perform a text-type or genre analysis.

The linguistic tagsets used in MAT follow the ones used in another tagger, i.e., the Stanford Tagger (2013), which is able to annotate the corpus with syntactic and part-of-speech information. MAT is able to automatically detect the extensive linguistic features as described in Biber (1988) to facilitate the preliminary linguistic analysis. As an attempt to replicate Biber's (1988) variation across speech and writing tagger for the multidimensional functional analysis of English texts, MAT is able to help visualise the *dimensional features* of an input text. The efficiency and accuracy of MAT has been tested and used by Jiang and Xu (2015), and Crosthwaite (2016).

### PROCEDURE

The multidimensional analysis, consisting of Dimensions 1, 2, 3, 4, 5 and 6 will be conducted using MAT. Discussions of dimensions will be extended from Dimension 1 to Dimension 5. Dimension 6 will not be included in the discussion because it has very few features with important loadings. First, the analysis was conducted by loading the corpora into MAT. MAT annotated the selected corpus with various linguistic features, such as private verbs (PRIV), demonstrative pronouns (DEMP), first-person pronouns (FPP1), etc. MAT then generated the relevant statistics, i.e., Z-score for the six dimensions and frequencies for each of the linguistic feature, normalised to per 100 word tokens. Apart from generating statistics, MAT also generated figures which help to approximate the genre/text type of *The Analects*. The classification of genre follows Biber's (1988) text type classification.

To sum up, MAT was used to carry out the quantitative analysis of the lexicogrammatical features of *The Analects of Confucius* (henceforth, Legge's AC) and *The Discourses and Sayings of Confucius* (henceforth, Ku's DSC), which was supplemented with a qualitative analysis as follows:

- 1) The factor scores on each of the first six dimensions were automatically computed. Comparisons were made across the two texts by following the text type classification developed by Biber (1988) (for example, see Figures 1 and 2).
- 2) The frequency of each linguistic characterisation of the two texts were automatically calculated and normalised to per 100 tokens. To identify features which might be overused or underused by Legge and Ku, features which were more statistically prevalent in the two sub-corpora were further analysed.
- 3) Finally, this study examined how the similarities and differences are influenced by some aspects of the translators' goals.

## RESULTS

The result of the MAT analysis consists of a number of output files. These files contain details of the comparisons between the two translated texts. The details include the dimension features following Biber's (1988) text type classification (see Figures 1 and 2), the frequencies as well as the Z-scores of all the linguistic features found in the texts (see Tables 3, 4, 5, 6 and 8). The frequencies and Z-scores were generated by normalising the word tokens to per 100 word tokens.

The Dimension scores of Legge's AC and Ku's DSC are reported in Table 2. In order to provide a clearer visualisation of the multidimensional analysis of the two translated texts, two figures displaying the possible genre of the texts based on Biber's (1988) eight text types (see Figures 1 and 2) are also included. Figures 1 and 2 below represent the 6 dimensional analyses of the two texts, respectively. Each dot represents the inclination of the particular dimension toward certain genre type.

TABLE 2. Dimension Scores for Legge's AC and Ku's DSC

	Legge's AC	Ku's DSC
Dimension 1	1.5	-3.62
Dimension 2	5.94	4.31
Dimension 3	3.24	4.68
Dimension 4	3.64	5.36
Dimension 5	0.28	0.16
Dimension 6	-1.76	-1.47

