Journal of Technology and Chinese Language Teaching

Volume 11 Number 1, June 2020
二〇二〇年六月　第十一卷第一期

ISSN: 1949-260X
http://www.tclt.us/journal

科技与中文教学

Journal of Technology and Chinese Language Teaching

A peer-reviewed online publication with in-print supplement
ISSN: 1949-260X http://www.tclt.us/journal

Volume 11 Number 1, June 2020

Articles

Columns

科技与中文教学

Journal of Technology and Chinese Language Teaching

A peer-reviewed online publication with in-print supplement
ISSN: 1949-260X http://www.tclt.us/journal

Contacts

URL: http://www.tclt.us/journal
Email: editor@tclt.us

Machine Translation for Editing Compositions in a Chinese Language Class: Task Design and Student Beliefs (应用机器翻译编辑中文作文：任务设计与学生信念)

Xu, Jun
(徐军)
Colorado State University
(科罗拉多州立大学)
jun.xu@colostate.edu

Abstract: The frequent use of machine translation (MT) in the daily lives of the digital generation presents challenges and opportunities for language teaching and learning. Rather than excluding MT from the classroom, educators have begun exploring various ways to integrate it into classroom instruction. While most studies ask students to post-edit a translation provided by MT, this study employed a different task design: having students post-edit self-written Chinese compositions with the help of MT. The study was conducted in a fourth-year Chinese language class at a public university. The beliefs of 12 students in the value of MT were investigated based on responses to a questionnaire and open-ended questions. The study found that students hold a positive attitude towards using MT in writing assignments. The students noted that MT helped them learn vocabulary and grammar, improve the quality of writing, boost confidence in Chinese use, and acquire autonomous learning skills. A comparison between this study and previous studies also revealed the critical role of task design in successfully implementing MT in classroom instruction.

摘要：机器翻译在日常生活中的广泛应用为语言教学带来了机遇与挑战。越来越多的研究在探索如何在课堂教学中应用，而不是排斥机器翻译。许多先行研究关注于学生如何通过机器翻译来编辑由机器翻译提供的写作版本，本研究采取了与其不同的任务设计。 12 位来自某公立大学四年级的中文学生使用机器翻译来编辑自己的中文作文。通过分析学生对问卷和开放式问题的回答，本研究发现学生对使用机器翻译编辑作文抱有积极的态度。 他们认为使用机器翻译可以帮助学习词汇和语法、提高作文质量、增强信心以及掌握自主学习策略。另外，本研究与先行研究比较结果也证明了任务设计在课程教学中应用机器翻译的重要性。

Keywords: Machine translation, L2 writing, Chinese language teaching, student beliefs, task design

关键词: 机器翻译、L2 写作、中文教学、学生信念、任务设计

1. Introduction

Nowadays, the digital generation frequently uses machine translation (MT) in their daily lives. The use of MT for language teaching and learning, however, has been controversial. The accuracy of MT products, academic dishonesty, and a possible impediment to language learning are the primary concerns identified in the literature (Benda, 2013; Case, 2015; Clifford, Merschel, & Munné, 2013; Correa, 2011, 2014; Ducar & Schocket, 2018; Garcia & Pena, 2011; Groves & Mundt, 2015; Jiménez-Crespo, 2017; Luton, 2003; Mundt & Groves, 2016; Stapleton & Kin, 2019). However, the quality of MT has improved significantly because of advances in artificial intelligence. For example, Google Translate (GT) launched a new GNMT (Google Neural Machine Translation) system in 2016. GNMT can learn from millions of examples and provide a significantly better quality of translation by encoding the semantics of sentences rather than merely memorizing phrase-to-phrase translation (Schuster, Johnson, & Thorat, 2016). Rapid improvement in MT resulted in the correction of many errors discussed in previous studies (Ducar & Schocket, 2018; Tian, 2018). Meanwhile, several studies reported that students still consult MT for assignments even though their instructors prohibit its use. (Correa 2011; Clifford et al., 2013; Tian, 2018). Also, pedagogical tools can include MT. Instead of being detrimental, MT use contributes to language learning from cognitive, linguistic, and affective perspectives (Correa, 2014; Enkin & Mejías-Bikandi, 2016; Garcia & Pena, 2011; Grove & Mundt, 2015; Jiménez-Crespo, 2017; Lee, 2019; Tsai, 2019; White & Heidrich, 2013).

Considering the widespread availability, easy accessibility, and the potential benefits of MT, educators no longer can merely defy the possibilities of MT in language learning and teaching by emphasizing its negative aspects. Instead, it is imperative to explore best practices to help students effectively and responsibly use MT to facilitate language learning (Benda, 2013; Correa, 2014; Ducar & Schocket, 2018; Groves & Mundt, 2015; Mundt & Groves, 2016; Jiménez-Crespo, 2017). However, only limited empirical research on this issue has been conducted (Lee, 2019; Tsai, 2019; White & Heidrich, 2013; Zhang, 2019).

Thus, using a task design that differs from previous studies, this preliminary study aims to investigate student beliefs regarding the use of MT as a language learning tool. Specifically, students first have an opportunity for discussion and instruction about using MT. Then students write compositions in their target languages (L2) without the help of MT, followed by corrections of their L2 writing using MT translation for comparison.

The reasons behind such a task design are twofold. First, Ducar and Schocket (2018) emphasized the importance of directly teaching learners how to use appropriate technology responsibly. However, previous studies failed to find instructions or discussions about using MT. Students in this study had an opportunity to receive guidance and discuss the use of MT with others before undertaking writing assignments; doing so seems to be an

indispensable component of the task design for MT as a tool for language teaching and learning. Second, MT is often treated as a "bad model" because, as Lee (2019) pointed out, most studies focused on students' post-editing of the MT translation. In this study, students post-edited their self-written compositions by using MT, a "peer" with intermediate level proficiency, as Correa (2014) and Ducar and Schocket (2018) suggested.

Meanwhile, this study also attempts to explore student beliefs about using MT with this task design. Students' experiences and expectations are essential factors in evaluating the effectiveness of language learning technology because the evaluation must "begin with the question 'effective for whom'" (Chun, Kern, & Smith, 2016, p.75).

2. Literature Review

While the significant potential of MT as a useful pedagogical tool of L2 writing has been described in previous studies, only limited empirical research exists. This section provides a brief overview of the task designs and the students' perceptions of the use of MT in such empirical research.

2.1 Task Design and the Use of MT

The application of MT for language learning and teaching has mainly focused on a process in which students write in their native languages (L1) first and then post-edit the translation provided by MT (Garcia & Pena 2011; Niño, 2009; White & Heidrich, 2013). For example, White and Heidrich (2013) asked students to write in their native language, English, to describe a picture prompt. The students were not told that the text would be used in a translation task later. Upon completion, they were instructed to use Google Translate (GT) to translate their L1 writing into German and edit that translation.

Using such a procedure, MT was treated as full of lexico-grammatical errors needing to be corrected, although students can benefit from the correction process (Garcia & Pena 2011; Lee, 2019; Niño, 2009). For example, Enkin & Mejías-Bikandi (2016) argued that MT can "help raise metalinguistic awareness of second language grammar and of the differences between grammatical constructions in the first and second language, which can help with the language learning process" (p.128). Besides, students can see the pitfalls of using MT. Ducar and Schocket (2018) pointed out that students can become aware of the fact that GT "does not take into consideration the roles that context, connotation, denotation, register, and culture play in language production and comprehension " (p.785).

Recent empirical MT studies modified such procedures by adding a step providing students with lexico-grammatical references in the target language to facilitate language learning (Lee, 2019). For example, Lee (2019) and Tsai (2019) investigated the effectiveness of such task design in the EFL context. Thirty-four Korean native-speaker students in a Korean university whose English proficiency was between intermediate and high-intermediate participated in Lee's (2019) study. Tsai (2019) explored the use of MT with Chinese EFL students from a university in Taiwan at three different levels (50

sophomores, 49 juniors, and 23 seniors) whose majors were English. Their English proficiency was between B1 and B2 of CEFR (Common European Framework of Reference for Languages). The students first wrote in their native language, Korean and Chinese, respectively (Step 1). Before using MT for editing, they translated those writings into English without the help of MT (Step 2). Then, students used MT to translate their L1 texts in Step 1 into English (Step 3). Comparing the MT versions and students' self-written versions, Tsai (2019) found that the quality of MT versions was significantly higher than those of students in terms of more words, fewer mistakes in spelling and grammar, and few errors per word. The MT versions also contained more advanced vocabulary. The results in Lee (2019) also indicated that the final versions had few lexico-grammatical errors and were of better quality. Although the students were satisfied using MT in their English writing and the quality of their papers improved significantly, one might be skeptical about whether and how such task design in which students used L1 to initiate the whole writing process facilitates language learning.

2.2 Student Beliefs About Using MT

Student beliefs about using MT have been examined utilizing questionnaires and interviews (Lee, 2019; Niño, 2009; Tsai, 2019; White & Heidrich, 2013). The studies showed that most students valued MT as a useful and supportive tool for writing assignments. In Niño's (2009) study, 75% of students reported MT as a helpful language tool, and 69% of them noted that they would use MT in the future. Furthermore, 75% of the students responded that practicing MT post-editing into the foreign language boosted their confidence in foreign language writing. Tsai (2019) reported that students were satisfied with their GT texts, and they believed that GT helped them complete the assigned writing task. Also, the studies showed that students thought that MT helps them find appropriate vocabulary. Tsai (2019) reported that most students thought the great benefit of GT was vocabulary use. Lee (2019) also found that 88% of students believed that MT was particularly helpful in helping them find more accurate words or authentic expressions for a given context. Besides, the studies also reported that students thought they became aware of potential grammatical errors in their writings and saw the limitations of MT by using MT (Lee, 2019; Niño, 2009; Tsai, 2019).

On the other hand, White and Heidrich's (2013) study reported students' unencouraging belief in MT based on pre-task, post-task questionnaires, and interviews. Students were confused about how to use MT in a sophisticated way. The questionnaire results showed that students agreed most with the statement, "I am wondering whether I used this resource sophisticatedly, i.e., whether it made my writing better or worse." Also, many students felt like using MT was cheating, even though they were told to use MT for the task. The lack of training about using MT may contribute to such unpromising results. Niño (2019) noted that not introducing MT properly to students is one example of bad practice when using MT. She argued that it is fundamental for language educators and students to foster awareness of the potential and limitations of MT in order to use MT in language learning.

Collectively, these studies outline a critical role for task design to play in using MT and influencing student perceptions of it. Thus, this study employs a different task design

with an emphasis on pre-task instructions and students editing self-written L2 writings with the help of MT. Based on such a task design, the study intends to address the following research questions:

1. Whether students perceive MT as a useful tool for L2 writing assignments?
2. What beliefs do students have regarding the use of MT under this task design?

3. Data and Methodology

3.1 Participants

Twelve students of a fourth-year Chinese class from a public university in the western US participated in this study. All were English native speakers with various Chinese learning experiences. Nine students started to learn Chinese after university matriculation, two students studied in high school, and one student was a heritage learner who spoke Cantonese at home. Three students had one semester or eight weeks of summer study abroad experience in China. Because of the variety of backgrounds, participants' Chinese proficiency varied from intermediate-low to advanced-low based on their homework and classroom performance evaluated by the researcher, who was the course instructor.

Table 1 Participants

Students	Learner type*	Study abroad	Proficiency
S1	High school	No	Intermediate-Mid
S2	University	Yes	Intermediate-Low
S3	University	No	Intermediate-Low
S4	Heritage	No	Advanced-Low
S5	University	No	Intermediate-Low
S6	University	Yes	Intermediate-Mid
S7	University	No	Intermediate-Mid
S8	University	No	Intermediate-Low
S9	University	No	Intermediate-Low
S10	University	Yes	Intermediate-High
S11	University	No	Intermediate-Low
S12	High school	No	Intermediate-Mid

*Learner type is defined as when the students initially began studying L2 Chinese.

3.2 Task Description

The participants were required to write two compositions and submit reflection papers after the completion of each draft. The two composition assignments were response essays to readings in class. The first task was about the "Tiger Mother," and the second task was about the issue of US gun control. The first task took place in October and the second task was conducted in December. The steps for each task followed the same procedure, although Step 1 was only applied to Task One (see Table 2). Contact hours for the course were three 50-minutes classes per week. Students completed Draft One

Monday, and Draft Two and Reflection One on Wednesday. Draft Three and Reflection Two were submitted on the Monday of the following week.

In Step 1, MT, including Google Translate, Baidu Translate, and Sogou Translate, was introduced to students in the class. Following Ducar and Schocket's (2018) suggestions, the instructor and students discussed ethical issues, strengths, and pitfalls regarding the use of MT. The following points were particularly emphasized: the use of MT as one autonomous learning strategy; writing as a process, not just a product (Williams, 2006); the internet as a corpus to explore the use of vocabulary suggested by MT; and MT, particularly Baidu, as an online dictionary. Students were instructed to carefully make decisions on whether to adopt the alternatives provided by MT. In addition, students were told that their grades were only based on the final draft to train students to "understand and practice writing as a growth and revision process rather than a short-term product" (Ducar & Schocket, 2018, p.792).

Table 2 The procedure of tasks in the writing assignment

Step 1	Instructions and discussion about using MT
Step 2	In-class writing in Chinese without any help (Draft One)
Step 3	In-class post-editing with machine translation tools (Draft Two)
Step 4	Submit Draft One and Draft Two as well as Reflection One
Step 5	Out-of-class revision
Step 6	Submit Draft Three and Reflection Two

In Step 2, students completed the writing assignment (Draft One) on computers without the help of MT at a lab during the regular class meeting time. After that, in Step 3, students were advised to choose one or all machine translation tools to complete Draft Two through a "translate-compare-detect errors-consider alternatives-rewrite" process (Lee, 2019). The students first translated the self-written Draft One into English with MT and edited the English to make it accurate and appropriate. The next step was to use machine translation tools to translate the revised English version back into Chinese. Students compared their self-written Chinese versions with the machine-translated Chinese versions and detected any errors. They edited their self-written Chinese versions by accepting or rejecting certain parts of the machine-translated Chinese version. Upon completing Draft Two, they must highlight any parts adopted from the machine-translated version.

In Step 4, students submitted their self-written Draft One and revision Draft Two as well as Reflection One. In Reflection One, students answered several questions about the use of MT, such as "what and how did you use MT in revision?" and "what did you gain from the revision process?" The questions were adopted from Zhang (2019).

In Step 5, students revised their Draft Two based on the instructor's comments and submitted their final draft, Draft Three, as well as Reflection Two in Step 6. Although the questions in Reflection Two for Task One were the same as those for Reflection One, students were guided to reflect on the whole writing process involving self-writing, revision with the help of MT, and instructor's comments. Reflection Two for Task Two, the issue of US gun control, which serves as data for the present study, was specially designed to understand students beliefs about using MT for writing after practicing two

tasks. It consisted of a questionnaire and open-ended questions adapted from Niño (2009), White and Heidrich (2013), and Zhang (2019) (see Appendix).

The excerpts in Table 3 show the same paragraph from student S7's first task. She completed Draft One in class on a computer without the help of MT. After that, she used Baidu to translate her self-written text into English. She underlined the parts to which she made changes to be acceptable English. The revised English translation was then translated into Chinese by Baidu. She also highlighted the changes between her Draft One and the Baidu Chinese translation. Finally, she compared her self-written Draft One and the Baidu Chinese translation and decided what to revise to complete Draft Two. She underlined all the items adopted from Baidu as well. It is important to note that several errors, such as "小时" in Draft One, were not corrected in Draft Two. Also, student S7 did not carefully underline all the parts she adopted from Baidu. For example, she changed "是因为她们没有努力画" to "而是因为她们没有努力画" without highlighting the "而是" part in Draft Two. What students *actually* changed and what errors students *should* correct were not examined in this study.

Table 3 Excerpts of student S7's writing

Draft One	Sophia 也说她和 LuLu画一篇 card 给妈妈，但是蔡美儿觉得不够好，还给女儿。很多西方家长觉得太 extreme ，但是女儿知道其实不是因为她们画的 card 不够好，是因为她们没有努力画。蔡美儿培养她们的努力。
Revised Baidu English translation	Sophia also said that she and Lulu drew a card to give to her mother when they were young, but Chua felt it wasn't good enough so give it back to her daughters. Many Western parents think it's too extreme, but her daughters know it's not because they didn't draw good cards, it's because they didn't work hard. Chua nurtured their work ethic.
Baidu Chinese translation	索菲亚还说，她和露露小时候画了一张卡片送给母亲，但蔡美儿觉得不够好，所以把卡片还给女儿。很多西方父母认为这太极端了，但她的女儿们知道，这不是因为她们没有画好牌，而是因为她们没有努力工作。蔡美儿培养了他们的职业道德。
Draft Two	Sophia还说她和 LuLu 小时画了一张卡片送给母亲，但是蔡美儿觉得不够好，所以把卡片还给女儿。很多西方家长觉得太极端了，但是女儿知道，不是因为她们没有画好牌，而是因为她们没有努力画。蔡美儿培养她们的职业道德。

Worth noting is the difference in task design between this study and previous studies. First, the present study included formal instruction and discussion about using MT. Second, students wrote compositions in the target language without any help (Step 2) in this study. This step is different from Lee (2019), Tsai (2019), and White and Heidrich (2013), in which the participants wrote compositions in their native languages first (Step 1) before translating them into L2 with MT.

Comprehensible output in L2 is essential for language learning. According to Swain and Lapkin (1995), "in producing the target language, learners may encounter a problem

leading them to recognize what they do not know, or know only partially. In other words, the activity of producing the target language may prompt second language learners to consciously recognize some of their problems, it may bring to their attention something they need to discover about their L2" (p.373). Thus, students were asked to write their Draft One in L2, Chinese, rather than their native language, English.

