The 34th Conference on Computational Linguistics and Speech Processing (ROCLING 2022)
Taipei, Taiwan, November 21-22, 2022. The Association for Computational Linguistics and Chinese Language Processing

Using Grammatical and Semantic Correction Model to Improve Chinese-to-Taiwanese Machine Translation Fluency

Yuan-Han Li ncku10509@gmail.com

Chung-Ping Young dryncku@gmail.com

Wen-Hsiang Lu whlu@mail.ncku.edu.tw

National Cheng Kung University
Department of Computer Science and Engineering

Abstract

Currently, there are three major issues to tackle in Chinese-to-Taiwanese machine translation: multi-pronunciation Taiwanese words, unknown words, and Chinese-to-Taiwanese grammatical and semantic transformation. Recent studies have mostly focused on the issues of multi-pronunciation Taiwanese words and unknown words, while very few research papers focus on grammatical and semantic transformation. However, there exist grammatical rules exclusive to Taiwanese that, if not translated properly, would cause the result to feel unnatural to native speakers and potentially twist the original meaning of the sentence, even with the right words and pronunciations. Therefore, this study collects and organizes a few common Taiwanese sentence structures and grammar rules, then creates a grammar and semantic correction model for Chinese-to-Taiwanese machine translation, which would detect and correct grammatical and semantic discrepancies between the two languages, thus improving translation fluency.

Keywords: Machine translation, Taiwanese grammatical rules, Lexical transformation, Syntactic transformation, Chinese-to-Taiwanese

1 Introduction

Machine translation systems are being increasingly used across multiple fields, such as businesses, tourism and medical industries. These systems can translate multiple languages like English, Chinese, Japanese and even obscure traditional languages such as Swahili and Croatian. Taiwanese machine translation is likewise gaining importance as the government becomes more aware of its historical and cultural importance, thus setting out to digitally preserve the language. There has been multiple research papers on Chinese-to-Taiwanese (henceforth referred to as C2T) machine translation, usually focusing on the issues of multi-pronunciation words and unknown words translation. However, despite the importance of grammatical and semantic differences between languages in machine translation, it has been observed that papers that integrate them into C2T translation systems are comparatively rare, which results in the output of C2T systems losing fluency and potentially the original meanings of the original Chinese input. As such, this paper proposes a grammatical and semantic error detection and correction model, which can improve C2T translation fluency by correcting the discrepancies between Chinese and Taiwanese grammar.

2 Related Work

Considering the importance of machine translation (Raad, 2020; Panayiotou et al., 2020; Kapoor et al., 2019), researchers begin to apply rule-based (Hurskainen and Tiedemann, 2017), statistical (Och et al., 1999; Koehn et al., 2003), and neural machine translation (Sutskever et al., 2014; Vaswani et al., 2017) technology to various languages, including C2T translation models (Lin and Chen, 1999), in order to tackle various recurring issues in this field, such as choosing the correct pronunciation for multi-pronunciation words and unknown words.

For multi-pronunciation words, (Wu, 2015) extracts the features of each word in the input sentence, such as part-of-speech (POS) and semantic meaning, then employ feature models

