様 式 C－19、F－19－1、Z－19 (共通)

科研費
KAKENHI

科学研究費助成事業　　研究成果報告書

令和 5 年 6 月 16 日現在

機関番号：２１６０２
研究種目：基盤研究(C)（一般）
研究期間：2018～2022
課題番号：１８Ｋ００６５６

研究課題名（和文）Using Google Translate for Academic English Writing Instruction

研究課題名（英文）Using Google Translate for Academic English Writing Instruction

研究代表者
Ｈｅｏ Ｙｏｕｎｇｈｙｏｎ（Heo, Younghyon）

会津大学・コンピュータ理工学部・上級准教授

研究者番号：１０６３１４７６

交付決定額（研究期間全体）：（直接経費）　　2,200,000円

研究成果の概要（和文）：このプロジェクトでは以下の3つの目的を達成した。１）Google翻訳された文章を人がどの程度探知できるかを調査した結果、機械翻訳文は人が区別できないほど自然であることが分かった。また、頻出単語上位50語が人の書いた文章とGoogle翻訳された文章とで異なっており、それぞれ特徴的であることが分かった。2）機械学習によって、Google翻訳されたテキストが検出可能か検証したところ、高い精度（89%~99%）で検出可能であった。３）Google翻訳の効果的な使い方を教える教材を作成することができた。

研究成果の学術的意義や社会的意義
Based on the result of our research (machine-translated writing being identifiable) and other studies, a system can be further developed and used to detect machine-translated writings (for L2 teaching). Also, we could instruct students how to use a machine translation tool to learn L2 writing.

研究成果の概要（英文）：There were three goals in this project. First, we investigated how well human readers can detect machine-translated text. We also performed a word analysis (n-gram analysis). Regarding the first goal, we found that the machine-translated text using Neural Machine Translation was so natural that human raters could not successfully detect them (51% accuracy). The results of the n-gram analysis showed that the top 50 most frequently used words were different between human-written and machine-translated texts, indicating some unique traits of each. Our second goal was to test whether machine learning can detect machine-translated texts. It was found that machine learning could detect machine-translated texts with high accuracy (accuracy rate 89% to 99%). The third goal was to create educational materials, and we were able to create materials teaching effective and appropriate use of Google Translate based on the results of our survey conducted on English instructors and our word analysis.

研究分野：Second Language Acquisition

キーワード：L2 writing　Machine translation　academic writing

１．研究開始当初の背景

Machine translation has constantly evolved along with increasing computation power since becoming widely available to the public by Google in 2006. In particular, the shift from Statistical Machine Translation (SMT) to Neural Machine Translation (NMT) in 2016 has significantly improved the quality of translated texts in major languages. NMT relies on a large neural network, enabling more natural translations when large data is available for a language pair (e.g., English-Japanese). Given the wide availability of good-quality machine translation tools, it is not realistic to expect L2 learners not to use them at all. Instead, EFL instructors should rather just focus on detecting the heavy use of machine translation without any modifications of the generated sentences by a learner or teach them how to use them effectively and properly to improve their writing, in other words, as a tool to "learn" L2 writing.

Regarding identifying the extensive use of it, it was relatively easy to detect machine-translated sentences with SMT-based systems, but recently, NMT has elevated translation quality to a level where the results are not easily distinguishable as machine-translated. To prevent students from fully dependent on machine-translating their L1 texts to L2 without modifying the text, it is crucial to introduce a system that can detect how much the text is machine translated. As a first step, it is important to confirm whether the machine-translated text is detectable, having its own style and traits as writing. After confirming machine-translated text as a specific genre of writing with its unique features and style, L2 writers can be introduced the traits of machine-translated texts so that they can use machine translation like Google Translate more effectively and appropriately when they learn L2 writing, specifically, academic writing in our project. We believe that using it properly with the review process followed after the machine generation can help L2 learners further learn about good English writing.