Also, the use of MT was limited to revision but not for the initial product. Revision is critical in L2 writing because it is unrealistic to expect error-free first drafts (Polio, Fleck, & Leder, 1998). By modifying writing outputs, learners can "test hypotheses about the second language, experiment with new structures and forms, and expand and exploit their interlanguage resources in a creative way" (Pica, Holliday, Lewis, & Morgenthaler, 1989, p.64).

In addition, the reason students were asked to translate back and forth and compare their self-written texts and corresponding English texts is that such a method can facilitate learning by noticing (Schmidt, 1990, 2010) and seems to be the standard practice in the literature (Lee, 2019; Tsai, 2019).

3.3 Data and Analysis

The data examined for this study was Reflection Two of Task Two (the issue of US gun control). Edwards and Liu (2018) propose that students should have multiple opportunities to experiment with any new method of teaching. Thus, this study focused on students' last reflections after completing two tasks. In addition, the Final Reflection was specially designed differently from the other three Reflections to allow students to reflect on the use of MT over the entire semester. Thus, the other three Reflections were excluded from this examination.

The Final Reflection consists of a 5-point Likert scale questionnaire and five open-ended questions. The questionnaire investigates students' general perceptions regarding the use of MT in writing exercises. It is identical to the one used in White & Heidrich (2013), which consists of 13 items ranging from 1 "strongly disagree" to 5 "strongly agree." The majority of the five open-ended questions are from Niño (2009) and Zhang (2019), which provided students opportunities to elaborate on their opinions about using MT. Following the methods in Baralt (2012) and Duff (2012), common themes, such as vocabulary, grammar, quality, confidence, and learning strategies were identified through multiple steps of coding the responses to open-ended questions.

4. Results

All students used Baidu, Sogou, or both to edit their writings. No student chose GT. Generally, students showed significantly positive attitudes towards the use of MT. This section first reports the results of the questionnaire regarding the participants' general perceptions of using MT. Then the responses to the open-ended questions will be discussed.

4.1 Results of the Questionnaire

Table 4 shows the participants' responses to the question: "When you used the machine translation, how did it make you feel?" from the questionnaire. Students indicated their level of agreement or disagreement with each statement on a five-point scale, with one showing "strongly disagree," and five indicating "strongly agree." The rank of each statement in Table 4 is based on the means of participants' responses. The top of the table suggests which statements students most strongly agreed with while the bottom of the table shows participants' strong disagreement with the statement. It is worth noting that the questionnaire consists of both positive and negative statements.

In general, students showed an extremely positive attitude towards the use of MT. The means of the top eight items were four and above; four indicated agreement (Q7, Q1, Q4, Q8, Q5, Q9, Q13, Q11). Students believed that vocabulary was the most beneficial outcome of using MT because MT helped them find the words to articulate what they wanted to say as well as new and sophisticated ones (Q7, Q8). Moreover, students responded that they were able to deliver their best work with the help of MT (Q4, Q5). It is important to note that most students did not agree that the use of MT was cheating (Q3). Only two students responded with "agree" to the statement, "Q3: I feel like I might have cheated." In contrast, eight students responded with "disagree" and "strongly disagree" to the statement.

Table 4 Students' beliefs about the use of MT

	Questions	Mean	SD
Q7	I feel like it helps me use words that fit what I want to say.	4.58	0.67
Q1	I feel I am giving my best effort by using this resource.	4.33	0.49
Q4	I feel like it helps me deliver my best work for my own satisfaction.	4.33	0.65
Q8	I feel like it helps me use words that are new and sophisticated.	4.33	0.65
Q5	I feel like it helps me deliver my best work for getting a good grade.	4.25	0.45
Q9	I feel like it helps me organize what I want to say more clearly.	4.17	0.94
Q13	I feel like it helps my voice emerge more distinctly.	4.08	0.90
Q11	I feel like it helps me develop better content.	4.00	0.95
Q12	I feel like it helps improve my style.	3.92	0.90
Q6	I feel like it helps me use more complex grammatical structures.	3.67	1.50
Q10	I feel like it helps me spell more sophisticatedly.	3.67	1.07
Q2	I am wondering whether I used this resource sophisticatedly, i.e., whether it made my writing better or worse.	3.42	0.90
Q3	I feel like I might have cheated.	2.08	1.16

*Scale:1=strongly disagree; 2=disagree; 3=Neither agree nor disagree; 4=agree; 5=strongly agree

4.2 Responses to Open-ended Questions

Student responses to the open-ended questions revealed more detail about their beliefs on using MT during post-editing. As indicated in Table 4, students appeared to

agree that MT was beneficial for vocabulary (Q7, Q8). To elaborate on this point, first, MT seemed to be able to help students find the words to express what they wanted to say. Student S3 noted that "using translation software has helped me develop good phrases and find helpful vocabulary for what I want to say." Another student, S4, responded that "using machine translation helps me find new, and often more sophisticated, vocabulary that can elevate my writing if I use it correctly." Second, students used MT as a useful tool to "fill in the gaps," as student S6 explained, between what they know and the unknown. The student further said that MT was useful when "I know how to structure a sentence, and I know most of the words, but a key phrase may be missing." Another student, S5, responded that "sometimes, I am close to saying what I really want to say but am not quite there yet, and the translation process helps me find the right words to better express my thought." Third, MT seemed to be useful in identifying appropriate words for various situations. Student S1 mentioned that MT "helps me choose a more appropriate word because there are different characters that are more appropriate for different situations."

Grammar is another aspect of the perceived usefulness of MT. Student S1 responded that MT helped her put things in the correct order. The student further explained that "sometimes I know the components of the sentences, but I'm not sure grammatically how to structure it." Moreover, by translating self-written sentences and corresponding English back and forth, MT functioned as a proof-reader so that students could *notice* grammatical errors. Student S4 commented that "machine translation has helped me realize when my grammar is completely wrong, which is paralleled by the sentence not reflecting what I intended to write. When I modify what I want to say in English and re-translate it back into Chinese, then I can see what grammar structure should be used." Another student, S9, noted that "after writing out sentences, sometimes I would translate my Chinese into English to make sure I did not have any grammatical mistakes."

Also, students believed that MT helped them deliver papers of better quality. First, MT seemed to be able to help articulate students' thoughts. By translating the self-written Chinese paragraphs to English, student S12 used MT to confirm whether her ideas were delivered without confusion. The student noted that "I mostly use translators to get an idea of what my work would sound like if translated back into English. By reading what the translator says when I place my Chinese writing in it, I get an idea of what the passage sounds like." Second, students could easily and quickly detect and fix errors with the help of MT. Student S10 explained that "I can quickly find mistakes which help me write better papers. I quickly find my mistakes by taking passages that I have written in Chinese and translating them into English. If I read English, I can quickly discover if I made a mistake, or if what I wrote in Chinese didn't make much sense." Another student, S4, compared MT with Microsoft Word and explained:

> *"Machine translation is like a more complex version of the grammar check feature in Microsoft Word. When I'm typing something in Word, there's the blue lines that indicate if the grammar is wrong or the red squiggly line if the word is spelled incorrectly. By using machine translation, I can see where I used improper grammar if the sentence doesn't translate well into English and I can also see where I accidentally typed the wrong character*

because it translated into a word that I've never learned before in Chinese (e.g., 枪击案 accidentally became 强奸)."

In addition to the benefits of MT concerning vocabulary, grammar, and better writing quality, students also elaborated on other strengths unlisted in Table 4. First, MT functioned as a reminder of what students have learned. Student S7 explained that she was "being reminded of vocab and grammar." Student S11 added that MT "reminds me of a way to say it differently than I have already learned but maybe forgot how to use well." Second, students used MT to confirm the correctness of what they wrote. Student S11 said that MT "validates the way I was going to say something." Third, MT also acted as a thesaurus not only for words but also for grammar structures. Student S4 noted that "re-translating what I wanted to say in English back into Chinese will often show a more sophisticated version of what I wanted to write." Last, students praised the practice of using MT itself because it is an autonomous learning skill. Student S7 noted that MT "helped me practice using a translator in the right way. I probably will never be fluent in Chinese; I'll always need some help from a translator. Practicing using the translator to incorporate into my own writing was the most useful aspect."

Eleven students responded that MT made them feel more confident in their Chinese or in the quality of their submissions. First, MT provided them an opportunity to confirm what they know, and consequently, boosted their confidence. Student S12 noted that "surprisingly when using the translators, I noticed that I already knew a decent amount of what I wanted to say and usually typed well. This made me feel more confident because the translators were telling me that what I had written was quite similar to what I had planned to say. In this sense, using translators has affirmed that I already know a decent amount of Chinese, but can occasionally refer to it for extra help." Second, students felt more confident in submitted assignments because MT functioned as a checker to identify mistakes in their writings. Student S5 explained that MT identified many "silly" mistakes she made. Another student, S9, responded that thanks to MT, "I have become more aware of my common grammar mistakes in Chinese." Student S10 noted that "I feel much more confident about being about to write an essay that lacks mistakes. It gives me confidence because I can look at my sentences and see if they have mistakes, whereas before using machine translation, I would frequently have many mistakes." On the other hand, however, Student S3 responded that MT did not improve his confidence in Chinese because he believed that confidence is "built up through speaking the language."

It is of importance to note that one student, S6, carefully distinguished between confidence in Chinese writing and confidence in the quality of the submitted writing assignment. He explained that MT "does make me more confident in pieces I am turning in because I believe they are of good quality due to a mix of machine translation and my personal skill helps improve the quality." However, "it's just that I become less confident on my own if I depend on these tools."

Students also explained whether their opinions on MT changed after completing the two writing assignments. Half of the students, six out of twelve, responded that their opinions remained the same because they were already aware of the benefits and potential risks of MT before the current study. Some of them used MT for various purposes since

high school. They had been cautiously using MT, despite their concerns, without becoming overly dependent on MT. Student S6 noted MT's potential to undermine student motivation for L2 learning. He explained that MT "can benefit and improve students writing. However, it will also cause some students who aren't as passionate about learning a foreign language to rely on it for an easy grade." Another student, S8, responded that "I also find it is really easy to cheat yourself out of knowledge in the language if you rely on the translator too much."

The other half of the students reported positive changes in their opinions about using MT. They previously avoided using MT because of concerns about academic dishonesty, inaccuracy in MT output, and instructions from former teachers. Two students, S1 and S12, mentioned they did not trust MT because their high school teachers completely forbade the use of MT for language learning because of inaccurate outputs and possible violations of academic integrity. Now, student S12 believed that "these tools provide extra support and assistance when learning a language." Student S10 mentioned that it felt like cheating to use MT at first, and then "it feels like a powerful tool." He further explained that "I think that my opinion changed because I learned how to use machine translation to its full advantage while avoiding its pitfalls."

Further, all students expressed willingness to continue using MT in the future for various purposes, from checking words and grammar to communication with family members. Six students mentioned that they would use MT to check their writing assignments; as student S5 wrote, "I will use it in the future to check over my writing once I have the first draft." Interestingly, student S3 decided to use MT to read documents in the future. Probably, the student had a positive experience with using WeChat online translation for reading in class during the in-class MT instruction.

5. Discussion

As mentioned in the literature review, students value MT as a useful tool for language learning (Lee, 2019; Niño, 2009; Tsai, 2019). The beliefs expressed by the students in this study were in line with those studies. The students in this study believed that MT was beneficial for their writing. Vocabulary improvement appeared to be the most favorable outcome of MT for students in this study. MT also helped them detect grammatical errors in their papers. Students were confident that MT helped elevate the quality of their submissions. All the students plan to continue using MT in the future.

One interesting finding was that students in this study distinguished between their confidence in products submitted to instructors and their confidence in Chinese proficiency in general. While student belief in MT's usefulness in improving the quality of their submissions have been identified in this and previous studies, students in this study revealed that MT also helped build up their confidence in Chinese. A possible explanation for this might be due to the task design. Contrary to previous studies in which students write in their L1 first, students in this study wrote in L2 *first* and then used MT to translate their writings to L1. The procedure, including the sequence of L2 writing, translation into L1, comparison, detecting errors, considering alternatives, and rewriting, provided them an

opportunity to validate what they knew regarding vocabulary and grammar, and what they wrote was correct. Consequently, their confidence in the Chinese language, in general, was established through such a process.

Another interesting finding was that students not only believed they would continue to use MT in the future but also thought they learned how to use it effectively and responsibly. Task design might also be related to such beliefs. Different from previous studies in which students' views were examined only after one task, here, students' opinions were investigated after training and two tasks. Edwards and Liu (2018) suggested the importance of multiple practice to fully take advantage of peer-response activity. Students in this study had two opportunities to practice MT to edit their writings, which might contribute to their overall positive attitude toward using MT.

The significant difference between the results of the questionnaire in this study and the identical one in White and Heidrich (2013) is surprising. First, the overall rate of approval of MT in the current study was higher than those in White and Heidrich (2013). As mentioned above, the means of the eight items in this study were four and above, while all items in White and Heidrich (2013) were under four. Second, the two statements with the highest means in White and Heidrich (2013) were "Q2: I am wondering whether I used this resource sophisticatedly, i.e., whether it made my writing better or worse" (Mean=3.59, SD=0.71) and "Q3: I feel like I might have cheated" (Mean=3.59, SD=1.18). In contrast, surprisingly, these two items were the two lowest in this study (Q2: Mean=3.42, SD=0.90; Q3 Mean=2.08, SD=1.16). Third, in White and Heidrich (2013), students showed the lowest agreement to the statement, "Q1: I feel I am giving my best effort by using this resource (Mean=2.06, SD=0.97). In the current study, this item ranked second-highest (Mean=4.33, SD=0.49).

Overall, the students in this study showed significantly more positive attitudes toward the use of MT compared with White and Heidrich (2013). They believed they were aware of a competent and responsible way to use MT while the students in White and Heidrich (2013) seemed very confused about how to use MT. Further, students in White and Heidrich (2013) considered the use of MT as a violation of academic integrity, while such was not the case in this study. The improved accuracy of MT might contribute to the different students' experiences in these two studies. Since Google Neural Machine Translation was launched in 2016, the quality of MT has significantly increased, which might ultimately positively affect students' user-experience. Another possible explanation for these discrepancies may be the lack of instruction on the use of MT and multiple opportunities to use MT. In White and Heidrich (2013), students of German were asked to describe a picture prompt in L1 (English) without knowing they were going to use Google Translate to translate their writings into German. Their opinions were asked immediately after the task. Students in Niño (2009) felt MT was useful was because "they reflect the views of a group of advanced students of Spanish who have received previous training in translation and who have been introduced of MT and in particular to MT post-editing" (p.249). Therefore, the task design in this study, including pre-task instruction and discussion as well as two opportunities to practice, might be one of the reasons for students positive attitudes towards the use of MT in writing assignments.

These findings, while preliminary, suggest the importance of task design in the practice of using MT in language teaching and learning. First, because Edwards and Liu (2018) pointed out the importance of instructions in order to fully take advantage of the benefits of peer response, it seems imperative to integrate instructions on the use of MT into writing class practice. During the instruction sessions, students should not only be informed about the strengths and pitfalls of MT but should also be made aware of how to use MT effectively and responsibly. It is also important to remind students of the error tolerance feature of MT, which can translate the original text with errors into correct outputs (Massardo et al., 2016). Translating the original text and corresponding text provided by MT back and forth is essential for the revision process so that students can avoid overlooking errors and recognize them in the original text. Second, students should learn how to use the internet as a corpus to examine critically and strategically alternate expressions suggested by MT. As student S1 noted, "when it suggests words I am not familiar with, I don't know what I want to do with them, or how to study them, so I build vocabulary." Third, students might need multiple opportunities to experience the excitement and frustrations through the complex revision process. Long-time training and practice of MT can help students eventually discern their autonomous learning strategies.

6. Conclusion

The present study aimed to examine student beliefs on using MT in post-editing self-written L2 Chinese writing assignments. This study showed students expressing more positive attitudes towards using MT when task design included pre-task instructions and discussions about using MT as well as multiple tasks. Students believed that MT helped them find new and appropriate vocabulary and grammar as well as expressions, improve the quality of writing, and boost their confidence in Chinese. Also, students felt that MT helped remind them of what they learned earlier and validate what they wrote. MT also functioned as a thesaurus for learning alternative expressions. Finally, students thought they had acquired an effective and responsible way to use MT in the future.

However, the small sample size of this study makes these findings less generalizable. Also, the results were solely based on students self-reporting. Other methods, such as interviews and think-aloud, might bring about different insights. Since the study was limited to student beliefs about using MT, whether their beliefs matched their behavior during the post-editing process remains unclear. Also, it is essential to test more profound and permanent learning with the use of MT in the future.

Despite its limitations, the study certainly added to our understanding of using MT for language learning and teaching. First, it confirmed that students believed MT could play a critical role in language learning. Second, it also suggested that students did not patently accept but employed various strategies to examine what MT provides. For this reason, instructors must reconsider their position about using MT in the classroom. Third, the findings also revealed the critical role of pre-task instruction regarding the use of MT. Altogether, the study expanded our understanding of the practice of MT for classroom instruction. Chun et al. (2016) pointed out that "the use of technology should not be seen as a panacea or a goal in and of itself, but rather as one means to support specific learning

goals" (p.77). Given the prevalence of MT in the digital generation's daily life and the unlikelihood that students will avoid taking advantage of MT, educators must carefully design a way of including MT to support language learning so that ultimately our students become "career, life, and world-ready" (ACTFL, 2017). The insights gained from this study may be of assistance for educators to adopt MT as a powerful pedagogical tool for language teaching and learning.

References

ACTFL. (2017). ACTFL position statement on the role of technology in language learning. Retrieved from https://www.actfl.org/news/position-statements/the-role-technology-language-learning

Baralt, M. (2012). Coding qualitative data. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 222-244). Hoboken, NJ: John Wiley & Sons.

