投稿類別：英文寫作類

篇名：

Informational Warfare in Taiwan: A Linguistic Study of Falsified Government Official Documents

作者：

廖珮妤。康橋國際學校林口校區。11 年 E 班

指導老師：

雷翔宇 老師

I. Introduction

(I) Information Warfare

Information warfare is a new type of warfare currently gaining more international attention. It is the procreate of the modern and technological era. As long as there is information, there is information warfare, its omnipresence making it incredibly diverse.

One typical type of information warfare is fake news. While most types of information warfare might be limited to military and economic uses, fake news, a new type of information warfare, can target any and all subjects, including civilians. Take Taiwan for example: there is a lot of suspicious information in a regular Taiwanese person's daily life. This phenomenon triggered the birth of Taiwan FactCheck Center (TFC), a platform that allows governmental professionals to analyze and fact-check suspicious news and helps people identify fake news. Many studies taking different approaches, especially from a social science and linguistic approach, have been trying to accomplish the same goal.

Yet there is one kind of fake news that has not been studied, one that is new and going rampant: falsified official documents. Because of the growing tension between China and Taiwan, these documents are getting more and more common. This situation calls for more attention and discussion.

(II) China-Taiwan Tension and Variants of Mandarin Chinese

Taiwan was first introduced to the Chinese language family around the seventeenth century, but its official language had been Japanese (due to the fifty-year Japanese colonialism) until 1945, when the Republic of China defeated the Japanese colonial empire. Standard Mandarin became Taiwan's new official language. The China-Taiwan tension started in 1949, when Taiwan gained independence from the mainland. Even with Taiwan being independent ever since, China still claims that Taiwan is a rebellious province under it that does not admit its loyalty, and vows to "unify" Taiwan one day with any necessary way. The current Taiwanese President's political stance leans more towards Taiwan independence than towards Chinese unification, which adds fuel to the fire.

Although the official language in both China and Taiwan is Mandarin, after being separated from the mainland, the Taiwanese residents developed their own variant of the Mandarin in Taiwan, called Taiwan Mandarin. Similar to how British and American English differ in terms of speech, Mandarin in Mainland China and Taiwan Mandarin have differences in syntax and phonology; the two Mandarins, however, have more distinct differences in their orthography than British and American English do. Most people thought traditional Chinese, or the orthography of Taiwan Mandarin, and Simplified Chinese, or the orthography of Mandarin in Mainland China, are only difference by how they look. However, other distinctions between the two writing systems are not to be disregarded. The present study aims to identify and study those differences in Traditional and Simplified Chinese in falsified official documents in this paper.

(III) Research Questions

1. What kinds of linguistic tokens can be found in falsified official government documents?
2. Which linguistic tokens appear the most frequently in falsified official government documents?
3. Which linguistic tokens are the most commonly found across falsified official government documents?

(IV) Significance of the Study

Fake news, falsified official government documents, and linguistics do not often appear in the same sentence, but because of how fake news are threatening the world's security and falsified official government documents are threatening Taiwan's, one should look for any and all possible solutions and techniques to protect themselves and their country. Linguistics might just be the unexpected but right tool to detect falsified official government documents and the solution to the issue that impacts not only Taiwan's, but the entire world's safety.

II. Literature Review

Information warfare is a global phenomenon that is defined as a physical or virtual attempt to manipulate a country, organization, or individual's actions (Libicki, 1995, as cited in Hewitt, 2009). Schwartau (1997) bases off of this idea to further type information warfare into three classes according to the scale and targets of the attack while assuming that the attack occurs in cyberspace. Class 1, or personal information warfare, is "**an attack against an individual's electronic privacy.**" Class 2, or corporate information warfare, targets "**industrial and economic data**" and often involves "**industrial espionage.**" Class 3, or global information warfare, is "**waged against ... political spheres of influence, global economic forces, and even ... entire countries**."