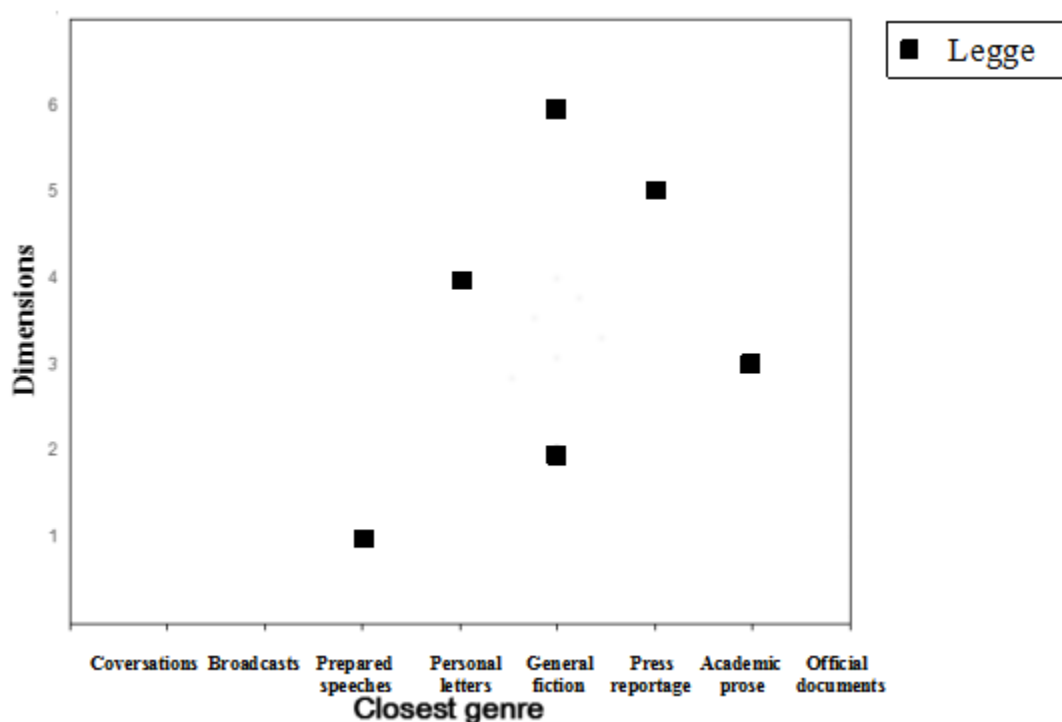


FIGURE 1. Mean Scores of Dimensions for Legge's AC along the Closest Genre

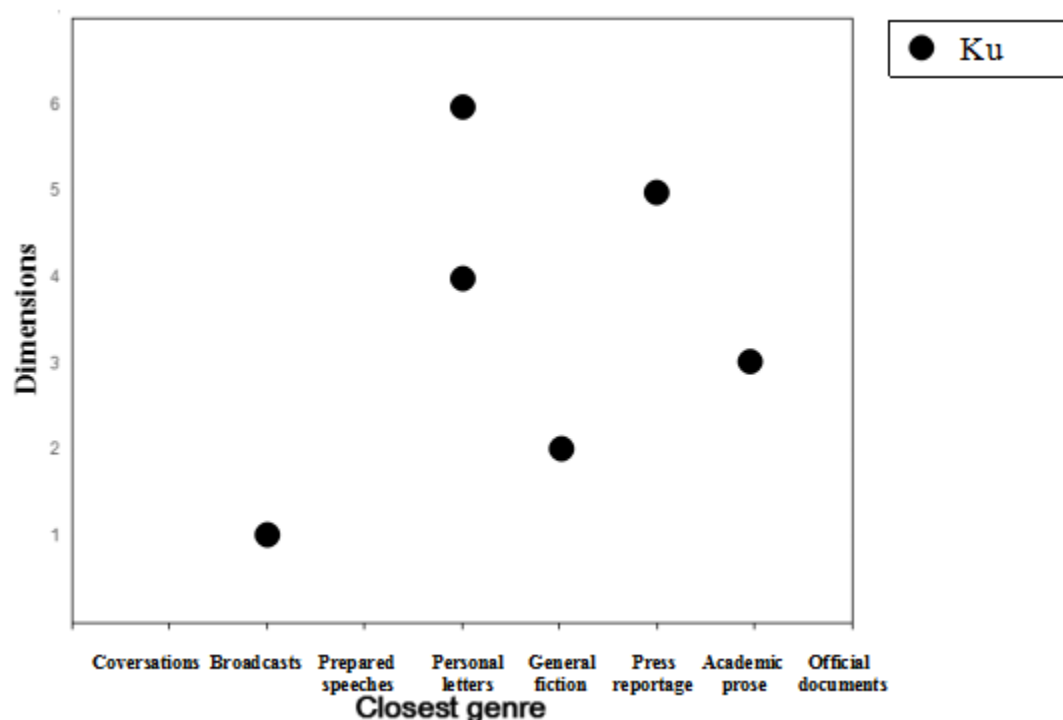


FIGURE 2. Mean Scores of Dimensions for Ku's DSC along the Closest Genre

The results of closest genre (approximation of genre type) can be seen in Figures 1 and 2. As the collection of sayings is attributed by Confucius, *The Analects* is shaped by indiscrete dialogues. However, in Figures 1 and 2, it could be seen that *The Analects* is distinct from ordinary conversations. Besides, of all the five dimensions, *The Analects* is seen distinct from the genre of conversation.

Overall, the MD results show that the two texts are similar in dimensions 2, 3, 4 and 5. This highly indicates that they were written similarly in style. Figures 1 and 2 also show that the two texts are different in dimensions 1 and 6 based on the mean scores. Dimension 1 will be delineated in the following section, while Dimension 6 will be excluded from discussion as there are very few features with important loadings in this dimension.

## DISCUSSION

The linguistic features along Biber's six dimensions together with their functional roles are presented and discussed in the following section. The possible relations between the quantitative results and the stances of the authors of the two translation versions are discussed in detail following the discussion of the linguistic features.

### DIMENSION 1

According to Biber (1988), Dimension 1 is defined as "high informational density and exact informational content versus affective, interactional, and generalised content" (p. 107). This dimension differentiates between *involved production* and *informational production*. The language of *involved production* has an affective/interactive focus. It could be detected by identifying the frequent use of verbs and pronouns in the texts. The language of *informational production* is of highly integrated propositional content. It is obvious when there are many nouns and adjectives used by the author (Biber, 1988). As shown in Table 2, Legge's AC has comparatively high involved scores on this dimension, while Ku's DSC has high



informational scores. Legge uses fewer linguistic features (e.g., *nouns*, *attributive*, etc.) which are functional for creating more information and more linguistic features (e.g., *present tense verbs*, pronoun *it*, etc.) concerning with the interpersonal involvement. Such expressions of verb tense, according to Reppen (1994), are the “immediacy of the issues addressed by the text” (p.62), demonstrating text involvement. Ku, on the other hand, tends to downplay involved production (e.g., less *present tense verbs*, pronoun *it*) and places greater stress on high amounts of information, for instance, the frequent use of nouns.

Table 3 compares the frequencies (normalised to per 100 words) of positive-loading (involvement) features and negative-loading (informational) features on Dimension 1 between Legge’s AC and Ku’s DSC’s.