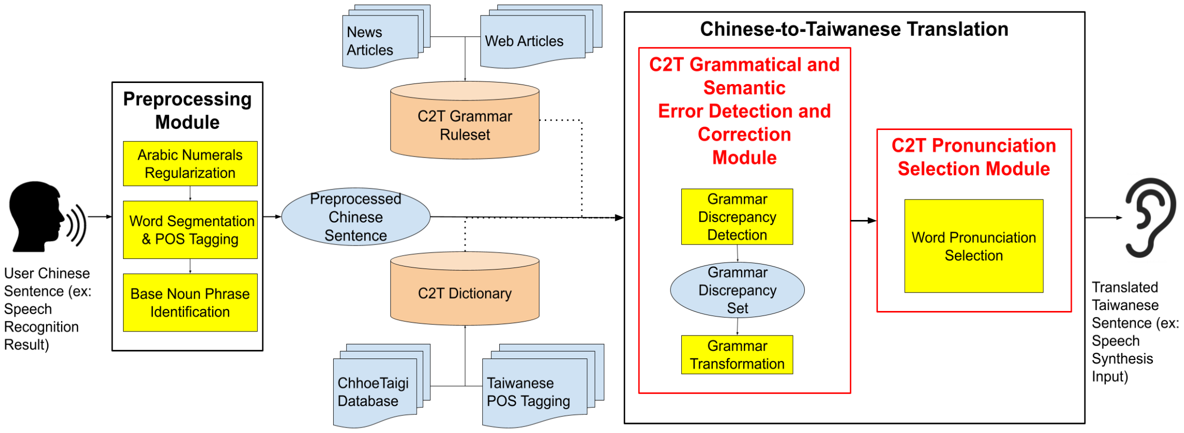

News Articles
Web Articles
C2T Grammar Ruleset
Preprocessing Module
Arabic Numerals Regularization
Word Segmentation & POS Tagging
Base Noun Phrase Identification
User Chinese Sentence (ex: Speech Recognition Result)
Preprocessed Chinese Sentence
C2T Dictionary
ChhoeTaigi Database
Taiwanese POS Tagging
Chinese-to-Taiwanese Translation
C2T Grammatical and Semantic Error Detection and Correction Module
Grammar Discrepancy Detection
Grammar Discrepancy Set
Grammar Transformation
C2T Pronunciation Selection Module
Word Pronunciation Selection
Translated Taiwanese Sentence (ex: Speech Synthesis Input)

Figure 1: System architecture of C2T machine translation using grammatical and semantic correction model

based on word features of co-occurring words and layered structure to select the most suitable translation rule for the Chinese words. For unknown words, (Chen, 2015) uses a mixed model which utilizes the prefix and suffix of words, pronunciation, word subsets, etc. to statistically analyze the corresponding Taiwanese pronunciation of unknown words in Chinese sentences. (Huang, 2015) also attempts to resolve insertion/deletion issues in C2T translation by compiling a set of insertion/deletion rules, calculate the confidence scores, then combine Naïve-Bayes and CRF statistical models to perform machine learning, improving the fluency of the Taiwanese translation output.

(Hsu et al., 2020), on the other hand, uses a Convolutional Neural Network (CNN) deep learning model, the C2T-pronunciation parallel corpus iCorpus and a precompiled Chinese-Taiwanese parallel dictionary to create a Chinese-character-to-Taiwanese-pinyin module and perform whole-sentence translation. However, the paper does not focus on the issues of multi-pronunciation words and unknown words encountered during C2T translation.

In comparison to the issues of multi-pronunciation words and unknown words pronunciation, C2T grammatical and semantic level issues, such as structural transformation or split Chinese words, are less commonly explored in these research papers. As such, this paper explores sentence structure transformation based on Taiwanese grammar.

3 Methods

3.1 System Architecture

Figure 1 shows the architecture of the C2T translation model using grammatical and semantic correction model. The system consists of two major modules: the preprocessing module, which regularizes numeral words in the Chinese sentence input, performs word segmentation and part-of-speech(POS) tagging, and identifies base noun phrases (BNP) in the input sentences, and the C2T translation module, which detects and corrects grammatical and semantic errors in Chinese sentences into their Taiwanese counterparts, then translates each word in the sentences. A detailed introduction will be listed in the following sections.

3.2 Preprocessing Module

To perform grammatical and semantic error correction (such as word-order transformation) and pronunciation selection in later modules, the C2T translation system in this study uses a preprocessing module that not only performs word segmentation and POS tagging, but also correct certain errors in user input sentences that are not strictly tied to grammatical errors, but nevertheless affect translation accuracy and fluency. These errors are described below.

(1) Arabic Numerals Regularization The preprocessing module regularizes Arabic numeral by transforming them into either traditional Chinese or modern Chinese depending on the semantic meaning. Traditional Chinese

Pronunciation type	Numeral type	Examples
Modern pronunciations	Number + Unit word	117 片 → 一百十七片、2 盒 → 二盒
	Time/Date	5 月 17 日 → 五月十七日
Classical pronunciations	Phone numbers	防疫專線 0800-001922 → 防疫專線控捌控控-控控一玖貳貳
Ordinal pronunciations	Ordinals (第 + Number + Unit word)	第 2 名 → 第貳名、第 5 位 → 第五位

Table 1: Common Arabic numerals examples and their corresponding pronunciation

(壹～玖，控) would be used for classical Taiwanese numerals, while modern Chinese (一～九，零) would be used for modern Taiwanese numerals. Table 1 shows some common Arabic numerals examples and their corresponding pronunciation.