Benda, J. (2013). Google Translate in the EFL classroom. *Writing & Pedagogy*, *5*(2), 317-332.

Blake, R. (2016). Technology and the four skills. *Language Learning & Technology*, *20*(2), 129-142.

Case, M. (2015). Machine translation and the disruption of foreign language learning activities. *eLearning Papers*, *45*, 4-16.

Chun, D., Kern, R., & Smith, B. (2016). Technology in language use, language teaching, and language learning. *The Modern Language Journal*, *100*(S1), 64-80.

Clifford, J., Merschel, L., & Munné, J. (2013). Surveying the landscape: What is the role of machine translation in language learning? *Revista d'innovació educativa*, *10*, 108-121.

Correa, M. (2011). Academic dishonesty in the second language classroom: Instructors' perspectives. *Modern Journal of Language Teaching Methods*, *1*(1), 65-79.

Correa, M. (2014). Leaving the "peer" out of peer-editing: Online translators as a pedagogical tool in the Spanish as a second language classroom. *Latin American Journal of Content & Language Integrated Learning*, *7*(1), 1-20.

Ducar, C., & Schocket, D. H. (2018). Machine translation and the L2 classroom: Pedagogical solutions for making peace with Google Translate. *Foreign Language Annals*, *51*(4), 779-795.

Duff, P. A. (2012). How to carry out case study research. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 95-116). Hoboken, NJ: John Wiley & Sons.

Edwards, J. H., & Liu, J. (2018). *Peer response in second language writing classrooms*. Ann Arbor, MI: University of Michigan Press.

Enkin, E., & Mejías-Bikandi, E. (2016). Using online translators in the second language classroom: Ideas for advanced-level Spanish. *Latin American Journal of Content & Language Integrated Learning*, *9*(1), 138-158.

Garcia, I., & Pena, M. I. (2011). Machine translation-assisted language learning: Writing for beginners. *Computer Assisted Language Learning*, *24*(5), 471-487.

Groves, M., & Mundt, K. (2015). Friend or foe? Google Translate in language for academic purposes. *English for Specific Purposes*, *37*, 112-121.

Jiménez-Crespo, M. A. (2017). The role of translation technologies in Spanish language learning. *Journal of Spanish Language Teaching*, *4*(2), 181-193.

Lee, S. M. (2019). The impact of using machine translation on EFL students' writing. *Computer Assisted Language Learning*, *33*(3), 157-175.

Luton, L. (2003). If the computer did my homework, how come I didn't get an "A"? *The French Review*, *76*(4), 766-770.

Massardo, I., van der Meer, J., O'Brien, S., Hollowood, F., Aranberri, N., & Drescher, K. (2016). MT post-editing guidelines. Retrieved from http://personel.klu.edu.tr/dosyalar/kullanicilar/sevdapekcoskunguner/dosyalar/dosya_ve_belgeler/taus_post-editing_guidelines_ders4.pdf

Mundt, K., & Groves, M. (2016). A double-edged sword: The merits and the policy implications of Google Translate in higher education. *European Journal of Higher Education*, *6*(4), 387-401.

Niño, A. (2009). Machine translation in foreign language learning: Language learners' and tutors' perceptions of its advantages and disadvantages. *ReCALL*, *21*(2), 241-258.

Pica, T., Holliday, L., Lewis, N., & Morgenthaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition*, *11*(1), 63-90.

Polio, C., Fleck, C., & Leder, N. (1998). "If I only had more time:" ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing*, *7*(1), 43-68.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129-158.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC 2010* (pp. 721–737). Singapore: National University of Singapore.

Schuster, M., Johnson, M., & Thorat, N. (2016). Zero-shot translation with Google's multilingual neural machine translation system. Retrieved from https://ai.googleblog.com/2016/11/zero-shot-translation-with-googles.html

Stapleton, P., & Kin, B. L. K. (2019). Assessing the accuracy and teachers' impressions of Google Translate: A study of primary L2 writers in Hong Kong. *English for Specific Purposes*, *56*, 18-34.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, *16*(3), 371-391.

Tian, Y. (2018). The challenge of machine translation to traditional translation homework in Chinese language learning. *Journal of Technology and Chinese Language Teaching*, *9*(1), 78-95.

Tsai, S. C. (2019). Using google translate in EFL drafts: A preliminary investigation. *Computer Assisted Language Learning*, *32*(5-6), 510-526.

White, K. D., & Heidrich, E. (2013). Our policies, their text: German language students' strategies with and beliefs about web‐based machine translation. *Die Unterrichtspraxis/Teaching German*, *46*(2), 230-250.

Williams, L. (2006). Web‐based machine translation as a tool for promoting electronic literacy and language awareness. *Foreign Language Annals*, *39*(4), 565-578.

Zhang, P. (2019, April). From input to output: A preliminary study on the effectiveness of intelligent tool-assisted reading and writing skill training. Paper presented at the 2019 Annual Conference Chinese Language Teachers Association, USA, Seattle, WA. [张霓. (2019, 4), 从输入到输出--智能工具辅助的读写技能训练成效初探. 美国中文教师学会2019大会. 西雅图，华盛顿州.]

Appendix

Final Reflection

1. When you used the machine translation, how did it make you feel? Please indicate your dis/agreement with the following statements:
   Strongly Agree=5
   Agree=4
   Neither Agree nor Disagree=3
   Disagree=2
   Strongly Disagree=1

   Q1: I feel I am giving my best effort by using this resource
   Q2: I am wondering whether I used this resource sophisticatedly, i.e., whether it made my writing better or worse.
   Q3: I feel like I might have cheated.
   Q4: I feel like it helps me deliver my best work for my own satisfaction.
   Q5: I feel like it helps me deliver my best work for getting a good grade.
   Q6: I feel like it helps me use more complex grammatical structures.
   Q7: I feel like it helps me use words that fit what I want to say.
   Q8: I feel like it helps me use words that are new and sophisticated.
   Q9: I feel like it helps me organize what I want to say more clearly.
   Q10: I feel like it helps me spell more sophisticatedly.
   Q11: I feel like it helps me develop better content.
   Q12: I feel like it helps improve my style.
   Q13: I feel like it helps my voice emerge more distinctly.

2. Do you think Machine Translation has helped you to improve your writing in Chinese? Why? In what sense?

3. Has Machine Translation given you more confidence in your foreign language written production? In which sense?

4. Are you going to use Machine Translation in the future? For what purpose?

5. What was your opinion on Machine Translation? What is your current opinion? Are they the same? Why?

6. What are your suggestions for future writing classes?

Error Tolerance of Machine Translation: Findings from Failed Teaching Design[1] (机器翻译的容错性：从一个失败的教学设计中得到意外发现)

Tian, Ye
(田野)
University of Pennsylvania
(宾夕法尼亚大学)
tianye1@sas.upenn.edu

Abstract: Science is constantly being revised, and failure is crucial to this process. This study is enlightened by the findings of a failed teaching design that aims to train intermediate level Chinese language learners to use Machine Translation as a self-editing tool to improve their writing proficiency. This study finds that Sogou Translate possesses a powerful error tolerance feature that can correctly translate students' Chinese sentences into correct English sentences, even though these sentences may contain various errors made by Chinese language learners, including misspellings, typos, misusing homonyms, misusing the part of speech of a word, misusing synonyms, ungrammatical phrases, and incorrect word or sentence order. This study also suggests three areas to further investigate and to apply error tolerance features of Machine Translation in the field of Chinese language teaching.

摘要：失败对科学的发展至关重要，本研究同样源自一项不成功的中文教学设计。该教学设计训练中级水平的中文学生使用机器翻译作为自主纠错工具，目的是为了提高他们的写作水平。然而本研究发现搜狗翻译包含强大的容错性，即使这些学生的中文写作中包含学习者常犯的各种错误，搜狗翻译仍然可以将其翻译成正确的英文。这些错误包括拼写错误，写错别字，误用同音异义词，误用词类，误用同义词，不合语法的表达，错误的词序语序等。本研究亦指出在中文教学领域进一步研究甚至应用机器翻译容错性的三个方向。

Keywords: Machine Translation, Sogou Translate, error tolerance, Chinese language teaching, self-editing

[1] This work was supported by a SAS Language Teaching Innovation Grant, Penn Language Center, University of Pennsylvania.

关键词：机器翻译、搜狗翻译、容错度、中文教学、自主纠错

1. Introduction

Computer-Assisted Language Learning (CALL) developed rapidly in the past two decades. Studies have shown that digital teaching tools, such as Machine Translation, Speech Synthesis, Speech Recognition, and Online Testing Solutions, have great potential to improve students' language proficiency levels in this mobile era. If used properly, Machine Translation can be used to enhance learners' awareness of self-correction on grammar, pronunciation, and word choice errors (Baker, 2013; Case, 2015; Correa, 2014; Groves & Mundt, 2015; Niño, 2009).

Machine Translation is not a new technology. The concept of Machine Translation emerged as early as the 1950s. The first successful large-scale application of Machine Translation was the Logos Machine Translation System, which translated English military manuals into Vietnamese during the Vietnam War in the 1970s. Unfortunately, restricted by unsatisfactory accuracy and expensive equipment, Machine Translation did not really impact foreign language teaching until Google Translate launched its phrase-based statistical machine translation (SMT) service in 2006 (Garcia & Pena, 2011; Somers, 2003).

The exact process is a patented secret but, roughly speaking, this is how Google's SMT-based algorithm works: Google collected billions of parallel documents that have been translated by human translators, such as official communications issued by the United Nations and the European Union, records of international tribunals, international company reports, and articles and books in bilingual forms. The principle is that the text one wants to translate may have been translated before by humans and can be found in their parallel text corpora. The algorithm of SMT analyzes the source language, detects the patterns, calculates the probabilities, and then generates the most likely translation in the target language (Benjamin, 2019; Grajales, 2015).

Although the accuracy of SMT has sometimes been questioned and mocked by human translators or the native speakers of the target languages, Google Translate evolved significantly along with the rapid development of information technology, network technology, and machine learning in the past decade. Google Translate soon developed into a neural network-powered machine translation engine, Neural Machine Translation (NMT), which is an end-to-end learning approach for automated translation. When it progressed into Google's Neural Machine Translation system (GNMT) in 2016, researchers claimed that:

> *Human evaluations show that GNMT has reduced translation errors by 60% compared to our previous phrase-based system on many pairs of languages: English ↔ French, English ↔ Spanish, and English ↔ Chinese. Additional experiments suggest the quality of the resulting translation system gets closer to that of average human translators* (Wu et al., 2016, p. 1).

The detailed algorithms of GNMT are beyond the scope of this study. Explaining their jargon, such as "LSTM network," "residual connections," "attention mechanism," "the decoder/encoder," "low-precision arithmetic," "inference computations," "beam search technique," "length-normalization procedure," or "coverage penalty" will not break open the black box of GNMT. But a statement made by Barak Turovsky (2016), the Product Lead of Google Translate, might shed some light on the mystery.

> *At a high level, the Neural system translates whole sentences at a time, rather than just piece by piece. It uses this broader context to help it figure out the most relevant translation, which it then rearranges and adjusts to be more like a human speaking with proper grammar. Since it's easier to understand each sentence, translated paragraphs and articles are a lot smoother and easier to read. And this is all possible because of end-to-end learning system built on Neural Machine Translation, which basically means that the system learns over time to create better, more natural translations* (para. 2).

Besides Google Translate, many technology companies have developed and launched their own Machine Translation services and products, such as Microsoft Translate Service, Siri Translate, Translate Facebook, DeepL Translator, Baidu Translate, Sogou Translate, and WeChat Translate. All of them not only offer website interfaces, but also mobile apps for Android and iOS. Thus, with the widespread use of smartphones, Machine Translation has become a double-edged sword in the field of foreign language education: while cheating is easier with the advent of the technology, it also has potential to become a valuable pedagogical tool.

While many foreign language instructors, as indicated in the following section, have claimed they successfully included Machine Translation into teaching practice, this study introduces a failed teaching design that aims to incorporate and evaluate Machine Translation as a means of Computer-Assisted Language Learning (CALL) for teaching in the Mandarin classroom. As with every major scientific innovation, there are countless failed projects that are equally important to the advancement of science. Although this teaching design did not achieve its original goals, it accidentally generated other insightful findings about the error tolerance of Machine Translation, which may enlighten future research and teaching practice.

2. Literature Review

Most of the previous research on Machine Translation in the field of foreign language teaching not surprisingly focuses on Google Translate, which is the most available Machine Translation tool for most languages. Related research mainly centers on the following three categories. The first investigates, of course, the accuracy of Google Translate. For example, Correa (2014) introduces some of the advantages and disadvantages of English ≒ Spanish translation in Google Translate. She states that Google Translate is "good at conjugating, spelling, basic agreement, and some common idioms"

(Correa, 2014, p. 7). But Google Translate often results in literal translation or unnatural writing that contains grammatical inaccuracies, discursive inaccuracies, and errors that humans do not commit. Google Translate is often "unable to account for cultural references and other extra-linguistic issues such as context, connotation, denotation or register" (Correa 2014, p. 7) and has difficulty with some idioms. Google Translate often does not translate misspelled words or proper nouns in the original text (they are just reproduced). Groves and Mundt (2015) point out that the accuracy of Google translations is close to the intermediate level of English learners when translating Malay and Chinese to English, and Google's accuracy improves over time.

The second category usually surveys the way in which foreign language learners use Machine Translation and the various (but more often diametrically opposed) opinions of foreign language instructors and learners toward the use of Machine Translation. Many studies have shown that students use Machine Translation even though their instructors explicitly prohibit it (Clifford, Merschel, & Munné, 2013; Correa, 2011, 2014; García & Pena, 2011; Niño, 2009). For example, Clifford, Merschel, and Munné (2013) use a questionnaire to survey instructors and students in the Department of Romance Studies at Duke University. They find that instructors and students have different attitudes toward Machine Translation. Although students are aware that Machine Translation is imperfect, more than 88% of respondents admit that they have used Machine Translation in their studies, and most students believe that Machine Translation helps them learn new languages. But most of the instructors they surveyed are skeptical about the effectiveness of Machine Translation. They believe that the emergence of Machine Translation interferes with traditional classroom teaching. Therefore, students are prohibited from using Machine Translation to finish assignments in the syllabus. It is worth noting that not all foreign language instructors object to the use of Machine Translation. For example, Niño (2009) and Baker (2013) show that there are still some foreign language instructors and students who maintain a positive attitude toward Machine Translation, especially instructors who are devoted to computer-assisted teaching and students in upper-level courses. They believe that the shortcomings of Machine Translation can even be used to improve students' understanding of the complexity of translation and language learning.

The third category discusses how to incorporate Machine Translation in foreign language classrooms to improve students' language proficiency. For instance, Correa (2014) states that since the output of Google Translate is far from perfect, Spanish instructors can have their students decipher the intended meaning and edit the output of Google Translate as a post-editing exercise. And as a pre-editing exercise, "students can pre-edit a text that they wrote in Spanish until they get an acceptable translation in English" (Correa, 2014, p. 11), which "can help with, among other issues, accent placement or gender assignment" (Correa, 2014, p. 12). In fact, Case (2015) summarizes that many studies demonstrate how to utilize the imperfect translation that is generated by Machine Translation for students to conduct pre-editing or error correction. Similar teaching designs can be found in Groves & Mundt (2015), Kliffer (2005), Niño (2008), Somers (2003), and in Zanettin (2009). This kind of research is more meaningful to foreign language teaching since it is directly related to classroom instruction.

Compared with other languages, studies on Machine Translation in Chinese language teaching and learning are rather inadequate in North America. Only Tian (2018) has published a study, which analyzes about 500 translation exercises that are collected from students' homework. He finds that students rely on Google Translate to do their homework in varying degrees. The accuracy of Google Translate is not unacceptable because Google Translate can do a better job than intermediate-level Chinese language learners, and it evolves very quickly. In terms of Chinese ≒ English translation, Sogou Translate [搜狗翻译] and Baidu Translate [百度翻译] are more accurate than Google Translate. This research shows that Machine Translation is accurate enough for students to use it as a shortcut to achieve a good grade.

3. An Interesting Failed Teaching Design

The failed teaching design introduced in this study is inspired by previous research that uses Machine Translate as a self-editing tool in foreign language classrooms. It is conducted in an intermediate (third-year) level Chinese language course at an Ivy League university, which uses *A New China* (《新的中国》) as its textbook. This design aims to help students gain experience in becoming self-assessors via Sogou Translate when they do their homework. Students are required to write an essay in Chinese first, then use Sogou Translate to translate their Chinese writings into English to check if their Chinese writings contain any obvious mistakes.[2] The intended objective is that since Sogou Translate is very accurate at least at the intermediate level, if the English translation looks incorrect to students, then students' original Chinese sentences are probably wrong. Students can go back to revise their Chinese essay until they get an acceptable English translation. Example 1 is instruction from one of the assignments. The writing sections of all twenty assignments for the entire academic year have been modified based on this design.

Example 1

Please write a short paragraph to introduce this image in Chinese, and then record your answer. Use as many new words or patterns as possible.

place+有+……; place+到处都 verb; 各种各样+的; 教+sb.+ verb /怎么 verb; ……，有的……，有的……，最常见的是……; 和/跟……有关的; 和/跟 A 比（起来）, B 更/比较 adj; 大大地+verb; 带着+noun; subj.（向 sb.）+提出（了）+noun; 改掉; 乱+verb; 随便+verb; 值得;

Requirements and Tips:

[2] This paper does not make a distinction between "error" and "mistake" from a linguistic viewpoint. It uses the them interchangeably.

A. Write down your answer in Chinese. (250 to 300 characters. **Tips: Please utilize Sogou Translate to check if your answer is correct**).
B. Record your answer. (Tips: You can utilize the embedded "read" function of Sogou Translate to practice your pronunciation).
C. Upload both your writing and your recording here.

