

The Feasibility of Chinese–English Machine Translation Applied to Academic Texts: Using Thesis Abstracts from National Digital Library of Theses and Dissertations (NDLTD) in Taiwan

Pin-Ling CHANG* and Chien-Hua HSU**

In 2016, Google upgraded its translation system Google Translate (GT) from statistical-based machine translation (SMT) to neural-based machine translation (NMT), the latest machine translation (MT) technology so far, resulting in faster translation speed and higher accuracy. However, according to Angkana Tongpoon-Patanasorn and Karl Griffith (2020), most previous studies focused on statistical-based GT, while few tested the quality of GT's latest NMT system. This study explores the feasibility of applying GT to Chinese–English academic texts and analyzes whether the English translation produced by GT meets the requirements of English academic writing (AW). The original data used in this study is comprised of Chinese abstracts collected from the website of National Digital Library of Theses and Dissertations in Taiwan. Specifically, twenty most clicked thesis abstracts from each of the five most clicked disciplines, i.e., Engineering, Business and Management, Society and Behavior, Education, and Humanities, are selected, amounting to a total of 100 abstracts. Through a qualitative coding method and content analysis, the results show that three AW features can be found in the English translation output, including this-format, mid-positioning of adverbs, and passive voice. In terms of verb choice, nearly half of the abstracts show use of phrasal verbs, indicating GT's inability to constantly adopt single-word verbs when translating academic texts from Chinese into English. Also, GT performs best on abstracts from Engineering discipline. By exploring the feasibility of applying GT to Chinese–English academic texts, this study may help contribute to a better understanding of and facilitating use of MT in academic circles.

Keywords: machine translation; Google Translate; academic texts; academic writing; Chinese–English translation

1. Research Background

Translation is one way to make communication easier and faster, which has been an indispensable activity since human beings were able to communicate; however, not everyone

* Associate professor at Chung Yuan Christian University (CYCU), Taoyuan.

E-mail: pinling.chang@cycu.edu.tw; ORCID ID: <https://orcid.org/0000-0001-7086-104X>.

** MA in Applied Linguistics and Language Studies, Chung Yuan Christian University (CYCU), Taoyuan.

E-mail: hsushahoa@gmail.com; ORCID ID: <https://orcid.org/0000-0001-9513-8616>.

(Received 10 October 2022; accepted 10 December 2022)

can afford human translation, which tends to be costly and time-consuming, leading to the invention of machine translation (MT). Since the 1990s, MT has been more rapidly developed with the advancement of computer technology, and many researchers have used statistical MT (SMT) system, which was released in 1991 and has boosted the accuracy and speed of MT (Yang 2018), to investigate more about MT. For instance, Peter F. Brown et al. (1990) propose a statistical approach called “parameter estimation” to test MT, while David Chiang (2005) presents a hierarchical phrase-based model for SMT. MT has been widely researched at both research centers and universities from the perspective of computational linguistics. Meanwhile, the technology of deep learning was involved in the MT system later in 2000, leading to the birth of neural MT (NMT), with both Baidu (a China’s version of Google) and Google launching their own NMT system (Yang 2018).

In the past decade, the research of MT has changed from pure study to practical application (Kao 2011), with different text genres used to test the quality of MT, such as medical texts (Yu 2019), technical texts (Yang 2018), and audiovisual texts (Burchardt et al. 2016). Haniyeh Sadeghi Azer and Mohammad Bagher Aghayi (2015) even compare the quality of different MT systems with six different text types. Despite the considerable number of MT-related studies from different perspectives, the quality of MT output from academic texts has yet to be extensively researched. Most of the previous studies into academic texts mainly focus on how to improve the quality of academic writing (AW) or academic abstracts in English (e.g., Klimova 2013; Sijia and Jie 2015) or how to avoid common writing problems of English academic abstracts in different disciplines (e.g., Zhang 2014; Zhong 2016). Literature review shows that there is still little research of the quality of MT used to produce academic abstracts. AW in English is relatively difficult for non-native speakers to manage. In order to master it, students and novice researchers normally need to be trained. As a result, to bridge the gap, this paper aims to explore the feasibility of MT applied to academic texts and to investigate whether MT could translate Chinese abstracts into English ones that meet the requirements of AW in English.

2. An Introduction to Academic Writing

AW is one of text genres that is worth exploring because of its practicality. Entering keywords such as ‘academic writing’ or ‘學術寫作’ *xué shù xiě zuò* on Google search results articles introducing its functions, features, or templates for reference. However, AW could differ in disciplines and languages. The AW in this paper is defined as general AW techniques, which means different writing methods or styles from different disciplines will not be compared. John M. Swales and Christine B. Feak specify that AW is a product of many considerations: “audience, purpose, organization, style, flow, and presentation” (2012, 3). It is hard for one to master AW without practice. Moreover, since English and Chinese are two very different linguistic systems, AW differs in both languages as well. English abstracts are required in most academic text cases, and MT could be one of the options for those whose native language is not English to translate their non-English thesis abstracts. Accordingly, it is worth exploring whether MT could generate adequate translation that meets the features of AW.

Michael Groves and Klaus Mundt (2015) conduct a quality assessment of Google Translate (GT) renditions of short AWs from Malay and Chinese into English. Their research materials are collected from students enrolling in a pre-university foundation course, containing a total of 1,523 words in Malay and 744 words in Chinese written in the students’ first language. Their findings show that GT is able to handle lexical and grammatical aspects. They argue that as Google’s database grows, the accuracy of GT could be improved. On the other hand, Angkana Tongpoon-Patanasorn and Karl Griffith (2020) use GT to test its quality and usability from Thai to English with academic texts, using 54 research abstracts from eight major disciplines of Humanities and Social Sciences. They suggest that the quality of MT is adequate and useful for language learners.

While still many more different text types have been researched in connection with MT, there is little research related to academic translation via MT from Chinese into English. Also, the previous MT studies that have examined the quality of MT used to translate academic texts

into English mainly focus on grammatical accuracy (Groves and Mundt 2015; Tongpoon-Patanasorn and Griffith 2020) instead of investigating whether MT could generate translation output of adequate AW style. Therefore, this study aims to bridge the research gap by exploring whether GT could generate academic text translations from Chinese to English that follow the rules of AW in the hope that Chinese–English MT might be useful for researchers who are non-native speakers of English.