(2) Word Segmentation and POS Tagging To select the correct pronunciation and retrieve the information needed for grammatical transformation, the module uses the CKIP word segmentation/ POS tagging system to segment the Chinese sentence input and perform POS tagging in preparation for the pronunciation selection process.

(3) Base Noun Phrase (BNP) Identification Some segments involved during sentence structure word reordering, especially noun subjects and objects, usually have phrases as their minimal unit. The Hanlp toolkit is able to use dependency parsing to obtain the noun phrases within a sentence. As such, the module also employs Hanlp toolkit to identify the noun phrases in a given Chinese input sentence so that the C2T grammatical and semantic error detection and correction module can successfully detect and revise the errors found during input. For example, in Taiwanese, "志明跑步比隔壁教室的阿甘跑得快" is translated as "志明走了 khah 隔壁教室 ê 阿甘緊". The phrase "隔壁教室的阿甘" is a complete object noun phrase that has its position swapped with "比" and "跑得".

3.3 Chinese-to-Taiwanese Translation

The translation module incorporates information from the precompiled C2T dictionary and grammar ruleset, and consists of two components: C2T grammatical and semantic error detection and correction module and C2T pronunciation selection module. Each component of the translation module will be detailed in the following sections.

3.3.1 C2T dictionary

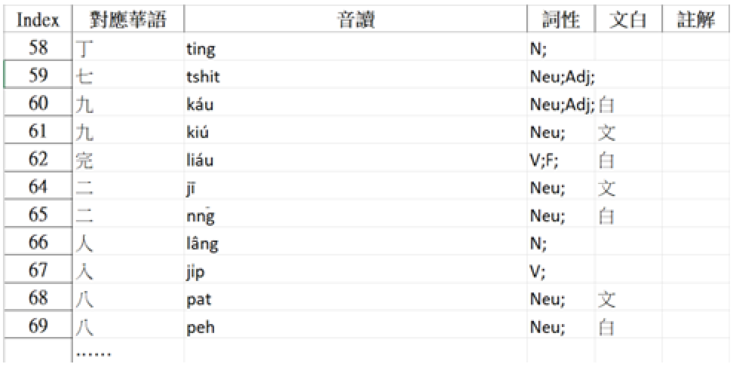
Figure 2 shows the structure of the dictionary used in this system. The dictionary uses the pronunciations taken from the Taiwanese Common Words Dictionary by MOE and the Chhoetaigi Taiwanese corpus organized by public sources. Since some Taiwanese pronunciations for words in Chhoetaigi Taiwanese corpus are not commonly used in the modern age, they are filtered out while the rest are compiled into the new C2T dictionary.

Each entry contains a Chinese word, the corresponding Taiwanese pronunciation, the part of speech (POS) of the Chinese word and whether the pronunciation is considered classical (文言) or modern (白話) (Table 2) .

Chinese word	Taiwanese Pronunciation	POS	Wenbai
香	hiong	N;	文
	hiunn	N;	白
	phang	Adj;	白
端午節	bah-tsàng-tseh ; gōo-gueh-tseh	N;	白

Table 2: Taiwanese words entries in C2T dictionary

For words with multiple accepted translations, such as "端午節" as either "bah-tsàng-tseh" or "gōo-gueh-tseh " (Table 2), all translations are compiled into the same entry.



Index	對應華語	音讀	詞性	文白	註解
58	丁	ting	N;		
59	七	tshit	Neu;Adj;		
60	九	káu	Neu;Adj;	白	
61	九	kiú	Neu;	文	
62	完	liáu	V;F;	白	
64	二	jī	Neu;	文	
65	二	nn̄g	Neu;	白	
66	人	lâng	N;		
67	入	jip	V;		
68	八	pat	Neu;	文	
69	八	peh	Neu;	白	
	......				

Figure 2: C2T dictionary corpus

The pronunciation selection module would use these criteria to decide on the translation of Chinese words in a given input sentence.

