Journal of Language & Translation 11-2
September 2010, 227-254

Shift in Controlled English Norms for Different Purposes and for Different Machine Translation Systems*

Chung-ling Shih
National Kaohsiung First University of Science and Technology

Abstract

This research identifies different controlled English (CE) norms to be followed in technical writing for a variety of purposes and for different machine translation (MT) systems. The results of the investigation show that CE norms for MT application are stricter than those for communicative reading. The primary inference here is that human beings can interpret the meanings of polysemous words, pronouns, prepositional phrases based on the context and easily detect the misspellings, but MT systems fail to do so. In addition, a comparison of CE norms for the application of two MT systems indicates that the corpus-based Google MT is less constrained than rule-based TransWhiz in the lexical area. This phenomenon is attributable to the selection of a highly probabilistic module as the semantic scoring preference for the suggested translation provided by Google MT, not word-for-word translation by TransWhiz. In contrast, Google MT is more constrained than TransWhiz in the syntactic area. The inference is that TransWhiz parses syntactic constructions and transfers the parsing result based on grammatical

* The author would like to thank the Taiwan National Science Council for a grant to complete this research.

rules stored in the MT system, so it may modify the original word sequence to make the translation conform to linguistic patterns in the target language. Contrary to this, Google MT depends on fuzzy or exact matches statistically retrieved from the labeled corpus. If no matches can be found, syntactically inappropriate translations will be produced. Seen in this regard, CE norms are never fixed and have to be modified through the evolution of time and MT technology.

Keywords: CE norms, man/reading, machine/translation, diachronic, synchronic, dynamic nature

1. Introduction

Nowadays, to meet the needs of real time communication and multilingual translation, machine translation (MT) starts to gain favor and attention again. Many professional translators who have turned to the help of translation memory tools return to use MT systems with a rekindled interest. Free use of online MT systems with no need for developing corpus is the core reason for its promised comeback. Nevertheless, to improve the quality of the MT output and to make it serve our greatest benefits, the issue of MT with controlled editing is worth an investigation.

This article identifies differences in controlled English norms in technical writing across time for different purposes and for different machine translation (MT) systems.[1] Either for effective communi-

[1] My teaching experience is the driving force for this research. Once in my class, 80% of students mistranslated the sentence Connect the telecommunications equipment into an outlet on a circuit different from that to which the receiving antenna is connected from English into Chinese. They told me that they had much difficulty analyzing the structure of the sentence. Thus, I edited the sentence as The outlets of the telecommunications equipment and the receiving antenna must be on the different circuits, and asked them to translate it again. This time all translations were correct. It occured to me that if technical texts are written in

cation or for appropriate MT application, controlled English (CE) plays a significant role because the simplified text in CE is conducive to the clarity of expression and helps non-native English audience or the MT system to catch the message easily and accurately. In addition, the transfer of the CE text into other target languages by the MT system creates the better quality translation than the natural English text does, and the better quality machine-produced translations require less post-MT editing to save time, costs and human labor. Due to the benefits of cost-effective MT application and easy communication in English reading, controlled English is an important issue worth our attention and investigation.

Few technical texts available on the market are written in plain English, and the syntax is sometimes too complicated to clearly convey the message. As Wojcik and Hoard (1996) have put, some technical writers use special vocabularies (jargon), personal styles and complicated grammatical construction to make the texts difficult to be understood by both non-technical audiences and technical experts. Mis-decoding of the text leads to wrong translation. If human translators cannot clearly understand the technical text, the MT system would find it more challenging. To reduce the difficulty for human comprehension and MT application, controlled language (CL) that restricts the vocabulary, grammar and style in technical writing is a feasible solution. Technical writers may use concise structures, and less-specialized, consistent vocabulary to improve the readability and machine translatability.

The use of CL in technical writing is not a new concept. In the regular column for the *International Journal for Language and Documentation*, Allen (2000) declared that the idea of CL was raised ten years ago, but it is given a new image in the technological era. The earlier CL in the industrial context follows some writing

---

simple English, the audience will catch the message without much difficulty and without misinterpretation.

rules including the use of restricted vocabularies and simple sentence structures in the 1970s and 1980s. Technical writers were asked to author texts in CL and famous examples included Caterpillar Fundamental English, Xerox's Multinational Customized English, AECMA Simplified English, and GIFAS Rationalised French (the mid-1980s) (Allen 2000). CL was then used to improve the readability and clear comprehension of documents, and it was irrelevant to MT implementation. Nowadays, due to business globalization and multilingual text development, technical writing in CL is not simply for improved readability, but for cost-effective MT application.