3. Research Design

This paper adopts GT as the research tool. GT is an online MT service known for its convenience and free of charge service. Its neural-based system has increased the accuracy and quality of translation output. However, little research has specifically examined this neural-based translation technology since its release in 2016. According to Tongpoon-Patanasorn and Griffith (2020), most studies that evaluate the quality of GT focus on its rule-based and statistical-based system. Therefore, this study aims to assess the updated GT and provide an in-depth investigation of academic texts.

The research materials are collected from the website of National Digital Library of Theses and Dissertations in Taiwan (NDLTD in TW).¹ All theses and dissertations written in Taiwan are documented on this website, which is available and accessible to the general public. A total of the abstracts of 100 most-clicked² theses from the most-clicked five disciplines were chosen (i.e., twenty abstracts per discipline), which are Engineering (EG), Business and Management (B&M), Society and Behavior (S&B), Education (ED), and Humanities (H) disciplines. The research questions are:

(1) What are the features of AW that can be found in the English translation of the selected thesis abstracts translated by GT?

¹ Available at <https://ndltd.ncl.edu.tw/cgi-bin/gs32/gswweb.cgi/ccd=k0s83F/webmge?mode=basic>.

² The definition of ‘most clicked’ theses: The system has operated since June 10th, 2000. This ranking can be sorted by institution, department, advisor, or graduate. To avoid cheating and hacking, every thesis can only be clicked once by the same person per day.

(2) In which discipline(s) (Engineering, Business and Management, Society and Behavior, Education, and Humanities) is MT suitable for Chinese–English thesis abstract translation?

As AW puts much focus on textual cohesion and coherence, it should be reasonable that the present study adopts human evaluation. Moreover, the coding themes are rather complex and may not be fully filtered out through software such as corpus tools. As a result, it is necessary to adopt manual qualitative coding and human evaluation. According to the literature review, the most salient features of AW at lexico-grammatical level are ‘this,’ ‘adv.,’ ‘passive,’ and ‘phrasal v.’ Other special features are also examined during the coding process. Furthermore, an online Google EXCEL is adopted, and a parallel corpus, AntPConc³ is constructed for further comparison among disciplines. It is hoped that the results could reveal significant discovery and the present study could make a contribution in academic circles.

4. Research Results

4.1 Data Analysis and Frequency of Appearing Academic Writing Features

The collected data are statistically analyzed through word count and frequency of appearing academic features. The data are segmented and aligned for visual convenience and for AntPConc analysis. The sentences are segmented as follows: (1) when a period is seen in the ST, the sentences before are counted as a segment; (2) if the sentences after a colon are bullet points, the sentences before the colon are counted as a segment; (3) sentences will not be segmented by a semicolon; (4) the sentences before a question mark are counted as a segment. Table 1 demonstrates the word count of Chinese source texts (STs) and target texts (TTs) from each discipline.

³ Available at <https://www.laurenceanthony.net/software/antpconc/>.

Table 1. Information of collected data

	ST word count	AVG word count per abstract (of ST)	TT word count	AVG word count per abstract (of TT)	Segment
EG	8845	442	5595	280	159
B&M	13186	659	7936	397	195
S&B	13192	660	8343	417	254
ED	13512	676	8200	410	292
H	12033	602	7744	387	203
SUM	60768	608	37818	378	1103

As can be seen, the average of ST word count is about 600, which corresponds to the suggested length (approximately within a page) of an abstract (Tsai 2014, 58). As for the average of TT word count, it also corresponds to the suggested length. EG appears to have the least word count. B&M, S&B, and ED seem to share the similar number of ST word counts; however, their number of segments differs. The gap of segment numbers between B&M and ED is about 100. This may be attributed to the use of bullet point format in ED. Bullet points are usually connected with a period, which separates two segments.

To answer research question one—‘What are the features of AW that can be found in the English translation of the selected thesis abstracts translated by GT?’—Table 2 shows the total of major appearing AW features in each discipline. The bracket in ‘adv.’ indicates the number of non-preceding adverbs. In average, both ‘this’ and ‘adv.’ are used twice in an abstract; ‘passive’ five times; and ‘phrasal v.’ less than once. Regardless of ST word count, each discipline shows signs of ‘this’ feature. This makes sense because in an abstract, the author should definitely refer to the study itself in order to introduce the study to readers.

Table 2. Frequency of major AW features from five disciplines

	EG	B&M	S&B	ED	H	Total	AVG per abstract
this	39	40	47	60	55	241	2.4
adv. (X)	45 (9)	31 (1)	49 (12)	36 (5)	49 (7)	210 (34)	2.1 (0.3)
passive	118	97	143	86	112	556	5.6
phrasal v.	4	13	18	14	15	64	0.6

In terms of the highest used themes, ‘this’ is frequently used in ED; ‘adv.’ is frequently used in S&B and H; ‘passive’ is frequently used in S&B; ‘phrasal v.’ is less used in EG. It is great to see that phrasal verbs are not used much because it is recommended to replace two-word verbs with a synonym single-word verb (Bailey 2003). In conclusion of the answer to research question one, ‘this,’ ‘adv.,’ and ‘passive’ are the three major features of AW that could be found in the English translation of the thesis abstracts translated by GT in the present study.

Table 3. Percentage of AW features appearing in each discipline

	EG	B&M	S&B	ED	H	AVG
this	25%	21%	19%	21%	27%	22%
adv. (X)	28% (20%)	16% (3%)	19% (24%)	12% (14%)	24% (14%)	20% (15%)
passive	74%	50%	56%	29%	55%	53%
phrasal v.	3%	7%	7%	5%	7%	6%

A look at the frequency of each AW feature, however, is not very credible. Table 3 also shows the percentage of appearing AW features in each discipline. This percentage is the result of the total count of appearing features (see table 2) divided by the count of segments (see table 1). Error rate for non-preceding adverbs is further counted, which is the result of the appearing non-preceding adverbs divided by the total appearing adverbs. In general, ‘this’ and ‘adv.’ are under 30%; ‘passive’ is from 29% to 74%; ‘phrasal v.’ accounts for 3% to 7%.