However, this study finds that the above teaching design does not really achieve its intended goal. Students' Chinese writings still contain many obvious errors, such as ungrammatical expressions, incorrect word order, misspellings, malapropisms, homonyms, and typos, even though students used Sogou Translate to check their answers before they submitted their homework. It turns out that the error tolerance feature embedded in Machine Translation is so advanced that it can translate those mistakes made in students' Chinese writing into acceptable English. Thus, students would assume that their original Chinese writings are also correct. In fact, when designing similar pre-editing exercises via Google Translate in her Spanish course, Correa (2014) warns that as Machine Translate tools

> *are getting better at guessing missing accents and reassigning gender, the possibility exists that an ungrammatical input … could produce a grammatical output. For this reason, it may be wise to remind students that having a good translation is never a guarantee that the original text is also error-free (or vice versa)* (p. 12).

However, this warning is overlooked by many previous, similar studies, including the aforementioned teaching design, because this phenomenon has never been systematically examined before.

4. Error Tolerance of Sogou Translation

Although the teaching design does not achieve its intended goal, it generates equally or maybe more interesting findings. This study discovers that when translating Chinese to English, Sogou Translate is very tolerant of errors in the original Chinese text. In other words, even though students' Chinese writings may contain assorted mistakes that Chinese language learners can make, Sogou Translate can still "guess" the "correct" meanings and translate them into acceptable English.

Error tolerance is the design of things via various technologies "to be resilient to human error." "When a human error does occur, error-tolerance design gracefully detects and handles it" (Spacy, 2017, p. 1). One of the most common error tolerance features might be input correction. For example, Microsoft Word's AutoCorrect feature can automatically fix misspelled words or correct capitalization of words. And Siri's voice commands may accept different variations of the same command. In the case of Sogou Translate, the error tolerance feature allows the machine to accept imperfect Chinese inputs and translate them into English. This is a basic and necessary design because in real-life scenarios even native Chinese speakers cannot guarantee that they will speak flawless or grammatically correct Chinese sentences all the time. This research reveals that the error tolerance feature of

Sogou Translate can ignore at least the following four types of common mistakes made by Chinese language learners, which usually would not be acceptable by native Chinese speakers or Chinese language instructors.

4.1 Misspelling Errors

When Chinese language learners type Chinese characters, one common error is that they misspell the Pinyin and then choose the wrong characters (or correctly type the pinyin but still choose the wrong characters). However, this study shows that Sogou Translate can detect the mistyped words due to similar sounds and guess the intended meaning based on the context and then generate the correct English translation. For instance, although a student misspelled 将来 (*jianglai*, which means "in the future") as 经来 (*jinglai*, which is not a Chinese word at all) in Example 4.1.1, Sogou Translate still translated it into "in the future" correctly. And in Example 4.1.2, when a student misspelled 奇怪 (*qiguai*, which means "strange") into 吸怪 (*xiguai*, which is not a Chinese word at all), Sogou Translate also correctly translated it. Similar phenomenon can also be found in Example 4.1.3 where 开个开放 was correctly (and surprisingly) translated into "opening up."

Example 4.1.1
Student's incorrect sentence:
我觉得**经来**，美国人应该欢迎移动支付，要不然不能跟中国竞争。
Sogou Translate result (August 1st, 2019):
I think Americans should welcome mobile payment **in the future**, otherwise they cannot compete with China.

Example 4.1.2
Student's incorrect sentence:
视频里的一个人告诉了主持用现金可能会给北京人一个**吸怪**的印象。
Sogou Translate result (August 1st, 2019):
One of the people in the video told the host that using cash might give Beijingers a **strange** impression.

Example 4.1.3
Student's incorrect sentence:
开个开放以后中国不再是铁饭碗的城市。
Sogou Translate result (January 8, 2020):
After **opening up**, China is no longer an iron rice bowl city.

4.2 Vocabulary Errors

Messing up a word's part of speech or using the wrong synonyms are also two common mistakes that Chinese language learners may struggle with when learning Chinese vocabulary. When dealing with Chinese sentences with such errors, Sogou Translate does not conduct a simple word-for-word translation and then produce an English sentence with

errors; instead, it fixes the errors "behind the scenes" first and then generates the correct translation. For instance, in Example 4.2.1, a student incorrectly wrote 浪费花钱 (because 浪费 should be followed by a noun, not a verb), but Sogou Translate still correctly translated into "is wasting its money" not "is wasting spending money." In example 4.2.2, a student confused the meanings of 肯定 (which expresses the certainty of speculation) and 当然 (which expresses the certainty of a fact) in his Chinese sentence, but Sogou Translate still correctly translated it. In example 4.2.3, Sogou Translate first corrects a student's misspelling error (it should be 名字 not 明子) by ignoring that mistake, and then correctly translates the misused noun (意思 which means "meaning") into the verb "means." And in example 4.2.4, a student mixed up the meanings of 认为 ("to believe") and 想起 ("to think of, to remind") in her Chinese sentence, but Sogou Translate still correctly translated it.

Example 4.2.1
Student's incorrect sentence:
现在，美国的交通和经济很不好。他说这是因为美国**浪费花钱**在中东，所以现在没有现代化的交通工具。
Sogou Translate result (August 8, 2019):
At present, the transportation and economy in the United States are very bad. He said this is because the United States is **wasting its money** in the Middle East, so there is no modern means of transportation.

Example 4.2.2
Student's incorrect sentence:
我去过中国也看到了他们的铁路，机场，和别的基础设施建设，**肯定**比美国漂亮和有效了。
Sogou Translate result (January 8, 2020):
I have been to China and have seen their railways, airports and other infrastructure construction, which are **certainly** more beautiful and effective than those in the United States.

Example 4.2.3
Student's incorrect sentence:
"Rust Belt" 的意思是很老的工厂地方。一方面这个**明子意思**工厂地方， 一方面这**意思**钢城市。
Sogou Translate result (January 14, 2020):
Rust Belt means a very old factory place. On the one hand, it **means** a factory, on the other hand, it **means** a steel city.

Example 4.2.4
Student's incorrect sentence:
以前在美国这个名字很有名。 现在这个名字让我**认为**幽灵城市。
Sogou Translate result (January 14, 2020):
The name used to be very famous in America. Now the name makes me **think of** ghost city.

4.3 Grammar Errors

This study shows that Sogou Translate can also "overlook" some of the grammatical errors that exist in students' Chinese writings, guess the intended meaning, and then produce the correct English translation, which is a capability that many novice Chinese language instructors might not even possess. For instance, in Example 4.3.1, although a student wrote an incorrect comparative bi-structure (比字句) that is, frankly speaking, difficult to understand even for an experienced Chinese language instructor, Sogou Translate produced the correct English sentence nonetheless. Similar grammar corrections can also be found in Example 4.3.2, Example 4.3.3, and Example 4.3.4, in which the student mistakenly used lian-structure (连字句), rang-structure (让字句), put a comma before the conjunction word 和 [and], and connected two adjectives with the conjunction word 和 respectively.

Example 4.3.1
Student's incorrect sentence:
但是在美国，**人比手机多用电脑**。
Sogou Translate result (July 14, 2019):
But in America, **people use computers more than mobile phones**.

Example 4.3.2
Student's incorrect sentence:
比如说新加坡的机场**连有**电影院。
Sogou Translate result (January 8, 2020):
For example, Singapore's airport **has** cinemas.

Example 4.3.3
Student's incorrect sentence:
可是在政府之内总是有冲突，让政府**不**取得很大进步
Sogou Translate result (January 8, 2020):
However, there are always conflicts within the government, which **prevents** the government from making great progress.

Example 4.3.4
Student's incorrect sentence:
我去过中国也看到了他们的铁路，机场，**和**别的基础设施建设，肯定比美国漂亮**和**有效了
Sogou Translate result (January 8, 2020):
I have been to China and have seen their railways, airports **and** other infrastructure construction, which are certainly more beautiful **and** effective than those in the United States.

4.4 Word and Sentence Order Errors

Another typical mistake often made by Chinese language learners who are native English speakers is getting word and sentence order wrong, which is caused by the negative transfer between Chinese and English. For instance, they tend to put time and location at the end of the sentence, which is incorrect in Chinese. In Example 4.4.1, a student mistakenly put the location words 在中东 [in the Middle East] at the end of a Chinese sentence, but Sogou Translate rendered it into correct English. Such is the case with Example 4.4.2 about the time phrase.

Example 4.4.1
Student's incorrect sentence:
现在，美国的交通和经济很不好。他说这是因为美国浪费花钱**在中东**，所以现在没有现代化的交通工具。
Sogou Translate result (August 8, 2019):
At present, the transportation and economy in the United States are very bad. He said this is because the United States is wasting its money **in the Middle East**, so there is no modern means of transportation.

Example 4.4.2
Student's incorrect sentence:
美国政府把许多的工厂设在东北部**在 60 年代**。
Sogou Translate result (January 8, 2020):
The U.S. government set up many factories in the northeast **in the 1960s**.

In addition, when composing a Predicate Adjectives sentence (形容词谓语句), which means using an adjective as a verb/predicate in a sentence (e.g. 学中文很麻烦), influenced by their mother tongue, English speakers sometimes mistakenly put the adjectives at the beginning of the Chinese sentence. But once again, Sogou Translate can produce a beautiful and grammatical English sentence from a student's ill-formed Chinese sentence. Example 4.4.3 and Example 4.4.4 illustrate this phenomenon.

Example 4.4.3
Student's incorrect sentence:
这是因为很贵建设新的基础设施建设
Sogou Translate result (July 14, 2019):
This is because it is very expensive to build new infrastructure

Example 4.4.4
Student's incorrect sentence:
很麻烦得到允许在私人的土地建设。
Sogou Translate result (July 14, 2019):
It is very troublesome to get permission to build on private land.

Sogou Translate can also fix similar word order mistakes when students put professional titles before a person's name in Chinese, such as the Example 4.4.5.

Example 4.4.5
Student's incorrect sentence:
总统特朗普
Sogou Translate result (July 14, 2019):
President Trump

5. Discussion

Besides the error tolerance of Sogou Translate, this study also discovers two more interesting findings that are worth further discussion.

5.1 Error Tolerance of Google Translate

Tian (2018) shows that Sogou Translate is more accurate than Google Translate in terms of English ≒ Chinese translation. This research provides new evidence for this conclusion because the function of error tolerance of Sogou Translate is more powerful than Google Translate. In other words, when the original Chinese input is imperfect, the same as Sogou Translate, Google Translate can also sometimes guess the intended meaning and generate the correct English translations, but Sogou Translate does a better job than Google Translate. For instance, Google Translate fails to decipher the real meaning of 经来 (*jinglai*, which is not a Chinese word) in Example 5.1.1. It chooses not to translate it at all. Similar evidence that Sogou Translate is better than Google Translate can also be found in Example 5.1.2, Example 5.1.3, and Example 5.1.4.

Example 5.1.1
Student's incorrect sentence:
我觉得**经来**，美国人应该欢迎移动支付，要不然不能跟中国竞争。
Sogou Translate result (August 1, 2019):
I think Americans should welcome mobile payment **in the future**, otherwise they cannot compete with China.
Google Translate result, (January 11, 2020):
In my opinion, Americans should welcome mobile payments, or they cannot compete with China.

Example 5.1.2:
Student's incorrect sentence:
开个开放以后中国不再是铁饭碗的城市。
Sogou Translate result (January 8, 2020):
After **opening up**, China is no longer an iron rice bowl city.
Google Translate result, (January 11, 2020):
After **opening**, China is no longer a city of iron rice bowls.

Example 5.1.3:
Student's incorrect sentence:
可是在政府之内总是有冲突，让政府不取得很大进步
Sogou Translate result (January 8, 2020):
However, there are always conflicts within the government, **which prevents** the government from making great progress.
Google Translate result, (January 11, 2020):
However, there are always conflicts within the government, **so that** the government does not make much progress

Example 5.1.4:
Student's incorrect sentence:
很麻烦得到允许在私人的土地建设。
Sogou Translate result (July 14, 2019):
It is very troublesome to get permission to build on private land.
Google Translate result, (January 11, 2020):
Very trouble getting permission to build on private land.

5.2 "Two-Step Translation" Method

Trying to guess students' intended meaning and/or correcting students' ill-formed Chinese sentences (e.g. Example 4.3.1 "但是在美国，人比手机多用电脑") can sometimes be very challenging even for experienced Chinese language instructors. But the error tolerance feature of Sogou Translate may help instructors by ignoring those mistakes and getting the correct English translation. Instructors can simply use Sogou Translate to translate that English sentence back into Chinese and possibly get a correct (and often a better) Chinese sentence that the student wants to express. Examples 5.2.1, 5.2.2, and 5.2.3 show how a student's incorrect and hard-to-understand Chinese sentence is translated into a flawless Chinese sentence via this "two-step translation" method.

Example 5.2.1
Student's incorrect sentence:
但是在美国，人比手机多用电脑。
Sogou Translate result (from Chinese sentence to English) (January 8, 2020):
But in America, people use computers more than mobile phones.
Sogou Translate result (from English sentence to Chinese) (January 8, 2020):
但是在美国，人们使用电脑多于手机。

Example 5.2.2
Student's incorrect sentence:
第二个事的引述是川普在发言说的。
Sogou Translate result (from Chinese sentence to English) (January 8, 2020):
The second thing is quoted by Trump in his speech.

Sogou Translate result (from English sentence to Chinese) (January 8, 2020):
川普在演讲中引用了第二点。

Example 5.2.3
Student's incorrect sentence:
曾经是我完全成为天主教徒的时候，我这样做了。
Sogou Translate result (from Chinese sentence to English) (January 30, 2020):
It was when I became a full Catholic that I did so.
Sogou Translate result (from English sentence to Chinese) (January 30, 2020):
我是在成为一名正式天主教徒后才这样做的。

Moreover, this study also finds that the error tolerance feature of Sogou Translate probably works better on longer passages than shorter sentences, which means the more context provided, the more accurate Sogou Translate is. For instance, Example 5.2.4 is a student's essay explaining the term "Rust Belt" in Chinese, which is poorly composed and consists of many errors. When using Sogou Translate to translate a single ill-formed sentence in that essay, "一方面这个明子意思工厂地方," it gets an equally strange English sentence "on the one hand, this place where Akiko means factory." However, when translating the whole paragraph, Sogou Translate renders a highly acceptable English paragraph, and many errors, including that sentence, in the original text, are correctly translated. And Sogou Translate can generate a much better Chinese paragraph (only containing a few minor errors) by translating this English paragraph back into Chinese again.

Example 5.2.4
Student's incorrect sentence:
一方面这个明子意思 工厂地方。
Sogou Translate result of this single sentence (January 15, 2020):
On the one hand, this place where Akiko means factory.
Student's poorly composed essay:
"Rust Belt" 的意思是很老的工厂地方。**一方面这个明子意思工厂地方**,一方面这意思钢城市。这个现象发生**了**从十九世纪中到二十世纪中 在美国的北部。 美国用很大的地方来造成一个产业地方。当时美国政府送很多技术给这个城市来造成一个重要的产业城市。 以前在美国这个名字很有名。 现在这个名字让我认为幽灵城市。以前这地方是一个成功的产业所在和当时产业上最发展的所在这主要是因为这个所在 既有丰富的资源又有廉价的劳动力。美国政府吧许多的工厂设在东北部的美国。 现在这个地方减少就业机会。 很多人失业了。 "Rust Belt", 一方面对经济做出了贡献，但是另 一方面创造工人失业。 那里的人们现在不享受美好的过去。
Sogou Translate result of the whole paragraph (January 15, 2020):

Rust Belt means a very old factory place. **On the one hand, it means a factory**, on the other hand, it means a steel city. This phenomenon occurred in the north of the United States from the middle of the 19th century to the middle of the 20th century. The United States uses a large area to create an industrial area. At that time, the U.S. government sent many technologies to the city to create an important industrial city. The name used to be very famous in America. Now the name makes me think of ghost city. In the past, this place was the place where a successful industry was located and where the industry was most developed at that time, mainly because this place has rich resources and cheap labor. The U.S. government has many factories in the northeastern United States. This place is now reducing employment opportunities. Many people have lost their jobs. Rust Belt, on the one hand, has contributed to the economy, but on the other hand, has created unemployment among workers. The people there do not enjoy the beautiful past now.

Sogou Translate result (from English sentence to Chinese) (January 15, 2020):

锈带意味着一个非常古老的工厂。一方面，它意味着一个工厂，另一方面，它意味着一个钢铁城市。这种现象发生在 19 世纪中叶至 20 世纪中叶的美国北部。美国使用大面积来创建工业区。当时，美国政府向这座城市发送了许多技术，以创建一个重要的工业城市。这个名字过去在美国非常有名。现在这个名字让我想起了鬼城。过去，这个地方是一个成功产业的所在地，也是当时该产业最发达的地方，主要是因为这个地方资源丰富，劳动力便宜。美国政府在美国东北部有许多工厂。这个地方现在正在减少就业机会。许多人失业了。锈带一方面促进了经济发展，但另一方面也造成了工人失业。那里的人们现在不喜欢美好的过去。

6. Pedagogical Implications and Recommendations for Future Research

Science cannot succeed without failures (Firestein, 2015). The findings from the aforementioned unsuccessful teaching design do not indicate that Machine Translate cannot be applied in the Chinese language classroom but only mean that using Machine Translate to design pre-editing exercises needs more sophisticated consideration. Moreover, this study suggests that there are at least three possibilities to further investigate how to utilize Machine Translate in Chinese language teaching.

1) This study is a qualitative case study of only one intermediate-level Chinese language course, which only investigates the error tolerance in Sogou Translate when conducting Chinese- to-English translation. Student samples and their Chinese writing examples do not reflect any quantitative statistical significance. Future researchers can further explore the phenomena of error tolerance of Machine Translation by expanding their research samples to include students at different language proficiency levels and

investigating similar error tolerance phenomena that may exist in English-to-Chinese translation.

