

Toward one-stop Information Mining: Tailoring Web Texts to Effective Machine Translation

Dr. Chung-ling Shih

Professor

National Kaohsiung First University of Science and Technology
Taiwan.

Abstract

This paper proposes the tailoring of web texts in controlled language (CL) to meet the audience's double expectations of "a-click-for-adequate-information" and "a-click-for-easy-understanding" through machine translation (MT) application, employing web texts on Taiwanese cuisine as a case study. Drawing on Grice's cooperative principles as the theoretical framework, the merits of tailored web texts are justified by conducting a study of contrasting two sets of ten MTs of web texts with and without meeting Grice's cooperative principles. The tailored web texts distinguish themselves from untailored web texts by presenting adequate information with thematic diversity, using headings as indexical references and clear, concise verbal presentations. Furthermore, a questionnaire-based survey shows that 85% of respondents favor MTs of tailored web texts and only 15% of them, MTs of untailored ones. The respondents dislike untailored texts for no clear, logical verbal presentation, lack of adequate headings as indexical references, many grammatical errors and ambiguous, incorrect words. In conclusion, to move toward "one-stop information mining" through MT application on the web, a suggestion is made that some existing web texts can be tailored in CL and be presented with special linguistic and pragmatic features as an effective and alternative way out.

1. Introduction

For many travelers, museum is a must-see place on their journey to foreign countries. The major reason is that the museum allows them to get all the necessary-to-know information within a short time at a place without consulting other data. Thus, the museum is often treated as a place to provide multi-thematic and multi-foci resources and meet the purpose of one-stop information access. Inspired by the museum, the researcher of this paper wonders if the web audience can read only a web text and gets all the necessary information they want without consulting additional data. What type of this web text would be? What special linguistic and stylistic characteristics does it embody? How can it achieve the intended purpose of effective communication of a good amount of information without language barrier? What is the web audience's reception of such a special web text? These questions motivate her to conduct this research and try to find the answers.

To answer the above question, many people think of online machine translation (MT) that has been anticipated to serve as an easy solution to effective transmission of information across borders and cultures on the web. However, the MT quality is often poor and unsatisfactory. Thus, to improve the readability and comprehension of MT, the verbal presentation of the source text is advised to use the controlled language (CL) that restricts the size and the scope of vocabulary and some syntactic and grammatical features. To date, a lot of research (Lee 1994; O'Brien 2003; O'Brien and Roturier 2007; J. Roturier 2004; Cardey, Greenfield & Wu 2004; Pym 1990) has been conducted to justify the effectiveness of using a controlled language (CL) to improve the semantic and grammatical accuracy of MT. However, the previous research focuses on the localization of technical documents for marketing and business purposes, not for instant communication of information on the web. Furthermore, the use of CL can only make web texts understood more easily and clearly. If we pursue the goal of "one stop information mining" through MT, we need to pay attention to other aspects, such as textual format and information amount.

In order for the web audience to get all the desirable information with only one click without the language barrier, the present research proposes that web texts are tailored not only in CL using the clear, concise and simple verbal presentations, but also provide a proper amount of information and heading-based indexical references.

The pragmatic conditions of proper information capacity, clear indexical reference and easy-to-understand information are often treated as the key factors that make a communication activity successful in daily life. The concerns with the strategies for attaining effective communication has been addressed in pragmatic studies and socio-linguistics (Bernstein 1971; Duff 2008; Sperber & Wilson 1986/1995; Grice, 1989/1975; Gumperz 1982a, 1982b, 1989a, 1989b, 1991; Eerdmans, Prevignano & Thibault 2002; Gutt, 1991, 1992, 2000). Among them, Grice's (1989/1975) cooperative principles fully support the special characteristics of a tailored web text this paper proposes, so they are used as the theoretical framework to justify the merits of tailored web texts. Either Grice's cooperative principles or the MT-oriented tailored web texts similarly pursue the goal of optimal communication that is grounded in a premise that without mutual cooperation on both parties, any communication cannot run successfully. Thus, an effective, cooperative interaction between the addressor/the text and the addressee/the machine is needed for the effective MT application.

Drawing on Grice's (1989/1975) manner, relevance and quantity maxims as the argumentative basis, the present research will probe how the tailored web texts can justify their potential benefits to help achieve the goal of effective and efficient MT-enabled information mining on the web. A comparative study of a set of English MTs of tailored and untailored web texts will be conducted by examining their distinct differences in the areas of verbal presentations, information capacity with thematic diversity, and textual format. The huge differences in these areas will make it easy to infer why tailored texts improve their communication effectiveness from the perspective of Grice's cooperative principles. However, to promote its empirical creditability, a questionnaire-based survey will be conducted and its results can help us understand the web audience's reception of the English MTs of tailored texts and untailored texts. In this sense, the present research aims at achieving two objectives: 1) to justify how the MTs of tailored web text allow for effective and efficient information access; 2) to investigate if the web audience favors the MT of tailored texts more than the MT of untailored texts. Based on the two concerns, three research questions are brought up to frame the investigation of the present research as follows.