The new concern gives a new direction for CL study. Gadaniec (1994), Bernth (1999), Bernth and Gdaniec (2001), and Underwood and Jongejan (2001) investigated how CL norms meet the requirements for machine translatability. By eliciting 28 negative translatability indicators from the study performed by Bernth and Gdaniec (2001), Sharon O'Brien and Johann Roturier (2007**)** conducted two empirical MT studies and then identified controlled English (CE) rules that had high impact, low impact and no impact.[2] Like O'Brien and Roturier (2007), I investigate CE norms for MT application. However, they analyzed English-to-German translations, and I investigate English-to-Chinese translations. Furthermore, I identify the differences in CE norms for different purposes such as man/reading and machine/translation, and for different MT systems such as rule-based TransWhiz and corpus-based Google MT. This research basically combines a diachronic study on variation in CE

[2] The findings of Sharon O'Brien and Johann Roturier's study (2007**)** indicate that high impact CE rules include misspelling, incorrect use of the full-stop, colon, semi-colon, double hyphen and the comma, and inappropriate use of long sentences and personal pronouns whose antecedents are not present**.** Low impact rules involve standalone personal pronoun, the use of parentheses, and the slash as separator, and no impact rules cover the use of noun clusters of three nouns or more, relative pronouns, passive voice, and incorrect use of subject-verb agreement and plural form of nouns.

norms in different times, and a synchronic study on variation in CE norms for two MT systems in the same era. The diachronic-synchronic research is expected to show the dynamic nature of CE norms through the close relevance of CE norms to the elements of time and technology. Figure 1 shows the structure of this research.

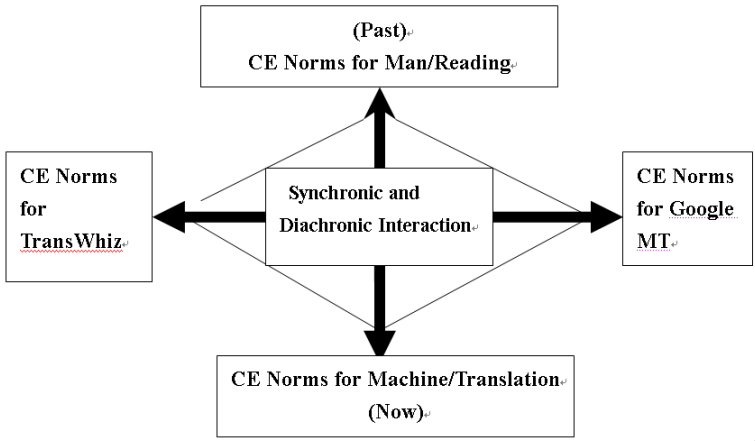

Fig. 1. The Diachronic-Synchronic Research.

This research emphasizes a shift in CE norms across time for different purposes and for the application of different MT systems. I believe that no universal sets of CE norms fit all types of MT systems for the appropriate automatic translation purpose. CE norms need to be customized to different parsing functions of MT systems. Keeping this assumption in mind, I raise some questions for investigation as follows:

(1) How do CE norms vary for effective reading and for appropriate MT application?
(2) How do CE norms shift for rule-based TransWhiz and

for corpus-based Google MT?

(3) How is the modification of CE norms relevant to the evolution of time and technology?

Theses questions investigate high and low restrictions of CE norms on technical writing for effective communication and for appropriate MT application in different times, and even for different MT systems that have different technological strengths. The mobile nature of CE norms will be confirmed with supportive statistical evidence.

2. Literature Review

This study deals with shifts in CE norms for different purposes and for two MT systems, so I would discuss the basic concepts of CE norms and MT systems at some length.

2.1. Controlled English Norms

Controlled English (CE) is a subset of controlled language (CL) that is restricted by specially selected vocabulary and simple syntax in a specific domain (Lehtola, Bounsaythip & Tenni 1998). As declared by Arnold et al., the controlled language is "a specially simplified version of a language" and it is conceived as "partial solution[s] to…perceived communication" (Arnold 1994: 211). A controlled language is a "variant of SL in which texts are composed according to a set of rules designed to enhance the clarity and readability of what is said" (Shuttleworth & Cowie 1997: 29). Both CL and CE impose special constraints on the usage of grammar, vocabulary and style in the text. In many cases, CE is presented as simplified English (SE).

The history of CE traces back to 1930 when basic English is

created by Charles Kay Ogden for the creation of a variant of English that can be easily learned and that allows English legal documents to be easily understood by non-native English speakers (Ceusters, et al. 1998). Later, the breakthrough of CE leads to the birth of simplified English in aircraft documentation and the development of other controlled English variants in various industries. CE of AECMA (European Association of Areospace Industries) is used as a worldwide standard for technical documents in the aerospace industry, following some basic writing rules such as "write one topic per sentence", "do not use gerunds", "avoid complex verbs", and "do not write a sentence more than 20 words" (Tedopres International BV 1974-2007). The US Small Business Administration (SBA)[3] also proposes the use of plain language to "write and deliver a clear message of what the government is doing, what it offers and what is asks of applications" (See the website of sbagov; qtd. in Shih 2002: 132). The SBA's Plain English Language program offers tips for writing plain English and explains specific features of controlled language, including "the use of active voice instead of passive voice, the use of clear words instead of less common phrases, the use of the same terms consistently, and avoidance of acronyms and confusing terms" (qtd. in Shih 2002: 132). Furthermore, "the Plain English Campaign of the United Kingdom promotes an efficient and friendly style of official writing in Plain English"[4] (see the website of PlainEnglish campaign; qtd. in

[3] "SBA is a leader in the plain language movement in the USA. SBA wrote the forms and instructions for its guaranteed loan program and won the coveted "Hammer Award" from the White House. SBA wants small business owners to understand better the paper work their government sends them. SBA is committed to easing the burden of reading the government's documents, and believes that writing in plain language should be as clear and understandable as the Liberty Bell" (see the website of sba.gov.; qtd. in Shih 2002: 155).