TABLE 3. Frequencies of Dimension 1 Features in Legge’s AC and Ku’s DSC (per 100 words)

	Legge’s AC	Ku’s DSC
Number of words (raw frequency)	28,074	37,466
<b>Positive-loading features</b>		
private verbs (PRIV)	1.66	1.62
that deletions (THATD)	0.09	0.18
present tense verbs (VPRT or VBP; VBZ)	5.8	5.41
second-person pronouns (SPP2)	0.79	1.32
demonstrative pronouns (DEMP)	0.56	0.42
emphatics (EMPH)	0.25	0.42
first-person pronouns (FPP1)	1.97	1.84
pronoun <i>it</i> (PIT)	1.16	0.92
<i>be</i> as main verb (BEMA)	2.75	2.42
causative adverbial subordinators (CAUS)	0.04	0.04
discourse particles (DPAR)	0.05	0.11
indefinite pronouns (INPR)	0.09	0.08
amplifiers (AMP)	0.07	0.11
<i>wh</i> questions (WHQU)	0.28	0.14
possibility modals (POMD)	1.03	0.73
<i>wh</i> clauses (WHCL)	0.15	0.1
<b>Negative features</b>		
nouns (NN)	18.86	19.67
prepositions (IN)	1.05	1.02
attributive adjectives (JJ)	4.71	5.56
average word length (AWL)	4.24	4.33

Table 3 shows that the dimension score of each text does not rely on one linguistic feature alone. For example, Legge’s AC and Ku’s DSC have similar frequencies of private verbs, which suggest some functional similarities between the two texts. Private verbs are observed based on the categorisations made by Quirk, Greenbaum, Leech, and Svartvik (1985). They are usually mental verbs (e.g., *think*, *believe*). The tendency to use private verbs by Legge and Ku is a reflection of spoken language as shown in *The Analects* (e.g., 子曰：“...我未见力不足者...” translated by Ku in English as Confucius remarked: “...I do not **believe** he will find that he has not the strength to do it...”). Although there is similar use of private verbs in both translated texts, the comparisons of other linguistic features between the two texts indicate the differences between the two texts.

As can be seen in Table 3, negative scores figure more prominently in Ku’s DSC than in Legge’s AC. Nouns, attributive adjectives, and average word length serve a similar function that pack the sentence densely with information about specific referents.

Nouns are the main means by which writers refer to specific entities or a set of concepts, and their high frequency reflects a high density of information. Using more nouns to convey information is an easily discernible feature in Ku’s DSC. Two samples extracted from Legge’s AC and Ku’s DSC are provided below:

Sample 1

1.7 子夏曰:“贤贤易色; 事父母, 能竭其力; ...”

**Tsze-hsia** said, “If a man withdraws his mind from the love of beauty, and applies it as sincerely to the love of the virtuous; if, in serving his parents, he can exert his utmost strength; ...”

EXCERPT FROM LEGGE’S AC

**A disciple of Confucius** remarked, “A man who can love worthiness in man as he loves beauty in woman; who in this duties to his parents is ready to do his utmost, ...”

EXCERPT FROM KU’S DSC

Sample 2

2.5 孟懿子问孝...

**Mang I** asked what filial piety was...

EXCERPT FROM LEGGE’S AC

**A noble of the Court in Confucius’s native State** asked him what constituted the duty of a good son.

EXCERPT FROM KU’S DSC

Sample 1 and Sample 2 reveal the differences (see noun phrases in bold) in Legge’s AC and Ku’s DSC. When translating the name of Confucius’ students, such as 子夏(Zixia), 孟懿子(Meng Yizi) in the above samples, Legge tends to transliterate them into **Tsze-hsia** and **Mang I**, while, Ku seeks to suggest a detailed description, i.e., **a disciple of Confucius** and **a noble of the Court in Confucius’s native State** by using more nouns. This is an overall trend across Legge’s AC and Ku’s DSC. The findings are in conformity with previous studies (e.g., Wang, 2008; Chen, 2009). Legge adopts a translation strategy of domestication by retaining information from the source text, while Ku follows the translation way of foreignisation by making *The Analects* closely conform to the Chinese culture.

The following samples also underscore the difference between Legge’s AC and Ku’s DSC.

Sample 3

5.25 颜渊, 季路侍...

Yen Yuan and Chi Lu being by his side,...

EXCERPT FROM LEGGE’S AC

**On one occasion**, when two of his disciples, the **favourite Yen Hui** and **Chung Yu the intrepid**, were in attendance on him,...

EXCERPT FROM KU’S DSC

In Sample 3, Ku provides more descriptions than Legge. “**On one occasion**” in DSC is additional information translated by Ku, which is different from word-for-word translation underlying in Legge’s AC. Ku frequently sets up scenarios between Confucius and the students in translation in order to be more precise; these scenarios are realised in the forms of nouns and attributive adjectives. A noun phrase containing an attributive adjective also displays a covert proposition. For example, the noun phrases “**favourite Yen Hui**” and “**Chung Yu the intrepid**” in Ku’s DSC presuppose the propositions “Yen Hui is favourite” and “Chung Yu is intrepid”. The relatively fewer noun phrases in Legge’s text could contribute to an air of objectiveness and conciseness about his translation.



Another recurrent feature characterising Dimension 1 is the word length. Biber (1988) suggests that longer words convey more specific, specialised meanings. Heightening a sense of information, Ku builds up his text with a relatively longer lexical length. This strategy is evident in Sample 4.

#### Sample 4

1.4 曾子曰：“吾日三省吾身，为人谋而不忠乎？与朋友交而不信乎？传不习乎？”

The philosopher Tsang said, “I daily *examine myself* on three points: whether, in transacting business for others, I may have been not **faithful**; whether, in **intercourse** with friends, I may have been not **sincere**; whether I may have not *mastered* and *practised* the **instructions** of my teacher.”

EXCERPT FROM LEGGE’S AC

A **disciple** of Confucius remarked, “I daily *examine* into my personal conduct on three points: First, whether in carrying out the **duties** entrusted to *me* by others, I have not *failed* in **conscientiousness**; Secondly, whether in **intercourse** with friends, I have not *failed* in **sincerity and trustworthiness**; Thirdly, whether I have not *failed to practise* what I profess in my teaching.”