2) Most of the Machine Translation tools, such as Google Translate, Sogou Translate, and Baidu Translate, offer speech recognition features, which are also tolerant of pronunciation errors (e.g. different accents). This means that the speech recognition function can correctly recognize speech with some flaws (Ruiz, Bertoldi, & Federico, 2019). In teaching Chinese language, instructors often highlight the importance of the accuracy of students' pronunciation, which is mainly judged by ear. How accurate is accurate enough? There is a lack of direct and objective standards. Is "machine-acceptable" accuracy good enough? In the future, a study of error tolerance that uses Chinese language learners' speech as research samples will be very promising.

3) Future research may focus on investigating how to train Chinese language learners to become self-assessors and make progress on their writing proficiency without the instructor's involvement via the aforementioned "two-step translation" method. Computer engineers may even develop an "automatic composition revision" program via this editing technique.

7. Conclusion

Science constantly benefits from failures. Failures often provoke very unpredictable insights and compel researchers to look at a problem differently. This study is inspired by an unsuccessful teaching design that uses Sogou Translate as a self-editing technique. It finds that when conducting Chinese-to-English translation, Sogou Translate possesses a powerful error tolerance function. It can correctly translate erroneous Chinese sentences composed by intermediate level Chinese language learners into grammatical English sentences. These errors include misspelling, misusing homonyms, typos, misusing the part of speech of a word, misusing synonyms, ungrammatical phrases, and incorrect word/sentence order. The error tolerance function in Sogou Translate is more powerful than Google Translate, which proves again that Sogou Translate is a better tool than Google Translate in terms of Chinese ≒ English translation. This study also suggests three possible areas for future research on Machine Translation. These include: investigating error tolerant algorithms that exist in English to Chinese translation; examining to what extent Machine Translation can tolerate a Chinese language learner's imperfect pronunciation, and; exploring how to train students to gain experience in self-assessing their oral and written Chinese via Machine Translation. These opportunities might raise students' metalinguistic awareness and help prepare them for lifelong learning.

References

Baker, C. L. (2013). *Student and instructor perceptions of the use of online translators in English composition* (Order No. 1536057). Available from ProQuest Dissertations & Theses Global. (1352759991).

Benjamin, M. (2019). Teach you backwards: An in-depth study of Google Translate for 103 languages. Retrieved from https://teachyoubackwards.com

Case, M. (2015). Machine translation and the disruption of foreign language learning activities. *eLearning Papers,* (45), 4-16.

Clifford, J., Merschel, L., & Munné, J. (2013). Surveying the landscape: What is the role of machine translation in language learning? *The Acquisition of Second Languages and Innovative Pedagogies,* (10), 14. doi:10.7203/attic.10.2228

Correa, M. (2011). Academic dishonesty in the second language classroom: Instructors' perspective. *Modern Journal of Language Teaching Methods, 1*(1), 14.

Correa, M. (2014). Leaving the "peer" out of peer-editing: Online translators as a pedagogical tool in the Spanish as a second language classroom. *Latin American Journal of Content & Language Integrated Learning., 7*(1), 1-20. doi:10.5294/laclil.2014.7.1.1

Firestein, S. (2015). *Failure: Why science is so successful.* Oxford University Press.

Garcia, I., & Pena, M. I. (2011). Machine translation-assisted language learning: Writing for beginners. *Computer Assisted Language Learning, 24*(5), 471-487. doi:10.1080/09588221.2011.582687

Grajales, C. A. G. (2015). The statistics behind Google Translate. Retrieved from https://www.statisticsviews.com/details/feature/8065581/The-statistics-behind-Google-Translate.html

Groves, M., & Mundt, K. (2015). Friend or foe? Google Translate in language for academic purposes. *English for Specific Purposes, 37*, 112-121. doi:10.1016/j.esp.2014.09.001

Kliffer, M. (2005). *An experiment in MT post-editing by a class of intermediate/advanced French majors*. 160-165. Paper presented at the 10th Annual Conference of the European Association for Machine Translation, Budapest, Hungary.

Niño, A. (2008). Evaluating the use of machine translation post-editing in the foreign language class. *Computer Assisted Language Learning, 21*(1), 29-49. doi:10.1080/09588220701865482

Niño, A. (2009). Machine translation in foreign language learning: Language learners' and tutors' perceptions of its advantages and disadvantages. *ReCALL, 21*(2), 241-258. doi:10.1017/S0958344009000172

Ruiz, N., Bertoldi, N., & Federico, M. (2019). Assessing the tolerance of neural machine translation systems against speech recognition errors. Retrieved from https://arxiv.org/pdf/1904.10997v1.pdf. doi:10.21437/Interspeech.2017-1690

Somers, H. L. (2003). *Computers and translation: A translator's guide*. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Spacey, J. (2017). 10 Examples of error tolerance. Retrieved from https://simplicable.com/new/error-tolerance

Turovsky, B. (2016). Found in translation: More accurate, fluent sentences in Google Translate. Retrieved from https://www.blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/

Tian, Y. (2018). The challenge of machine translation to traditional translation homework in Chinese language learning. *Journal of Technology and Chinese Language Teaching, 9*(1), 78-95. Retrived from

http://www.tclt.us/journal/2018v9n1/tian.pdf. [田野. 2018. 机器翻译对中文传统翻译作业的挑战. 科技与中文教学, 9(1), 78-95.]

Wu, Y., Schuster, M., Chen, Z., Le, Q., Norouzi, M., Macherey, W., . . . Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. https://arxiv.org/abs/1609.08144

Zanettin, F. (2009). Corpus-based translation activities for language learners. *The Interpreter and Translator Trainer, 3*(2), 209-224. doi:10.1080/1750399X.2009.10798789

A Corpus-driven Contrastive Study of the Top 100 Content Words in English and Chinese (中英最常用 100 个内容单词：一项基于语料库的对比研究)

Kang, Tingting
(康婷婷)
Lafayette College
(拉斐特学院)
kangt@lafayette.edu

Luo, Han
(骆涵)
Lafayette College
(拉斐特学院)
luoh@lafayette.edu

Abstract: This corpus-driven study examines the construction and frequency distribution of the top 100 most frequently used content words in American English, Chinese, and American and Chinese first-year students' compositions. First, this paper presents the top 100 most frequently used content words in American English and Chinese from two comparable corpora, i.e., the Corpus of Contemporary American English (COCA) and the Chinese National Corpus (CNC). Second, the top 100 most frequently used content words were drawn from two specific corpora consisting of American and Chinese freshmen's English compositions. Linguistic similarities and differences in terms of the usage of content words across the two sets of comparable corpora were identified. For example, the results showed that people from both American and Chinese cultural backgrounds relied heavily on verbs and nouns in their languages. However, Chinese people tended to prefer using direction-oriented nouns and food-related words, which were nearly absent in the COCA and American freshmen's compositions. The cultural implications associated with the linguistic similarities and differences are discussed and pedagogical implications of the findings are also offered.

提要：本文研究了当代美国英语语料库（COCA）、中国国家语料库（CNC）以及大一中美学生写作语料库中前 100 个最常用内容词的比例和分布，并详细分析了常用内容词在不同文化背景和语境中的相似和不同之处。结果表明，美国英语和汉语都大量使用动词和名词，但中国人更倾向于使用与方向相关的名词和与食物相关的内容词，而这在 COCA 和美国新生的作文中却很罕见。本文还讨论了中英内容词使用中的相似性和差异性的相关文化内涵，以及研究结果对教学实践的指导性意义。

Keywords: Chinese language, content words, contrastive analysis, corpus, culture, English language

关键词：中文、内容词、对比研究、语料库、文化、英语

1. Introduction

Anthropologist-linguist Edward Sapir (1929, 1966) argued that language is the perfect symbolic system to describe the content of every culture. Different cultures tend to conceptualize the world differently and such differences are reflected in language forms (Lakoff & Johnson, 1980; Langacker, 1987, 1990; Talmy, 2000). Not surprisingly, Benjamin Whorf (1956) proposed that studies on not only vocabulary but also grammatical structures, such as word-classes, lexical word inflection, and derivation, provide a window into the mind of people from different cultures.

Inspired by these theories on the relationships among language, thinking, and culture, also known as the Sapir-Whorf hypothesis, a large number of pioneering contrastive studies have been conducted to examine color terminologies (e.g., Berlin & Kay, 1991), contrastive rhetoric (e.g., Kaplan, 1966), space concepts (e.g., Brown, 1994), and metaphors (e.g., Lakoff & Johnson, 1980) in different languages. As a matter of fact, the Sapir-Whorf hypothesis was a well-discussed topic in the 1990s and 2000s (Regier & Xu, 2017), and the pendulum has begun to swing back in recent years (e.g., Cibelli, Xu, Austerweil, Griffiths, & Regier, 2016; Kadarisman, 2015; Neuliep, 2017; Tseng, Carstensen, Regier, & Xu, 2016; Wang, 2016). Most of these recent studies have further supported the close relationship between language and culture (i.e., the Sapir-Whorf hypothesis) by using sophisticated data analysis tools.

Moreover, with the advancement of corpus linguistic research, scholars have been able to conduct contrastive studies and compare different languages through massive linguistic data obtained from various corpora (McEnery, Xiao, & Tono, 2006). Due to the extent of similarities among alphabetic languages, existing corpus-based contrastive research, to date, has primarily focused on comparing such languages as Spanish, French, and Dutch to English (e.g., Butler, 2008; Defrancq & De Sutter, 2010; Gladkova, 2010). This is because it is easier to find formal or translational equivalents between English and other alphabetic languages, which is not the case for a non-alphabetic, character-based language like Chinese. In contrast, contrastive studies in English and Chinese are relatively more difficult to carry out, and thus corpus-driven research in this area is still rather meager despite increasing attention from a number of scholars (McEnery, Xiao, & Mo, 2003; Xiao & McEnery, 2005; Chung, 2008; Qian & Piao, 2009; Chen, 2010).

In particular, no corpus-based research has been conducted to examine the most frequently used content words in comparable English and Chinese corpora. This study attempts to fill in this research gap by analyzing the top 100 most frequently used content words as displayed in four corpora, i.e., the Corpus of Contemporary American English (COCA)[1], the Chinese National Corpus (CNC)[2], and two specific corpora of American and Chinese freshmen's English compositions.

[1] c.f. http://corpus.byu.edu/coca/.
[2] c.f. http://www.cncorpus.org/

2. Literature Review

2.1 Corpus-based Contrastive Studies

In the past few decades, corpus-based analysis has become an important method for comparing different languages by utilizing "a large and principled collection of natural texts" (Biber, Conrad, & Reppen, 1998, p. 4). With the development of English language corpora, large corpora have also become available in other languages, such as Spanish, French, German, Portuguese, Japanese, and Chinese (McEnery, Xiao, & Tono, 2006). Consequently, a good number of corpus-based contrastive studies have emerged with the aim of comparing different languages.

2.2 Corpus-based Contrastive Studies among Alphabetic Languages

In order to make linguistic features comparable across languages, words used in different languages that share similar parts of speech, meanings, and forms are most frequently examined in the field of lexical corpus-based contrastive studies. Therefore, contrastive linguistic studies among alphabetic languages are relatively easier to carry out due to the extent of similarities among these languages.

Hudson (1994) compared the percentage of nouns (i.e. common nouns, proper nouns, and pronouns) in the Brown and LOB (Lancaster-Oslo-Bergen) corpora across different genres and proposed a striking constancy that the noun ratios in written English were always between 33% and 42%. With regard to other word-classes and other languages, he concluded a trend that among written English, written Swedish, New Testament Greek, written Welsh, spoken English, and children's English, there was a negative relationship between prepositions/common nouns and verbs/pronouns.

The scope of the research subjects in contrastive corpus-based lexical studies has become narrower in recent years. For example, Butler (2008) focused on the idea, concept, and notion in English and their formal equivalents in Spanish, examining their frequencies, adjectival collocations, and idiomatic contractions within two comparable corpora, the British National Corpus (World Edition) and the Corpus del Español. The results indicated that, overall, there was a striking similarity between the use of idea, concept, and notion in English and Spanish, with some minor differences. Molina-Plaza and de Gregorio-Godeo (2010) analyzed the stretched verb collections with *give* in English and *dar* in Spanish. With regard to the different structures of verb collocations, they provided substantial pedagogical applications for the L2 learners of English and Spanish.

In addition to English and Spanish, other alphabetic languages have also been compared through corpus analysis. For instance, Gladkova (2010) explored the linguistic and cultural variations of *sympathy*, *compassion* and *empathy* in English and their translational equivalents in Russian words *socˇuvstvie*, *sostradanie*, and *soperezˇivanie*. By applying the natural semantic meta-language research method into this study, the researcher successfully explained the semantic and conceptual differences of using these emotional words in two cultures. Defrancq and De Sutter (2010) compared the contingency

hedges of English *depend*, French *dépendre* and Dutch *afhangen*, *liggen* and *zien*, and discovered some consistent linguistic features for the contingency hedges in English, French, and Dutch.

2.3 Corpus-based Contrastive Studies between English and Chinese

Despite less conceivable linguistic similarity between Chinese and alphabetic languages, corpus-based contrastive studies between English and Chinese have also started to catch up. These studies have focused on analyzing aspect markers, tenses, passive constructions, kinship terms, word metaphors, and borrowed words (e.g., Chen, 2010; Chung, 2008; Qian & Piao, 2009; Xiao & McEnery, 2005; Yu, Yu, & Lee, 2017).

In one of the first of its kind, Xiao and McEnery (2005) used a corpus of written British English, the Freiburg-Lancaster-Oslo-Bergen corpus (FLOB), and the Lancaster Corpus of Mandarin Chinese (LCMC) (i.e., a comparable corpus to FLOB), to identify some of the basic grammatical features across English and Chinese. The results showed that "English is predominantly a tense language, whereas Chinese is exclusively an aspect language" (Xiao & McEnery, 2005, p. 1). In other words, English marks tense and aspect, but there are no morphology-like devices in Chinese to mark tense, number, gender, or case but only aspect markers (e.g. *–le*, *-guo*, *-zai*, and *-zhe*) to represent differences in time and situation. By using the Chinese-English matched corpora, FLOB and LCMC, they further explored the aspect-marking differences between Chinese and British English and how British English aspect marking was translated into Chinese.

Moreover, McEnery, Xiao, and Mo (2003) demonstrated the differences and similarities of aspect markers among not only Chinese and British English but also American English by using the LCMC, Frown, and FLOB corpora. Along the same lines, Xiao, McEnery, and Qian (2006) compared the characteristics of passive constructions in British English (*be/get* + past participle) and Mandarin Chinese (*bei*/*jiao*/*rang*/*gei*) by using data of the FLOB, LCMC, and two other spoken corpora. Their findings insightfully demonstrated that passive constructions are more frequently used in English than in Chinese due to the unpleasant and undesirable semantic prosody in Chinese passives.

In addition to explorations of grammatical differences, cultural influences in word selection have also been a focus of discussion in English and Chinese corpus-based contrastive studies. Influenced by Lakoff and Johnson's (1980) conceptual metaphor theory, a number of corpus-based contrastive studies have been conducted (e.g., Chung, 2008; Chen, 2010; Qian & Piao, 2009) to examine to what degree types of metaphors and their collocations can mirror cultural similarities and differences. For example, Chen (2010) concluded that Chinese is a very typical metaphorical language that tends to link physical experience with various subjective notions due to the influence of Confucianism and Taoism. Qian and Piao (2009) compared kinship taggers in LCMC and FLOB and revealed a scheme of annotating Chinese kinship into LCMC, but due to the complex meanings of some Chinese kinship terms, tagging them all in LCMC was problematic. The two researchers explained that this is because the concept of family has always been an

important aspect of Chinese life and thus the Chinese language has a much richer cluster of words describing family relations than English does.

Finally, researchers have also used various types of learner corpora to compare the nature of English and Chinese language. For example, Tardif, Fletcher, Liang, Zhang, Kaciroti, and Marchman (2008) analyzed babies' first 10 words in their first language among English-, Mandarin- and Cantonese-speaking children. The findings showed that Chinese babies obtained more words than American babies and especially more words within the category of people terms, which echoes Qian and Piao's (2009) study. Chan (2010) looked at Chinese learners' English written errors to elaborate on the difference between English and Chinese. Based on data gathered from 387 ESL learners' free writings, Chan argued that mother tongue influence was the most important factor that leads to ESL learners' written errors.

2.4 Research Gaps & Research Questions

As discussed previously, although corpus-based contrastive studies between English and Chinese have started to emerge, the number is still rather meager partly due to the linguistic distance between Chinese and other alphabetic languages. In addition, most existing corpus-driven contrastive research comparing Chinese and English has mainly focused on specific lexical or grammatical features such as aspect markers, tenses, passive constructions, and kinship terms. Moreover, the majority of previous English and Chinese corpus-based contrastive studies tended to only address the linguistic differences surrounding certain features across the two languages, without further exploring the potential cultural connotations indicated by certain linguistic forms. As Aijmer, Altenberg, and Johansson (1996) noted, comparable corpora could possibly increase our knowledge of cultural differences in many different ways. It thus might be interesting and worthwhile to delve into the cultural explanations behind linguistic differences.

Notwithstanding many differences between English and Chinese, one linguistic similarity between the two languages is that morphologically both of them are analytical languages where lexical meanings are expressed by using separate words, so comparisons of content words across these two languages are practicable. In addition, analyzing content words expands the research from focusing on a single lexical or grammatical feature to larger numbers of lexical items. Moreover, content words, which are often used to convey intended messages, may be more appropriate for interpreting culture compared to other closed word groups, such as prepositions, pronouns, articles, and so forth. However, no research, thus far, has examined and compared the most frequently used content words in English and Chinese by looking at massive linguistic data gathered from comparable corpora.