Although ‘this’ is most frequently used in ED, it turns out that H sees the highest proportion. It is evident that ‘this’ is used by all abstracts in H, and abstracts in H tend to describe more about their study, while those in ED tend to display results with bullet points. Although ‘adv.’ and ‘passive’ appear to be the most frequently used AW features in S&B (see table 3), EG sees the highest proportion in both features among the others. However, EG also shows high tendency of informal positioning of adverbs. Also, the proportion of non-preceding adverbs seems to be the highest in S&B. In contrast, adverbs are highly accurately placed in B&M. As for ‘passive,’ the result is corresponding to Ting-An Chen’s (2018) finding that passive voice is highly used in engineering-related texts. It is also possible that in EG, the process of

experiment or research design is often described. The better way to describe process and procedure is to use passive voice in AW. In average, the use of ‘passive’ accounts for 53%, which is the most used AW feature among other three features. Nevertheless, the quality of GT’s translation remains unverified as this is not the objective of this study.

In summary, GT does generate output that could follow the style of AW in terms of ‘this,’ ‘adv.’ and ‘passive.’ Since AW in this study is defined as general AW techniques, it is believed that through GT, the English output could generally meet the requirement of AW regardless of the five disciplines tested in this study. However, judging from the analysis results, it seems that abstracts from EG may perform better than the other disciplines. Although adverbs in EG tend not to be placed in mid-position, their meanings remain the same. Moreover, with a rather simple post-editing strategy (i.e., switch the position of the adverb), abstracts from EG could also be academically written. In addition, ‘passive’ tends to appear more than in other disciplines. Furthermore, literature review has proved that passive voice is commonly used in the contexts of engineering. As a result, with relatively higher frequency of AW features, it is believed that the discipline of Engineering is suitable for using GT to conduct Chinese–English thesis abstract translation.

As for the performance of other disciplines, the second-best discipline is H, followed by B&M, S&B, and ED based on the use of ‘this’ and ‘adv.’ features. H has great performance in terms of ‘this’ and ‘adv.’ because both features account over the average. B&M, S&B, and ED all have general AW feature of using this-format. Since B&M has the most accurate use of ‘adv.’, it is ranked as the third best performance discipline. Although non-preceding adverbs are used in S&B as well, the ‘adv.’ is least used in ED. Thus, S&B is considered to perform better than ED. The biggest issue found in ED is that bullet point formats are used nearly in each abstract, which will be discussed later in other findings. However, it should be noted that the accuracy of GT’s output is not discussed in the present study. As a result, the numbers of found AW features do not translate to high accuracy of AW. The focus of this study is on whether GT could generate translations that meet the features/styles of AW.

4.2 Findings of the Four Salient Academic Writing Features

4.2.1 *This-Format / Ben-Format*. Three collocations in the ST are first focused on, including 本文 *běn wén* (this article/paper), 本研究 *běn yán jiù* (this research), and 本論文 *běn lùn wén* (this paper/thesis/dissertation); then how these words are translated into English is checked, and further organization is made. As can be seen from table 4, the most important element is to have the Chinese word ‘本’ *běn* (this). According to the Revised Mandarin Chinese Dictionary released by the Ministry of Education in Taiwan, this word could be considered an adjective, meaning ‘current’ or ‘present,’ which shares the same function with this-format in English AW, where the phrases such as ‘the current study’ or ‘the present study’ are possible to be used as well.

Table 4. ‘This’ in ST and its TT from each discipline

	ST	TT			ST	TT		
EG	本文 <i>běn wén</i> (this article)	this article	10	B&M	本文 <i>běn wén</i> (this article)	this article	1	
		this paper	1		本研究 <i>běn yán jiù</i> (this research)	this research	21	
	本研究 <i>běn yán jiù</i> (this research)	this research	7			this study	18	
		this study	4	sum		40		
	本論文 <i>běn lùn wén</i> (this paper)	this thesis	2	ED	ST	TT		
		this paper	10		本文 <i>běn wén</i> (this article)	this article	1	
	這篇論文 <i>zhè piān lùn wén</i> (this paper)	this paper	2		本研究 <i>běn yán jiù</i> (this research)	this study	19	
	本篇論文 <i>běn piān lùn wén</i> (this paper)	this paper	2			this research	40	
	sum			38	sum			60
	S&B	ST	TT		H	ST	TT	
本文 <i>běn wén</i> (this article)		this article	6	本文 <i>běn wén</i> (this article)		this article	16	
本研究 <i>běn yán jiù</i> (this research)		this study	21	本研究 <i>běn yán jiù</i> (this research)		this study	5	
		this research	18			this research	10	
本論文 <i>běn lùn wén</i> (this paper)		this dissertation	1	本論文 <i>běn lùn wén</i> (this paper)		this paper	6	
		this thesis	1			this dissertation	1	

sum	47		this thesis	10
\		此論文 <i>cǐ lùn wén</i> (this paper)	this thesis	1
			this paper	1
		此研究 <i>cǐ yán jiù</i> (this research)	this research	2
		本論論文 <i>běn piān lùn wén</i> (this paper)	this thesis	1
		此篇論文 <i>cǐ piān lùn wén</i> (this paper)	this paper	1
		這篇論文 <i>zhè piān lùn wén</i> (this paper)	this paper	1
sum				55

It is noted that 篇 *piān* (a quantifier) is sometimes added between 本/這/此 *běn/zhè/cǐ* (this) and 文/論文/研究 *wén/lùn wén/yán jiù* (article/paper/research/study). The results show that whether 篇 *piān* is added or not, GT still generates this-format in English.

There is much difference among such words as article, study, research, paper, thesis, and dissertation in academic texts. The words ‘article’ and ‘paper’ are often used to refer to a study published in academic journals. In this research, all of the selected Chinese texts are abstracts of master’s theses or PhD dissertations, and, therefore, use of the words ‘article’ and ‘paper’ are not appropriate in their English translations. As shown in table 4, GT uses ‘thesis’ much more often than ‘dissertation’ in translating 本論文 *běn lùn wén*. Also, GT produces the word ‘dissertation’ in two abstracts, both of which are for master’s theses.⁴ Additionally, the term 本論文 *běn lùn wén* is not consistently translated into a certain equivalent but into several different terms regardless of the disciplines. This indicates that GT may not be stable enough to produce the same results under similar circumstances and not intelligent enough to choose appropriately from more than one option.