There is another type of classification of information warfare: Martin C. Libicki (1995, as cited in Damjanović, 2017) focuses on means instead of subjects. Under Libicki's classification, information warfare can be studied from several aspects, including military, electronics, surveillance, economy, technology, and psychology. "Psychological warfare" is where fake news best belong: by definition, psychological warfare uses information against the human mind to "**influence [people's] emotions, motives and objective reasoning and, ultimately, the behavior of foreign governments, organizations, groups and individuals to achieve** [the attacker's] **own interests and goals,**" which is exactly what fake news do. It is quoted from Noam Chomsky, a philosopher and political scientist, states that psychological warfare "**affects about 20% of the population, relatively educated, who takes part in making certain decisions,**" making even the knowledgeable people vulnerable to fake news and other psychological attacks. Psychological attacks, including fake news, are considered dangerous and permanent in contemporary conflicts.

Because fake news has the potential to be extremely harmful and should be taken with utmost seriousness, researchers have been studying and analyzing it from various approaches including linguistics. Marquardt (2019) had tried studying the correlation between a news article's accuracy and the connotations and emotions evoked by the words in the title. She also attempted to find a connection between news accuracy and 1) the title's noun-to-verb ratio, 2) mean sentence length in words, 3) frequency of keywords appearing in the article, and 4) text complexity level. Her finding proves four of the five factors (connotations words, noun-to-verb ratio, frequency of keywords, and sentence length) unreliable since she could not find any explainable relationship between them and validity of the news article. Marquardt does, however, consider text complexity level as a reliable indicator; she concludes that news articles with highly complex titles are more likely to be true. She also determines in extension of keyword frequency and word connotations that the more keyword connotation is related to the overall mood of the article, the more likely the news is true.

Mahyoob et al. (2020) took a similar approach to study fake news, but instead of studying the language of news article titles to predict news accuracy, the they studied the linguistic features (or more specifically, grammatical features) of news on social media platforms. This study is quite successful, as it produced an abundance of authentic results. It concluded that authentic news has more passive voice, reported speech, negation, and proper noun use, but a lot less conjunctions, to-infinitives, modals, and long sentences than fake news do. Moreover, fake news has more superlatives and excessive emotional adverbs. These indicators are, according to Mahyoob et al., all reliable enough to predict a social media news' accuracy.

Using linguistic knowledge to identify fake news is a relatively new field of study, therefore there aren't enough papers done to go more in depth with it. The author has not found any research papers done to spot falsified official government documents, so this might be the pioneer for future research of similar topics.

III. Research Methods

The present study collects a total of nine Taiwan's government official documents (OGDs) that are already marked as fake as the subject of this paper, all of them printed between 2015 to 2021. They are all the documents the author can find that are publicized and clear enough to be studied, as the sources are extremely but unsurprisingly scarce possibly due to the government's likely attempt to shield the public from any falsified documents. Documents that are potentially existent spreading within the government are not open to the public, so the author decides against studying them. All but one of the documents are short, with contents no more than one A4-sized page. All of the documents are collected from the Internet.

These documents vary in genre, from the Taiwanese President's letter to a seizure order, and from a city government's public announcement to the police department's report. The wide variation provides the author ample information about how China attempts to interfere with not only Taiwan's politics, but also its military, health, economy, and social conditions.

The detailed descriptions of the nine documents are listed below. Document A is a seizure order from the National Ministry of Justice made effective on August 11th, 2015. Document B is an announcement from the Wugu District Office about public database upgrades made public on October 27th, 2017. Document C is a message from the Police Department to the public about the safety of social media LINE publicized on February 25th, 2019. Document D is a letter written on February 17, 2020, from the National Ministry of Health and Welfare to the Tainan major. Document E is an announcement from the Executive Yuan made public on February 26th, 2020 about policies regarding facial masks. Document F is a letter written on February 28th, 2020 from President Tsai to the Premier of Executive Yuan. Document G is an announcement from the Taoyuan City Government regarding the COVID-19 global pandemic made public on February 29th, 2020. Document H, printed is a letter written on October 2nd, 2020 from the Ministry of Justice Investigation Bureau to the Presidential Office Building. Document I is a letter written on April 16th, 2021 from the Presidential Office Building to the Ministry of Defense.