To fill in these research gaps, this study attempts to expand the scope of earlier English and Chinese corpus-based contrastive investigations by exploring the 100 most frequent content words in American English, Chinese, and American and Chinese freshmen English compositions. It also aims to enhance knowledge of the interrelationships between language and culture. More specifically, the research questions for this study are:

(1) What are the 100 most frequent content words (i.e., noun, verb, adjective, and adverb) in American English, Chinese, and American and Chinese freshmen English compositions? (2) How are the 100 most frequent content words in American English, Chinese, and American and Chinese freshmen English compositions different or alike? The cultural implications of the research results will also be discussed whenever possible.

3. Method

3.1 Corpora Used in This Research

In order to extract the top 100 most frequently used content words in English and Chinese, this study selected four corpora: i.e., the Corpus of Contemporary American English (COCA), the Chinese National Corpus (CNC), and the two specific corpora consisting of American and Chinese freshmen's English compositions. This is because these four chosen corpora represent two main cultures (American & Chinese) and two subcultures (American first-year students & Chinese first-year students at college). In addition, including four instead of two related corpora could likely enhance the validity of this research and provide more concrete evidence to elaborate upon the relationships between language and culture. It should be noted that Chinese first-year students' compositions were written in their second language, i.e., English rather than their native language, Chinese. It is interesting to see to what degree Chinese students' first year ESL writing is influenced by their L1 and Chinese culture.

The COCA was the first large and diverse corpus of American English, and the CNC was the largest balanced corpus of Chinese. The COCA contained more than 450 million words from 1990-2012; the CNC provided a balanced collection of texts from 1919-2012. The corpus has nearly 100 million characters out of about 50 million characters are tagged. Even though the two corpora were different in size, both of them included not only written but also spoken resources in America and China. Furthermore, these two corpora contain similar text types. In the COCA, "texts are evenly divided between spoken (20%), fiction (20%), popular magazines (20%), newspapers (20%), and academic journals (20%)" (Davies, 2009). In the CNC, all the resources are in Chinese, and approximately 50% of the texts come from arts and social sciences (politics and law: philosophy, politics, religion, and law; history: history, archaeology, and nationalities; society: sociology, psychology, linguists, education, literary theory, news, and folk-customs; economy: industrial economy, agricultural economy, political economy, and economics of finance and trade; art: music, essay, biography, reportage, fiction, and spoken; military and sports; living), 30% from natural science (mathematics, biochemistry, astronomical geography, maritime meteorology, agriculture and forestry, and medical), and 20% from general fields (administrative documents, statutes, judicial documents, business proclamations, protocol speech, and expository writing). Additionally, the COCA had a function to search for the most frequently used words by part of speech, and the CNC website had a most frequently used word list annotated with part of speech.

The data of American and Chinese freshmen compositions were collected from the freshmen composition classes at an American public university. This course was a required

course for all first-year students, and it offered special sections for the international students. Therefore, the texts of American freshmen compositions were selected from 34 American students who were in the regular freshmen composition classes, and the texts of Chinese freshmen compositions were selected from the 32 Chinese students in the international students' composition sections. The minimum language requirement for them to take this course was to have internet-based TOEFL scores higher than 59 or to be currently placed in Level 5 in the intensive English program at the university. As listed in their course syllabus, all the students in the freshmen composition classes needed to complete six writing projects. The two writing texts that we chose to use in this study were a short informational argumentative essay and an extended argumentative essay. In total, this study included 68 writing texts from Americans and 64 writing texts from Chinese ESL students.

3.2 Procedures

The preliminary work in this study was to build up the top 100 most frequently used content word lists in the COCA, the CNC, and the American and Chinese freshmen compositions. For the top 100 content word list in American English, as mentioned in the previous section, the COCA website provides a search function for extracting words by part of speech and ranking them based on frequency. Therefore, the researchers searched for the top 100 most frequent nouns, verbs, adjectives, and adverbs individually to include in a master list of the 400 most frequently used content words. Then, this 400-word list was ranked by total raw frequencies (TOT). The next step was to clean data to make sure all the words that appear in this list belong to the appropriate content word categories. To define the content word categories, this research used the definition of noun, verb, adjective, and adverb in English in Longman Student Grammar of Spoken and Written English (Biber, Johansson, Leech, Conrad, & Finegan, 2002). Additionally, with the aim of including the words that carry meaningful information, the nouns in this research refer to both common nouns and proper nouns, and the verbs include lexical verbs, primary verbs (e.g. *be* and *have*), and auxiliary verbs (e.g., *can* and *will*). Also, it is worth noting that to capture the complexity of natural language and how language is used in real life, the frequency list was generated based on word tokens.

In the CNC, there was no feature as in the COCA that can search for word frequencies among different word classes, but on the CNC website, a most frequent word list, annotated with parts of speech, was available. Therefore, the researchers extracted the top 100 most frequent content words, including nouns, verbs, adjectives, and adverbs, from the master list.

To extract the top 100 content words in American and Chinese freshmen compositions, a free concordance program, AntConc 3.3.5 (Anthony, 2012), was used to count the word token frequencies among the texts from American and Chinese freshmen compositions. Each word's part of speech was labeled along with the rules that had been used in the content word list in the COCA. For the words that could have more than one part of speech, the original sentences that contained the target words were checked to mark the frequencies under appropriate content word category.

After establishing these four top 100 content word lists, their content word ratios and distribution were analyzed. First, the number of nouns, verbs, adjectives, and adverbs in each top 100 content word list was calculated. Second, the 100 ranks were further divided into 10 frequency bands to analyze their distribution. In other words, lexical items from rank 1 to 10 belong to frequency band 1, words appear from rank 11 to 20 have been grouped into frequency band 2, and the like. By doing so, the distribution of different content word classes across the four lists can be visually represented. Additionally, in order to obtain the relationships of content word distribution among the four lists, the statistical method, Spearman's rho, was utilized. Spearman's rho can range in value from -1 to +1. An absolute value of one indicates a perfect linear relationship and a value of zero indicates the absence of a linear relationship.

4. Results

4.1 Construction of the Top 100 Most Frequently Used Content Words

The top 100 most frequently used content words in American English, Chinese, and American and Chinese freshmen compositions are presented in Lists 1, 2, 3, and 4 in Appendix A. Table 1 summarizes the construction of the top 100 content words across these four corpora. Even though the noun, verb, adjective, and adverb ratios varied, some patterns and trends could be discovered.

First, the construction of the four lists was taken up mostly by nouns and verbs (total number of verb = 117; noun = 90). Second, Chinese speakers in the CNC and Chinese freshmen compositions corpora tended to use more nouns than English speakers and fewer verbs and adverbs. Third, when comparing the COCA and the CNC to the freshman compositions, there were more nouns and adjectives and fewer verbs and adverbs used in compositions than in general communication.

Table 1. Constructions of content words across American English, Chinese, and American and Chinese freshmen compositions

Corpora	Noun	Verb	Adjective	Adverb	Total
COCA	22	46	8	24	100
CNC	33	38	11	18	100
American freshmen compositions	35	33	16	16	100
Chinese freshmen compositions	47	29	13	11	100
Total	90	117	35	58	400

4.2 Frequency Distribution of the Top 100 Most Frequently Used Content Words

The results of content word distribution based on the 10 frequency bands are included in Table 2. The four tables in Appendix B demonstrated the correlation of the content words among four corpora. The absolute Spearman's rho scores greater than 0.5

were the verb distribution between American English and American freshmen writing (0.845), American English and Chinese noun distribution (0.599), and American English and Chinese freshmen compositions noun distribution (0.567). These strong correlations seem to indicate that in addition to cultural contexts, rhetorical contexts (i.e., general communication vs academic writing) also played an important role in the observed distinct content words distributions regardless of L1 or cultural backgrounds.

Table 2. Distribution of content words
(FB=frequency band; v=noun; v=verb; adj=adjective; adv=adverb)

		American English				Chinese				American freshmen compositions				Chinese freshmen compositions			
FB	Rank	n	v	adj	ad	n	v	adj	adv	n	v	adj	adv	n	v	adj	adv
1	1-10	0	10	0	0	3	3	0	4	1	8	0	1	3	6	0	1
2	11-20	2	6	0	2	3	4	1	2	1	4	2	3	5	3	1	1
3	21-30	0	5	1	4	0	7	2	1	3	4	1	2	4	2	4	0
4	31-40	2	2	1	5	6	3	1	0	5	3	1	1	4	2	0	4
5	41-50	0	9	1	0	2	6	1	1	4	4	1	1	4	3	1	2
6	51-60	3	5	0	2	1	2	2	5	5	3	1	1	6	2	2	0
7	61-70	4	2	0	4	6	1	1	2	5	1	3	1	5	4	0	1
8	71-80	5	2	1	2	4	5	0	1	5	1	2	2	5	4	1	0
9	81-90	5	3	0	2	5	3	0	2	2	2	3	3	5	2	2	1
10	91-100	1	2	4	3	3	4	3	0	4	3	2	1	6	1	2	1

5. Discussion

Unlike previous contrastive corpus-based research, which focused on one type of part of speech (e.g., Hudson, 1994) or specific lexical items (e.g., Butler, 2008), results discovered from this study provided a macro level of analysis on the use of content words in American English and Chinese.

First, the results showed that people from both American and Chinese cultures rely heavily on verbs and nouns in their languages. Specifically, American English and Chinese noun distribution based on the 10 frequency bands were quite similar, which may indicate that the two cultures conceptualize many aspects of the world in similar ways (Yu, 1995, 1998; Yu, Yu, & Lee, 2017). For example, the concept of "time" appeared among the top 100 most frequently used content words in these four corpora, which included words like *time*, *years*, *year*, and *day* in List 1, 年[year], 时[time], 现在[now], 月[month], and 时间[time] in List 2, *time*, *year*, and *years* in List 3, and *time* in List 4. Although this result may be subject to other interpretations, it seems that the value of time tends to be universal in both American and Chinese cultures (Lakoff & Johnson, 1980; Yu, 1998, 2012). The value of time in both cultures is also reflected in a number of proverbs in English and Chinese (e.g., *A stitch in time saves nine. Time and tide wait for no man. Time flies. Time will tell.* 时间就是金钱[Time is money]. 光阴似箭[Time flies like an arrow]. 岁月不待人[Time waits for no man]. 时间检验真理[Time will tell the truth]).

Substantial similarities in terms of frequently used content words were also discovered between Chinese and American freshmen's compositions. For example, the noun *parents* appeared in the top 100 content words in both of the freshmen composition

corpora. Combined with a close examination of the contents of first-year compositions, this result seems to indicate that parents still tended to play an active and important role in both groups of students' first year of college life. Other common nouns across the two corpora included *education*, *school*, and *college*, which were related to their student status. Interestingly, both groups of students frequently used words related to playing computer games, such as *games*, *video*, and *game* in List 3 for American students and *internet*, *web*, and *game* in List 4 for Chinese students, reflecting American and Chinese freshmen's common interests as peers irrespective of their different cultural backgrounds. These seemingly intuitive findings provide convincing evidence for one of the basic assumptions of the cognitive linguistic framework that language reflects human conceptualizations of world experiences (Lakoff & Johnson, 1980, 1999; Langacker, 1987, 1990, 2008).

Since language is a result of conceptualization (Lakoff, 1987; Talmy, 2000) and different cultures tend to perceive the world and human life experiences differently to various degrees (Yu, 2009), this study, not surprisingly, discovered a number of differences among the top 100 most frequently used content words between American English and Chinese. For instance, Chinese speakers in the CNC tended to use more nouns than English speakers and fewer verbs and adverbs. Interestingly, even in Chinese students' first year ESL writing, we found similar patterns when compared with American students' first-year compositions. It seems that Chinese students' first year ESL writing tended to be influenced by their L1. This difference between American and Chinese speakers in findings may be even traced back to Chinese and American people's different ways of conceptualizing nouns and verbs (Shu, Zhang, & Zhang, 2019). For example, Shen (2019) insightfully pointed out that compared to the English word class construction in which English nouns and verbs are two separate categories, nouns in Chinese constitute a superordinate category that includes the verb category, a view echoed by Wang's (2019) analysis of the conceptual spatialization of actions or activities in Chinese. In other words, according to Shen (2019) and Wang (2019), a noun-verb distinction should not be assumed in the study of Chinese grammar. Shen (2019) also provided an elaborate discussion of the cognitive and philosophical roots of this difference between English and Chinese.

In spite of some interesting similarities as discussed above, the top 100 content words used in American and Chinese freshmen's compositions also showed a number of culture-specific differences. For example, American freshmen frequently used nouns such as *age*, *alcohol*, *sex*, *drug*, *war*, *violence*, *health*, and *(stem) cell*, which were almost absent in Chinese international students' writings. The Chinese students who were studying abroad also used a large number of exclusive nouns which were less present in American students' writings. Take Chinese students' exclusive nouns that appeared in List 4 as an example. They included words such as *internet*, *money*, *food*, *guns*, *phone*, *English*, *language*, *right*, *Chinese*, *law*, *penalty*, *abortion*, *(cosmetic/plastic) surgery*, *euthanasia*, and so on. These differences are not surprising and do not seem to be too hard to explain as both sets of exclusive nouns reflect those aspects of life heavily discussed or experienced in American and Chinese students' respective cultures. Similarly, these findings corroborated nicely with the cognitive-linguistic notion of the human conceptualization of life experiences; thus, the language forms used to reflect these conceptualizations are culturally shaped (Lakoff & Johnson, 1980, 1999; Yu, 2009, 2017).

A closer examination of the results on the top 100 content words revealed some other interesting differences in the process of corpora analysis. For example, Chinese people tended to prefer using direction-related nouns than American speakers, such as words 中[middle], 上[up], 里[inside], 后[back], 下[under], and 内[inside] in List 2. Another difference revolved around the concept of "food," reflecting the significance of food in the Chinese culture. More specifically, the distinct verb, 吃[eat], appeared in the top 100 content words in the CNC; *food* and *foods* appeared in the top 100 most frequently used content words in the Chinese freshmen corpus, whereas none of the top 100 content words in the two corpora produced by Americans was related to food.

The past decade has witnessed a growing body of research on the relationship between language, culture, and cognition (Chen, 2010; Maalej & Yu, 2011; Yu, 2009, 2017). For example, Yu (2009) examined the Chinese word 心[xīn] and provided a cognitive linguistic study of the Chinese conceptualization of the heart, revealing that the word 心[xīn] covers the meanings of both "heart" and "mind" as understood in English. He further traced the roots of the conception of the heart in ancient Chinese philosophy and traditional Chinese medicine, arguing that a holistic view that sees the heart as the center of both emotions and thought lies at the core of Chinese thinking and culture. Inspired by this line of research, the authors of this study speculate that the above-mentioned differences with regard to the preference of using direction-related nouns in Chinese as well as the dense usage of words associated with food may also be explained culturally. In ancient China, people in dread of nature attempted to use certain hypotheses to explain various phenomena. One common superstition is that *center* [中], *up* [上], *north* [北], and *left* [左] symbolize unchallengeable power and nobility. For example, *China* [中国] is literally translated as "central country;" *emperor* [皇上] is literally translated as "royal up," and in Chinese architecture, the most exalted people should live in the north of an architectural complex. Traditionally, on a formal occasion, males should stand to the left of females, and the host always lets the most honorable guests sit on his/her left. For example, there is a four-character saying in Chinese, 虚左以待 (Sima, 91BC), which means "emptying my left seat to wait for my honorable guest." Chinese people's preference of using direction-related nouns may be traced back to these cultural traditions in ancient China. Similarly, the significance of food in Chinese culture also has a long history and its importance to Chinese culture is extensively manifested in the Chinese language. As the Chinese saying goes, 民以食为天, food is valued as highly as the sky in people's lives. People even ask others whether they have eaten to greet each other in Chinese daily life. For example, 你吃了吗?[Did you eat?] is equivalent to "How are you doing?" in English. All these linguistic examples show that food is an essential part of Chinese people's life and an important aspect of Chinese culture.

6. Conclusion

Supported by the Sapir-Whorf hypothesis, this study was conducted with the aim of discovering the linguistic and cultural regularities of the top 100 content words across

American English, Chinese, and American and Chinese freshmen compositions. A corpus-based method was used to analyze the relationship between language and culture. The results demonstrated that both similarities and differences in terms of frequently used content words between American English and Chinese and between American and Chinese freshmen compositions may be attributed to the cultural contexts in which speakers experience and conceptualize the world. The findings provided potential evidence for the interrelated relationship among language, culture, and cognition (Lakoff & Johnson, 1980; Yu, 2009).

This study has a number of limitations. First, the traditional word class classification (i.e., noun, verb, adjective, adverb) was adopted to define and categorize content words. However, as some linguists have pointed out, the noun-verb distinction may not apply to the Chinese language (Shen, 2019; Wang, 2019). Second, only a very limited number of freshmen compositions on two argumentative essays assignments from Chinese and American students were used for analysis. The content words used in these writings may be affected by the specific topics and thus may not be representative of word usage in freshmen's academic writing. Third, this study focused on a macro-level analysis of the construction and frequency distribution of the top 100 content words across four corpora, which made in-depth linguistic analyses of specific lexical or grammatical items impossible. Therefore, future corpus-based contrastive studies may need to reconsider the appropriateness of traditional word class labels such as nouns and verbs when it comes to languages drastically different from English (e.g., Chinese). It is also important for future studies to include a wider range of corpora to enhance representativeness. For example, it would be interesting to examine American English and Chinese corpora representing a variety of registers or subcultures, such as spoken, popular magazines, newspapers, and academic journals. Moreover, future contrastive studies may also identify specific lexical or grammatical items that are comparable in English and Chinese and conduct more in-depth linguistic and cultural analysis. Finally, this study focuses on the comparison of American English and Mandarin. Future studies may expand research along this line to other languages, such as Japanese, Korean, and Spanish, providing further evidence for the relationship between language, culture, and conceptualization.