4.2.2 *Mid-Positioning and Non-Preceding Adverbs.* To investigate adverbs, some preparations need to be done first. 地 *de* is an explicit adverb marker in Chinese, in which adverbs of

⁴ Only eight of the collected abstracts are used for PhD dissertations.

manner usually come with 地 *de*. However, throughout the collected data, the word 地 *de* has been used for adverbial function eight times only. According to the Revised Mandarin Chinese Dictionary, the use of 地 *de* as an adverbial function is equal to the character 的 *de*. As a result, investigation into adverbs in Chinese will not be conducted in this study. Instead, since the ‘adv.’ rule in English AW mainly focuses on whether adverbs are placed in mid-position, it should be adequate to examine the English TTs before discussing their ST.

A total of 210 adverbs that modify verbs are found in the data, and 34 out of them are considered non-preceding adverbs. The researchers further analyze whether the adverbs are placed at the sentence-final (end)⁵ position and find 22 out of 34 non-preceding adverbs, which is about 65%, appearing at the end of a sentence or a clause. It is possible that when adverbs in ST are written near a comma or a period, where GT may consider to segment, GT tends to place such adverbs in back of their modified verbs in English translations. Also, 25 out of the 34 non-preceding adverbs occur immediately adjacent to the main verb; hence, they do not largely affect the meaning of the sentences concerned. One of the possible explanations for these few discrepancies could be attributed to the fact that GT’s database may include more cases where adverbs are commonly placed in the sentence-final position. In summary, the total found adverbs are 210, and 34 of them belong to non-preceding adverbs, which is about 16%. Despite the fact that some non-preceding adverbs are found from the data, GT still has a tendency to place an adverb before its modified verb in most cases.

4.2.3 Passive Voice. In Chinese STs, locating passive voice is as difficult as finding adverbs. Some explicit passive markers in Chinese are 被 *bèi*, 受 *shòu*, 為...所 *wéi...suǒ*, etc. However, similar to Chinese adverb markers, passive markers in Chinese are barely used, and there is “much less any doer of the action” (Ye and Shi 2014, 87; our translation). A search of the word 被 *bèi*, the most explicit marker, shows 16 hits through AntPConc, but only 14 out

⁵ When a non-preceding adverb appears before a comma or a period, it is regarded as being in the sentence-final position or ‘end’ position.

of them are used for passive voice; moreover, 3 out of 14 被 *bèi* are not translated into explicit passive voice in English (i.e., typical be verb plus a past participle). As a result, the focus of this ‘passive’ feature is mainly on the examination of the English translated texts. Some Chinese passive markers will still be discussed so as to show GT’s ability to recognize passive voice.

Table 5 presents the location of the explicit Chinese passive marker 被 *bèi* in the data. 被 *bèi* is magnified and in bold type. The use of passive voice is underlined in both texts. Irreverent texts are removed due to space constraints.

Table 5. Appearances of 被 *bèi* in the Chinese ST and their TT

Location	ST	TT
EG segment 7-2	其中以人臉辨識最廣為被使用和研究，因為 [...]. <i>qí zhōng yǐ rén liǎn biàn shí zuì guǎng wéi bèi shǐ yòng hé yán jiū, yīn wéi [...].</i> (Among them, face recognition is the most widely used and researched, because [...].)	Among them, face recognition <u>is</u> the most widely <u>used</u> and researched, because [...].
EG segment 11-1	[...], 但圍棋程式卻從未曾被人們認為是強大的, [...] [...], <i>dàn wéi qí chéng shì què cóng wèi céng bèi rén men rèn wéi shì qiáng dà de, [...].</i> ([...], but the Go program was never considered powerful by people [...].)	[...], but the Go program <u>was</u> never <u>considered</u> powerful [...].
EG segment 16-1	電子領域中產品的發展趨勢被要求尺寸縮減、[...] <i>diàn zǐ lǐng yù zhōng chǎn pǐn de fā zhǎn qū shì bèi yào qiú chǐ cùn suō jiǎn [...].</i> (The development trend of products in the electronic field is required to reduce size, [...].)	The development trend of products in the electronic field <u>is required</u> to reduce size, [...].
B&M segment 10-1	近年來知識管理的話題雖不斷被討論，[...] <i>jìn nián lái zhī shí guǎn lǐ de huà tí suī bù duàn bèi tāo lùn, [...].</i> (In recent years, the topic of knowledge management has been continuously discussed, [...].)	Although the topic of knowledge management has <u>been</u> continuously <u>discussed</u> in recent years, [...].

<p>S&B segment 6-5</p>	<p>一、「自尊的滿足」：如對被別人尊重的需求 <i>yī , /zì zūn de mǎn zú/ : rú duì bèi bié rén zūn zhòng de xū qiú.</i> (1. “Satisfaction of self-esteem”: such as the need to be respected by others.)</p>	<p>1. “Satisfaction of self-esteem”: such as the need to <u>be respected</u> by others.</p>
<p>S&B segment 9-2</p>	<p>咖啡在 1884 年被引進台灣之前，早已廣泛流傳於世界各地， [...] <i>kā fēi zài 1884 nián bèi yǐn jìn tái wān zhī qián , zǎo yǐ guǎng fān liú chuán yú shì jiè gè dì , [...].</i> (Before Coffee was introduced to Taiwan in 1884, it had been widely circulated around the world. [...])</p>	<p>Coffee had been widely circulated throughout the world before it <u>was introduced</u> to Taiwan in 1884. [...].</p>
<p>S&B segment 13-7</p>	<p>[...] (3) 運用多元方式提高霸凌行為被發現的機率; [...] [...] (3) <i>yùn yòng duō yuán fāng shì tí gāo bà líng háng wéi bèi fā xiàn de jī lǜ ; [...].</i> [...] (3) Use diverse ways to increase the probability of bullying behavior being found; [...].)</p>	<p>[...] (3) Use diversity Ways to increase the probability of bullying <u>being detected</u>; [...].</p>
<p>S&B segment 19-2</p>	<p>故「再利用」一詞，於近幾年間被各界提出與熱烈討論， [...] <i>gù /zài lì yòng/ yī cí , yú jìn jǐ nián jiān bèi gè jiè tí chū yǔ rè liè tāo lùn , [...].</i> (Therefore, the term “reuse” in recent years has been put forward and discussed enthusiastically by all walks of life [...].)</p>	<p>Therefore, the term “reuse” has <u>been put forward</u> and discussed enthusiastically by all walks of life in recent years [...].</p>
<p>ED segment 8-13</p>	<p>(2) 自我因素：為自我概念、興趣、宗教信仰、價值觀、及被尊重的感覺 <i>(2) zì wǒ yīn sù : wéi zì wǒ gài niàn , xìng qù , zōng jiāo xìn yǎng , jià zhí guān , jí bèi zūn zhòng de gǎn jiào .</i> ([2] Self factors: self-concepts, interests, religious beliefs, values, and feelings of being respected.)</p>	<p>(2) Self factors: self-concepts, interests, religious beliefs, values, and feelings of <u>being respected</u>.</p>
<p>H segment 6-3</p>	<p>隨著大批日本人來台，日人的風俗民情、生活習慣被引進台灣， [...] <i>suí zhe dà pī rì běn rén lái tái , rì rén de fēng sú mín qíng , shēng huó xí guàn bèi yǐn jìn tái wān , [...].</i> (As a large number of Japanese people came to Taiwan, Japanese customs and traditions and living habits were introduced to Taiwan, [...].)</p>	<p>With the arrival of a large number of Japanese people in Taiwan, Japanese customs, customs, and living habits <u>were introduced</u> to Taiwan, [...].</p>