RQ1: What are the differences between tailored and untailored web texts in their verbal presentations, thematic Diversity and textual format?

RQ2: Why can the tailored web texts in CL achieve the goal of MT-enabled "one-stop information access" on the Web from the perspective of Gricean cooperative principles?

RQ3: What is the web audience's reception of the MTs of tailored and untailored web texts?

To answer RQ1, we need to measure tailored web texts against untailored ones. RQ2 is concerned with the theoretical justification of communicative effectiveness of tailored texts in CL from the perspective of Gricean cooperative principles. The answer to RQ3 will help us know if the web audience favors the MT of tailored texts better than untailored texts because it helps the audience pursue the goal of one-click-for-adequate-information access with easy and clear comprehension.

II. Controlled Language and Gricean Cooperative Principles

Technology has affected the way we interact with others and has also changed the way we communicate with people globally. It pushes us to use a new type of web writing because the online MT system can automatically translate web texts. To optimize the effectiveness of MT application, the web text can be authored using a controlled language. CL is a kind of artificial language or a sub-language and is different from natural or ordinary language. It needs to shake off the distinctive linguistic and culture-specific attributes of the source text and follows some rules. An universal rule is the use of shorter sentence because shorter and simple sentences lead to more grammatically accurate renditions produced by the machine, be it functionally rule-based or statistics/examples-based. Furthermore, all the cultural references and metaphorical expressions that have ambiguous and multiple meanings should be avoided. To be sure, CL needs to follow a set of grammatical rules. Take controlled Chinese for example; it needs to use articles, quantifiers, possessive cases, SVO structure, verbal tense markers, conjunctions, and transitional words if it wants to create more correct MTs in English and other Indo-European languages (Shih 2013).

Another theory that needs to introduce is Grice's (1989/1975) cooperative principles because it is used to justify the advantages of the tailored web texts in CL for efficient and effective information access. The Gricean principles consist of maxims of quality, quantity, relation/relevance and manner, and they are initially used to support the effectiveness of a successful conversation.

From Grice's viewpoint, a smooth interaction between addressers and addressees depends on their cooperation by following a set of principles. The quantity maxim requires the speaker to use a right amount of information. Too

much or too little information is not appropriate. The relation/relevance maxim asks the speaker to relate his/her conversation or speech to the topic or/and the subject so that the hearer/audience can easily infer the meanings of given messages and comprehend the meanings. The manner maxim refers to the avoidance of unclear, ambiguous expressions. Presenting semantically vague information will increase the hearer's interpreting load unless the speaker has a special intention and purpose. The maxim of quality emphasizes the conveyance of true and genuine information, not providing a false source of information. If the speaker/addressor does not tell the truth, the hearer/addressee will not trust him/her. According to Grice, if the speaker violates one or two maxims above, their communication is likely to fail and the conversation cannot continue.

The present research is based on an assumption that if the MTs of web texts want to communicate their messages clearly, the web texts need to seek cooperation with the machine and the audience. Thus, web texts can be tailored following Grice's cooperative principles. However, the maxim of quality is not so important for MT-enabled communication, so it is not taken into account when Grician cooperative maxims are used to support the present proposal of tailoring web texts. The other three maxims are used to examine how a web text needs to be tailored in compliance with them. Table 1 shows the relevance of tailored web texts in CL to Grice's three maxims.

Table 1: The connection of tailored web texts in CL to Grician maxims

	Cooperative Principles	Tailored web texts
Quantity	To provide required information (neither too much nor too little; just right). To make the strongest statement.	To present information without redundant and specific details. To avoid simplifying the presentations too much.
Relation/ Relevance	To provide information related to the topic.	To present all the messages related to the six themes under three headings
Manner	To avoid obscure expressions. To avoid ambiguous words.	To eliminate unclear, multi-meaning words. To paraphrase all cultural references. To use clear, concise sentence structures.

(Note: Produced by the researcher of this paper)

Since tailored web texts in CL meet Grice's cooperative maxims, their MTs are assumed to be able to effectively communicate diverse thematic messages to the web audience. To seek empirical justification, a questionnaire-based survey is conducted and investigates the audience's reception.