[4] "Plain English Campaign is an independent organization fighting for crystal-clear language and against jargon, gobbledygook and other confusing language. It is based in New Mills, Derbyshire in England. They define plain English a

Shih 2002: 133). This campaign proposes the replacement of bureaucratic words and jargon with plain English alternatives, and the reduced use of active voice and nominalizations. These principles for either simplified English or plain English are regularly consulted in technical writing and are defined as CE norms in this research.

Norms in sociology and social psychology are "general values or ideas shared by a community" and are later transformed into "performance instructions appropriate for and applicable to particular situation" (Toury 1995: 55). Norms specify "what is prescribed and forbidden as well as what is tolerated and permitted in a certain behavioral dimension" (55). According to Toury (1995), there are two sets of norms applicable to translation: preliminary vs. operational. Preliminary norms relate to "translation policy", either determined by the government, the publication company or the translation industry within a culture, and operational norms, including matricial norms and textual-linguistic norms that govern the degree of fullness of translation (which sections or segmentation are deleted) and the selection of particular text types and mode of translation (59). In this research, when set within Toury's (1995) theoretical frame, CE norms are perceived as the language policy of the government and the industry in the primary area, and serve as textual-linguistic standards in the authoring of technical documentation in the operative area. However, CE norms shift as the linguistic policy and textual linguistic standards change. One example is that Caterpillar Incorporation uses Caterpillar Fundamental English (CFE) in the mid-60s to the 70s for the

---

something that the intended audience can read, understand and act upon the first time they read it. Plain English is needed in all kinds of public information, such as forms, leaflets, agreements and contracts. The gold rule advocated by the campaign is that plain English should be used in any information that ordinary people rely on when they make decisions" (see the website of plainenglish.com.uk; qtd. in Shih 2002: 155).

improved readability of product documents, but now CFE is replaced with Caterpillar Technical English (CTE) to produce satisfactory automatic translation by the KANT system (Kaji 1999). In brief, CE norms vary in practice to meet the changed language policies raised by the government and the industry for different purposes in different times.

2.2. Machine Translation (MT) Systems

Machine translation generally means the automatic transfer of the source language (SL) into the target language (TL) by machine/computer. The development of MT has gone through more than 50 years and has not achieved the goal of high-quality automatic translation. However, the upgrade of computing power has rekindled people's interest in MT application in recent years, and an increasing population worldwide has handled massive information posted on the Internet with the aid of MT on the daily basis. MT is viewed as a tentative solution to real-time translation of online information in foreign languages. Many MT products have been commercialized and released on the market since the mid 80s although they can not produce 100% perfect translation and post-MT editing is required for improved readability. Nowadays, some MT products are available on the Internet for free. Two well-known and satisfying MT systems in English-to-Chinese translation are TransWhiz and Google MT. Thus, this research tests CE norms on these two MT systems and investigates whether CE norms are applicable to all MT systems without any difference.

TransWhiz was developed and released by Taiwan's Otek International Company in the early 20s and has provided the free online version for the public use in recent years. The online TransWhiz uses the rule/grammar-based approach, allows the user to choose few specialized dictionaries, and has increased the size of grammar rules, so that it produces the better syntactic parsing result

than the old TransWhiz does. However, the lack of knowledge-based semantic analysis still causes some difficulties in eliminating needless parsing results, and therefore ridiculous translations are often produced. The choice of specialized dictionary may restrict the scope of semantic analysis and improves the quality of the translation. But, the limited vocabulary size of current TransWhiz still cannot solve the translation problem. Thus, we expect to reduce syntactic complexity and semantic ambiguity through the use of the source text written in CE.

Google MT uses the corpus/statistics-based approach, and is more effective in the semantic analysis than the syntactic parsing. The preference score for semantic analysis and syntactic parsing is "generally calculated in terms of statistics with a strict mathematical founding" (Tanaka 1999: 5). This system requires "large-sized labeled corpora" to "train the probabilistic models" (5). Fuzzy and exact matches retrieved from the corpus are used as suggested translations. However, the corpora size of the existing Google MT is limited, so it often produces incomplete and awkward translations. The tentative remedy remains the editing of the source text following CE norms for appropriate MT application.

3. Methodology

3.1. Texts for CE Norms Research

In this research, I develop an English-Chinese parallel corpus that consists of two sub-corpuses: one containing technical English texts written in CE and their translations by TransWhiz; the other containing technical English texts written in CE and their translations by Google MT. Table 1 shows the internal structure of this parallel corpus. ST represents the source text, and TT, the target text.