3.3.2 Chinese-Taiwanese Grammatical Ruleset

Figure 3 shows the types and amount of Chinese-Taiwanese grammatical and semantic differences occurring in news articles. This study collects Taiwanese grammatical rules from news articles, web articles, and Wikipedia, and compiles them into eight major categories for the C2T grammatical error detection and correction module ruleset. Furthermore, 16259 Chinese sentences are extracted from 1851 news articles, with empty strings and repeated sentences removed, in order to analyze the appearance frequency of each type of grammatical differences between Chinese and Taiwanese sentences.

3.3.3 C2T Grammatical and Semantic Error Detection and Correction Module

The preprocessed Chinese sentence would be sent to the grammatical and semantic error detection and correction module to undergo grammatical transformation. The module would output a set of grammatical discrepancies found in the Chinese sentence based on the compiled Taiwanese sentence structure and grammatical rules, then perform word switching or word order revision depending on the corresponding error revision method for each rule. The output of the module is a sentence that complies to Taiwanese grammar. A few of these grammatical rules are listed in Table 3.

3.3.4 C2T Pronunciation Selection Module

After the grammatical and semantic error detection and correction module transforms the Chinese sentence to better fit Taiwanese grammar, the sentence would be inputted into C2T pronunciation selection module. The module would then select the correct pronunciation of each Chinese word by looking up the dictionary for its POS, corresponding Chinese entry and Wenbai pronunciation, and output the translated sentence.

(1) Abbreviation Word Restoration: In news articles, certain Chinese words tend to be abbreviated, like "因爲" being abbreviated as "因", "但是" as "但" and "可以" as "可". However, since native Taiwanese speakers usually speak the whole original word, the module would restore them into their original forms based on the semantic context so that they can be properly translated.

(2) Word Pronunciation Selection: For each word in the input sentence, if there are multiple Taiwanese translation entries that fit the POS of the original word, the module would prioritize entries with modern pronunciation. Otherwise, it chooses the entry with the smallest index number. For word entries with multiple pronunciations, the module chooses the first pronunciation. Unknown words are translated by dissecting them into characters and translating them separately. In addition, some words have suffixes with unique meanings, and the module would translate them separately from the main word. The algorithm of the pronunciation selection module is described below in Algorithm 1.