III. Methodology

3.1 Collected Samples

For a comparative study, ten untailored web texts are collected from the website of Encyclopedia of Taiwan, and ten tailored web texts are retrieved from Shih's (2014) government-granted project (see Appendix 1). The two sets of ten web texts deal with the same topic of Taiwanese foods, including 1) stinky tofu; 2) zongzi (sticky rice dumpling); 3) steamed sandwich; 4) braised pork rice; 5) ba-wan; 6) oyster omelet; 7) Gangshan mutton hotpot; 8) oyster vermicelli; 9) bread bowl and 10) beef noodles. These Taiwanese foods are chosen because they are extremely popular in Taiwan, and have become the must-eat foods for international visitors to Taiwan. Noticeably, the ten tailored web texts are written in controlled Chinese by the researcher of this paper after she consults online texts. The controlled Chinese might not sound natural but the web texts in controlled Chinese do not target at the Chinese audience. They are only used for effective MT application and target at the global audience who cannot read Chinese.

3.2 Methods: Comparative Study and Questionnaire-based Survey

To identify some linguistic, thematic and format differences between tailored and untailored texts, a comparative study is conducted by contrasting ten sets of web texts with and without meeting Grician maxims of quantity, relation/relevance and manner. Their verbal presentations, thematic messages and heading-based formats are investigated. Since the tailored texts are written in controlled Chinese, they do not use many special expressions such as classical words, fixed Chinese phrases, Chengyus and relevant others.

Contrary to them, untailored texts often use these special words and cultural references because they are not intended for MT application. Their linguistic difference is obvious, so some special words are used as one of analytical criteria to distinguish tailored web texts from untailored ones.

On the other hand, to show their difference in thematic diversity, six themes are chosen to be the second analytical criteria, including history, legend, cooking steps, ingredients, events and comments. It is assumed that the more themes a text addresses, the higher thematic diversity it shows, and the more convenient for the audience's information access. Thus, we investigate if the tailored web texts deal with more themes than the untailored texts. Furthermore, the heading-based format will be investigated by checking the frequencies of headings in tailored and untailored web texts. The investigation is grounded on an assumption that if relevant information is gathered under a specific heading, the thematic coherence is created and more effective communication can be attained. Functionally, the heading helps frame the relevant information and focuses the audience's attention to a specific theme. Some headings are therefore used as the third analytical criteria and those under investigation include "origin", "cooking method", and "cultural value". In short, special words, cultural references, six themes and three headings are used as three sets of analytical criteria in the present comparative research.

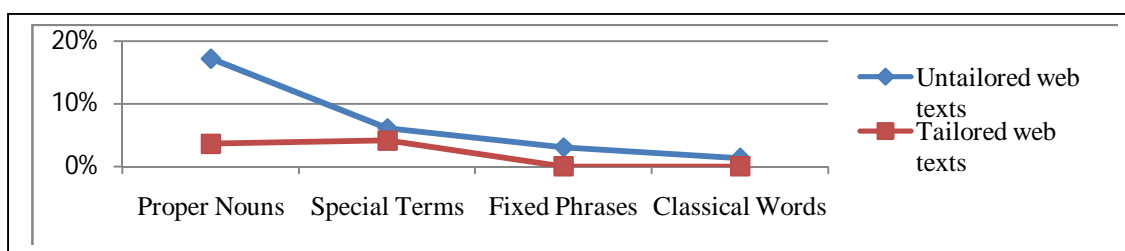
In the online questionnaire-based survey, twenty international university students are invited to fill out the questionnaire on the Internet. They do not contact the researcher in person and therefore their opinions are not affected by her. Since young people are frequent web users, their reception of the MTs of tailored web texts should be prioritized. For this reason, the respondents of this questionnaire are mainly young people. The content of the questionnaire consists of two parts—I) respondents' profile, II) automated English MTs for choice. In part II, participants are asked to read the English MTs of Version A (untailored text) and Version B (tailored text), and then make a choice. Some selection questions are answered so that we may know the reasons for their choice (see Appendix 2). It is noted that since English is recognized a lingua franca and can be understood by many international people, the MT tests use English as the target language. The statistical results are calculated manually after the questionnaire data are downloaded from the web.

IV. Findings and Discussions

4.1 Tailored vs. Untailored Web Texts

In response to RQ1 about the linguistic, contextual differences between tailored and untailored web texts, the findings indicate that there are some lexical, thematic and heading-relevant differences between them. In the area of lexical items, statistical results show that untailored web texts demonstrate the type/token ratio of proper nouns as 17.2 % (9.1/52.85 words on average), much higher than that of tailored web texts, only 3.57 % (5.2/14.53 words on average). The ratio of fixed Chinese phrases remains higher in the untailored web texts (3.02 %; 3.2/52.85 words on average) than in tailored texts (0%). Additionally, the ratio of special terms is slightly higher in untailored web texts (6.1 %; 3.2/52.85 words on average) than in tailored web texts (4.12 %; 0.6/14.53 words on average). One more interesting point is that untailored web texts are often peppered with classical words, showing the ratio of 1.32 % (0.7/52.85 words on average), but tailored web texts do not use any classical words. Figure 1 shows the difference in the type/token ratios of proper nouns, special terms, fixed phrases and classical words between tailored and untailored web texts.