EXCERPT FROM KU’S DSC

Of the sample analysed here, Ku positions frequent nouns and long words, such as **disciple**, **duties**, **conscientiousness**, **intercourse**, **sincerity** and **trustworthiness** to attain a highly informational feature. Long words refer to words which have more letters/syllables. In contrast, Legge uses fewer nouns and long words. Legge’s translation is more “involved” or more context-dependent. This quality is achieved mainly through pronouns. The positive features in Legge’s text are seen conveying a sense of involvement as well as creating a bond of solidarity between his translation and the reader. For example, when translating “吾日三省吾身”(wú rì sān shěng wú shēn), Legge uses *myself* in “I daily examine myself on three points”, while Ku’s “I daily examine into my personal conduct on three points” maintains distance from Ku’s perspective on this sentence. The use of *myself* seems to create a more affective response in relation to how Legge felt and thought about the compulsory daily deeds, on the other hand, the use of “my personal conduct” by Ku sounds more behavioristic in nature.

To summarise, for this Dimension 1, the scores reveal the different linguistic features used by Legge and Ku. One of the most distinctive features of Legge’s translation is the high frequency of verbs and pronouns and low frequency of nouns, long words and adjectives. The relative infrequent informational forms, such as the use of fewer nouns and attributive nouns, contribute to the sense of interpersonal involvement in his translation. In other words, Legge’s translation does not seem to take into account the context where the morals and advices were taught. This is evident as there are word-by-word translations found in Legge’s text. Meanwhile, Ku’s translation exhibits more features associated with informational production. These features are realised by nouns, attributive adjectives and long words.

## DIMENSION 2

Dimension 2 is the visualisation between Narrative and Non-Narrative discourse. Legge’s AC and Ku’s DSC both valorise high positive scores on Dimension 2, with a concomitant of narrativity. The closest genre of the two texts is general fiction, which demonstrates high narrativity. The features contributing to a positive score on Dimension 2 are *past tenses*, *third person pronouns* (e.g., *replied*; *he*), and *public verbs* among other features. Public verbs are

those verbs which can be seen (e.g., *write, tell*) (Quirk, Greenbaum, Leech, & Svartvik, 1985). They are usually known as activity verbs. The negative features of Dimension 2 are present tense verbs and attributive adjectives. Table 4 provides the mean scores for Legge’s AC and Ku’s DSC along Dimension 2.

TABLE 4. Frequencies of Dimension 2 Features in Legge’s AC and Ku’s DSC (per 100 words)

	Legge’s AC	Ku’s DSC
Number of words (raw frequency)	28,074	37,466
<b><i>Positive-loading features</i></b>		
past tense verbs (VBD)	5.4	4.59
third-person pronouns (TPP3)	5.6	4.82
perfect aspect verbs (PEAS)	0.59	0.56
public verbs (PUBV)	2.88	2.31
synthetic negation (SYNE)	0.31	0.27

As can be seen from Table 4, Legge’s AC and Ku’s DSC are similarly narrative in focus. Legge’s AC has a comparatively higher concentration of narrative features. The same subject matter of *The Analects* corresponds to the similar frequencies of these features in the two texts. *The Analects* is much like a fiction in narrating Confucian doctrine. Being involved in dialogues, essentially in *The Analects*, Confucius transmits Confucian values and exhorts his students to be ethically well-cultivated men in a story-telling way. Table 4 also identifies a difference in overall mean scores, with a lower frequency of positive features found in Ku’s DSC than in Legge’s AC.

Sample 5 presents a pattern of linguistic features between Legge’s AC and Ku’s DSC on Dimension 2. The positive linguistic features for Dimension 2 are italicised and in bold, while negative Dimension 2 features are only in bold.

#### Sample 5

2.5 孟懿子问孝。子曰：“无违。”樊迟御，子告之曰：“孟孙问孝于我，我对曰：‘无违。’”樊迟曰：“何谓也。”子曰：“生，事之以礼；死，葬之以礼，祭之以礼。”

Mang I ***asked*** what filial piety ***was***. The Master ***said***, “It is not being disobedient.” Soon after, as Fan Ch’ih ***was driving him***, the Master ***told him***, saying, “Mang-sun ***asked*** me what filial piety ***was***, and I ***answered him***,- ‘not being disobedient.’” Fan Ch’ih ***said***, “What ***did*** you mean?” The Master ***replied***, “That parents, when alive, be served according to propriety; that, when dead, ***they*** should be buried according to propriety; and that ***they*** should be sacrificed to according to propriety.”

EXCERPT FROM LEGGE’S AC

A noble of the Court in Confucius’s native State ***asked him*** what ***constituted*** the duty of a good son. Confucius ***answered***, “Do not fail in what is required of you.”

Afterwards, as a disciple ***was driving him*** in his carriage, Confucius ***told*** the disciple, saying, “My Lord M – ***asked*** me what constituted the duty of a good son, and I ***answered***, ‘Do not fail in what is required of you.’”

“What ***did*** you mean by that ?” asked the disciple.

“I meant,” replied Confucius, “when his parents ***are*** living, a good son should do his duties to them according to the usage prescribed by propriety; when ***they are*** dead, ***he*** should bury ***them*** and honour their memory according to the rites prescribed by propriety.”

EXCERPT FROM KU’S DSC

Both of the two passages in Sample 2 show a large number of narrative features such as employing verbs of motion (e.g., *was driving him*) to describe a series of events. A lot of past tense verbs highlighted in the samples above indicate the performers' action. However, a greater difference resides in the fact that virtually Ku uses more present tense verbs compared to a few present tense verbs found in Legge's translation. For example, when referring to the translation of the sentence in Sample 5, the main linguistic features distinguishing Legge's and Ku's texts are present tense verbs.