Table 1. The Internal Structure of the Technical Parallel Corpus

Operational Manuals	STs in CE for TransWhiz	TTs (I) by TransWhiz	STs in CE for Google MT	TTs (II) by Google MT
Nero user's manual (2001)	English (2,128wds.)	Chinese (3,679wds.)	English (2,301wds.)	Chinese (3,415wds.)
Honda owner's manual (2003)	English (2,479wds.)	Chinese (4,682wds.)	English (2,117wds.)	Chinese (4,268 wds.)
Sony Ericsson user's manual (2001)	English (1,251wds.)	Chinese (2,338wds.)	English (1,209wds.)	Chinese (2,036wds.)
iPod user's manual (2002)	English (2,415wds.)	Chinese (4,300wds.)	English (2,285wds.)	Chinese (4,075wds.)
Instructions of Daikin air-conditioner (2008)	English (507wds.)	Chinese (770wds.)	English (481wds.)	Chinese (710wds.)
	Total Wds: 24,099		Total Wds: 22,897	

The time span of technical texts ranges from 2001 to 2008 and the disciplines include information technology (e.g. Nero Backup device, Sony Ericsson mobile phone), automobile industry (Honda) and electric appliances (e.g. air-conditioner). The technical texts are released by the companies that produce the products, so the writers/authors are assumed to be different although they are anonymous. The texts to be tested on MT systems are extracted from the corpus. These texts cover a variety of linguistic patterns, discursive modes and stylistic representations due to different subjects and different writers, so the results of MT tests would be more complete and more comprehensive than a few texts of the same subject written by the same writer/author. A great diversity in specialized discourses and stylistic representations in the collected texts has enhanced the validity of the research result.

3.2. The Method

The method of descriptive translation study (DTS)[5] is used to identify different CE norms that have to be followed in the editing of technical texts for different purposes and for the application of two MT systems. Unlike the prescriptive translation study (PTS) that moves from theory to practice, this DTS moves from practice to theory. The PTS first raises a set of norms and then uses some examples to support them. In contrast, the DTS analyzes a huge volume of data and then identifies some norms through observations and inference. As Toury (1995) has put, the DTS uses "empirical criteria" to explore what is existing as the translation phenomenon, whereas the theoretical translation study (i.e. PTS) uses the "theoretical, conditional criteria" to predict possible translation phenomenon (19).

The collected operational manuals are written in natural English, so I spend much time editing them into CE. After I test the CE texts on two MT systems, I check if the MT outputs have reached the perfect and good levels based on O'Brien and Roturier's (2007) MT assessment criteria (discussed later). Nowadays, some companies have used CE checkers[6] to edit the source text for MT application,

[5] Laviosa (2002) pinpoints that the methodological procedures of DTS usually involve three stages. The first stage is to identify "the object of study", and then moves to the selection of SL and TL languages (13). The final stage is to "generalize some norms governing equivalence for the selected pairs of texts" (14). Thus, after the goal is set up, the researcher first collects and edits technical texts from natural English into controlled English, and then submits them to TransWhiz and Google MT for translation into Chinese. Finally, the frequently used CE norms are detected from the satisfactory MT outputs.

[6] A controlled-language checker is a program that detects the violation of restrictions and gives alarm messages for MT outputs. Famous controlled language checkers include "The Simplified English Grammar and Style Checker/Corrector" (SECC), developed by the European Commission on the basis of the METAL machine translation system, "LANT MASTER", developed by LANT Ltd., "MAXit" by Smart Communications, Inc., "ClearCheck" by Logica Carnegie Group, Inc., and

such as Boeing Simplified English checker (BSEC), the Simplified English Grammar and Style Checker/Corrector (SECC) and others. These checking tools are not released on Taiwan's market, and they are used for the translation from English into other Indo-European languages, not for English-to-Chinese translation. Thus, the editing of all the source texts from natural English into controlled English in this research is manually completed.

3.3. The Assessment Criteria for Readability and for Machine Translatability

The criteria for readability are set up by following three Cs: clarity, conciseness and consistency. All CE norms that make the simplified text meet this triple-C standard are identified as effective for man's reading. In addition, to check if the CE norms that are applied to edit source texts lead to cost-effective MT application, I set the three-level criteria for the quality assessment of MT outputs. Table 2 shows the assessment criteria through adaptation of the model raised by O'Brien and Roturier (2007).

Table 2. The Assessment Criteria for MT Outputs

Levels	Specifications	SL Sentences and MT Outputs
Perfect	The MT output is perfect and does not need to be edited. The end-user does not have to cross-refer to the SL text and could understand the MT output easily.	SL: As the fan spins fast, it will cause injury. MT: 因為風扇快速地旋轉，它將引起傷害。[Because the wind fan rapidly spins, it will cause harm].

"Boeing Simplified English Checker" (BSEC) by Boeing (Kaji 1999). A number of companies have also been developing CE authoring tools for in-house use.

Levels	Specifications	SL Sentences and MT Outputs
Good	The MT output is acceptable and readable though it has some minor grammatical or lexical mistakes. The end-user can still understand the MT output without consulting the SL text.	SL: As the fan of the air conditioner is rotating at a high speed, it will cause injury. MT: 由於風扇的空調旋轉高速，它將造成傷害。[Because the air conditioner of the wind fan spins rapidly, it will cause harm]
Poor	The MT output is unreadable and incomprehensible. It contains serious errors. The end-user is unable to catch any message from the MT output without reading the SL text.	SL: As the fan of the air inlet is rotating rapidly, it will bring about injury. MT: 作為球迷的進氣道是快速旋轉，它會帶來傷害。[As the air inlet of a ball game fan rapidly rotates, it will bring harm].