Figure 1: The distribution of special linguistic items in the tailored and untailored web texts



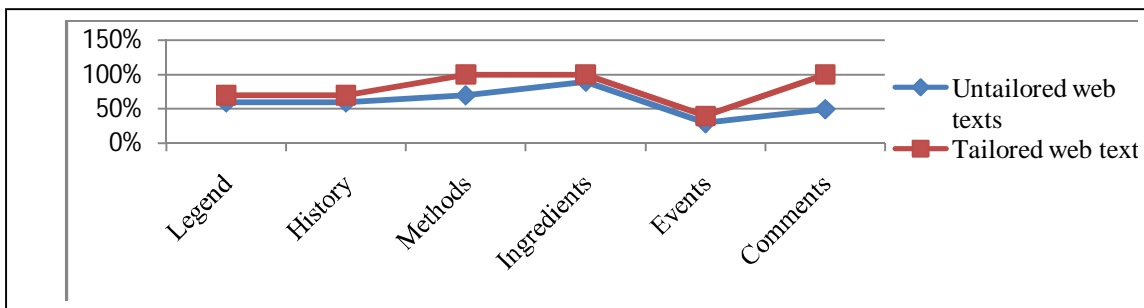
(Note: Produced by the researcher of the present paper)

Regarding thematic diversity, tailored web texts deal with more themes than non-tailored web texts do. Statistical results show that seven out of ten (70 %) tailored texts address the themes of legend and history; all of them (100 %), the themes of cooking methods and ingredients; four out of ten (40 %), the theme of cuisine-relevant events, and all of them (100 %), the theme of comments on the culture value of cuisine.

In contrast, only six out of ten (60 %) untailored texts deal with the theme of legend; six out of ten (60 %), the theme of history; seven out of ten (70 %), the theme of cooking methods; nine out of ten (90 %), the theme of ingredients; three out of ten (30 %), the theme of cuisine-relevant events; and five out of ten (50 %), the theme of

comments on the cultural value of cuisine. An overall comparison leads to a conclusion that 80% of tailored texts address six themes, but only 60 % of untailored texts attain this goal. Figure 2 displays the frequencies of six themes in the tailored and untailored web texts.

Figure 2: The frequencies of six themes in the tailored and untailored web texts



(Note: Produced by the researcher of the present paper)

It is also interesting to find that the themes of cooking method and comments on cultural value of cuisine present the largest gap (50%) between tailored and untailored web texts. Some of non-tailored web texts introduce neither cooking procedures nor a critique of cuisine value. For me, the web writers might think that other web texts on recipes would introduce the information, so they do not have to introduce it. On the other hand, I think the reason for the lack of comments is that some web writers might feel that web texts on cuisine aim to transmit factual information and so personal judgmental comments are not necessary. However, to expand the scope of information input, to attract the audience’s interest and to save the audience’s time for consulting other data, it is better to provide cooking procedures and giving some commentary views within one text.

In addition to thematic diversity, the difference in the heading-based format deserves our attention. The findings show that all the tailored web texts (100%) use three headings, including “origin”, “cooking methods” and “cultural value”. In contrast, only two out of ten (2/10=20%) untailored web texts present the headings. Take the untailored web text on “steamed sandwich” for example; the ingredients and the background of the cuisine are mixed together without a heading to frame them within one paragraph. Furthermore, in the untailored web text on “braised pork rice”, the cooking procedures, ingredients, relevant events and cultural value are all introduced under the same heading. Lack of headings or use of no enough heading easily makes the audiences lose the focus in the reading process. Actually, the more headings are used, the higher level of informative relevance is achieved. All the tailored texts present at least three headings to frame relevant information for thematic coherence and to help the audience easily interpret messages by referring back to the heading.