<p>H segment 10-1</p>	<p>大甲媽祖往新港遶境進香是國內規模最大的宗教活動，亦被美國探索頻道列為「世界三大宗教活動」之一。</p> <p><i>dà jiǎ mā zǔ wǎng xīn gǎng rǎo jìng jìn xiāng shì guó nèi guī mó dà de zōng jiāo huó dòng, yì bèi měi guó tàn suǒ pín dào liè wéi [shì jiè sān dà zōng jiāo huó dòng] zhī yī。</i></p> <p>([The Goddess of] Dajia Mazu Pilgrimage to Xingang for incense is the largest religious activity in the country, and it is also listed by the US Discovery Channel as one of the “three large religious activities in the world.”)</p>	<p>Dajia Mazu’s detour to Xingang for incense is the largest religious activity in China, and it is also listed as one of the “Three Religious Activities in the World” by the Discovery Channel of the United States.</p>
--------------------------------------	--	--

Clearly, 被 *bèi* is always translated into passive voice in English. However, as 被 *bèi* is seldom observed in Chinese, table 6 shows some examples of implicit passive sentences in Chinese STs and their explicit English passive voice in the corresponding TTs.

Table 6. Examples of ‘passive’ from each discipline

Location	ST	TT
<p>EG segment 2-5</p>	<p>本論文透過實驗方式驗證減少電源路徑上的漣波雜訊將能有效降低積層陶瓷電容器所引起的高頻顫噪現象, [...]。</p> <p><i>běn lùn wén tòu guò shí yàn fāng shì yàn zhèng jiǎn shǎo diàn yuán lù jīng shàng de lián bō zá xùn jiāng néng yǒu xiào jiàng dī jī céng táo cí diàn róng qì suǒ yǐn qǐ de gāo pín chàn zào xiàn xiàng, [...]</i></p> <p>(This paper verifies through experiments that reducing the ripple noise on the power path will be able to effectively reduce the high-frequency microphonic phenomenon caused by multilayer ceramic capacitors, [...])</p>	<p>This paper verifies through experiments that reducing the ripple noise on the power path will effectively reduce the high-frequency microphonic phenomenon caused by multilayer ceramic capacitors. [...].</p>
<p>B&M segment 12-3</p>	<p>本研究資料來源為直接取自各花卉網站內容，以及花卉網站個案訪談資料。</p> <p><i>běn yán jiū zī liào lái yuán wéi zhí jiē zì qǐ zì gè huā huì wǎng zhàn nèi róng, yǐ jí huā huì wǎng zhàn gè àn fāng tán zī liào。</i></p> <p>(The source of the data for this study is directly taken from the content of various flower websites and data of case interviews of flower websites.)</p>	<p>The source of the data for this study is directly taken from the content of various flower websites, as well as case interview data from flower websites.</p>

<p>S&B segment 11-3</p>	<p>利用實驗經濟學的方法，並加入心理學的人格特質分析，本研究共進行九場實驗，實驗受試者八十四人，得到之研究結果發現：[...].</p> <p><i>lì yòng shí yàn jīng jì xué de fāng fǎ, bìng jiā rù xīn lǐ xué de rén gé tè zhì fèn xī, běn yán jiū gòng jìn háng jiǔ chǎng shí yàn, shí yàn shòu shì zhě bā shí sì rén, dé dào zhī yán jiū jié guǒ fā xiàn:</i> [...].</p> <p>(By using the method of experimental economics and adding the analysis of personality traits in psychology, this study conducted nine experiments. The experimental subjects: 84 people. The obtained research results found: [...].)</p>	<p>Using the method of experimental economics and adding the analysis of personality traits in psychology, this study conducted nine experiments with 84 experimental subjects. The results obtained showed that: [...].</p>
<p>ED segment 20-2</p>	<p>本研究採用準實驗研究設計，教學活動設計是以「延伸式探討」為主，選擇「水裡的生物」和「燈泡亮了」兩個單元。</p> <p><i>běn yán jiū cǎi yòng zhǔn shí yàn yán jiū shè jì, jiāo xué huó dòng shè jì shì yǐ [yán shēn shì tàn tǎo] wéi zhǔ, xuǎn zé [shuǐ lǐ de shēng wù] hé [dēng pào liàng le] liǎng gè dān yuán.</i></p> <p>(This research adopts a quasi-experimental research design. The teaching activity design is based on “extended discussion.” Two units, “living things in the water” and “the light bulb is on,” were selected.)</p>	<p>This research adopts a quasi-experimental research design. The teaching activity design is mainly based on “extended discussion”. Two units are selected: “Biology in the Water” and “Light Bulb”.</p>
<p>H segment 14-8</p>	<p>本文的研究方法主要建構在文獻分析和質性的訪談法兩方面。</p> <p><i>běn wén de yán jiū fāng fǎ zhǔ yào jiàn gòu zài wén xiàn fēn xī hé zhì xìng de fǎng tán fǎ liǎng fāng miàn.</i></p> <p>(The research method of this article is mainly constructed in two aspects: literature analysis and qualitative interview method.)</p>	<p>The research method of this article is mainly constructed in two aspects: literature analysis and qualitative interview method.</p>

In EG, “所引起的” *suǒ yīn qǐ de* is translated into “[which is] caused by.” 所 *suǒ* could be considered a passive voice marker in Chinese. The phrase “積層陶瓷電容器” *jī céng táo cí diàn róng qì* (multilayer ceramic capacitors) is a doer, and “高頻顫噪現象” *gāo pín zhàn zào xiàn xiàng* (the high-frequency microphonic phenomenon) an action. Both of them are objects rather than persons, so it is possible that GT adopts passive construction for this sentence.