A comparison of the three levels of MT outputs demonstrates that the more concise the SL sentence is, the better quality the MT output creates. The perfect-level MT output is produced as the result of using the single verb “spin”, not the verb phrase “bring about”, the single adverb “fast”, and the prepositional phrase “at the high speed”. Additionally, to clarify the message, the SL sentence with the perfect-level MT output has used a single noun, “the fan”, not the complicated noun structure that consists of two nouns connected by a preposition “of”, such as “the fan of the air conditioner” and “the fan of the air inlet”. Thus, a tentative conclusion is made that to make the MT output achieve the perfect-level of readability, the SL sentence must, at least, conform to the CE norms of clarity and conciseness.

4. Findings

In response to Question 1, how CE norms shift for reading and for MT application, the answer is that technical writing for MT application is more constrained by CE norms (eight out of ten in both lexical and syntactic areas) than for easy comprehension (six out of ten in both lexical and syntactic areas). In other words, more CE norms have to be followed in technical writing for machine/translation than for man/reading. Table 3 shows the difference in CE norms for these two purposes in the lexical area. The signal √ stands for acceptance; X, rejection.

Table 3. Different CE Norms for Man/Reading and for Machine/ Translation in the Lexical Area

CE Norms in the Lexical Area	Reading (6/10)	MT (8/10)
Use of determiners	√	X
Use of modifiers before general nouns	X	√
Avoidance of pronouns	X	√
Avoidance of more than four nouns in sequence	√	X
Avoidance of acronyms or abbreviations	√	√
Avoidance of polysemous words	X	√
Avoidance of more-than-three-word phrases	√	√
Avoidance of gerunds	√	√
Use of plain words, not specialized ones	√	√
Avoidance of misspellings and wrong punctuations	X	√

Syntactic structures in operational menus also reveal a higher degree of CE norms restriction for MT application than for man's

reading. Table 4 shows a comparison of CE norms for two different purposes in the syntactic area.

Table 4. Different CE Norms for Man/Reading and for Machine/ Translation in the Syntax Area

CE Norms in the Syntactic Area	Reading (6/10)	MT (8/10)
Use of active voice instead of passive voice	√	√
Avoidance of prepositional phrases as adverbs at the end of sentences	X	√
Avoidance of elliptical constructions	√	√
Avoidance of long sentences	√	√
Avoidance of which/that/who-led clauses	√	√
Avoidance of nominalizations	√	X
Avoidance of participle-led clauses	√	X
Avoidance of prepositional phrases as adjectives	X	√
Avoidance of conjoined prepositional phrases	X	√
Avoidance of putting subordinate clauses after main clauses	X	√

In response to Question 2, How CE norms shift for two MT systems, the finding shows that Google MT is constrained by six out of ten (60%) CE norms, but TransWhiz, by nine out of ten (90%) in the lexical area. However, Google MT is constrained by seven out of ten (80%), but TransWhiz, simply by four out of ten (30%) in the syntactic area. Table 5 shows the difference in CE norms in technical writing for TransWhiz and for Google MT in the lexical area.

Table 5. Differences in CE Norms for TransWhiz and for Google MT in the Lexical Area

CE Norms (in the lexical area)	TransWhiz (9/10)	Google MT (6/10)
Avoidance of a/an	√	X
Use of modifiers before nouns	√	√
Avoidance of specialized, technical terms	√	√
Avoidance of polysemous words	√	√
Avoidance of acronyms	√	√
Avoidance of three-or-over-three-word phrases	√	X
Use of some punctuations for markups, but avoidance of wrong punctuation	√	X
Avoidance of pronouns	√	√
Avoidance of gerunds	X	X
Avoidance of misspelling	√	√

In contrast, Google MT is more constrained by CE norms in the syntactic area than TransWhiz. Table 6 shows this difference.

Table 6. Differences in CE Norm for TransWhiz and for Google MT in the Syntactic Area

CE Norms (in the syntactic area)	TransWhiz (3/10)	Google MT (8/10)
Avoidance of prepositional phrases as adverbs	√	√
Avoidance of prepositional phrases ad adjectives	X	√
Avoidance of participle-led incomplete sentences	X	X

CE Norms (in the syntactic area)	TransWhiz (3/10)	Google MT (8/10)
Avoidance of complicated sentence structures that contain more than 25 words	√	√
Avoidance of putting adverbs at the end of the sentence	X	√
Avoidance of active voice	X	√
Avoidance of putting subordinate clauses after main clauses	X	√
Avoidance of which/that/who-led clauses	X	√
Avoidance of nominalized structures	√	X
Avoidance of noun groups connected by "of"	X	√

The above findings indicate that TransWhiz and Google MT have their strengths and weaknesses. Google MT cannot automatically modify some English structures into the linguistic conventions specific to the Chinese audience, but TransWhiz can do. However, Google MT has a better performance in the Chinese translations of English specialized terms and some lexical items than TransWhiz does. After learning their functional constraints, the editors would know what CE norms should be selected and what CE norms can be omitted for the application of different MT systems. The above CE norms serve as guidelines for the editors to customize controlled source texts to different MT systems.

5. Discussions

The implications of above findings in response to the three questions are discussed as follows.