A number of pedagogical applications can be drawn from this study on account of the linguistic and cultural similarities and differences discovered across American English and Chinese. First of all, since both American Chinese and English relied heavily on nouns and verbs, American language learners of Chinese would benefit from early instruction on frequently used nouns and verbs in Chinese. Second, since language, culture, and cognition are interrelated, Chinese instructors may encourage learners to compare and contrast Chinese and their native language and guide the students to trace the roots of the identified linguistic similarities and differences to the levels of culture and cognition. For example, explaining the cultural implications behind linguistic phenomena such as the preference of using direction-related nouns or the high frequency of food-related words in Chinese may not only help the students understand the language better, but also can potentially boost their motivation in learning Chinese language and culture. Finally, teaching culture is now considered an integral part of language instruction. Various methods and strategies have been explored to enhance students' intercultural knowledge. This study shows that teaching

culture through analyzing language can be a viable and effective channel as pervasive evidence has been established for the relationship between language and culture. It is thus important to develop students' awareness of linguistic differences between Chinese and English and seek cultural explanations for such differences.

References

Aijmer, K., Altenberg, B., & Johansson, M. (1996). *Languages in contrast*: *Papers from a symposium on text-based cross-linguistic studies*. Lund, Sweden: Lund University Press.

Anthony, L. 2012. *AntConc 3.3.5.* Retrieved from http://www.antlab.sci.waseda.ac.jp/software.html

Berlin, B., & Kay, P. (1991). *Basic color terms: Their universality and evolution*. Berkeley, CA: University of California Press.

Biber, D., Conrad, S., and Reppen, R. 1998. Co*rpus linguistics: Investigating language structure and use.* Cambridge, UK: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (2002). *Longman student grammar of spoken and written English*. Harlow, England: Pearson Education Limited.

Brown, P. (1994). The INs and ONs of Tzeltal locative expressions: The semantics of static descriptions of location. *Linguistics, 32*, 743-790.

Butler, C. S. (2008). The very idea! A corpus-based comparison of idea, concept, and notion and their formal equivalents in Spanish. *Atlantis, 30*, 59-77.

Chan, A. (2010). Toward a taxonomy of written errors: Investigation into written errors of Hong Kong Cantonese ESL learners. *TESOL Quarterly, 44*, 295-319.

Chen, A. (2010). A conceptual understanding of bodily orientation in Mandarin: A quantitative corpus perspective. *Corpus Linguistics & Linguistic Theory, 6,* 1-28.

Chung, S. (2008). Cross-linguistic comparisons of the MARKET metaphors. *Corpus Linguistics & Linguistic Theory, 4*, 141-175.

Cibelli, E., Xu, Y., Austerweil, J. L., Griffiths, T. L., & Regier, T. (2016). The Sapir-Whorf hypothesis and probabilistic inference: Evidence from the domain of color. *PloS One, 11*(7), e0158725.

Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics, 14*, 159-190.

Defrancq, B., & De Sutter, G. (2010). Contingency hedges in Dutch, French and English: A corpus-based contrastive analysis of the language-internal and -external properties of English depend, French dépendre and Dutch afhangen, liggen and zien. *International Journal Of Corpus Linguistics, 15*, 183-213.

Gladkova, A. (2010). Sympathy, compassion, and empathy in English and Russian: A linguistic and cultural analysis. *Culture & Psychology, 16*, 267-285.

Hudson, R. (1994). About 37% of word-tokens are nouns. *Language, 70*, 331-339.

Kadarisman, A. E. (2015). Linguistic relativity, cultural relativity, and foreign language teaching. *TEFLIN Journal, 16*, 1-25.

Kaplan, R. B. (1966). Cultural thought patterns in intercultural education. *Language Learning, 16*, 1-20.
Lakoff, G. (1987). *Women, fire, and dangerous things.* Chicago, IL: University of Chicago Press.
Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York, NY: Basic Books.
Lakoff, G., & Johonson, M. (1980). *Metaphors we live by*. Chicago, IL: The University of Chicago Press.
Langacker, R. (1987). *Foundations of cognitive grammar* (Vol. 1). Stanford, CA: Stanford University Press.
Langacker, R. (1990). *Concept, image and symbol*. Berlin, Germany: Mouton de Gruyter.
Langacker, Ronald. (2008). *Cognitive grammar: A basic introduction.* Oxford, UK: Oxford University Press.
Maalej, Z. A., & Yu, N. (Eds.). (2011). *Embodiment via body parts: Studies from various languages and cultures* (Vol. 31). Amsterdam, the Netherlands and Philadelphia, PA: John Benjamins Publishing.
McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. New York, NY: Taylor & Francis.
McEnety, A., Xiao, Z., & Mo, L. (2003). Aspect marking in English and Chinese. *Literary and linguistic Computing, 18*, 361-378.
Molina-Plaza, S., & de Gregorio-Godeo, E. (2010). Stretched verb collocations with give: Their use and translation into Spanish using the BNC and CREA corpora. *Recall, 22*, 191-211.
Neuliep, J. W. (2017). Sapir-Whorf hypothesis. *The International Encyclopedia of Intercultural Communication*, 1-5.
Qian, Y., & Piao, S. (2009). The development of a semantic annotation scheme for Chinese kinship. *Corpora, 4*, 189-208.
Regier, T., & Xu, Y. (2017). The Sapir-Whorf hypothesis and inference under uncertainty. *Wiley Interdisciplinary Reviews: Cognitive Science, 8*, e1440.
Sapir, E. (1929). The study of language as a science. *Language*, *5*, 207-214.
Sapir, E. (1966). The status of linguistics and a science. In D. G. Mandelbaum (Ed.), *The selected writings of Edward Sapir in culture language and personality*. Berkeley, CA: University of California Press.
Shen, J. (2019). Types of negatives and the noun-verb distinction in English and Chinese. In D. Shu, H. Zhang, & L. Zhang (Eds.), *Cognitive linguistics and the study of Chinese* (pp. 121-155). Amsterdam, the Netherlands and Philadelphia, PA: John Benjamins Publishing.
Shu, D., Zhang, H., & Zhang, L. (Eds.) (2019*). Cognitive linguistics and the study of Chinese*. Amsterdam, the Netherlands and Philadelphia, PA: John Benjamins Publishing.
Sima, Q. (91BC). *Shi ji.* Retrieved from https://ctext.org/shiji/zhs. [司马迁. 91BC. 史记].
Talmy, L. (2000). *Toward a cognitive semantics* (Vol. II). Cambridge, MA: MIT Press.
Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & Marchman, V. A. (2008). Baby's first 10 words. *Developmental Psychology, 44*, 929–938.
Tseng, C., Carstensen, A. B., Regier, T., & Xu, Y. (2016). A computational investigation of the Sapir-Whorf hypothesis: The case of spatial relations. In A. Papafragou, D.

Grodner, & D. Mirman (Ed.), *Proceedings of the 38th annual conference of the cognitive science society* (pp. 2231-2236). Austin, TX: Cognitive Science Society.
Wang, W. (2019). The conceptual spatialization of actions or activities in Chinese: The Adjective + Verb construction. In D. Shu, H. Zhang, & L. Zhang (Eds.), *Cognitive linguistics and the study of Chinese* (pp. 157-182). Amsterdam, the Netherlands and Philadelphia, PA: John Benjamins Publishing.
Wang, Y. (2016). Do we know more about Whorf? *International Journal of Applied Linguistics and English Literature, 5*, 215-223.
Whorf, B. (1956). Grammatical categories. In J. B. Carroll (Ed.), *Language, thought and reality* (pp. 87-101). Cambridge, MA: MIT Press.
Xiao, R., McEnery, T., & Qian, Y. (2006). Passive constructions in English and Chinese: A corpus-based contrastive study. *Languages in Contrast, 6*, 109-149.
Xiao, Z., & McEnery, A. M. (2005). A corpus-based approach to tense and aspect in English-Chinese translation. In W. Pan, H. Fu, X. Luo, M. Chase, & J. Walls (Eds.), T*ranslation and contrastive studies* (pp. 114-157). Shanghai, China: Shanghai Foreign Language Education Press.
Yu, N. (1995). Metaphorical expression of anger and happiness in English and Chinese. *Metaphor and Symbolic Activity, 10*, 59–92.
Yu, N. (1998). *The contemporary theory of metaphor: A perspective from Chinese* (Vol. 1). Amsterdam, the Netherlands and Philadelphia, PA: John Benjamins Publishing.
Yu, N. (2009). *The Chinese HEART in a cognitive perspective: Culture, body, and language* (Vol. 12). Berlin, Germany and New York, NY: Walter de Gruyter.
Yu, N. (2012). The metaphorical orientation of time in Chinese. *Journal of Pragmatics*, *44*, 1335-1354.
Yu, N. (2017). Life as opera: A cultural metaphor in Chinese. In Farzad Sharifian (Ed.), *Advances in cultural linguistics* (pp. 65-87). Singapore: Springer.
Yu, N., Yu, L., & Lee, Y. C. (2017). Primary metaphors: Importance as size and weight in a comparative perspective. *Metaphor and Symbol*, *32*, 231-249.

Appendix A

List 1: Top100 most frequent content words in COCA
(v=noun; v=verb; adj=adjective; adv=adverb; POS=part of speech; TOT=total frequency)

Rank	Word	POS	TOT	Rank	Word	POS	TOT
1	is	v	4210980	51	year	n	355417
2	was	v	3384970	52	should	v	354854
3	be	v	2118761	53	still	adv	341596
4	are	v	2104489	54	got	v	341321
5	have	v	2095904	55	made	v	337895
6	do	v	1520663	56	world	n	337050
7	had	v	1507568	57	take	v	332656
8	were	v	1240986	58	day	n	329131
9	has	v	1192469	59	'll	v	326576
10	said	v	1100532	60	too	adv	322122
11	would	v	1057713	61	life	n	319753
12	can	v	996271	62	come	v	311036
13	been	v	900791	63	when	adv	310792
14	so	adv	894227	64	really	adv	308855
15	will	v	862031	65	man	n	305588
16	just	adv	789921	66	never	adv	301090
17	people	n	787379	67	being	v	294906
18	did	v	772448	68	most	adv	280882
19	know	v	729773	69	school	n	277227
20	time	n	722079	70	Mr	n	276925
21	could	v	711598	71	president	n	274418
22	now	adv	695534	72	why	adv	272605
23	're	v	680528	73	right	adv	268966
24	think	v	636774	74	things	n	254785
25	how	adv	627139	75	state	n	253571
26	then	adv	623932	76	children	n	253054
27	other	adj	621657	77	house	n	252421
28	more	adv	594410	78	let	v	251330
29	get	v	585015	79	American	adj	243007
30	says	v	570281	80	might	v	239682
31	also	adv	537124	81	women	n	237129
32	going	v	535002	82	again	adv	237035
33	years	n	527554	83	percent	n	226447
34	new	adj	492596	84	where	adv	225492
35	see	v	482363	85	students	n	224843
36	here	adv	475701	86	family	n	220769
37	well	adv	472664	87	look	v	219273
38	way	n	464767	88	put	v	215548
39	very	adv	445333	89	work	n	215544

40	only	adv	429745	90	found	v	212226
41	'm	v	428957	91	thing	n	211525
42	go	v	423453	92	today	adv	210795
43	say	v	422223	93	great	adj	209705
44	make	v	410072	94	big	adj	207732
45	good	adj	409930	95	always	adv	207114
46	want	v	375134	96	old	adj	206748
47	does	v	367909	97	used	v	203493
48	've	v	366671	98	high	adj	202617
49	may	v	363636	99	came	v	202288
50	'd	v	356231	100	all	adv	201195

List 2: Top100 most frequent content words in CNC
(v=noun; v=verb; adj=adjective; adv=adverb; POS=part of speech; TOT=total frequency)

Rank	Word	Translation	POS	TOT	Rank	Word	Translation	POS	TOT
1	是	verb be	v	118382	51	研究	study/research	v	8627
2	有	have	v	53522	52	更	more	adv	8602
3	也	also/too	adv	47034	53	已	already	adv	8600
4	不	no/not	adv	46950	54	却	but	adv	8253
5	就	about/at once	adv	44145	55	再	again	adv	8199
6	中	middle	n	40105	56	最	the most	adv	7957
7	说	say	v	35047	57	主要	main	adj	7879
8	上	up	n	34850	58	不同	different	adj	7822
9	都	all	adv	34261	59	不是	verb be not	v	7765
10	人	people	n	33915	60	中国	China	n	7721
11	要	demand/want	v	27324	61	关系	relation	n	7715
12	又	again	adv	25682	62	人们	people	n	7702
13	来	come	v	25410	63	才	just	adv	7634
14	年	year	n	21818	64	作用	affect	n	7548
15	到	arrive/go to/reach	v	21665	65	现在	now	n	7527
16	还	still	adv	20735	66	已经	already	adv	7358
17	大	big	adj	20050	67	重要	important	adj	7135
18	时	time	n	17995	68	我国	our country	n	6948
19	里	inside	n	17774	69	情况	circumstance	n	6922
20	发展	develop	v	17307	70	知道	know	v	6773
21	很	very	adv	16774	71	出	out	v	6742
22	可以	can/may	v	16724	72	社会主义	socialism	n	6711
23	使	make	v	16470	73	做	do/make	v	6708
24	去	go	v	14914	74	必须	must	adv	6701
25	没有	don’t have	v	14544	75	人民	people	n	6669
26	为	become	v	14499	76	成	become	v	6592
27	能	can	v	13781	77	走	go/walk	v	6589

28	看	look	v	13755	78	月	month	n	6530
29	小	small	adj	12687	79	方面	aspect	n	6518
30	多	many	adj	12028	80	需要	need	v	6511
31	后	back/behind	n	12026	81	便	therefore	adv	6351
32	会	will/can	v	11782	82	出来	come out	v	6335
33	好	good	adj	11743	83	发生	happen	v	6315
34	社会	society	n	11461	84	水	water	n	6283
35	进行	carry on	v	11085	85	过程	process	n	6231
36	问题	question/problem	n	10899	86	只	only	adv	6142
37	下	under	n	10737	87	科学	science	n	6101
38	如	like/as	v	10312	88	方法	way	n	6098
39	国家	country	n	10138	89	叫	call/shout/name	v	6041
40	工作	job	n	9655	90	内	inside	n	6002
41	起来	rise up	v	9588	91	技术	technology	n	5978
42	生产	produce	v	9419	92	一般	common	adj	5928
43	可	can/very/but	v	9361	93	许多	many	adj	5904
44	就是	even/quite right	adv	9195	94	吃	eat	v	5893
45	新	new	adj	9157	95	具有	have	v	5870
46	用	use	v	9062	96	高	tall/high	adj	5864
47	想	think	v	9028	97	形成	form	v	5850
48	不能	can’t	v	8834	98	影响	influence	v	5786
49	生活	life	n	8694	99	时间	time	n	5736
50	经济	economy	n	8680	100	事	thing	n	5731

List 3: Top100 most frequent content words in American freshmen compositions (v=noun; v=verb; adj=adjective; adv=adverb; POS=part of speech; TOT=total frequency)

Rank	Word	POS	TOT	Rank	Word	POS	TOT
1	is	v	1995	51	take	v	122
2	are	v	1288	52	young	adj	118
3	be	v	1026	53	get	v	117
4	as	adv	758	54	sex	n	117
5	have	v	726	55	however	adv	117
6	was	v	462	56	any	adj	116
7	has	v	441	57	years	n	116
8	would	v	437	58	United States	n	111
9	can	v	417	59	cells	n	110
10	many	adj	358	60	video	n	110
11	people	n	353	61	world	n	110
12	will	v	348	62	still	adv	108
13	all	adv	291	63	parents	n	106
14	been	v	271	64	used	v	106
15	were	v	253	65	different	adj	105

16	do	v	250	66	violent	adj	105
17	other	adj	247	67	believe	v	103
18	water	n	245	68	same	adj	103
19	also	adv	242	69	school	n	103
20	only	adv	237	70	where	adv	103
21	should	v	222	71	animals	n	101
22	being	v	219	72	dental	adj	99
23	when	adv	217	73	drug	n	99
24	women	n	210	74	education	n	99
25	how	adv	201	75	stem	n	99
26	research	n	193	76	then	adv	99
27	age	n	190	77	American	adj	98
28	drinking	v	182	78	very	adv	95
29	may	v	181	79	why	adv	95
30	could	v	172	80	new	adj	93
31	time	n	165	81	become	v	92
32	just	adv	163	82	health	n	92
33	life	n	157	83	men	n	92
34	make	v	152	84	fact	n	91
35	children	n	146	85	use	n	90
36	government	n	145	86	game	n	89
37	some	adj	145	87	now	adv	89
38	games	n	143	88	war	n	89
39	help	v	142	89	good	adj	88
40	had	v	141	90	made	v	88
41	way	n	137	91	go	v	87
42	well	adv	136	92	high	adj	87
43	violence	n	133	93	did	v	86
44	students	n	132	94	example	n	84
45	cell	n	130	95	money	n	84
46	does	v	130	96	society	n	83
47	need	v	130	97	laws	n	82
48	year	n	126	98	often	adv	82
49	alcohol	n	123	99	person	n	82
50	child	n	122	100	able	adj	81

List 4: Top100 most frequent content words in Chinese freshmen compositions
(v=noun; v=verb; adj=adjective; adv=adverb; POS=part of speech; TOT=total frequency)

Rank	Words	POS	TOT	Rank	Words	POS	TOT
1	is	v	1855	51	information	n	132
2	people	n	1192	52	new	adj	132
3	can	v	943	53	According	v	131
4	are	v	895	54	been	v	131
5	have	v	718	55	technology	n	131