In B&M, “來源為直接取自各花卉網站內容” *lái yuán wéi zhí jiē qǔ zì gè huā huì wǎng zhàn nèi róng* (The source is directly taken from the content of various flower websites)

is translated into “data . . . is directly taken from the content of various flower websites.” 為 *wéi* is considered a passive marker. Meanwhile, it can be noticed that the adverb is placed before “taken,” which meets the requirement of AW.

In S&B, “得到” *dé dào* (get/obtain) is translated into “[which are] obtained.” In fact, 所 *suǒ* could precede 得到 *dé dào*, forming a sentence such as 本研 究所得 到 *běn yán jiù suǒ dé dào* (what this research obtained). However, 所 *suǒ* can be omitted without losing the passive expression.

In ED, “is based on” is used. Interestingly, in ST, “教學活動設計是以...為主, 選擇 [水...和...兩個單元” *jiào xué huó dòng shè jì shì yǐ...wéi zhǔ, xuǎn zé [shuǐ...hàn... liǎng gè dān yuán* (The teaching activity design is based on ... Two units, “... water” and ... were selected.) could be regarded as a complex sentence. The subject of the second clause, “教學活動” *jiào xué huó dòng* (teaching activity), is omitted, which is often seen in Chinese. Nevertheless, GT separates the sentence and makes the latter part another complete sentence. What’s more, it is translated into a passive sentence as in “two units are selected.”

Last, in H, “方法主要建構在” *fāng fǎ zhǔ yào jiàn gòu zài* (method mainly construct in) is translated into “method ... is mainly constructed.” If Chinese passive marker is inserted, it can be rendered as “方法主要是建構在...的” *fāng fǎ zhǔ yào shì jiàn gòu zài...de* (method is mainly constructed in).

To sum up, GT seems to recognize Chinese passive voice. Even though Chinese sentences may not include a clear passive marker 被 *bèi*, GT can still generate passive voice through subtle clues. It has to be reminded that the quality of the translated output remains unverified. This study is aimed at investigating whether English translations are of AW style.

4.2.4 Use of Phrasal Verbs. The percentage of appearing phrasal verbs analyzed in previous sections reveals low usage of phrasal verbs throughout the data (see tables 2 and 3); however, the collected data are ‘abstracts,’ which should be written in a logical and concise way. This means if GT produces phrasal verbs in such context, it is possible that GT may use more phrasal

verbs when being applied to a full academic text. Therefore, a further look at the number of abstracts where phrasal verbs are used should be discussed. Table 7 shows some statistics on the ‘phrasal v.’ feature.

Table 7. Statistics regarding ‘phrasal v.’

	Abstracts involving phrasal verbs	Percentage of abstracts involving phrasal verbs	Frequency of phrasal verbs from each discipline	Percentage of phrasal verbs in segment
EG	4	0.2	3	15%
B&M	13	0.7	9	45%
S&B	18	0.9	13	65%
ED	14	0.7	12	60%
H	15	0.8	7	35%
Total	64		44	

The statistics in the first column can also be seen in table 2, where phrasal verbs are found less used by GT compared with the other salient AW features found to be used by GT in this study. Although phrasal verbs only appear 64 times of all the collected data and seem to be the least frequently appearing feature from the data, they should be avoided as much as possible when it comes to AW. In terms of the percentage of abstracts where phrasal verbs are found, EG remains the lowest, while S&B and ED appear to be the first and second highest disciplines that cause GT to adopt phrasal verbs. On the whole, nearly half of the data contain at least one use of phrasal verbs. To avoid use of phrasal verbs, a little post-editing of the translation output should probably be adopted.

Table 8. Appearing phrasal verbs and their single-word verb synonyms

Phrasal verbs	Frequency	Synonym or related words
put forward	21	accomplish, execute
carry out	9	overtake, surpass
find out	8	decrease, decline
pay attention to	8	challenge, encounter
make up	2	deliver, bear

make up for	2	agree with, similar with
sort out	2	get, reach
take care of	2	compose, constitute
take into consideration	2	compensate, offset
catch up with	1	concentrate on
die out	1	suggest, indicate
face up to	1	advance, propose
give birth to	1	see, visit
keep up with	1	understand, solve
point out	1	administer, care
seek out	1	consider, regard
take into account	1	consider, regard
sum	64	

Table 8 shows the frequency of appearing phrasal verbs. All the synonyms for phrasal verbs are provided according to Merriam-Webster Dictionary.⁶ It can be seen that ‘put forward’ is highly used. The source term of ‘put forward’ in all of the abstracts is 提出 *tí chū*. The second and the third highest used phrasal verbs are ‘carry out’ and ‘pay attention to.’ While ‘pay attention to’ might be a frequent phrase to use in academic texts, ‘carry out’ could be replaced by other words such as ‘accomplish’ and ‘execute.’ A further analysis also finds that 提出 *tí chū* is used in ST 49 times, and 22 of them are translated into “propose,” causing a possibility that in terms of 提出 *tí chū*, GT has a 50% chance to adopt phrasal verbs. To conclude, even though the data collected show low frequency of use of phrasal verbs, GT has yet fully met the requirements of AW in relation to verb choice.

4.3 Other Findings

4.3.1 Bullet Point Format. The use of bullet points is also found in the collected abstracts. Table 9 shows the number of abstracts where bullet points are used in each discipline.