To analyze the linguistic tokens in a more unbiased and complete manner, the author ranks the tokens based on two assessment criteria: occurrence frequency and commonality. A token type's occurrence frequency is the total number of the token type in all OGDs, and a high occurrence frequency indicates a high chance of finding multiple instances of the said type of token in FOGDs. A token type's commonality is the number of FOGDs that have instances of the token type. A high commonality indicates a high chance of spotting the said type of token in any given FOGD.

IV. Analysis and Results

There are a total of ninety-four linguistic tokens found in these nine FOGDs. They are put into three main categories—format errors, font/form discrepancies, and cultural differences—that represent the three ways one can recognize fake government documents from China.

FOGDs often fail to replicate a real document's format, so format errors are the first ones people notice in an FOGD. In a linguistic sense, format errors are equivalent to syntactic and social genre differences between Traditional Chinese and Simplified Chinese that will erroneously alter the

sentence structure. These format errors' presence indicate that the document is not official, but they may not be enough to tell "who" is behind the act. Spotting format differences is the first step to recognize FOGDs, therefore "Format Errors" has to be the first category.

Font/form discrepancies are an important indicator that tell researchers that Chinese people are the ones who faked the government documents. These discrepancies can be considered as "visible" accents. As accent in phonetics and phonology is widely studied, font/form discrepancies are equally worthy of attention. Font/form discrepancies can be spotted without needing to understand any of the document's context or meaning; they can be discovered just by observing the characters' shapes, strokes, and overall appearance.

Cultural differences will then be the last category. A FOGD's "social habits"—or sociolinguistics—differentiates Taiwanese OGDs from Chinese ones. Cultural differences are therefore the third category to be discussed because they also play an important role in identifying Chinese FOGDs. We will go into more detail on the classifications in the following studies.

(I) Format Errors

Format errors are incorrect formatting and language of an FOGD. Out of 94 linguistic tokens found in all nine FOGD, 57 of them are format errors, which is 60.6% of all linguistic tokens. After further inspections, the author discovers that some format error tokens share certain similarities that allow the tokens to be grouped into more detailed subcategories: type one incorrect formatting (IF1); type two incorrect formatting (IF2); inconsistency (IC); incorrect grammar (IG); pleonasm (P); unclear content and confusing language (UCCL); indentation, spacing, and alignment error (ISAE); colloquial language and inappropriate abbreviation (CLIA); wrongly written characters (WWC); and legislation format error (LCE).

The first subcategory, type one incorrect formatting (IF1), focuses on any inconsistency with specific OGD format. To be put simply, there are a set of formatting rules that guide government officials to write OGDs, and any violations in these rules are considered IF1 errors. Take Document A for example: while the subject line is supposed to go before the instruction subtitle, Document A swaps the order and puts the instruction subtitle in front (see Figure 1). Also, Document C is missing its subject line and instruction subtitle altogether. Six out of nine FOGDs have IF1 errors, and that is 17.54% of format errors.

Type two incorrect formatting (IF2) has a larger target range than IF1: while IF1 applies to OGD writing, incorrect formatting is the violation of formatting rules for every formal document. One example can be found in Document A, where the personal information protection circles are hand-drawn and messy, not printed and neat like they should be. Also, in Document C, the page number at the footer of the document is surrounded by brackets that would normally be nonexistent in any formal document (see Figure 1). There are three out of nine FOGDs with incorrect formatting and a total of eight error examples, taking up 14.04% of the total number of format errors. Incorrect formatting is a relatively reliable guideline to detect FOGDs.

Inconsistency (IC) refers to any inconsistent language within the same FOGD. Document H demonstrates an example of inconsistency: the recipient of this document in the heading section writes "Presidential Office Building," but at the bottom of the document it writes "Presidential Office Building, President Tsai Ying-Wen's Office." Consistency is the most basic rule that all government officials should be familiar with and follow without doubt, so violations in consistency shows extreme unprofessionalism. Six out of nine documents have inconsistencies, hence inconsistency can now be considered a shared trait across most FOGD. There are a total of seven instances of inconsistency, taking up 12.28% of format error tokens.