２．研究の目的

The goal of this project was threefold. First, we investigated how well human readers (experienced L2 instructors) can detect machine-translated text. With the introduction of NMT in 2016, L2 instructors have been anecdotally reporting how natural machine-generated (translated) sentences sound and how difficult it is now to spot Google-translated writings. We also performed the n-gram analysis of both types of text (human-written vs. machine-translated) to further identify the features of each type of text. The second goal of this project was to test whether machine learning can learn the traits of machine-translated academic texts and classify data into machine-translated and human-written texts. If machine-translated texts are found to be identifiable, L2 instructors can further attempt to learn the style of machine writing to detect them in student writing, or a system can be developed to detect machine writing in the future. The third goal of this project was to create educational materials teaching effective and appropriate usage of Google Translate based on the results of our survey conducted on EFL instructors in Japan and our n-gram analysis on Human-written and machine-written academic texts.

３．研究の方法

Methodology: Goal 1

We examined 1) whether experienced EFL (English as a Foreign Language) instructors in Japan can successfully identify machine-translated texts and 2) what the distinguishing features of machine-translated texts and texts written by L2 Japanese learners of English are via the survey and also the n-gram analysis. A Lime Survey was conducted asking participants to distinguish machine-translated from L2 human-written scientific abstracts from Japan. Three surveys were made, each with five machine-translated and five human-written abstracts. Twenty-four participants provided judgments on ten abstracts each and also provided the basis for their judgment. We also conducted an n-gram analysis of machine-translated texts and human-written texts. The top 50 most frequent unigram, bigram, trigram and quadragram were analyzed separately for both types of texts and also the comparison was made between the two text types to see how the ranking of the same n-gram is different in the two types of texts.

Methodology: Goal 2

The way machine learning was used for our analysis of machine-translated texts was twofold. One is more intuitive/self-learning with the multi-layered deeper structural examination of texts, and the other is a more widely adopted supervised learning technique, SVM, with the keyword analysis using TF-IDF (Term Frequency-Inverse Document Frequency) and Feature Vector. We compared the two techniques in terms of the accuracy rate in detecting machine-translated texts. With regard to the deep learning analysis, we employed the framework called TensorFlow developed by Google for its extensive built-in support for deep learning.

Methodology: Goal 3

Based on the result of our survey on EFL instructors and the n-gram analysis, we created teaching materials instructing students enrolled in an English thesis writing class in Japan, instructing how to use machine translation effectively for their research writing.

４．研究成果

Results: Goal 1

We examined whether 24 English instructors teaching in Japan can successfully distinguish machine-translated text from human writing (L2 writing). The result indicates that they were essentially just guessing (51% accuracy), not being able to tell the difference much between human-written and machine-translated text. Human-written texts were reported to contain simple grammar errors; misspellings; punctuation/spacing errors; passive overuse; incorrect passive use; typical L2 expressions. Regarding the features of machine-translated sentences, they noted that machine-translated texts contained long sentences with many subordinate clauses; incoherent sentences/paragraphs; stylistic problems (not in keeping with conventions of scientific writing); incorrect passive use.

In addition to the observations of the experienced EFL instructors, we could explore more traits of the machine and human writing through n-gram analysis. The top 50 most common words/phrases were analyzed using two settings: (+punctuation) and (-punctuation). The result of the n-gram analysis is shown in Figure 1, including only the top 10 in this report.

Unigram (Top 10)

Human (+punc)	Machin (+punc)	Human (-punc)	Machine (-punc)
The	the	The	the
This	,	This	and
are	and	are	of
paper	of	paper	a
for	a	for	to
A	to	A	it
which	it	which	In
proposed	In	proposed	this
systems	this	systems	method
new	method	new	is

Bigram (Top 10)

Human (+punc)	Machin (+punc)	Human (-punc)	Machine (-punc)
. The	of the	This paper	of the
. This	. In	is proposed	and the
This paper	, and	for the	it is
. A	, we	paper describes	in the
is proposed	, it	has been	In this
for the	and the	a new	to the
paper describes	it is	paper presents	As a
. It	, the	paper proposes	as a
, which	in the	for a	on the
has been	In this	the proposed	a result