4.2 Tailoring Web Texts in Compliance with Grician Cooperative Maxims

In answer to RQ2 regarding effective communication of MT from Grician cooperative perspective, the inference we arrive at is that the MTs of tailored web texts in CL meet Grice’s maxims of quantity, relation and manner, and therefore they are effective for the target audience’s one-stop information mining on the web. For example, in discussing the cooking method, the tailored text on “Zhongzi” uses seven short sentences, concurring with Grice’s maxim of quantity—neither too much nor too little information. In contrast, the untailored text on the same topic uses only one sentence “using leaves to wrap rice and bound with colored silk”, not giving adequate information and henceforth not meeting Grice’s maxim of quantity. Additionally, the tailored text uses three headings as indexes to direct and guide the audience’s reading. The headings help integrate relevant information into a coherent whole, meeting Grice’s relation/relevance maxim. However, the untailored text mixes all the information without headings. More importantly, the tailored text in CL uses short sentences, common words, markers of functional, grammatical words, and fewer cultural references, meeting Grice’s maxim of manner. Table 2 shows how tailored and untailored texts meet or do not meet Grice’s maxims.

Table 2 Untailored vs. tailored web texts from the perspective of Grice’s maxims

Untailored Texts	Tailored Texts
------------------	----------------

Cooperative Principles	Maxim of quantity (a good amount of information)	Not to meet the maxim of quantity due to an introduction of inadequate or too much information about a theme.	To meet the maxim of quantity due to the input of adequate information about a theme.
	Maxim of relation/relevance (thematic coherence)	Not to accord with the relation/relevance maxim due to the lack of headings to frame relevant information.	To accord with the relation/relevance maxim by using three headings to frame relevant information for guided reading.
	Maxim of manner (clear, logical presentation)	Not to concur with the manner maxim due to the use of many proper nouns, fixed phrases and long sentences.	To concur with the manner maxim due to the use of simple, easy-to-understand words, and short, concise, clear sentences.

(Note: Produced by the researcher of this paper)

The tailored texts put all the thematic messages and relevant information under a thematic heading, so it creates thematic coherence and helps the audience easily understand the message. In addition, each tailored text presents three headings and addresses six themes, so it allows the audience to enjoy diverse information. A more significant point is that tailored texts use a clear and concise presentation and therefore greatly improve the grammatical and semantic accuracy of their MT. In contrast, untailored texts provide specific details and distract the audience’s attention. A worse thing is that they use many cultural references and proper nouns, and the existing SMT (eg. Google Translate) still cannot render them accurately.

4.3 The Audience’s Reception

In reply to RQ3 as to the audience’s reception, the findings show that seventeen out of twenty respondents (85%) like the English MT of Version B (tailored text) and their reasons involve adequate information to acquire, three headings to group relevant information, fewer grammatical errors, fewer incorrect words and fewer ridiculous translations. In contrast, only three out of twenty respondents (15%) like Version A (untailored text). Their reasons for disfavoring Version A include no clear, logical presentation (eight out of twenty; 40%), lack of headings as indexes (eight out of twenty; 40%), many grammatical errors (eight out of twenty; 40%), incorrect words (four out of twenty; 20%), ambiguous words (five out of twenty; 25%), inappropriate words (seven out of twenty, 35%).

In light of the statistical results, we may check some examples to understand why the overwhelming majority of respondents favor the MTs of tailored texts than untailored texts. Take their difference in the ratio of special linguistic items for example. The ratio of proper nouns in untailored texts is four times that of tailored texts and so its English MT tends to be unreadable and incomprehensible. Table 3 is an example of the English MT of “Zongzi”. Version A is the MT of the untailored text, and Version B, that of the tailored text. The MT is created by Google Translate.

Table 3: The English MTs of untailored and tailored texts

Version A	Version B
<p>As early as the Xia and Shang era began Dragon Boat Festival, dragon boat racing and eating zongzi but indeed Qu Yuan's death was derived from the folk customs. Qu Yuan was concern for the fate of being a traitor framed, not in power but has been repeatedly reused Demotion, King Xiang finally moved back, I wrote bitterly of "Huaisha" will Kaiseki hanging Miluo River, people who heard All the news we have dispatched rescue boats around, searching in vain fears Qu Yuan after the body was eaten fish, so people put the wrapped dumplings thrown into the river, they will not harm fish feed Qu Yuan's body. The earliest dumplings called "Bing tank" (There is another angle to say millet, just mention the Dragon Boat Festival thought angle millet. They eat millet angle when the Dragon Boat Festival every summer solstice when the Jin dynasty, the representative point turn year of life.), is the rice into the bamboo tube to allow soldiers are convenient to carry. For the expression of reverence and nostalgia Qu Yuan, a year on this day people will "soldier tank" into the river, then someone dreamed Qu Yuan said the people were all cast dragon eat food, such as using leaves to wrap rice colored silk and bound after, do not fear the dragon eats, so it evolved into rice dumplings with palm fronds. Many changes dumplings until today, not only varied fillings, and even ice dumplings, sweet dumplings, dumplings and other kinds of purple rice dumplings.</p>	<p>Origins A famous Chinese poet Qu Yuan committed suicide in the Miluo River. On hearing the news, people began their search for him. However, they did not find anything. So they prepared a lot of "zongzi" (sticky rice dumplings). They then throw the zongzi into the river. They were expecting that the fish will eat the zongzi, instead of Qu Yuan's body.</p> <p>Cooking methods 1) glutinous rice, mushrooms, chestnuts, peanuts and dried shrimp are soaked for one night. 2) mushrooms, pork and "fried shallot flakes" are mixed and fried. 3) Bamboo leaves are cleaned, and is rolled into a funnel shape. 4) "sticky rice" is wrapped into bamboo leaves. Meat, mushrooms are all added. 5) Zongzi are wrapped with cotton thread. 6) Zongzi is boiled with water, high heat, for half an hour. Medium heat for another one hour. 7) Turn off the stove and leave it there for a while.</p> <p>Cultural Implications In Taiwan, we have the northern zongzi, and southern zongzi. Northern zongzi is more solid because it is steamed. Southern zongzi is more viscous, as it is boiled in water. On the Dragon Boat Festival, Taiwanese eat zongzi, to commemorate Qu Yuan.</p>