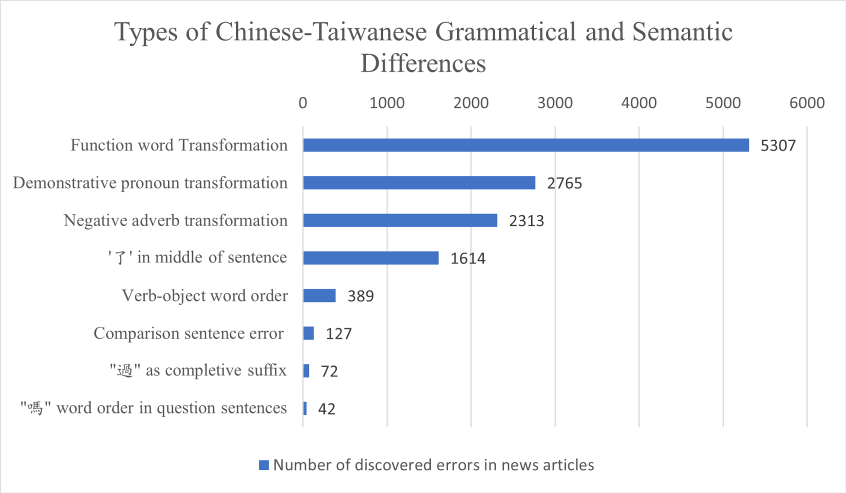


Figure 3: Types of Chinese-Taiwanese grammatical and semantic differences

Grammatical differences	Transformation rule(s)	Examples
“共” sentences (kā sentence)	function words such as “把, 向, 跟” → “kā”	• 他向老爸借兩百萬 → I kā lāu-pē tsioh nn̄g pah-bān • 醫生跟你說要多喝水 → I-sing kā lí kóng ài ke lim-tsuí • 大家把小偷抓到警察局了 →Tak-ke kā tshat-á liah-khì kíng-tshat-kiok--ah
“了”at the end of sentences	“了” → “--ah”	• 燈光照到他了 → Ting-hué tsiò-tioh i--ah • 我找到她了 → Guá tshuē-tioh i--ah • 他把飲料買回來了 → I kā liâng-tsuí bé-tńg-lâi--ah
Negative adverb translation	不 +V. → m̄ (毋)+V. 不 +Adj. → bô (無) +Adj.	• 這裡是不穩定的環境 → Tsia sī bô ún-tīng ê khuân-kíng • 我不知道 → Guá m̄-tsai-iánn • 他不聽我的話 → I m̄ thiann guá ê uē
Basic comparison sentences	A + 比 +B + adj → A + khah + adj + B	• 女兒比兒子貼心 →Tsa-bóo-kiánn khah tah-sim tsa-poo-kiánn • 他比你高一點兒 → I khah lò lí tsit-sut-á • 你比他好不到哪兒去 → Lí khah-hó i bô-guā-tsē
“了” preceded by verb	V.+ 了 → ū (有) + V.	• 弟弟買了一台機車 → Sió-tī ū bé tsit tâi oo-tóo-bái • 媽媽吃了一粒蘋果 → A-bú ū tsiah tsit-liap phōng-kó • 她贏了五百元 → I ū iânn gōo-pah khoo
“嗎” word order in question sentences	Translate “嗎”as “kám” (敢), and advance its position to right after the subject	• 他知道這件事嗎?→ I kám tsai-iánn tsit-khuán tāi-tsì? • 你找到它了嗎? → Lí kám tshuē-tioh i--ah? • 那件事情很困難嗎? → Hit-kiānn tāi-tsì kám tsiok khùn-lân?

Table 3: C2T grammatical differences and transformation rules

Algorithm 1 Pronunciation Selection

```
for Every Chinese word in segmented sentence input do
    if Word has special suffix then
        Translate the suffix separately
    else
        Check the number of entries with the same Chinese word and POS
        if Word has 1 corresponding entry then
            Apply word as translation
        else if Word has multiple corresponding entries then
            Select entry with smallest index number
        else
            Translate each character separately
        end if
    end if
end for
```

4 Experimental Results

4.1 Dataset and Evaluation Metrics

To evaluate the system built for this study, thirty news articles are selected and divided into 5 categories: social, lifestyle, economics, weather and technology, each with 6 articles. In total, 265 sentences and 8667 words from the articles are used to test the C2T machine translation system. For the grammatical and semantic error detection and correction module, grammatical correction rate is utilized as evaluation criteria, which is defined as:

> Grammatical correction rate = a/b, where a is the number of grammatical errors detected and corrected by the module, and b is the number of grammatical errors in news article Chinese sentences

For the pronunciation selection module, Word Error Rate (WER) is utilized as the evaluation criteria. Its formula is defined below:

> WER = x/y, where x is the number of erroneously translated words, and y is the number of total words

The results of each experiment are listed in Table 4 and Table 5.

In addition, 10 sentences from the dataset are selected to evaluate the C2T system in this paper (henceforth referred to as 公跨麥) against 3 other baseline systems: the C2T machine translation system developed by NCKU (Pan, 2021), the popular Ithuan Dopaiji Taiwanese translation system [1], and the translation system developed by Hsu et al., available at National Chiao Tung University Speech Communication Lab (NCTU SCL) website[2]. The comparison results are shown in Table 6.

Amongst the example sentences chosen, it is observed that the NCKU, Ithuan and NCTU SCL systems are all unable to translate arabic numerals correctly, in addition to not being able to correct grammatical errors, in particular errors that involve word order transformation, such as the positional differences between the question particle “嗎” and its Taiwanese counterpart “敢”. For example, in the sentence “5000 元的藍牙耳機讓我眞的會好奇，眞的有這個價值嗎”, 公跨麥 translates “5000” into “gōo-tshing” and successfully moves the question particle “嗎” to immediately after the invisible subject “這”, both of which the other three systems failed to revise.