### DIMENSION 3

The occurring positive features characterising Dimension 3 (Explicit versus Situation-dependent Reference) are *Wh* relative clauses on object positions (e.g., the man *who Lily likes*), pied piping constructions (any preposition followed by *who*, *whose*, or *which*. e.g., the manner *in which he was told*), *Wh* relative clauses on subject positions (e.g., the girl *who likes popcorn*), phrasal coordination and nominalisations (any noun ending in *-tion*, *-ment*, *-ness*, or *-ity*, plus the plural forms). The negative features on Dimension 3 are time adverbials, place adverbials and adverbs.

This dimension can also be referred to as Elaborated versus Situation-dependent Reference (Biber, 1995). Conrad and Biber (2001) write that "Dimension 3 represents a continuum between texts that have elaborated, explicit reference, and reference that is more dependent on the situational context" (p.33). For example, the positive features of *Wh* relative clauses and nominalisations have an affinity with densely integrated and informational elaborated discourse. On the contrary, the negative features commonly refer to times and places outside of the text.

The Dimension 3 results (see Table 5) can be understood in relation to the factor scores on Dimension 1 (Involved versus Informational Production), on which Ku's text also contains more features associated with information production. Table 5 reports the results of mean scores of Legge's AC and Ku's DSC.

TABLE 5. Frequencies of Dimension 3 Features in Legge's AC and Ku's DSC (per 100 words)

	Legge's AC	Ku's DSC
Number of words (raw frequency)	28,074	37,466
<b><i>Positive-loading features</i></b>		
<i>Wh</i> relative clauses on object positions (WHOBJ)	0.06	0.11
pied piping constructions (PIRE)	0.16	0.12
<i>Wh</i> relative clauses on subject positions (WHSUB)	0.17	0.37
phrasal coordination (PHC)	0.59	0.88
nominalisations (NOMZ)	1.73	1.79
<b><i>Negative features</i></b>		
time adverbials (TIME)	0.18	0.45
place adverbials (PLACE)	0.21	0.2
adverbs (RB)	3.22	3.6

As shown in Table 5, Ku is seen attempting to strike a balance in the use of elaboration and reference techniques as both features gain high scores, respectively.

The contrasting poles of Dimension 3 are exemplified by Samples 6 and 7. The positive linguistic features for Dimension 3 are italicised and in bold, while negative Dimension 3 features are only in bold.

Sample 6

8.10 子曰：“好勇疾贫，乱也。人而不仁，疾之已甚，乱也。”

The Master said, “The man **who is fond of** daring **and** is dissatisfied with poverty, will proceed to **insubordination**. So will the man who is not virtuous, when you carry your dislike of him to an extreme.”

EXCERPT FROM LEGGE’S AC

Confucius remarked, “A man of courage **who hates to** be poor will be sure to commit a crime. A man without moral characters, if too much hates, will also be sure to commit a crime.”

EXCERPT FROM KU’S DSC

Sample 7

3.18 子曰：“事君尽礼，人以为谄也。”

The Master said, “The full observance of the rules of propriety in serving one’s prince is accounted by people to be flattery.”

EXCERPT FROM LEGGE’S AC

Confucius remarked, “Men **now** account in servile to pay to their prince all the honours due to him.”

EXCERPT FROM KU’S DSC

In Sample 6, Legge’s AC and Ku’s DSC both illustrate well the use of *wh*-relative clauses to elaborate noun referents. *Wh*-relative clauses on subject position in Sample 6, such as **who is fond of** and **who hates** provide a transparent description for readers, implying the exact objects in the sentence. In contrast, in Sample 7, the negative feature of time adverbials is also evident in Ku’s DSC, such as the time adverbial **now** is used to refer to times outside of the text itself in an imaginary world created by the text.

**DIMENSION 4**

Dimension 4 only has positive features: infinitives (e.g., *to...*), prediction modals (e.g., *will, would*), suasive verbs (e.g., *agree, decide, recommend*), conditional subordination, necessity modals (e.g., *should, ought*), split auxiliaries (e.g., *they are objectively shown that...*), and possibility modals (e.g., *might, may*). This dimension is interpreted as marking overt persuasion or argumentation of the author’s perspective or indexing the author’s construction of relations between propositions in the discourse. The factor scores of the two text samples for Dimension 4 are shown in Table 6.

TABLE 6. Frequencies of Dimension 4 Features in Legge’s AC and Ku’s DSC (per 100 words)

	Legge’s AC	Ku’s DSC
Number of words (raw frequency)	28,074	37,466
<b>Positive-loading features</b>		
infinitives (TO)	1.98	2.19
prediction modals (PRMD)	0.78	0.89
suasive verbs (SUAV)	0.68	0.5
conditional subordination (COND)	0.53	0.49
necessity modals (NEMD)	0.42	0.57
split auxiliaries (SPAU)	0.23	0.51
possibility modals (POMD)	1.03	0.73

The Dimension 4 factor scores of Legge’s AC and Ku’s DSC are similar to personal letters proposed by Biber (1988), suggesting an average level of over persuasive effort. In overall, Ku’s text has higher scores than that of Legge, suggesting Ku’s more overt persuasive effort. Of the six features which index ‘overt persuasive effort’ in the two texts, infinitives occur with the highest frequency in Legge’s AC and Ku’s DSC. Leech and Short (1981) state that “infinitives are generally non-factive...[indicating] possibilities and hypotheses rather than known facts” (p. 104). Therefore, infinitives are likely to heighten the sense of overtness of persuasive effort.

Table 6 also depicts the differences of features on Dimension 4 between Legge’s AC and Ku’s DSC. The possibility modals listed by Biber (1988) are *can*, *may*, *might*, and *could*. One of the differences is that Legge’s AC features a statistically greater use of possibility modal verbs than Ku’s DSC (see Table 7). For instance, the use of modal verb “can” in Legge’s AC is noteworthy. This modal verb is mostly used to describe ability in Legge’s translation, in which it is contextually used to describe the ability of a noble man. It occurs three times more often in Legge’s AC than in Ku’s DSC. The infrequent use of *can* in Ku’s translation could possibly be due to Ku’s preferences in using different modality verbs to indicate his stance and attitude, such as *may*, *could*.