Incorrect grammar (IG) are mistakes in combination or sequence of words that interfere with a person's understanding of the content. For example, inside the brackets in Document B is an incomplete sentence without a verb. There are a total of only four incorrect grammar in the FOGD. This type of error takes up just 7.02% of format error tokens.

Pleonasm (P) is the use of redundant words, phrases, or sentences in one description. Senders of OGDs avoid pleonasm because of its inability to convey information with concision. Document C, for example, unnecessarily repeats the word "user" in the sender company's terms of service. Three out of nine FOGDs have pleonasm, which take up 5.26% of all format errors. Although pleonasm is rare, it is a great way to detect FOGDs because it shows extreme unprofessionalism that is never allowed in OGD writing.

Unclear content and confusing language (UCCL) is another type of format error because of its vagueness and potential to mislead readers. One should bear in mind that UCCL is a different category from IG; the difference is that a document with UCCL does not necessarily have grammar mistakes but does have apparently fabricated file names and unclear references. Document D, for example, mentions an "estimation of potential population report" that isn't elaborated on. Similarly, Document H mentions a "secret service text" that is not directly related to the document's content. Five out of nine documents have UCCL. There are a total of nine occurrences, which is 15.79% of all format error tokens. According to the data, it is safe to assume that an OGD with unclear content and confusing language is highly likely to be a falsified one.

Layout Errors (LE) are the erroneous arrangement of the content that don't follow the standard layout of an OGD. Document D demonstrates three example of LEs: the content under the instruction subheading is not aligned neatly as it should; there is an extra space between a bullet point and the following sentence (see Figure 1); and the spacing between sentences is too wide. Five out of nine documents have LE, but the frequency of these errors' appearances is quite high: there are a total of ten instances, taking up 17.45% of all the format errors. The data makes ISA errors one of the most common and useful format errors to identify FOGD.

Colloquial language and inappropriate abbreviation (CLIA) are perhaps one of the most easily found format errors because they are incredibly out of place in the formal OGD. In Document H's heading section marks the document as "Top Secret," but it has been inappropriately abbreviated to the point that it loses its emphasis to the recipient. Likewise, Document G phrases its request for citizens to stay in Taoyuan City in a colloquial and blunt manner (see Figure 1). 5.26% of format error tokens are under the CLIA subcategory and three out of nine FOGDs share this issue.

Wrongly written characters (WWC) are characters that aren't the simplified Chinese version of the correct character, just ones that are used wrong. There is only one such instance in all nine FOGDs: in Document A, a character is mistyped into a wrong character. WWCs only take up 1.75% of all format errors.

Legislation formatting errors (LFE) are errors regarding the formatting or naming of a law. There are specific rules dedicated to making sure all legislations are following a certain format. LFE occurrences are a special case because of how it can overlap with other categories (e.g. type one formatting errors), but because it involves legal concerns it should be an individual category. One example can be found in Document G, whose cited law is missing its particle for the ordinal numbers. There are LFE cases appearing in two FOGDs, taking up 3.51% of format error tokens.

All in all, IF1 and LE are the two subcategorized linguistic tokens with the highest occurrence frequency under the format errors main category, appearing ten times each. UCCL has the second

highest occurrence frequency of nine. WWC has the lowest occurrence frequency, only appearing once in all nine FOGDs.

In terms of commonality, IF1 and IC are the most commonly found format errors across the FOGDs, both spotted in six out of nine FOGDs. WWC proves yet again to be the least common format error and is only spotted in one FOGD. It can be concluded that IF1 is the most useful detector of OGD out of the ten format error subcategories because of its high scores under both assessment techniques.

Figure 1. Examples of format errors. (a). Swapped instruction subtitle and subject line (IF1). (b). Extra space after bullet points (LE). (c). Improper brackets (IF2). (d). Colloquial language that translates to "citizens don't leave Taoyuan" (CLIA). (e.) Missing particle in legislation (LFE).
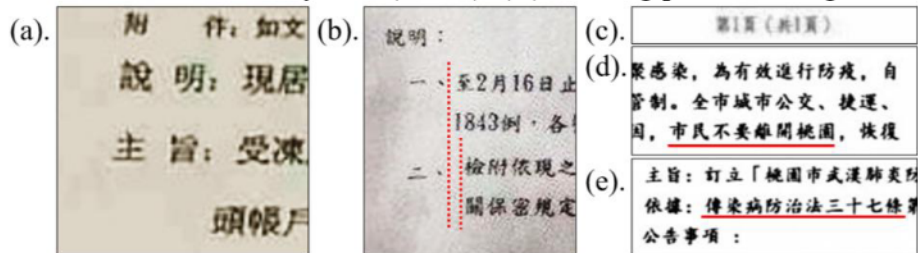
(II) Font/Form Discrepancies