(Note: Produced by the researcher of this paper)

As shown above, the untailored text (Version A) uses eight proper nouns, three fixed phrases, three special terms, and two classical words, but the tailored text uses only six proper nouns without fixed phrases and classical

words. Thus, the English MT of Version A is difficult for understanding. For example, the fixed phrase 憂國憂民(you-guo-you-min; lit: concern with his people and country) is mistranslated as “was concern for the fate of being”, and the special words such as 兵罐(bing-guan; lit: canned food for soldiers) and 轉捩點(zhuan-lie-dian; lit: the turning point) are mistranslated as “Bing tank/soldier tank” and “point turn”. A classical word 自縊(ziyi; lit: commit suicide) is mistranslated as “hanging”. In contrast, there are only some grammatical errors, such as inconsistent tense verbs and the lack of a preposition, in Version B. The two verbs “throw” and “will” should use the past tense verbs, and “high heat” should be corrected as “at high heat”. But generally viewed, the Version B has a plain and easy-to-understand verbal presentation in compliance with Grice’s maxim of manner, and therefore it is more easily understood and more favored by the respondents.

One more thing we need to notice is that Version A does not have sub-headings so the audience might be confused about which parts present primary information and which parts, secondary one. Unlike Version A, there are three headings in Version B, and so relevant messages can be clearly communicated. For example, the subheading “cooking method” helps the audience immediately know that cooking procedures of rice dumplings are going to be introduced. When seeing the subheading “origin”, the audience knows they are going to read some background information about the food. Three headings in the tailored text comply with Grice’s maxim of relation/relevance. Furthermore, three headings suggest three key themes, and the audience may choose what they want to read only by looking at the headings.

Another example may be drawn from the web text on “Gaangshan mutton firepot”. In the English MT of the untailed text, 岡山三寶 (Gangshan san bao; lit: three valuable things in Gangshan) is mistranslated as “Okayama Sambo” by Google Translate. In a different manner, this noun phrase is correctly translated as “three specialties in Gangshan” in the English MT of the tailored text. Also we see that in the English MT of the untailed text, the names of two places 小崗山 (Xiaogangshan; lit: small Gangshan) and 燕巢(Yanchao; lit: swallow nest), and some special terms, such as 熱補 (re-bu; lit: hot tonic food), 燥熱現象(zaore xianxiang; lit: hot and dry feelings), and 清血功能 (qing xue gongneng; lit: blood purification function), are mistranslated as “small Hill Gang”, “nests”, “hot fill up”, “eating hot phenomenon” and “clear blood function”. These mistranslations are fully incomprehensible to native English speakers. However, these MT errors are not found in the English MT of the tailored text. Overall, a comparison between the English MTs of tailored and untailed texts shows that the tailored texts demonstrate more favorable conditions for effective communication such as diverse themes, more headings, fewer grammatical errors, and fewer unclear and incorrect words, so their English MTs are more favored by the web audience.

4.4 Pragmatic Significance

This research has identified some linguistic, thematic and format differences between tailored and untailed texts. The international respondents’ reception of the English MTs of tailored and untailed web texts is also investigated and reported. The findings have shed light on some pragmatic significance as follows.

- Due to the CL requirements for clear, concise, simple verbal presentations, tailored web texts in CL are more suited to MT application than untailed web texts.
- The MTs of tailored texts are more favored because they are in compliance with Grice’s cooperative maxims, providing the comprehensible content, adequate information input and easy references based on the headings.
- Tailored web texts suggest a possible way of achieving the goal of one-stop information mining/scanning due to their verbal simplicity, right information capacity, thematic diversity and informative relevance.