TABLE 7. Frequency of Possibility Modals in Legge’s AC and Ku’s DSC

possibility modals	Legge’s AC		Ku’s DSC	
	Raw frequency	Frequency per 100 words	Raw frequency	Frequency per 100 words
<i>can</i>	139	0.50	61	0.16
<i>may</i>	66	0.24	79	0.21
<i>might</i>	17	0.06	10	0.03
<i>could</i>	27	0.10	44	0.12

Another difference is that Ku uses more frequently split auxiliaries than Legge. A split auxiliary places an adverbial in a salient position. The adverbial often indicates the writer’s stance toward the predicate or its complement. An example of this feature is shown in Sample 8. The positive linguistic features for Dimension 4 are italicised and in bold.

#### Sample 8

1.2 有子曰：“其为人也孝弟，而好犯上者，鲜矣；...”

The philosopher Yu said, “They *are few* who, being filial and fraternal, are fond of offending against their superiors....”

EXCERPT FROM LEGGE’S AC

A disciple of Confucius remarked, “*A man* who is a good son and a good citizen *will seldom be* found to be a man disposed to quarrel with those in authority over him;...”

EXCERPT FROM KU’S DSC

In Sample 8, “其...鲜矣 (it/he/she/they...is/ are few)” is translated directly by Legge, who uses ‘be’ as a main verb (*They are*) followed by an adjective (*few*). However, Ku adopts ‘a split auxiliary’ to translate it as “*A man...will seldom be*”. The adverbial ‘*seldom*’ in this split auxiliary reveals Ku’s stance towards ‘*A man*’, suggesting such ‘*A man*’ occurs rarely. The split auxiliaries reflect that Ku and Legge tend to employ different strategy in the translation. The strategy lies in the emphasis the writer wishes to give and the rhetorical effects he hopes to create.

### DIMENSION 5

Dimension 5 differentiates between the abstract and non-abstract styles, where ‘abstract style’ approximates ‘passivised style’ since passive constructions are the main underlying linguistic feature. Dimension 5 has only positive features. These features are conjuncts (e.g., *thus*, *however*), agentless passives (e.g., *he was acclaimed a hero*), past participial (passive) adverbial clauses (e.g., *Given these characteristics, it is not surprising that...*), *by-* phrases, past participial (passive) postnominal clauses (also called past participial WHIZ deletions e.g., *a picture taken by him*), and other adverbial subordinators (that is, those with multiple functions, not consistently causative, concessive or conditional). Most of these features are passives, presenting information with little or no emphasis on the agent. Biber (1988) interprets relatively high scores on Dimension 5 as marking abstractness of information. Table 8 shows the mean scores of positive features on Dimension 5 in Legge’s AC and Ku’s DSC.

TABLE 8. Frequencies of Dimension 5 Features in Legge’s AC and Ku’s DSC (per 100 words)

	Legge’s AC	Ku’s DSC
Number of words (raw frequency)	28,074	37,466
<b><i>Positive-loading features</i></b>		
conjuncts (CONJ)	0.15	0.23
agentless passives (PASS)	1.34	0.91
past participial adverbial clauses (WZPAST)	0.07	0.11
<i>by-</i> passives (BYPA)	0.07	0.04
past participial postnominal clauses (PASTP)	0.04	0.05
other adverbial subordinators (OSUB)	0.11	0.1

Of the six features, the difference lies in *agentless passives* as employed in Legge’s AC and Ku’s DSC. Legge uses higher frequency of *agentless passives* in AC than Ku in DSC. The sample below illustrates the typical use of *agentless passives* by Legge. The positive linguistic features for Dimension 5 are italicised and in bold.

#### Sample 9

1.12 有子曰：“礼之用，和为贵...”

The philosopher Yu said, “In practising the rules of propriety, a natural ease is ***to be prized***”.  
 EXCERPT FROM LEGGE’S AC

A disciple of Confucius remarked, “In the practice of art, what is valuable is natural spontaneity.

EXCERPT FROM KU’S DSC

In Sample 9, ‘为’ (*wéi*) in the source text can be interpreted as ‘*by*’. Legge translates it as ***to be prized*** by using *agentless passive*, following the source language. In contrast, Ku seems to seek for sense-to-sense translation by using adjective *natural* and nominalisation *spontaneity*. “Agentless passive” involves promotion of an indirect object to subject, suggesting that Legge drew on the same syntagmatic combination as the original passive sentence in Chinese, while Ku translates it in an active voice by “is natural spontaneity”. Adjective and nominalisation are almost relentless in Ku’s translation, as seen in the aforementioned Dimension 1. *The Analects* is composed of dialogues, presenting an ostensible conversation. The closest genre of *The Analects* on Dimension 5, however, is press reportage, which indicates its non-abstract style.



### REASONS: THE SIMILARITIES AND DIFFERENCES

The reported results have revealed both similarities and differences in the translation of *The Analects* between Legge and Ku's English versions. While they are similar in Dimensions 2, 3, 4, and 5, suggesting Legge's AC and Ku's DSC belong to the similar genre type on these four dimensions, Ku uses more nouns, adjectives, long words, nominalisations, time adverbials, adverbs, prediction modals, necessity modals but fewer private verbs, present tense verbs, *be* as main verb, past tense verbs, third-person pronouns, public verbs, possibility modals and agentless passives than Legge. The linguistic features revealed in the similarities and differences are typically concerned with the strategies used by Legge and Ku in their translation. Two strategies of translation are observed through the analyses of generated linguistic features in the texts, one strategy centred on domestication by Legge and another centred on foreignisation by Ku. The first strategy adopted by Legge realises a form of translation maintaining compliance with the original Chinese syntactic patterns and another by Ku centred on transforming the source language into a different syntactic pattern during the translation process. The different strategies in their translation could be ascribed partly to their goals of translation.