Font/form discrepancy refers to the different appearance of the characters that can be used to indicate the creator of the FOGDs. 18 out of 94 tokens are font/form discrepancies, and that is 19% of all linguistic tokens. This study observes that four groups of tokens possess common traits and that they can be sorted into four smaller subcategories: simplified and variant Chinese characters (SVCC), punctuation mark font discrepancies (PMFD), character style discrepancies (CSD), and number font discrepancies (NFD). The four subcategories will be examined in the following studies.

To standardize the style of script of all fields of formal writing, Taiwan sets a standardized traditional Chinese font for Chinese characters that is used in all OGDs. All characters in formal writing must be standardized, therefore any other variant of of Chinese characters is normally unacceptable. However, some variants are still appearing in formal documents including FOGDs. Simplified and variant Chinese characters (SVCC) refer to the different forms of a character that are not considered formal in Taiwan. The informal Chinese characters found in FOGDs are mainly vulgar variants (so-called Sutizi) and alternatively styled variant (so-called Yitizi). Keep in mind that while vulgar variants are sometimes the simplified version of their traditional Chinese counterparts, they are different from Simplified Chinese characters because vulgar variants can both simplify and complicate a character. Simplified Chinese can be found almost effortlessly by Taiwanese people because of its characters' simpler writing and fewer strokes, while variant Chinese can be found with a little more effort. In Document C, for example, there is an alternatively styled variant character (see figure 2). Because many words in simplified Chinese are written exactly the same as the words in traditional Chinese, the present study determines the frequency of simplified Chinese's occurrence by counting the number of characters written in simplified Chinese. Four out of nine FOGDs with a total of six simplified Chinese characters are found with such instances. They take up 37.50% of all font/form discrepancy tokens.

Note that not all variant characters are technically erroneous and not all are exclusively from China; it depends on the specific characters because Taiwan actually uses and accepts some of them. One example of vulgar variant is the "Tai" in Taiwan; it can be written in two different forms. In many cases, these two forms of "Tai" are interchangeable and the vulgar variant "Tai" is accepted by the public: the names of the stops in Taiwan's Taipei Metro transit system use the vulgar variant "Tai"

instead of the standardized "Tai" (see Figure 3). The two typhoons that Taiwan named are proof that Taiwan uses alternatively styled characters on rare occasions: when Taiwanese climatologists were naming the two typhoons—Yanhua, meaning fireworks, and Xiehu, meaning gecko (see Figure 3)—no professional linguists were there to assist with naming, hence the accidental usage of alternatively styled characters.

Punctuation mark font discrepancies (PMFD) refer to incorrect punctuation marks in OGDs. One can tell that the content of an OGD has a different font by paying attention to the location of the common punctuation marks such as periods, commas, and colons. Normally they would appear in the middle of the black space between two characters, but they're often placed in the left-bottom corner in Chinese FOGDs (see Figure 2). Unlike the case of simplified Chinese, PMFD affect all punctuation marks within the same FOGDs, the author only counts each FOGD with punctuation mark font discrepancies once. Five out of nine FOGDs have punctuation mark font discrepancies. There are a total of five instances, making up 31.25% of font/form discrepancy tokens.