⁶ Available at <https://www.merriam-webster.com>.

Table 9. Frequency of bullet points appearing in each discipline

	EG	B&M	S&B	ED	H	sum
Frequency of bullet points	1	12	13	17	7	50

Half of the collected abstracts show the use of bullet point format. It is interesting to note that nearly every abstract in ED adopts this strategy. While a list of bullet points may facilitate a quick understanding of one’s research, it is not recommended to use in AW. Research has revealed that spatial design, such as non-hierarchical bullet point form for a poster, could help readers better perceive the relevant research (O’Halloran, Tan, and Smith 2016). However, Larissa D’Angelo (2016) indicates that typical bullet point format may be inappropriate in a presentation. Therefore, bullet point format might be more often seen in a visual design such as posters. As a whole, bullet point format should be avoided in AW, and it is more academically appropriate to use signal words such as ‘first,’ ‘second,’ and ‘third’ and other connecting words in academic texts.

From the conclusion drawn from research question two that abstracts from EG are probably more suitable to be translated from Chinese to English via GT followed by H, B&M, S&B, and ED, the finding of bullet point format also shows consistency. There is only one abstract found to adopt bullet point format in EG, and 17 abstracts in ED. As a result, it is possible that when the use of bullet point format is reduced, the quality of translated output in terms of AW features could be increased.

4.3.2 The Use of Notional First-Person Pronouns. The use of ‘I’ in AW is inappropriate (except for certain fields). However, during the coding process, some examples where ‘I’ is used are found. To further explore this phenomenon, table 10 is created for discussion. Chinese 我 *wǒ* (I), 我們 *wǒ men* (we), 筆者 *bǐ zhě* (the author), and 研究者 *yán jiù zhě* (the researcher) are to be searched in the Chinese corpus built via AntPConc; meanwhile, English ‘I’ and ‘we’ will be searched in the English corpus for cross examination.

Table 9. Searched key word(s) in bilingual corpora: Use of notional first-person pronouns

Searched key word(s) in bilingual corpora	EG	B&M	S&B	ED	H
我 wǒ / I	0/0	0/0	0/2	0/0	0/7
我們 wǒ men / we	10/10	1/4	1/3	0/2	5/13
筆者 bǐ zhě / the author	0/0	0/0	2(1)/3	0/0	4(1)/5
研究者 yán jiù zhě / the researcher(s)	0/0	1/1	3/3	4/4	1/1

In EG, neither ‘I’ nor 我 wǒ is found throughout the texts. However, 我們 wǒ men (we) is found to be used 10 times. Since 筆者 bǐ zhě (the author) and 研究者 yán jiù zhě (the researcher) are formal terms used to refer to the author him/herself, these two terms are also analyzed in this section. Nevertheless, there is no use of these two terms in EG.

In B&M, the use of 我 wǒ (I) for referring to the author is not found. 我們 wǒ men (we) is used once, but it is not used for self reference. However, interestingly, ‘we’ is used four times by GT, but none of its equivalence of 我們 wǒ men (we) is found in the ST. GT seems to generate a subject for its sentence. 研究者 yán jiù zhě (the researcher) is used once, and its translation becomes ‘the researchers.’ As a singular or plural form of a noun can barely be recognized in Chinese, it is not surprising that GT has rendered the singular noun 研究者 yán jiù zhě (the researcher) into the plural noun ‘researchers.’

None of the first-person pronouns 我 wǒ (I) in S&B refers to the author him/herself, but ‘I’ still occurs twice. 我 wǒ occurs frequently in ED, but none of them is referring to the author. No use of ‘I’ is found, either. However, ‘we’ is used twice throughout the texts. Finally, none of 我 wǒ in H is used as self reference, but ‘I’ is used in TT for seven times. Only one 我 wǒ is equivalent to ‘I’ (again, it is not used for self reference). The other six uses of ‘I’ are automatically generated by GT. 我們 wǒ men (we) is found to be used 5 times, but ‘we’ appears 13 times. Among them, 8 ‘we’ are automatically generated by GT.

These findings are worth mentioning because (1) problem of segmentation may reveal GT’s drawbacks, (2) use of bullet point format could lead to informal AW style, and (3) while the subject of a sentence is often omitted or implied in Chinese, GT is intelligent enough to

recreate a subject in its English translation; however, the frequent use of the first-person pronoun in academic abstracts as found in this study does not follow the general style of AW in English. The three findings discussed here may help provide insights into how to produce a well-organized and high-quality English academic text translated from Chinese via GT.

5. Conclusion

The analysis results show that ‘this’ is generally used in each discipline; however, two of 本研究 *běn yán jiù* (this study/research) are not translated into this-format possibly due to syntactic ambiguity of Chinese. Basically, 本文 *běn wén* (this article) / 本研究 *běn yán jiù* (this research/this study) / 本論文 *běn lùn wén* (this paper or this thesis or this dissertation) can be translated into this article/this study/this paper/this thesis by GT. As for adverb positioning, most adverbs are placed in mid-position with a few exceptions, most of which are descriptive adverbs such as ‘rapidly’ and ‘vigorously.’ The results of the study suggest that when adverbs in ST are written near a comma or a period, where GT may consider to segment, GT possibly tends to place such adverbs in back of their modified verbs in English translations. Passive voice is widely used in English and the obtained data also show signs of this feature. Even though passive markers are rarely seen in Chinese, GT still has the ability to identify implicit Chinese passive voice and then generate adequate English output. EG shows the highest frequency of the usage of passive voice, which corresponds to Chen’s (2018) report that passive voice is widely adopted in the field of engineering. Regardless of the overall quality of translated abstracts, AW features could be found. While GT can follow some principles of AW and produce translation in an academic style, it may still adopt a less formal style by using phrasal verbs instead of single-word verbs that are preferably used in English AW. The most frequently used phrasal verbs by GT throughout the data is ‘put forward,’ which could be replaced by ‘propose’ or ‘suggest’ in AW according to the Merriam-Webster Dictionary. The present study discusses whether GT meets the ‘general’ principles of AW; hence, as long as GT

English output shows a certain proportion of discussed AW features, GT may be considered to have reached a certain stage of generating expected translations. Since EG shows high frequency of passive voice and low usage of phrasal verbs, it is believed that EG performs better than other disciplines discussed in this study when GT is used for academic abstract translation from Chinese into English.