A similar finding was revealed in Wang's (2008) comparative study of the English translation of Confucian *Analects* of Legge's version and Ku's version, in which he found that the length of Ku's version is much longer than Legge's version. Wang (2008) superficially compared the lengths of both texts. This seems insufficient to differentiate between the two translated texts. The current study bridged the gap by performing a much deeper and detailed analysis of both texts using MD approach.

As Liao (2009) suggests, the meaning of an utterance is related to the goals pursued and the most important element of context resides in a consideration of the participants' goals. Legge was a missionary. He translated *The Analects* as he was convinced of the need for missionaries to be able "to comprehend the ideas and culture of the Chinese and thus in 1841 he began the long task of translating the Confucian Classics into English" (Terry, 2016, p.99). His translation work is for the benefit of the missionaries in China. For example, the use of 'agentless passive' and the modality 'can' has the effects of simplifying. Direct word-for-word translation with appropriate simplification is necessary as it helps other missionaries in comprehending *The Analects* in a more efficient way. This is thus beneficial for the missionaries back then.

In contrast to the missionary goals underlying Legge's translation practices, Ku intended to "redefine the original meaning of Chinese Confucian classics and eradicate the prejudices caused by the translations of James Legge and other missionaries" (Wang, 2008, p.48). Translation and culture are interrelated (Fatemeh, Imran-Ho-Abdullah, & Norsimah, 2013). Cohering well with Chinese virtues and Western cultures, Ku translated Confucian classics in order to change the westerners' stereotypical stance towards Chinese people and alleviate the tense relations between China and England after westerners' reading. Ku contextualised the morals and advices so that the Confucian teaching is accurately and comprehensively presented in the translated text. For example, Ku tended to use the noun phrase containing an attributive adjective to translate names in *The Analects*. As shown in Sample 3, the noun phrases "favourite Yen Hui" and "Chung Yu the intrepid" in Ku's DSC imply that "Yen Hui is favourite" and "Chung Yu is intrepid". Ku's translation indicates that Yen Hui is Confucian's favourite student. The contextualisation of moral advice is aptly presented in Ku's translation, showing that a man should be righteous. Given the underlying goals, it is sensible to see word-for-word translation in AC by Legge and sense-for-sense translation in DSC by Ku. It is worth noting that meanings are comprehended when the word senses diffused across the text are extracted within the context (Fakharzadeh & Mahdavi,

2017). Ku's effort in maintaining the translated meanings within the context seemed exemplary of the translated texts of *The Analects*.

## CONCLUSION

This study has identified complex lexico-grammatical patterns in *The Analects of Confucius (AC)* and *The Discourses and Sayings of Confucius (DSC)*. Although they do not differ in the overall dimensions of multidimensional analysis, the two translated texts are distinct from each other in the specific types of linguistic features. Legge's *The Analects of Confucius* deploys a domesticating method by using more private verbs, present tense verbs, *be* as main verb, past tense verbs, third-person pronouns, public verbs, etc., whereas Ku's *The Discourses and Sayings of Confucius* uses nouns, adjectives, long words, nominalisations, time adverbials more often under the strategy of foreignisation. The identified differences in lexico-grammatical patterns could possibly be related to the distinct goals of the two translators.

Several broad implications follow from the findings for studying and translating *The Analects*. First, *The Analects* as a Chinese classic has been taught and learned in China and worldwide. An obscure understanding of the classic is likely to misrepresent the nature of Chinese culture and fail to capture its diversity and complexity. Therefore, in order to learn and comprehend *The Analects*, the translation and the analysis of linguistic features of the translated texts provide the readers and researchers a clearer understanding of what *The Analects* are and how *The Analects* are perceived by the readers and researchers. Second, it is pedagogically sound to raise teachers and students' awareness of the linguistic features in translation practices to facilitate lexico-grammatical understanding. Such awareness-raising can be achieved through teaching which links translation practices to lexico-grammatical translation practices. Teachers may introduce the relevant software such as "MAT tagger" in the classroom when they discuss linguistic features of various texts. This would expose the students to the way linguistic features are analysed by adopting MD approach. Furthermore, the current study has demonstrated the usefulness of MD approach in analysing translated texts. It has shown that MD approach is able to provide more thorough perspectives on textual variation across and within the classic literary texts.

To sum up, in order to have a better and thorough understanding of the English translations of *The Analects*, one has to examine the various linguistic features and perspectives associated with the texts. Any single perspective in analysing the texts could be insufficient to provide valuable insights into the nature of the texts. A comprehensive and thorough examination using MD approach is therefore useful in identifying the similarities and differences between the similar and comparative contexts.

This study is essentially exploratory in nature, as it is the first using MD approach in analysing translated texts. As the study was based on only two translated versions without the use of inferential statistics, the findings and observations are subject to further statistical analysis by future research. Besides, future research could investigate if MD approach is suitable to be used in a more comprehensive translated text. It might be possible to extend MD approach by looking into other specialised discourse.

## NOTES

1. The description of the above variables can be found here: <https://sites.google.com/site/multidimensionaltagger/Manual.pdf?attredirects=1>
2. A list of the Stanford Tagger tags and the description of how they are identified can be found here: <http://catalog.ldc.upenn.edu/docs/LDC99T42/tagguid1.pdf>

## ACKNOWLEDGMENTS

The author would like to thank the Institute of Postgraduate Studies of Universiti Sains Malaysia for the postgraduate fellowship award which helped to fund this research.. The author would also like to thank her supervisor Assoc. Prof. Dr. Hajar Abdul Rahim for her endless encouragement and support. Also special thanks go to the anonymous reviewers for their insightful comments.