Character style discrepancies (CSDs) are harder to spot because they are the small details, such as the length or the direction of a certain stroke, within characters that indicate the OGD's unreliability. The reason why character style discrepancies occur is because of the Chinese FOGD fabricators' failed attempt to make the characters of the FOGD appear to be traditional characters. Since the fabricators are converting simplified Chinese characters that are already in the wrong font, called "new character forms," to traditional Chinese characters, the converted characters will be in the wrong font as well, causing the observed CSDs that affect all the characters in the same FOGD. This study therefore counts only once for each FOGD with CSDs. In Document E, for example, has CSDs for all the characters, but the most noticeable character infected with this type of discrepancy is the one whose second-to-last horizontal stroke is shorter than the last horizontal stroke when it should be the other way around. Two out of nine FOGDs have CSD. Character style discrepancies take up 12.50% of font/form discrepancy tokens.

Number font discrepancies (NFD) refer to the incorrect font of numbers. One can tell a number's font is different by paying close attention to a number's position within a character's space and the size of the number. In an OGD, numbers should be located in the middle of a character's space and have the same height as the Chinese characters. Similar to PMFD and CSD, the incorrect font of an FOGD affects all numbers in the same FOGD, therefore the author counts each FOGD with number font discrepancies only once. The numbers' sizes in Document A, however, are irregular and smaller than the Chinese characters'. Number font discrepancies take up 18.75% of font/form discrepancy tokens.

In terms of occurrence frequency, SVCC has the highest score of 6 but is followed closely by PMFD, whose occurrence frequency is 5. This phenomenon can be explained by acknowledging the fact that while it takes a certain level of sensitivity to words for one to spot CSD and NFD, the Taiwanese public does not need linguistic knowledge to identify PMFD and even more so in the case of SVCC. This study discovers that the tokens cannot be compared simply by counting the frequencies of each subcategory, however, since SVCC are counted by number of characters and the other three subcategories are counted by number of FOGDs affected by the discrepancies. Because of the mathematical data collection difficulties and different comparing standards for each discrepancy, this study cannot study and rank the discrepancies in terms of occurrence frequency.

Comparison in terms of commonality then becomes the only assessment technique available for font/form discrepancies. PMFD is the most commonly found font/form discrepancy token as it is present in five out of nine FOGDs while SVCC is present in four out of nine. The two remaining subcategories (CSD and NFD) have slightly lower scores (two and three, respectively), suggesting that they are slightly less suitable to be used to detect FOGDs in comparison to SC and PMFD.

Figure 2. Examples of font/form discrepancies. (a). (b). Alternatively styled character (SVCC) marked with red underline and (c). its standardized traditional character counterpart. (d).(e.) Wrong punctuation position (PMFD) as marked with red underline and (f). its correct counterpart.
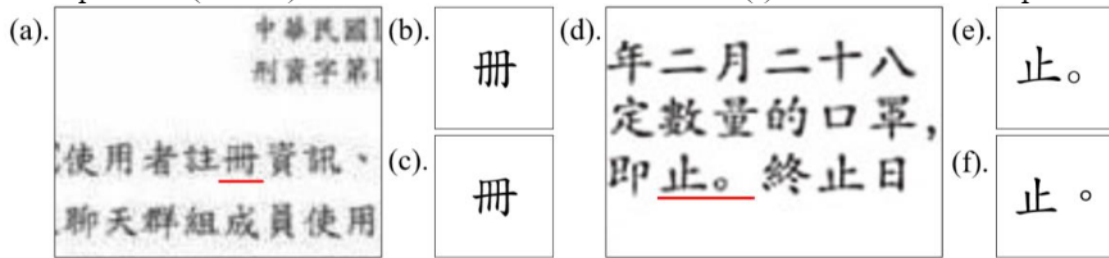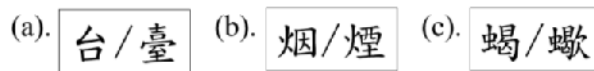

Figure 3. Real-life examples of variant characters and their traditional Chinese counterparts. (a). Vulgar variant "Tai" and standardized traditional "Tai." (b). Alternatively styled "Yan" and standardized traditional "Yan." (c). Alternatively styled "Xie" and standardized traditional "Xie."

(III) Cultural Differences

Cultural differences between China and Taiwan are so prominent that they are frequently observed in FOGDs. They refer to any mentions of objects and wording that are used in Mandarin in Mainland China but not Taiwan Mandarin. They are often only distinguishable when a person is familiar with both Taiwanese and Chinese culture. 19 out of 93 tokens show cultural differences, and that's 20% of all linguistic tokens. This study discovers common traits shared by certain tokens and deems it appropriate to sort them into two subcategories: wording cultural difference (WCD) and factual cultural difference (FCD).