4.2 Error Analysis

4.2.1 C2T Grammatical and Semantic Error Detection and Correction Module

(1) “共” Sentences Translation Error In the C2T error correction module, prepositions like “把、跟、向” are transformed into their corresponding Taiwanese word “共”(kā), however there is one exception for “向”: if there is a directional word following “向”, such as “上” and “下”, then even though “向” is still a preposition, it also contains the semantic meaning “朝... 方向” alongside the properties of a verb, therefore it should be translated as “hiòng” rather than “kā”.

(2) “了” Particle Transformation Error The particle “了” has different transformation rules depending on its position. For instance, when “了” is placed in the middle of the sentence and after a verb, it may semantically mean “completion of an action”, in which case

[1] https://suisiann.ithuan.tw/

[2] http://tts001.iptcloud.net:8804/

Grammatical rules	**Function word transformation**	**Negative Adverb Translation**	**Demonstrative pronoun Translation**	**Comparison sentences word order revision**
Number of appearances	85	36	44	2
Number of correctly revised errors	78	33	43	1
Grammatical rules	**“了” in middle of sentence**	**“嗎” word order in question sentences**	**“過” as verb suffix**	**Verb-object word order revision**
Number of appearances	14	3	1	3
Number of correctly revised errors	9	1	1	0
Grammatical correction rate = 87.6%				

Table 4: Grammatical and semantic error detection and correction module experiment result

Total Words	**Number of erroneously translated words**	**Number of missing words**	**Number of untranslated words**	**Number of extra inserted words**	**Number of erroneously positioned words**
8667	865	30	20	69	26
WER ≈ 11.7%					

Table 5: Pronunciation selection module experiment result

“了” is translated as “有” (ū) and switches positions with the verb before it. In verbs with an adverb inserted in the middle like “上錯了菜”, however, “了” can be seen as a particle without meaning and be omitted, yet our correction module only swaps “了” with the word before it without considering its POS, creating erroneous transformations like “上有錯菜”.

4.2.2 C2T Pronunciation Selection Module

In the C2T pronunciation selection module, aside from failing to translate words not included in the dictionary corpus, there are also instances observed in which new words not in the original Chinese sentence are supposed to be inserted into the Taiwanese translation to improve semantic fluency. Examples include “見巡邏車經過時”, in which an extra word “到” is inserted after “見” in the Taiwanese translation as “看到巡邏車經過時”, and instances in which an additional unit word is inserted between a Chinese number directly followed by a noun, such as “四人” being translated into Taiwanese as “四個人”, with an additional unit word “個” inserted between “四” and “人”.

5 Conclusion and Future Work

This paper focuses on correcting grammatical and syntactic differences between Chinese and Taiwanese encountered during C2T machine translation by building a grammatical and semantic error detection and correction module, which can transform the grammar of the Chinese sentence inputs into their corresponding Taiwanese sentence structures in accordance

System name	**WER**	**Translation sample sentence 1:** 男子酒測值高達**0.83**毫克，被依公共危險罪送辦。	**Translation sample sentence 2:** **5000** 元的藍牙耳機讓我真的會好奇，真的有這個價值嗎？
NCKU	39.2%	lâm-tsú tsiú tshik tat kuân-kàu X.XX hô-khik ，pī i kong-kiōng guî-hiám tsuē sàng pān 。	XXXX guân ê nâ gê hī ki niū guá ū-iánn ē-hiáu hònn-kî ，ū-iánn iú tse ê kè-tat kiám？
Ithuan	32.8%	lâm-tsú tsiú tshik tat ko tat 0.83 hô-khik,pī i kong-kiōng guî-hiám tsuē sàng pān.	5000 guân ê nâ gê hīnn-ki niū guá tsin tik ē hònn-kî,tsin ê Ū tsit kò kè-tat má?
NCTU SCL	20%	tsa1 poo1 tsiu2 tshik4 tat8 kuan5 kau3 phi5 khi3 sam1 ho5 khik4 , pi7 an3 kong1 kiong7 hui5 hiam2 tsue7 sang3 pan7 .	tiat8 si7 tsin1 guan5 e5 lam5 ge5 hi7 ki1 hoo7 gua2 tsin1 e5 hue7 honn3 ki5 , tsin1 e5 u7 tsit4 e5 ke3 tat8 ma1？
公跨麥	8%	tsa-poo-tsiú-tshik-tit kuân-kàu **không-tiám-pat-sam**-hô-khik，**hōo** i kong-kiōng-guî-hiám-tsuē sàng-pān 。	**gōo-tshing-khoo**-ê lâm-gâ-ní-ki hōo guá ū-iánn ē hònn-kî, **kám** ū-iánn ū tsit tsit-ê-kè-tat ?