Trigram (Top 10)

Human (+punc)	Machin (+punc)	Human (-punc)	Machine (-punc)
. This paper	. In this	This paper describes	As a result
This paper describes	, it is	This paper presents	In order to
. It is	. As a	This paper proposes	In this paper
This paper presents	, and the	is shown that	is possible to
is proposed .	As a result	has been developed	it is possible
This paper proposes	paper , we	paper describes the	a method of
. A new	. Therefore ,	paper proposes a	is necessary to
is shown that	this paper ,	paper describes a	it is necessary
has been developed	a result ,	paper presents a	In this report
paper describes the	. In the	It is shown	the time of

Quadragram (Top 10)

Human (+punc)	Machin (+punc)	Human (-punc)	Machine (-punc)
. This paper describes	. As a result	This paper describes a	it is possible to
. This paper presents	this paper , we	This paper presents a	it is necessary to
. This paper proposes	As a result ,	This paper describes the	at the time of
This paper describes a	In this paper ,	This paper proposes a	it was found that
This paper presents a	. In order to	It is shown that	As a result of
This paper describes the	. In this paper	In this paper we	was found that the
. It is shown	. In addition ,	is shown that the	it was confirmed that
This paper proposes a	it is possible to	This paper deals with	propose a method to
It is shown that	it is necessary to	is based on the	from the viewpoint of
this paper , a	In this report ,	it is shown that	we propose a method

Figure 1. Top 10 results of each n-gram analysis

Based on the result of the survey and n-gram analysis, some unique features of each type of text were listed. For example, it was evident that the direct translation of the fixed Japanese phrase in formal writing, such as “本論文では” appears more frequently in machine writing (#4) (direct translations of the Japanese text) while using “paper” as a subject (authentic English expression) is far more frequently found in human writing (#1~#6, and #1~#4). “This paper…” phrase was not found in the top 50 of the machine writing list.

Results: Goal 2

While the human rater had great difficulty in identifying NMT machine-translated text, machine could still detect machine-translated texts. Models were created through deep learning using Multilayer Perceptron (MLP) neural network by TensorFlow to establish a system detecting machine-translated texts. From 385,184 machine-translated and 193,922 human-written English, we selected 376,000 sentences for machine-translated data and 193,000 sentences for human-written data. The accuracy rate ranged from 49.9% to 66.7% when we experimented the data with TensorFlow. This result is clearly not an indication of any kind of learning on the attributes of machine-translated texts.

In the second experiment, we ran the learning session with SVM using TF-IDF. The examination on the identification of machine-translated academic text using SVM with TF-IDF was conducted using the same set of data used in the first experiment with deep learning. Contrary to the case of the previous experiment with MPN, we got a consistently high accuracy rate, ranging from 89% to 99%, across the rounds. This high accuracy result, in comparison to the overall low accuracy rates of the previous experiment (49.9%~ 66.7%), indicates that supervised learning focusing on the use of keywords within and across the corpus is more suitable machine learning technique in identifying machine-translated texts.

The results show that machine-translated text is identifiable with its unique traits and style. It is promising that the high accuracy rate of detecting machine-translated text can help researchers to develop a system that can help us spot machine-translated writing, and EFL instructors can make use of it to instruct L2 writing. One thing to note about the traits of machine-translated text is though that it is not clear whether those traits are something humans can recognize or spot. Regarding this, we concluded that the result of the survey we conducted on EFL instructors and the n-gram analysis can help us better with the ideas for methods distinguishing the two types of writing for human raters and with creating educational materials for students.