V. Conclusion

In conclusion, the present research aims to deliver an important message that when a web text introduces a theme or a subject, how much information it should communicate and in what way it should present the information is more important than what information it gives. The present research justifies that MT communication can be highly improved due to the input of a good amount of information, thematic diversity, informative relevance to indexical headings and the use of simple, clear, concise verbal presentations in CL. This suggests that the design of web texts should meet Grice’s cooperative maxims.

Thus, a suggestion is made that to move toward the goal of “one-stop information mining” through MT application on the web, existing web texts may try to be tailored in CL and be presented with special linguistic and pragmatic features as described in the present research as an effective and alternative way out.

Acknowledgements

This work was supported by the Ministry of Science and Technology in Taiwan under Grant [103-2410-H-327-006]. The author shows her much appreciation here.

References

- Bernstein, B. (1971). *Class, codes and control*. Vol. I. London: Routledge & Kegan Paul.
- Cardey, S., Greenfield, P. & Wu, X. (2004). Designing a controlled language for the machine translation of medical protocols: The case of English to Chinese. In *Proceedings of AMTA 2004* (pp.37-47). LNAI, Springer-Verlag.
- Duff, P. (2008). *Case study research in applied linguistics*. New York: Taylor & Francis Group.
- Eerdmans, S., Prevignano, C., & Thibault, P. (2002). *Language and interaction. Discussions with J. J. Gumperz*. Amsterdam: Benjamins
- Lee, A. (1994). Controlled English with and without machine translation. In *Aslib proceedings* (pp. 131-133). London: MCB UP Ltd.
- O'Brien, S. (2003). Controlling controlled English. In *Controlled language translation, EAMT-CLAW-03* (pp. 105-114). U.K.: Dublin City University.
- O'Brien, S. & Roturier, J. (2007). How portable are controlled language rules? A comparison of two empirical MT studies. In B. Maegaard (Ed.). *Proceedings of machine translation summit XI* (pp. 345-352). Copenhagen: Centre for Language Technology.
- Pym, P. J. (1990). Pre-editing and the use of simplified writing for MT: An engineer's experience of operating an MT system. In P. Mayorcas (Ed.). *Translating and the computer 10* (pp. 80-96). London: Aslib.
- Roturier, J. (2004). Assessing a set of controlled language rules: Can they improve the performance of commercial machine translation systems? In *Proceedings of the international conference translating and the computer 26* (pp. 18-19). London: Aslib.
- Grice, P. (1989/1975). *Logic and conversation*. In G. Paul (Ed.). *Studies in the way of words* (pp. 22-40). Cambridge MA: Harvard University Press.
- Gumperz, J. (1982a). *Discourse strategies*. Cambridge: Cambridge University Press.
- Gumperz, J. (Ed.). (1982b). *Language and social identity*. Cambridge: Cambridge University Press.
- Gumperz, J. (1991). Contextualization and understanding. In Duranti, A. and Goodwin, Ch. (Eds.). *Rethinking context*. Cambridge: Cambridge University Press.
- Gumperz, J. (1989a). *Contextualization cues and metapragmatics: the retrieval of cultural knowledge*. MS: University of Berkeley.
- Gumperz, J. (1989b). *Linguistic variability in interactional perspective*. MS: University of Berkeley.
- Gutt, E.-A. (2000). Issues of translation research in the inferential paradigm of communication. In M. Olohan (Ed.). *Intercultural faultlines—Research models in translation studies 1: Textual and cognitive aspect* (pp. 161-79). Manchester: St. Jerome.
- Gutt, E.-A. (1992). On the nature and treatment of implicit information in literary translation: A relevance-theoretic perspective. Retrieved Dec. 22, 2014 from <http://cogprints.org/2494/2/IMPINF8.htm>
- Gutt, E.-A. (1991). *Translation and relevance: Cognition and context*. Oxford: Basil Blackwell.
- Shih, C.-L. (2013). *New web cultural writing: Fast communication across borders*. Taichung: Bai-xiang Cultural Co. Ltd.
- Shih, C.-L. (2014). A new de-territorizing webscape: The MT-tailored multi-dimension/foci controlled web writing on Taiwanese cuisine. A MOST-granted project. Available at <http://www2.nkfust.edu.tw/~clshih/plan/plan9/plan9.php>
- Sperber, D. & Wilson, D. (1986/1995). *Relevance: Communication and cognition*. Oxford: Blackwell.