Wording cultural differences are the vocabularies China uses differently from Taiwan to refer to the same object. This is similar to a British person calling the American elevator a "lift." The easiest to explain and the most noticeable instance of wording cultural differences is in Document I: in the document, Kaohsiung City in Taiwan is referred as "Taiwan Kaohsiung City, Republic of China" whereas Taiwanese people normally just call it "Kaohsiung City." Six out of nine FOGDs have wording cultural differences, making up 52.63% of all cultural difference tokens.

Factual cultural differences are the different customs and realities in China and Taiwan. These differences occur because of the FOGD forgers' unfamiliarity with Taiwanese culture: they can occur quite frequently since the FOGD forgers can be so accustomed to their own culture that they forget that it is not the same case in Taiwan or that they are unaware that the information on these FOGDs are simply wrong. In Document C the telephone number the document claims to be the police force's number simply does not exist, and in Document A the number of digits in a person's identity is erroneous. Five out of nine FOGDs show instances of factual cultural differences, making up 47.73% of all cultural difference tokens.

In short, WCD is spotted only a little more frequently than FCD; WCD scores a ten and FCD scores a nine on the occurrence frequency assessment technique. The results for the two subcategories are extremely close, so neither one can be assumed to be a better FOGD indicator than the other.

The result is similar in terms of commonality, where WCD is found in six out of nine (66.67%) FOGDs and FCD is found in five out of nine (55.56%) FOGDs. It is proven yet again that WCD and FCD seem to be equally useful tools to detect FOGDs.

(IV) General Discussion

Despite the fact that this study chooses different assessing techniques to assess the three categories of tokens, this study still attempts to compare them using the same technique one at a time. When this study ranks the three categories in terms of token occurrence frequency, the order from highest frequency to lowest is format errors, cultural differences, and font/form discrepancies. Of all 93 linguistic tokens, the number of format errors exceeds the other two categories', having a total of 57 tokens. This phenomenon does not indicate that format errors are the most potentially effective detector of FOGDs, however, since there are a lot more subcategories under the format error main category than the other two categories, the result is predictable and explainable. Font/form discrepancies have a total of 18 tokens and cultural differences have 19 (see Figure 4). It can be inferred that format errors are the most frequently spotted linguistic tokens among FOGDs.

In terms of commonality, the cultural difference category is ranked the highest; format errors and font/form discrepancies follow in that order. Nine out of nine (100%) FOGDs are recorded to have instances of cultural differences. In comparison, eight out of nine (88.89%) FOGDs have format errors and font/form discrepancies (see Figure 4). All three categories are extremely common across FOGDs and can act as potentially great FOGD indicators. The result under this assessing technique is different from the result from the occurrence frequency technique.

Figure 4. Comparison of three categories. (a). Token occurrence frequency by categories (b). Token commonality by categories.
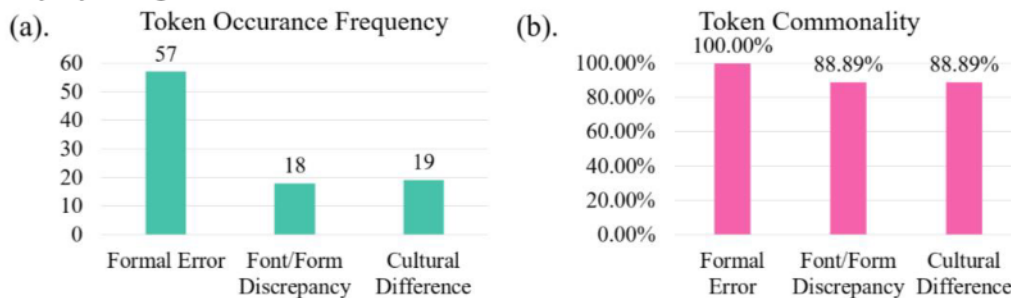
V. Conclusion and Suggestion

This study explores the linguistic tokens of FOGDs and attempts to find how these tokens have the potential to be useful FOGD detectors in the future. The author discovers three main categories of linguistic tokens—format errors, font/form discrepancies, and cultural differences—and a total of sixteen subcategories under them, with four under format errors, ten under font/format discrepancies, and two under cultural differences. The categories and subcategories within the same categories in terms of token occurrence frequency and commonality across FOGDs are further compared.