Results: Goal 3

Based on our observations, the results of our survey on EFL instructors, and n-gram analysis, educational materials were created teaching effective and appropriate usage of Google Translate to generate more natural and accurate academic English sentences. Then, it was distributed to students enrolled in Thesis Writing at a Japanese university. Our material focus on the following two main areas: 1) Editing the original Japanese text (input), and 2) points to consider when revising machine-translated outputs. Regarding the Japanese input, we have included subsections such as "Avoiding Japanese idiomatic expressions" and "Avoiding long/complex sentences." We have provided examples of sentences that, when fed into machine translation, produce incorrect outputs. Regarding the machine-translated outputs, we have created subsections that demonstrate how to revise typical machine-translated expressions to polish them according to the conventions of academic writing. For example, regarding the use of the Japanese formal writing expression “その結果”, we showed how it is often machine-translated into “as a result,” which is okay but not quite natural in most contexts. With a sample sentence, we suggested a revision of the given machine-translated text. This educational material instructing students how to use Google Translate effectively and appropriately for English academic writing was shared with students enrolled in Thesis Writing via the university LMS as shown in Figure 2.

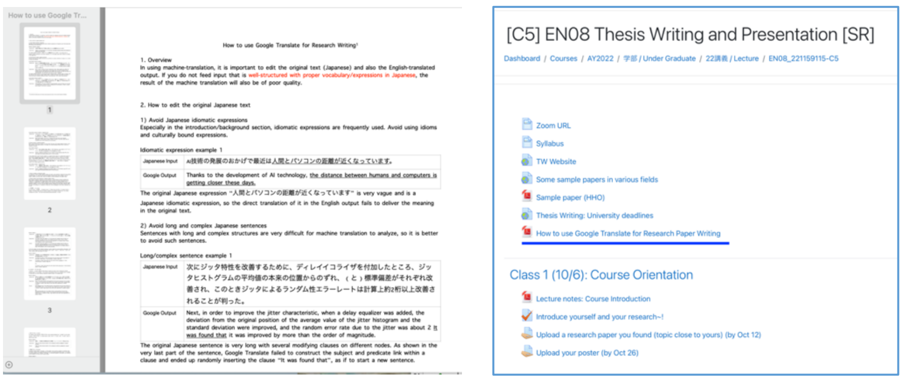

Figure 2. Materials uploaded on the university LMS

5．主な発表論文等

〔雑誌論文〕　計2件（うち査読付論文　0件／うち国際共著　2件／うちオープンアクセス　0件）

1．著者名 Younghyon Heo, Jeremy Perkins and Hyowon Song	4．巻 24
2．論文標題 Discrimination between Machine-translated and L2 Human-written Text: Features Identified by English Teachers	5．発行年 2019年
3．雑誌名 Proceedings of Pan-Pacific Association of Applied Linguistics	6．最初と最後の頁 92-93
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1．著者名 Younghyon Heo, Jeremy Perkins and Incheon Paik	4．巻 1
2．論文標題 Identification of Machine-Translated Texts Using Machine Learning	5．発行年 2018年
3．雑誌名 Proceedings of Discourse and Cognitive Linguistics Society of Korea	6．最初と最後の頁 147-151
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

〔学会発表〕　計3件（うち招待講演　0件／うち国際学会　3件）

1．発表者名 Younghyon Heo, Jeremy Perkins and Hyowon Song
2．発表標題 Discrimination between Machine-translated and L2 Human-written Text: Features Identified by English Teachers
3．学会等名 Pan-Pacific Association of Applied Linguistics（国際学会）
4．発表年 2019年

1．発表者名 Younghyon Heo, Jeremy Perkins and Incheon Paik
2．発表標題 Identification of Machine-Translated Texts Using Machine Learning
3．学会等名 Discourse and Cognitive Linguistics Society: Spring Conference（国際学会）
4．発表年 2018年

１．発表者名 Younghyon Heo
２．発表標題 Instructing how to use Google Translate properly
３．学会等名 Korean Association of Language Sciences（国際学会）
４．発表年 2023年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

－

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	Ｐｅｒｋｉｎｓ　Ｊｅｒｅｍｙ (Perkins Jeremy) (30725635)	会津大学・コンピュータ理工学部・上級准教授 (21602)	
研究分担者	Ｐａｉｋ　Ｉｎｃｈｅｏｎ (Paik Incheon) (70336478)	会津大学・コンピュータ理工学部・教授 (21602)	

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関