5.1. Shift in CE Norms Across Time for Reading and for MT Application

After comparing the CE norms for reading and for MT application, I find that CE norms for the former purpose are less strict than those for the latter. In the lexical area, there are no restrictions on the use of determiners, modifiers, pronouns, polysemous words, and misspellings for the reading purpose, but these restrictions are reserved for MT application. My inference is that man can correctly interpret the meanings of polysemous words based on the context, but MT systems are not able to do so. For example, the audience can figure out the meaning of the word "run" as "execute" in the sentence *Nero BackItUp can set up jobs to run automatically*, but the MT system like TransWhiz interprets the word "run" as "the action of moving fast" and produces the inappropriate Chinese translation *尼羅 BackItUp 能建立工作自動地跑* [Nero BackItUP can establish tasks to automatically run].

Furthermore, people can understand the meanings of pronouns by referring back to the precedents, but MT systems translate all pronouns based on surface meanings and then produce unclear Chinese translations. For example, the audience may guess the meaning of "it" as "this section" in the sentence *It shows you how to use seat belts properly* by consulting previous sentences, but the MT system translates "it" as "它" (ta) in Chinese. This message is not clear, so I recommend the replacement of "it" with its specific reference, "this section", and then a more communicative Chinese translation "本章節" (ben-jhang-jie) is produced. Finally, man easily detects misspelled words and may grasp their meanings based on the context, but current MT systems remain unable to solve this problem. For example, the word "break" is misspelled in the sentence *Before cleaning, be sure to stop the break or pull out electric wire*. However, the audience may detect the misspelling and interpret it as something relevant to "switch". In contrast, the MT

system fails to identify the misspelling and translates it into Chinese as “打破” (da-po). This translation does not make any sense, so it is necessary to use correct spellings in the source text for MT application.

In the syntactic area, to achieve the purpose of appropriate MT application, CE norms must be followed in technical writing such as avoidance of prepositional phrases as adverbs/adjectives, conjoined prepositional phrases and subordinate clauses placed after main clauses. Unlike this, these norms are not required for man’s reading. My inference is that the audience is quite aware of syntactic differences between English and Chinese. For example, when the prepositional phrase “in all types of collisions” is used as an adverb/a place marker, it is normally put at the end of the English sentence like *A set belt is your best protection in all types of collisions*. However, the place marker is often put at the beginning of a Chinese sentence. If the MT system like Google MT translates the sentence into Chinese without modifying the position of the prepositional phrase, the awkward Chinese translation will be produced. In this case, we usually recommend adaptation of the prepositional phrase into a subordinate clause like “when all types of collisions happen”, so the appropriate Chinese translation will be produced. Furthermore, man can disambiguate the meanings of the conjoined prepositional phrase such as “a system for backing up and restoring information” based on the context, and correctly interprets it as “a system that backs up files and that restores information”. However, Google MT mistranslates it into Chinese as “系統的備份和恢復信息” [“a system for backing up” and “restoring information”].

It is worth notice that there are restrictions of CE norms such as avoidance of participle forms and wh/who/that-led clauses for man’s reading, but these restrictions are lifted for MT application. My inference is that non-native English audiences have difficulty analyzing the incomplete constructions of particle (-ing and –ed)

forms and relative-led clauses because these constructions have no explicit subjects. To increase syntactic clarity, the use of independent clauses is recommended. However, TransWhiz has stored many grammatical rules, so it can appropriately process and translate these constructions. Only corpus-based Google MT remains incapable of doing so.

In general, there is a higher degree of CE restrictions for MT application than for man's reading because man has the knowledge to interpret semantic and syntactic information within the context, but current MT systems are still functionally limited. This finding supports Toury's (1995) concept that norms are governed by a specific language policy within the industry. CE norms inevitably shift when its purpose shifts from effective communication in the past to appropriate MT application in present days.

5.2. Shift in CE Norms for Two MT Systems

A comparison of MT outputs leads to a finding that rule-based TransWhiz is more constrained by CE norms in the lexical area, but less constrained in the syntactic area than corpus-based Google MT. One overt difference is that TransWhiz cannot appropriately handle a/an, more-than-three-word verb and prepositional phrases, and words connected with inappropriate punctuations, but Google MT can do so. The reason for this is that TransWhiz semantically seeks word-for-word correspondence between ST and TT, and easily produces awkward translations. For example, TransWhiz eternally translates "a/an" as "一個" (yi-ge) in Chinese without changing quantifiers for different modified nouns. Actually, the translation of "a" as "一個" in "一個女人" (yi-ge-nyu-ren/a woman) is acceptable, but the translation of "a" as "一個" in "一個安全帶" (yi-ge-an-cyuan-dai/ a seat belt) is inappropriate. The correct translation is "一條安全帶" (yi-tiao-an-cyuan-dai). In contrast, Goolge MT automatically eliminates the translation of "a/an" in some cases and

therefore no similar translation error occurs.

As stated earlier, Google MT picks the highly probabilistic module through statistical analysis, so lengthy verb and prepositional phrases can be appropriately translated if exact or fuzzy translation matches are retrieved from the labeled corpus. Unlike this, TransWhiz analyzes lengthy verb and prepositional phrases word for word, so awkward and unnatural translations are easily produced. For example, TransWhiz translates the three-word verb phrase, "make contact with", as "作與…的接觸" (zuo-yu…de-jie-chu/do with….contact), but Google MT translates it as "接觸" (jie-chu/contact). Only when the dictionaries supported by TransWhiz already save the lengthy verb phrases can appropriate translations be produced.