According to the result, the format errors category are both the most frequently spotted linguistic tokens and the most commonly found across all nine FOGDs the author examines. Furthermore, type one incorrect formatting and wording cultural differences are both the most frequently spotted and the most commonly found subcategories. From here one can deduct that out of all the approaches (categories and subcategories) this study presents, he or she has the highest chance of successfully identifying FOGDs when they look for type one incorrect formatting and wording cultural differences. That being said, it is important to understand how one can spot these linguistic tokens.

It is often presumed that cultural differences are relatively obvious to the Taiwanese people because they are already familiar with Taiwan and China's vocabulary differences. Most Taiwanese people are decently exposed to enough Chinese culture to tell if a certain word is in Mandarin from

Mainland China. They are, presumably, less familiar to format errors and font/form discrepancies since only a few pay close attention to the exact formatting of real OGDs. Since these presumptions are not yet proven by research, this situation sparks a new research question regarding the variating effectiveness these linguistic tokens are for the public to utilize and detect FOGDs that future researchers can further delve into.

      This study suggests the Taiwanese government strengthen educational programs to the public so that the Taiwanese people can familiarize with the categories in this study and have a better chance of successfully detecting OGDs. Results of this study show that despite scoring lower than format errors do, font/form discrepancies and cultural differences are also valuable and worthy of becoming part of the education programs. This conclusion leads to a new opportunity for future researchers to investigate which linguistic token categories are the most beneficial for governments to add into educational programs.

VI. References

Damjanović, D. Z. (2017). Types of Information Warfare and Examples of Malicious Programs of Information Warfare. *Vojnotehnički Glasnik (Military Technical Courier)*, *65*(4), 1044-1059.

Hewitt, O. (2009). Information Warfare -- Doing Battle in the 21st Century. *Dalhousie Journal of Interdisciplinary Management*, *5,* 1-13.

Libicki, M. (1995). *What is Information Warfare?* National Defense University, Institute for National Strategic Studies.

Mahyoob, M., Algaraady, J. & Alrahaili, M. (2020). Linguistic-Based Detection of Fake News in Social Media. *International Journal of English Linguistics*, *1*(1), 99-109.

Marquardt, D. (2019). Linguistic Indicator in the Identification of Fake News. *Mediatization Studies*, *3,* 95-114.

Schwartau, W. (1997). What Exactly is Information Warfare? —Part 1. *Network Security, 9,* 17-19.

Appendix

Doc.	Printed Date	Spreading Domain	Source	URL
A	11 Aug 2015	Physical flyer	Liberty Times Net	https://news.ltn.com.tw/news/society/breakingnews/1927234
B	27 Oct 2017	Internet	Banqiao Precinct, New Taipei City Police Dept.	https://www.banqiao.police.ntpc.gov.tw/cp-200-38776-11.html
C	25 Feb 2019	LINE	Yahoo News	https://reurl.cc/RjmoG6
D	17 Feb 2020	Weibo	LINE TODAY	https://today.line.me/tw/v2/article/xL18Qw
E	26 Feb 2020	Weibo	Mandarin Daily News	https://www.mdnkids.com/2020COVID-19/index/?Sn_=162
F	28 Feb 2020	Internet	Mandarin Daily News	https://www.mdnkids.com/2020COVID-19/index/?Sn_=162
G	29 Feb 2020	Facebook	Taoyuan City Gov.	https://reurl.cc/8W8VZj
H	2 Oct 2020	Twitter	CNA News	https://www.cna.com.tw/news/firstnews/202012110028.aspx
I	16 Apr 2021	Twitter, Facebook	Taiwan FactCheck Center	https://tfc-taiwan.org.tw/articles/5298