Table 6: C2T translation system performance evaluation. The NCTU SCL system uses numbers to denote Taiwanese tone, while the other three systems use Tailuo tone symbols. (Blue words are translations of numeral words, and red words denote grammatical discrepancies between Chinese and Taiwanese. We denote the revised errors in bold words.)

with the grammatical rules of the target language.

Experiment proves that a grammatical and semantic error detection and correction module can successfully improve translation fluency of C2T machine translation. The correction module can be widely used in areas that require machine translation, and would greatly contribute to meetings, tourism, language education, elder care, AI, etc.

Future work would include expanding the dictionary to include more words exclusive to Chinese which are found in news articles, such as technical terms, and idioms. In addition, for some Chinese words and sentence structures, translating them into Taiwanese and transforming the grammatical structure requires deeper knowledge of the original semantic meaning. Sometimes the translation result may be structured differently from the original, and this issue would also be explored in the future. Lastly, the grammatical transformation rules and words used in this paper may not be commonly used by native Taiwanese speakers, since translating from Chinese to Taiwanese is relatively lax. As such, a field survey may be conducted in the future to collect native speakers' opinions and feedback on the results of this translation system to help it output sentences that better fit the users' daily usage habits.

References

Shih-Hsiang Chen. 2015. Decision for pronunciation of out-of-vocabularies in a mandarin to taiwanese text-to-speech system. Master's thesis, National Chung Hsing University.

Wen-Han Hsu, Cheng-Jung Tseng, Yuan-Fu Liao, Wern-Jun Wang, and Chen-Ming Pan. 2020. A preliminary study on deep learning-based Chinese text to Taiwanese speech synthesis system. In *International Journal of Computational Linguistics & Chinese Language Processing, Volume 25, Number 2, December 2020*, Taipei, Taiwan. Association for Computational Linguistics and Chinese Language Processing.

Chih-Chao Huang. 2015. A study on example-based mandarin-taiwanese machine translation. Master's thesis, National Taiwan Ocean University.

Arvi Hurskainen and Jörg Tiedemann. 2017. Rule-based machine translation from English to Finnish. In *Proceedings of the Second Conference on Machine Translation*, pages 323–329, Copenhagen, Denmark. Association for Computational Linguistics.

Ravish Kapoor, Angela Truong, Catherine Vu, and Dam-Thuy Truong. 2019. Successful verbal communication using google translate to facilitate awake intubation of a patient with a language barrier: A case report. *A A Practice*, 14:1.

Philipp Koehn, Franz J. Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pages 127–133.

Chuan-Jie Lin and Hsin-Hsi Chen. 1999. A Mandarin to Taiwanese Min Nan machine translation system with speech synthesis of Taiwanese Min Nan. In *International Journal of Computational Linguistics & Chinese Language Processing, Volume 4, Number 1, February 1999*, pages 59–84.

Franz Josef Och, Christoph Tillmann, and Hermann Ney. 1999. Improved alignment models for statistical machine translation. In *1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*.

Guan-Xun Pan. 2021. Taiwanese speech synthesis system based on taiwanese tone sandhi and implementation of chinese to taiwanese translation. Master's thesis, National Cheng Kung University. Unpublished thesis.

Anita Panayiotou, Kerry Hwang, Sue Williams, Terence Chong, Dina Logiudice, Betty Haralambous, Xiaoping Lin, Emiliano Zucchi, Monita Mascitti, Anita Goh, Emily You, and Frances Batchelor. 2020. The perceptions of translation apps for everyday health care in healthcare workers and older people: A multi-method study. *Journal of Clinical Nursing*, 29.

Bareq Raad. 2020. The role of machine translation in language learning. *Academic Research International*, 7:2348–7666.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. *CoRR*, abs/1409.3215.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, L ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Shi-Yao Wu. 2015. Combing a multi-feature model and a layer approach in solving the polysemy problem in a chinese to taiwanese tts system. Master's thesis, National Chung Hsing University.