Appendix 1**Collected texts on Taiwan cuisine from the website of *Encyclopedia of Taiwan***

	Authors/Time	Topics	URL
1	Wu, M.-Q.(2009)	Stinky tofu	http://nrch.culture.tw/twpedia.php?id=11834
2	Gu, W.-T.(2009)	Zongzi	http://nrch.culture.tw/twpedia.php?id=11700
3	Yan, R.-X.(2009)	Steamed sandwich	http://nrch.culture.tw/twpedia.php?id=11780
4	Liu, Z.-R.(2010)	Braised pork rice	http://nrch.culture.tw/twpedia.php?id=23427
5	Zhou, B.-Z.(2009)	Ba-wan	http://nrch.culture.tw/twpedia.php?id=11786
6	Lin, W.-H. (2009)	Oyster omelet	http://nrch.culture.tw/twpedia.php?id=16007
7	Che, Y.-F.(2009)	Gangshan mutton hotpot	http://nrch.culture.tw/twpedia.php?id=11813
8	Huang, Z.-Q. (2009)	Oyster vermicelli	http://nrch.culture.tw/twpedia.php?id=16008
9	Chen, J.-S.(2009)	Bread bowl	http://nrch.culture.tw/twpedia.php?id=20551
10	Lin, X.-Y. (2009)	Beef noodles	http://nrch.culture.tw/twpedia.php?id=100159

Appendix 2**A questionnaire on tailored and untailored texts on Taiwanese cuisine****Part I. Respondents' profile**

<p>1. Name _____</p> <p>2. Gender: __ male; ___ female</p> <p>3 Education: __ Ph.D. ___ MA; ___ BA; ___ High School</p>

Part II. Reception of two sets of English MTs

You are asked to make a choice after you read Versions A and B. The two versions are English translations that are automatically created by online Google Translate without human editing. The source text of Version A is derived from Encyclopedia of Taiwan; that of Version B, from Shih's research project.

“Oyster omelet”

Version A

Omelets common snack in Taiwan, this single product price is about tens of dollars. In addition to the snack food commonly known as oyster after the shelling oysters, but also contains water Fan Shufen, cornstarch, eggs, vegetables and other little. The snack area improved from Tainan Anping "fried food chase", the legendary 17th century, the Dutch attack, according to the Southern Ming Zheng war when Taiwan, soldiers under the command of several kinds of mixed grain with these ingredients act as war, after becoming popular dishes.

Although families can boiling omelets, but this is still the most common stalls. Depending on the amount of customer sharing home usually prepare two feet to three feet flat wok station, placed under the gas fire. Boiling food, flat pan put a little oil and place evenly. After a few oysters placed from several to ten, a little fried under. Less than a minute, pour evenly mixing ratio of water, Fan Shufen, cornstarch mixed batter. Etc. batter slowly dries, add eggs, then you can turn it over after one minute, add vegetables, pot. Finally, a vendor will be topped with various soy sauce, tomato juice and sauce mixture to the exclusive Sami. While the sauce tastes very important, but still depends on whether oyster omelets delicious new is not fresh, not juicy fat. In today's trend of words, and finally welcomed oyster bags from Chiayi, Changhua East Stone and King Gong.

In addition to common Taiwanese oyster omelet is one of the well-known popular snacks, tourists also come to Taiwan's representatives will eat snacks. Shilin Night Market in holiday light focused vendors, for example, in a single day can omelets million copies sold, the total turnover of hundreds of thousands more.

Version B**Origin**

During the Dutch occupation, Zheng Chenggong led an army to attack the Dutch, in Taiwan. Dutch cut off their food supply. Koxinga's army came up with an idea. They mixed oysters, sweet potato flour and water and fried omelet. This is the oyster omelet, a very popular food in Taiwan.

Cooking methods

- 1) Sweet potato flour and chopped chives are added to the water. We stir them evenly.
- 2) Heat the wok. Oysters are added to fry.
- 3) Egg and cabbage are added. In addition, sweet potato flour and water are mixed. They are joined.
- 4) Fry until both sides are golden brown.
- 5) Chili sauce and soy sauce are added.

Cultural value

For Taiwanese, oyster omelet is one of the most popular foods. For international tourists, they want to eat the oyster omelet, when they visit Taiwan. This means that local food ingredients can be turned into a cheap, delicious food.

Multiple Choices

1. I like the content and the format of ____ Version A or ____ Version B because
____(a) the content provides adequate information about a theme.
____(b) the content is neither too detailed nor too simplified.
____(c) the format with some headings helps frame information for thematic reference.
____(d) the format with some headings makes me easily identify different thematic messages.
2. I like the verbal presentations of ____ Version A or ____ Version B more because of
____(a) easy understanding.
____(b) only few grammatical errors.
____(c) simple sentence structures.
____(d) clear meanings.
5. I dislike ____ Version A or ____ Version B because it has
____(a) many grammatical errors.
____(b) some incorrect words.
____(c) a lot of vague meanings.
____(d) some inappropriate words.