The suggestions for future research are: (1) enlarging the scale, (2) investigating other AW features, (3) testing different language pairs, (4) exploring AW in Chinese, (5) using updated data. This study covers a small scale of data, which may reduce the validity and reliability of the results. By enlarging the scale or adopting an existing much larger corpus, the statistics could be more accurate. Next, there are still other AW features to be noted (e.g., the use of ‘I’). However, it is acknowledgeable that passive construction in translated English abstracts is not thoroughly discussed due to time and length limitation. Since passive voice accounts for a rather high frequency in translated English abstracts, future research can investigate this phenomenon further. In addition, since GT supports more than a hundred languages, exploring academic texts via different language pairs could be another interesting study. Also, further analysis of Chinese AW could be done in the future. Although the quality of translated abstracts in this study (and the quality of the collected STs) has not been investigated, it may be likely that adopting Chinese AW style appropriately in Chinese STs may lead to better English results from GT. As a result, in terms of this issue, either collecting a rather recent data within academic contexts or exploring pre-editing strategy on Chinese could be done in the future.

References

- Azer, Haniyeh Sadeghi, and Mohammad Bagher Aghayi. 2015. “An Evaluation of Output Quality of Machine Translation (Padideh Software vs. Google Translate).” *Advances in Language and Literary Studies* 6 (4): 226–237. doi:10.7575/aiac.all.v.6n.4p.226.
- Bailey, Stephen. 2003. *Academic Writing: A Practical Guide for Students*. London: RoutledgeFalmer.
- Brown, Peter F., John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin. 1990. “A Statistical Approach to Machine Translation.” *Computational Linguistics* 16 (2): 79–85. <https://aclanthology.org/J90-2002>.
- Burchardt, Aljoscha, Arle Lommel, Lindsay Bywood, Kim Harris, and Maja Popović. 2016. “Machine Translation Quality in an Audiovisual Context.” *Target* 28 (2): 206–221. doi:10.1075/target.28.2.03bur.
- Chen, Ting-An. 2018. *Yīnghàn bǐjiào yǔ fānyì (zuìxīn bǎn)* [English and Chinese translation: A comparative study (New edition)]. Taipei: Bookman Books.
- Chiang, David. 2005. “A Hierarchical Phrase-Based Model for Statistical Machine Translation.” In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, 263–270. doi:10.3115/1219840.1219873.
- D’Angelo, Larissa. 2016. “The Academic Poster Genre: Friend or Foe?” In *The Routledge Handbook of English for Academic Purposes*, edited by Ken Hyland and Philip Shaw, 392–402. London: Routledge.
- Groves, Michael, and Klaus Mundt. 2015. “Friend or Foe? Google Translate in Language for Academic Purposes.” *English for Specific Purposes* 37:112–121. doi:10.1016/j.esp.2014.09.001.
- Kao, Hsiao-Ting. 2011. “Training in Pre-MT Editing with the Controlled Language: A Pilot Study.” Master’s thesis, National Kaohsiung First University of Science and Technology.
- Klimova, Blanka Frydrychova. 2013. “Common Mistakes in Writing Abstracts in English.” *Procedia - Social and Behavioral Sciences* 93:512–516. doi:10.1016/j.sbspro.2013.09.230.
- O’Halloran, Kay L., Sabine Tan, and Bradley A. Smith. 2016. “Multimodal Approaches to English for Academic Purposes.” In *The Routledge Handbook of English for Academic*

Purposes, edited by Ken Hyland and Philip Shaw, 256–269. London: Routledge.

- Sijia, Dong, and Hong Jie. 2015. “An Analysis of Problems in Writing Research Paper English Abstracts.” *Journal of Language and Literature Studies*, no. 5, 146–147, 163. <https://www.cnki.com.cn/Article/CJFDTotal-YWWY201505060.htm>.
- Swales, John M., and Christine B. Feak. 2012. *Academic Writing for Graduate Students: Essential Tasks and Skills*. 3rd ed. Michigan: University of Michigan Press.
- Tongpoon-Patanasorn, Angkana, and Karl Griffith. 2020. “Google Translate and Translation Quality: A Case of Translating Academic Abstracts from Thai to English.” *PASAA: Journal of Language Teaching and Learning in Thailand* 60:134–163. <https://files.eric.ed.gov/fulltext/EJ1259132.pdf>.
- Tsai, Po-Ying. 2014. *Cóng zìjù dào jiégòu: Xuéshù lùnwén xiězuò zhǐyǐn (dì èr bǎn)* [A structural guide to academic writing (Second edition)]. Taipei: National Taiwan University Press.
- Yang, Ding-Hua. 2018. “A Comparative Study of Strategic Differences in Chinese-English Bidirectional Post-MT Editing: Use of 3C User Manual as Samples.” Master’s thesis, National Kaohsiung University of Science and Technology.
- Ye, Zinan, and Lynette Xiaojing Shi. 2014. *Introduction to Chinese-English Translation*. Taipei: Bookman Books.
- Yu, Mu-Yen. 2019. “Cóng gōng néng fān yì lǐ lùn tàn tǎo jī qì fān yì zài yī xué wén tǐ dē shí yòng xìng- Yì Wéi Jī Bǎi Kē Yī Xué Fān Yì Jì Huà wéi lì.” [Exploring the practicality of machine translation in medical articles from the perspective of functional theories of translation: A case study on WikiProject Medicine.] Master’s thesis, Wenzao Ursuline University of Language.
- Zhang, Xiao-Rei. 2014. “Lùn tǐ yù lèi lùn wén yīng wén zhāi yào dē xiě zuò.” [On writing of English abstract of sports papers.] *Journal of Guangxi Normal University for Nationalities* 31 (3): 62–64. doi:10.3969/j.issn.1674-8891.2014.03.019.
- Zhong, Guo Xiang. 2016. “Kē jì qí kān yīng wén zhāi yào dē xiě zuò jì qiǎo tàn suǒ.” [Writing techniques in English abstracts of scientific and technical journals.] *Journal of Minnan Normal University (Natural Science)* 29 (2): 124–126. doi:10.3969/j.issn.1008-7826.2016.02.021.