## REFERENCES

- Bloch, J., & Chi, L. (1995). A Comparison of the Use of Citations in Chinese and English Academic Dis-course. In D. D. Belcher & G. Braine (Eds.). *Academic Writing in a Second Language: Essays on Research and Pedagogy* (pp. 231-274). Norwood, NJ: Ablex.
- Biber, D. (1984). *A Model of Textual Relations within the Written and Spoken Modes* (Unpublished doctoral dissertation). The United States, University of Southern California.
- Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, D. (1995). *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge: Cambridge University Press.
- Chen, Y. (2009). A Functional Analysis of Quotations from *Lun Yu*. *Foreign Languages and Their Teaching*. Vol. 2, 49-52.
- Conrad, S. (1996). Investigating Academic Texts with Corpus-based Techniques: An Example from Biology. *Linguistics and Education*. Vol. 8, 299-326.
- Conrad, S., & Biber, D. (2001). *Variation in English: Multi-dimensional Studies*. *Studies in Language and Linguistics*. Harlow: Pearson Education/Longman.
- Crosthwaite, P. (2016). A Longitudinal Multidimensional Analysis of EAP Writing: Determining EAP Course Effectiveness. *Journal for English for Academic Purposes*. Vol. 22, 166-178.
- Daniel, F-C.T. (2001). How Should Doctors Approach Patients? A Confucian Reflection on Personhood. *Journal of Medical Ethics*. Vol. 27, 44-50.
- Fakharzadeh, M., & Mahdavi, B. (2017). Full Sentence vs. Substitutable Defining Formats: A Study of Translation Equivalents. *3L: Language Linguistics Literature®*, *Southeast Asian Journal of English Language Studies*. Vol. 23(2), 167-179.
- Fatemeh, S., Imran-Ho-Abdullah, & Norsimah, M. A. (2013). Rendering Happiness Metaphors: A Cognitive Analysis From Persian Into English. *GEMA Online® Journal of Language Studies*. Vol. 13 (2), 193-205.
- Fung, Y. M. (2014). *Ren* As A Heavy Concept in *The Analects*. *Journal of Chinese Philosophy*. Vol. 41(1-2), 91-113.
- Grant, L., & Ginther, A. (2000). Using Computer-Tagged Linguistic Features to Describe L2 Writing Differences. *Journal of Second Language Writing*. Vol. 9(2), 123-145.
- Halliday, M. A. K. (1994). *An Introduction to Functional Grammar (2nd edn)*. London: Edward Arnold.
- Halliday, M. A. K. (1988). On the Language of Physical Science. In M. Ghadessy (Ed.). *Registers of written English*. London: Pinter Publishers.
- Hu, X. Y. (2010). A Corpus-based Multi-dimensional Analysis of the Stylistic Features of Translated Chinese. *Foreign Language Teaching and Research*. Vol. 42(6), 451-458.

- Huang, G. W. (2012). A Unique Feature of Translating Ancient Chinese Works: From Intralingual Translation to Interlingual Translation. *Foreign Languages in China*. Vol. 9(6), 64-71.
- Hymes, D. (1974). *Foundations in Sociolinguistics: An Ethnographic Approach*. Philadelphia: University of Philadelphia Press.
- Jiang, J. L., & Xu, J. J. (2015). A corpus-based multi-dimensional analysis of business English registers. *Foreign Language Teaching and Research (bimonthly)*. Vol. 47(2), 225-236.
- Leech, G. N., & Short, M. H. (1981). *Style in Fiction: A Linguistic Introduction to English Fictional Prose*. London: Longman.
- Liao, M. Z. (2009). The Principle of Goal Direction and Communication. *Foreign Language Research*. Vol. 4, 62-64,
- Liu, L.F. (2014). The Translation of *Lun Yu (The Analects of Confucius)* in Russia. *Foreign Languages in China*. Vol. 5, 96-103.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based Language Studies: An Advanced Resource Book*. New York: Routledge.
- Nini, A. (2015). *Multidimensional Analysis Tagger (Version 1.3)*. Available at: <http://sites.google.com/site/multidimensionaltagger>
- Precht, K. (2000). *Patterns of stance in English* (Unpublished Ph.D. dissertation). Northern Arizona University, Flagstaff, AZ.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.
- Reppen, R. (1994). *Variation in elementary student language: A multi-dimensional perspective* (Unpublished doctoral dissertation). Northern Arizona University, Flagstaff.
- Roger, T. A., & Henry, R. J. (1998). *The Analects of Confucius: A Philosophical Translation*. The United States: Ballantine Books.
- Rui, Y. (2010). Soft Power and Higher Education: An Examination of China's Confucius Institutes. *Globalisation, Societies and Education*. Vol. 8(2), 235-245.
- Stanford Tagger. (2013). *Stanford Tagger v. 3.1.5*. Retrieved from: <http://nlp.stanford.edu/software/tagger.shtml>
- Terry, C. M. (2016). *Why Study Religion?: Understanding Humanity's Pursuit of the Divine*. Washington: Baker Academic Press.
- Wang, D. B. (2008). *A Comparative Study of the English Translation of Confucian Analects- With Special Emphasis on James Legge's version and Gu Hongming's version* (Unpublished doctoral dissertation). ShangDong University, ShanDong.

#### ABOUT THE AUTHOR

Mengyu HE is currently a PhD candidate at the English Language Studies section, School of Humanities, Universiti Sains Malaysia. Her PhD research is funded by USM Fellowship. Her research interests include discourse analysis, corpus linguistics and phraseology.