In the syntactic area, more CE norms need to be followed in technical writing for Google MT than for TransWhiz. The significant difference is that Google MT has to avoid the use of prepositional phrases as adjectives, which/that/who-led clauses, adverbs next to modified verbs, and main clauses placed before subordinate clauses. My inference is that in the parsing of syntactic construction, Google MT mainly depends on retrieved fuzzy or exact matches and if no appropriate matches can be found, it will do word-for-word literal translation and produces the error of syntactically inappropriate translation. In contrast, TransWhiz transfers syntactic parsing results based on grammatical rules, so satisfactory translations of these syntactic constructions are produced. For example, the prepositional phrase "in your vehicle" is used to modify "infant and children", so it is put to the right of the modified noun in the English sentence *The seat belt can properly protect infants and children* <u>*in your vehicle*</u>. Unlike this, the Chinese sentence conventionally uses the left branching structure, so the prepositional phrase should be put to the left of the modified noun. Without changing the original position of the phrase, Google MT literally mistranslates the sentence into Chinese as *在安全帶可以妥*

善保護嬰兒和兒童在您的汽車 [In the seat belt may protect infants and children in your car]. Differently, TransWhiz automatically modifies the original word sequence and produces a more satisfactory translation*安全帶能適當地在你的車輛中保護嬰兒和孩子* [The seat belt can properly in your car protect infants and children].

In another example, the that-led clause, "that must be away from moisture", is used as an adjective to modify the noun, "objects", so it is put to the right of the modified noun in the sentence *Do not put objects that must be away from moisture*. Unlike this, the phrase should be put to the left of the modified noun in the Chinese sentence. Google MT neither retrieves the exact or fuzzy translation from the corpus nor modifies the original word sequence following the English-Chinese syntactic difference, so the literal mistranslation, *不要讓物體，必須遠離水分* [Do not let the object have to be away from water], is produced. However, TransWhiz parses and transfers the clause based on different English and Chinese syntactic structures, so the appropriate Chinese translation is produced like*不要放一定要遠離水分的物體* [Do not put the object that must be away from water].

Additionally, it is noted that Chinese sentences conventionally follow the cause-effect order while many English sentences show the opposite structure. Due to this syntactic difference, the subordinate clauses led by such conjunctions as "if", "when", "because" and others must be modified and be placed before main clauses in Chinese translations. TransWhiz can handle this problem well since it automatically changes the sequence of main and subordinate clauses in the translation process. Unfortunately, Google MT literally mistranslates sentences follwoing the original effect-cause order when the system fails to retrieve exact or fuzzy matches from the corpus. For example, after modifying the original word sequence, TransWhiz appropriately translate the English sentence *A pregnant woman should always wear a seat belt when*

she drives or rides in a vehicle into Chinese as*當她開車或搭乘一輛車輛的時候，一個懷孕的女人應該總是穿著一個安全帶* [When she drives or rides in a car, a pregnant woman should always wear a seal belt]. In contrast, Google MT literally mistranslates it as*孕婦應始終佩戴安全帶時，她駕駛或乘坐的車輛* [When the pregnant woman should always wear seat belt, the car she drives or rides]. Thus, avoidance of putting main clauses before subordinate clauses is identified as one of CE norms restrictions for Google MT, but not for TransWhiz.

In light of different strengths and weaknesses of TransWhiz and Google MT, CE norms have to be customized to different MT systems. This finding has supported Toury's (1995) argument that textual linguistic standards are subject to modification in the process of operation, and so are these CE norms for the application of different MT systems.

5.3. The Dynamic Nature of CE Norms

There is no doubt that CE norms in technical writing vary over time in response to different language and translation policies implemented in the industry for different purposes. In the past, CE norms were used for non-native English audience to easily comprehend technical documents. Nowadays, to develop multilingual translations for products and to increase the company's competitive edge, all technical texts such as operational manuals are written in CE for cost-effective MT application. Since machines are not as smart as human beings, CE norms for MT application are more demanding than those for man's reading. The modification of CE norms over time comes in line with Toury's (1995) view that translation norms should be examined not only in terms of the policy of the government and the industry that determines the translation purpose, but also in terms of the linguistic features for effective execution. CE norms have evolved over time to meet

varied degrees of simplification for different purposes in different times. Time has become a key element to govern shifts in CE norms. Additionally, to cope with different technological features of MT systems, CE norms shift between Google MT and TransWhiz. Thus, technology also plays a key role in adapting CE norms.

In short, CE norms shift in two directions: a diachronic shift from reading-oriented to MT-oriented ones, and a synchronic shift from TransWhiz-oriented to Google MT-oriented ones. The former tries to achieve the purposes of effective communication and cost-effective MT application, and the latter complies with different parsing engines of MT systems. Since the invention of new MT systems renders some CE norms ineffective, CE norms needs on-going modifications. The target environments in which CE norms are designed and used are themselves dynamic and changing, so CE norms are never complete and finalized. Seen in this regard, the dynamic nature of CE norms is confirmed. CE norms become increasingly complex when the trend of business globalization and MT technological development continuously evolve.

6. Conclusion

In this diachronic-synchronic research, the results of the investigation show that CE norms change not only within different industrial and business contexts in different times, but also for the application of two different MT systems in the same epoch. CE norms for easy comprehension are less strict than those for cost-effective translation produced by MT systems in both lexical and syntactic areas. Furthermore, rule-based TransWhiz is more constrained by CE norms in the lexical area, but less constrained in the syntactic area than corpus-based Google MT. These findings support the dynamic nature of CE norms and justify the close relevance of CE norms to the evolution of time and technology.

At present, CE norms should be viewed as an ongoing project, not a static one. They need to be upgraded and modified when the new MT system with a new parsing function is developed. We cannot rely on CE checking tools for MT application because there are too many ambiguities that cannot be detected by the machine. The SECC checker, for example, can only attain 87% precision levels (Adriaens & Macken 1997). Unlike the CE checker, technical authors/editors may detect and eliminate as many ambiguous syntactic constructions as possible following the guidelines of CE norms. However, CE norms cannot be generalized for once-for-all solution. Customized ones must be provided to meet the needs of different MT systems in the CE guidebook. This research is expected to help fulfill this goal, but due to the on-going technological evolution of MT technology, we can never reach "the terminal station" in the long journey of CE norms research. In general, CE norms that have been designed for English-Chinese translation serve as guidelines to control source texts for effective MT application. Although MT systems are not a panacea for translation, they can be a useful, helpful aid when used in the right way for the right purpose.

References

Adrianens, G. & L. Macken. 1997. Technological Evolution of a Controlled Language Application: Precision, Recall and Convergence Tests for SECC. In *Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation* (TMI 95). Leuven, Belgium: Katholieke Universiteit. 123-141

Allen, J. 2000. Controlled Language -- Changing Faces. *International Journal for Language and Documentation (IJLD)* 3, 20-21.

Arnold, D., *et al*. 1993. *Machine Translation: An Introductory*

Guide. London: Blackwells.

Bernth, A. 1999. Controlling Input and Output of MT for Greater Acceptance. In *The Proceedings of The 21st ASLIB Conference*. London: Aslib Proceedings.

Bernth, A. & C. Gdaniec. 2001. MTranslatability. *Machine Translation* 59.1, 175-218.

Ceusters, F., *et al.* 1998. From a Time Standard for Medical Informatics to a Controlled Language for Health. *J. Med Inform*, 48, 1-3, 85-101.

Gdaniec, C. The Logos Translatability Index in Technology Partnerships for Crossing the Language Barrier. In *Proceedings of the First Conference of the Association for Machine Translation in the Americas*. Washington, DC: AMTA. 97-105.

Kaji, H. Controlled Languages for Machine Translation: State of the Art. In *Proceedings of MT Summit VII.* Singapore: Kent Ridge Digital Labs. 3-8.

Laviosa, S. 2002. *Corpus-Based Translation Studies, Theory, Findings, Applications*. Amsterdam: Editions Rodopi BV.

Lehtola, C. *et al.* Controlled Language Technology in Multilingual User Interfaces. A paper presented at the 4th ERCIM Workshop on "User Interfaces for All" 19-21 *October*, 1998, Stockholm, Sweden. Available at URL <http://www.ui4all.gr/UI4ALL-98/lehtola.pdf> Nov. 25, 2008.

O'Brien, S & J. Roturier, How Portable are Controlled Language Rules? A Comparison of Two Empirical MT Studies, *In:* Maegaard, Bente ed. *Machine Translation Summit XI,* 10-14 September 2007, Copenhagen: Centre for Language Technology. 345-352.

Shih, C. 2002. *Theory and Application of MT/MAHT Pedagogy*. Taipei: The Crane Publishing Co. Ltd.

Shuttleworth, M. & M. Cowie. 1997. *Dictionary of Translation Studies*. Manchester, UK: St. Jerome Publishing.

Tanaka, H. 1999. What Would We Do Next for MT System

Development? In *Proceedings of MT Summit VII,* 13-17 September 1999, Singapore: Centre for Language Technology. 3-8.

Tedopres International BV. 1974-2007. Simplified Technical English Software and Services. Available at URL <http://www.simplifiedenglish.net/en/controlled_english/> Nov. 2, 2008

Toury, G. 1995. *Descriptive Translation Studies and Beyond*. Amsterdam: John Benjamins.

Underwood, L. & B. Jongejan. 2001. Translatabiltiy Checker: A Tool to Help Decide Whether to Use MT. In *Proceedings of MT Summit VIII.* Santiago de Compostela, Spain: (No Publisher). 363-368.

Wojcik, H. & E. Hoard. 1996. Controlled Languages in Industry. Available at URL <http://cslu.cse.ogi.edu/HLTsurvey/ch7node8.html> Dec. 2, 2008.

Chung-ling Shih
Department of English
Graduate Institute of Interpreting and Translation
National Kaohsiung First University of Science and Technology
No.2, Cho-yueh Rd., Nanzih District., Kaohsiung City 811, Taiwan (R.O.C)
Phone: +886-7-6011000 ext.5121; Email: clshih@ccms.nkfust.edu.tw

Received Mar. 2010; Reviewed Apr. 2010; Revised version received May. 2010.