6	be	v	557	56	school	n	129
7	will	v	519	57	human	n	128
8	children	n	495	58	cannot	v	127
9	death	n	446	59	learn	v	127
10	some	adj	437	60	food	n	127
11	time	n	425	61	English	n	126
12	students	n	404	62	need	v	125
13	penalty	n	383	63	development	n	125
14	also	adv	341	64	find	v	124
15	has	v	333	65	Web	n	124
16	should	v	309	66	women	n	124
17	life	n	299	67	how	adv	124
18	do	v	284	68	language	n	123
19	parents	n	275	69	jobs	n	122
20	many	adj	275	70	game	n	120
21	use	n	266	71	college	n	118
22	other	adj	254	72	important	adj	118
23	online	adj	232	73	believe	v	114
24	think	v	215	74	get	v	114
25	make	v	210	75	problems	n	111
26	education	n	209	76	could	v	110
27	China	n	202	77	lot	n	110
28	world	n	187	78	environment	n	109
29	good	adj	186	79	right	n	107
30	when	adv	182	80	cell	n	106
31	part	n	180	81	become	v	105
32	all	adv	174	82	better	adj	105
33	Internet	n	167	83	want	v	103
34	was	v	163	84	just	adv	102
35	abortion	n	163	85	some	adj	102
36	very	adv	162	86	foods	n	100
37	countries	n	156	87	Chinese	adj	100
38	way	n	147	88	euthanasia	n	99
39	different	adj	147	89	young	adj	99
40	society	n	146	90	guns	n	98
41	abroad	adv	144	91	phone	n	98
42	know	v	143	92	shows	v	97
43	may	v	142	93	law	n	97
44	surgery	n	142	94	oil	n	96
45	only	adv	142	95	why	adv	95
46	however	adv	138	96	study	v	93
47	government	n	137	97	global	adj	93
48	public	adj	136	98	country	n	92
49	help	v	134	99	crime	n	92
50	money	n	134	100	years	n	92

Appendix B

Table 1: Spearman's rho scores of noun distribution across the top 100 content word lists in English, Chinese, and American and Chinese freshmen compositions (n1=noun distribution in List 1; n2=noun distribution in List 2; n3=noun distribution in List 3; n4=noun distribution in List 4)

	n1	n2	n3	n4
n1		0.599	0.426	0.567
n2			0.288	0.016
n3				0.340

Table 2: Spearman's rho scores of verb distribution across the top 100 content word lists in English, Chinese, and American and Chinese freshmen compositions (v1=verb distribution in List 1; v2=verb distribution in List 2; v3=verb distribution in List 3; v4=verb distribution in List 4)

	v1	v2	v3	v4
v1		0.220	0.845	0.322
v2			0.373	-0.112
v3				0.071

Table 3: Spearman's rho scores of adjective distribution across the top 100 content word lists in English, Chinese, and American and Chinese freshmen compositions (adj1=adjective distribution in List 1; adj2=adjective distribution in List 2; adj3=adjective distribution in List 3; adj4=adjective distribution in List 4)

	adj1	adj2	adj3	adj4
adj1		0.411	-0.137	0.244
adj2			-0.117	0.495
adj3				0.106

Table 4: Spearman's rho scores of adverb distribution across the top 100 content word lists in English, Chinese, and American and Chinese freshmen compositions (adv1=adverb distribution in List 1; adv2=adverb distribution in List 2; adv3=adverb distribution in List 3; adv4=adverb distribution in List 4)

	adv1	adv2	adv3	adv4
adv1		-0.492	-0.057	0.027
adv2			0.085	-0.357
adv3				-0.328

Linguistic Feature Analysis of CEFR Labeling Reliability and Validity in Language Textbooks
(以語言特徵為本的教材分級難度及有效性之驗證)

Hong, Jia-Fei
(洪嘉馡)
National Taiwan Normal University
(國立臺灣師範大學)
jiafeihong@ntnu.edu.tw

Peng, Chun-Yi[1]
(彭駿逸)
Borough of Manhattan Community College, CUNY
(曼哈頓社區學院)
cpeng@bmcc.cuny.edu

Tseng, Hou-Chiang
(曾厚強)
National Taiwan Normal University
(國立臺灣師範大學)
ouartz99@ntnu.edu.tw

Sung, Yao-Ting
(宋曜廷)
National Taiwan Normal University
(國立臺灣師範大學)
sungtc@ntnu.edu.tw

Abstract: Despite the importance of grading language textbooks for teaching and learning, few studies have addressed the issues of reliability, validity, and efficiency of grading texts. This study adopted an automated textbook grading system to examine the grading consistency of five L2 Chinese textbook series labeled with CEFR difficulty levels. Twelve linguistic features were selected to represent the most crucial aspects of text readability: lexicon, semantics, syntax, and cohesion. Both the validity and reliability of grading assignments were tested between and within textbook series. The results suggested that 4 out of the selected 5 textbook series did not assign grading levels accurately reflective of actual text difficulty.

摘要：語言教科書的分級不管對於教學或是學習都是非常重要的一環，但是卻很少討論文本分級的可靠性、有效性的研究。本研究以自動分析教科書等級系統，檢測以 CEFR 作為標示難度的五套華語文教科書的等級一致性。在本研究中，選取了 12 種不同的語言特徵作為最具分級影響力的關鍵指標，分別取自詞彙類、語義類、語法類及篇章凝聚類等四大語言層面。本研究主要探究在不同教科書之間的分級一致性與相同教科書不同等級的分級，其有效性和可靠性。研究結果顯示，在本研究所選定的五套華語文教科書當中，四套華語文教科書沒有依照實際文本難度進行等級分級。

Keywords: Text Readability, reliability, validity, linguistic features, CEFR, language textbooks

[1] Peng, Chun-Yi is the corresponding author.

關鍵詞: 文本可讀性、信度、效度、語言特徵、歐洲共同語文參考標準、語言教科書

1. Introduction

The applications of machine learning have become increasingly important across various disciplines, such as health care (Caruana et al., 2015), education (Chang & Sung, 2019; Hsu et al., 2018; Lin et al., 2019; Lu & Chen, 2019; Lee et al., 2016), and speech recognition (Chen & Hsu, 2019). A crucial application of machine learning in education is the assigning of grade levels to textbooks for adaptive learning (Tseng et al., 2019). With correctly graded materials, educators can better select or even edit existing resources to cater to learners' changing proficiency levels. For learners, the use of appropriately graded materials is also important. It assists them in identifying their proficiency levels, allows them to check their progress, and enhances their learning efficiency. Thus, a standardized text grading system is beneficial for both educators and learners.

L2 Chinese textbooks prove to be a useful example of the necessity for a standardized text grading system. Although many L2 Chinese textbook materials are graded in terms of standards such as the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR; Council of Europe, 2001), the assignment of grade levels is mostly, if not all, carried out by developers based on their own expertise and professional experiences. The variation in expertise and professional experiences runs the risk of inconsistency when standards such as CEFR are applied to language materials grading. That is, the same materials may be assigned by different developers to different difficulty levels within the same set of standards. Such inconsistency creates potential problems when those materials are adopted for teaching and learning. In order to ensure the accuracy and consistency in textbook grading, many highly experienced language educators must be involved in the compilation and grading process. This is often time-consuming and labor-intensive. Furthermore, it can prove to be rather difficult to reach a consensus among educators on a consistent grading scheme.

To this end, this study introduces the use of a standardized textbook grading protocol proposed by Sung et al. (2015b) as a tool for CFL textbook grading. More specifically, we use the CRIE-CFL system, a tool based on Sung et al.'s (2015b) grading protocol, to analyze 5 textbook series that have been graded manually by their developers. Sung et al.'s (2015b) model and the CRIE-CFL system have been shown to be a valid tool in language materials grading. By comparing readings from the tool and the grading levels assigned manually by their developers both within and across those five textbook series, we hope to illustrate the usefulness of such a tool in measuring the accuracy and maintaining consistency of manual gradings to the actual difficulty of language learning materials included in the textbooks.

2. Overview of the CEFR

The CEFR was created by the Council of Europe in 2001 with the aim of providing a unified framework for the teaching, learning, and assessment of all of the languages used within Europe (Fulcher, 2004). The principles of the CEFR framework implies that the 'can-do' statements are unitarily understandable and can be interpreted in only one way which will be the same for everyone in every European country (Vinther, 2013). It provides a set of guidelines for language teaching materials and language evaluation, as well as a point of reference for grading learner levels in order to reduce the barriers of interaction between people speaking different languages within different European countries (Council of Europe, 2001; Little, 2006, 2007). The CEFR has had a profound influence on the design of teaching materials, curriculum planning, and language proficiency testing in several European countries (Hulstijn, 2007). Its role in Europe has evolved from a supportive education tool to a tool used to shape language education policies (Bonnet, 2007; Fulcher, 2007).

The CEFR is a detailed and complex system for evaluating language proficiency levels. It uses "horizontal" and "vertical" dimensions to describe a particular learner's ability to communicate. The horizontal dimension provides a general description of communicative language competency; it consists of several scales that describe various language activities that a learner may encounter, such as context, topic, and purpose (Council of Europe, 2001; Hulstijn, Aldersen, & Schoonen, 2010). The vertical dimension categorizes the language proficiency (i.e. statements of learning objectives) of a learner by using six levels which are organized into three divisions: A1 and A2 (basic users), B1 and B2 (independent users), and C1 and C2 (proficient users). The vertical dimension has various practical applications such as curriculum design and the creation of qualifying examinations (Council of Europe, 2001). The combination of these two dimensions, and their varying definitions, results in communicative language being understood as an amalgamation of the scope of language use (horizontal dimension) and the manifestation of language proficiency (vertical dimension) (Hulstijn et al., 2010). Using both of these dimensions, the CEFR is able to describe and outline the expected reading, listening, speaking, and writing abilities of a learner at each level of proficiency.

The CEFR was officially published in 2001 in both English and French (Little, 2006). In November 2001 a European Union Council resolution recommended using the CEFR as the common system for the recognition of language proficiency. Subsequently, the CEFR became an important system for providing criteria for the validation of foreign language abilities, including Chinese teaching (Figueras, 2012) and second-language teaching in many regions (Hulstijn, Aldersen, & Schoonen, 2010). It also provides reference indicators for second-language learning, assists in the compilation of teaching materials, and supports the assessment of language proficiency (Little, 2006).

Beyond using two distinct dimensions to describe communicative language competency, the CEFR deliberately avoids describing language proficiency in theoretical terms. Instead the CEFR provides general descriptions; this means that its scales for the scope of language use are short, easy to use, and applicable to many different languages (Little, 2007). In addition to providing guidance for the appropriate level of a teaching

material or text, the CEFR can also be used to label the difficulty level of language assessments. The ease of use and applicability of the CEFR labeling system to different languages as a common standard has resulted in it being used for defining the difficulty level of language tests developed by various institutions (Alderson, 2007).

3. Feature-based Tools for Grading L2 Teaching Materials

Readability research can be a useful point of departure for L2 text grading. Readability is often understood as *text comprehensibility*, or how well a text can be comprehended by the reader (Klare, 1984). Methods for measuring text readability have long been widely available for alphabetic languages (Dale & Chall, 1948), as are the readability formulas for grading textbooks (Faison, 1951). Traditional readability research assumes that the difficulty level of a text is determined by its semantics and syntax (Collins-Thompson, 2014), and that it is possible to create formulas to predict the difficulty level of a given text based on those two elements. For example, Flesch–Kincaid (1948) readability tests for English, which make use of the number of syllables and words in a sentence to assign grades to English books.

Recently, however, researchers have begun to challenge the way that text difficulty is determined. Collins-Thompson (2014), for instance, points out that only a few shallow linguistic features are actually used in order to estimate text difficulty; these features do not reflect the actual reading process and overly simplify the assessment of text difficulty. As a result, various attempts have been made to approximate the complex process of text understanding (Graesser, McNamara, Louwerse, & Cai, 2004; McNamara, Louwerse, McCarthy, & Graesser, 2010), such as exploring the relationship between text cohesion indicators and other indicators (Benjamin, 2012), and using computational cohesion and coherence metrics (Crossley & McNamara, 2008; Crossley, Louwerse, McCarthy, & McNamara, 2007; Graesser et al., 2004).

For non-alphabetic languages, Sung et al. (2013; 2015a) developed multi-level Chinese readability models, taking into account features at lexical, syntactic, semantic, and cohesive levels. These models were subsequently extended to determine the difficulty levels of L2 Chinese texts (CRIE-CFL readability model) (Sung et al., 2015b). Sung et al. (2015b) proposed a CRIE-CFL system combining the CEFR grading criteria with the readability assessment methods trained by the support-vector-machine (SVM) technology (Vapnik, 1995). The training data of CRIE-CFL consist of 1,578 texts from 28 CFL textbook series published across 23 countries and regions such as the United Kingdom, Germany, France, Italy, Australia, Mainland China, and Taiwan, etc. where the CEFR standard is often used for learning material grading purposes. The CEFR-graded materials include *Practical Audio-Visual Chinese* (2nd Edition)[2], *Far East Everyday Chinese*[3], and *New Practical Chinese Reader*[4], etc. In order to ascertain the appropriate CEFR level for each text in the training data, expert educators, who had been teaching

[2] National Taiwan Normal University (Eds). 2008. *Practical audio-visual Chinese* (2nd Edition). Taipei: Cheng Chung Bookstore.
[3] Yeh, T. M. (Ed). 2008. *Far East everyday Chinese.* Taipei: Far East Book Company.
[4] Liu, X. (2007). *New practical Chinese reader.* Beijing: Beijing Language and Culture University Press.

CFL for more than 10 years and were familiar with CEFR level grading, read the selected materials and then assigned the corresponding CEFR level. Information on each level is provided in Table 1.

Table 1 Information on Each Level of CRIE-CFL Built-in Texts

CEFR level	No. of texts	No. of characters	No. of characters, mean (*SD*)	No. of words, mean (*SD*)
A1	155	9888	64 (37)	45 (25)
A2	337	48060	143 (64)	101 (45)
B1	470	145006	309 (198)	211 (139)
B2	345	165807	481 (221)	322 (152)
C1	190	122025	642 (358)	425 (253)
C2	81	121900	1505 (978)	1019 (695)
Total	1578	612686	388 (432)	263 (297)

The CRIE-CFL system takes into consideration a variety of text features so that the model is not biased toward a small number of features (McNamara et al., 2002). Sung et al. (2015b) utilized the F-score (Chen & Lin, 2006; Chang & Lin, 2008; Ding, 2009), a commonly used algorithm for selecting relevant features, to determine which features would improve the readability model most significantly. The F-score allows for the predicting power of the model. According to Chen & Lin (2006), the larger the F-score is, the more likely this feature is discriminative. In Sung et al.'s (2015b) study, each text is represented by a series of feature values based on textual complexity. Ideally, texts within the same level should have similar feature values. The algorithm compares those values between and within levels (e.g. the CEFR A1 vs. A2). Features with a high F-score are more useful for assigning grade level.

Eventually, Sung, et al. (2015b) verified the performance of the CRIE-CFL system, which yielded exact-level, adjacent-level, and division accuracies of 75%, 99%, and 90%, respectively. In addition, a trend analysis showed that the values of the 30 indicators that determine the CFL text difficulty level changed significantly with the CEFR levels. This means that the linguistic features data in the current CRIE-CFL corpus have rational validity; moreover, as discussed in Sung et al. (2015b), since the selection of teaching materials for CRIE-CFL is representative of texts from all levels, the quantitative features are valid. The CRIE-CFL itself can, therefore, be considered an anchored teaching material and the data of its various linguistic features can be used as a benchmark for comparison with other teaching materials (Sung, et al., 2015b).

Sung et al. (2016) made use of protocols presented in Sung, et al. (2015a; 2015b) and released a web-based CRIE system[5]. It provides four subsystems: CRIE (Analysis of texts written for native Chinese readers), CRIE-CFL (Analysis of texts written for

[5] c.f. http://www.chinesereadability.net/CRIE/index.aspx?LANG=CHT

learners of Chinese), CRIE-DK (Assesses the knowledge content levels of texts), and WECAn & HanParser (Word segmentation and part-of-speech tagging tools). The CRIE also provided 82 multilevel linguistic features, segmentation, syntactic parsing, and feature extraction. In this study the CRIE-CFL system is applied to examine the grading of five CFL textbook series.

4. Methods

4.1 Instruments

This study utilized the readability analysis system CRIE-CFL developed by Sung et al. (2016) to analyze textbook content. The CRIE-CFL automatically captures the linguistic features of Chinese texts and provides an objective numeric value for each linguistic feature found in the texts. In this study, the CRIE-CFL system is used to obtain quantitative values for each linguistic feature from five L2 Chinese textbook series to examine the consistency of their CEFR grading by their developers.

4.2 Materials for Analysis

In order to maintain consistency in the comparison and interpretation of result data using the CRIE-CFL system, this study selected five CFL textbook series (c.f., Table 2) that have been assigned CEFR proficiency levels. Three of the textbook series were published in the Greater China region because the Chinese-speaking area offers a wide range of CFL materials to select from. The rest two were selected from Europe (i.e., France and Germany) where CEFR was established. The fact that the five textbook series are from different publishers ensures that they are not subject to similar publishing guidelines, which might not represent the actual developments of CFL textbooks in different regions.

As a first attempt to compare grading consistency among different CFL textbooks using the CRIE-CFL system, this study was limited to those where manual CEFR grading by their developers are readily available. It did not include popular textbooks from other regions such as the Integrated Chinese series in North America (Li, Wen, & Xie, 2012), though future research could extend to include textbooks from more regions.

The CEFR scale (A1, A2, B1, B2, C1, and C2) contains six levels; however, Chinese language teaching materials at level C are very rarely seen on the textbook market, and textbook publishers do not tend to give classification to such materials. In addition, texts in *Chinesisch ohne Mühe* (hereafter *Chinesisch*; published in Germany) and *Le chinois par boules de neige* (hereafter *Boules de neige*; published in France) are labeled exclusively with levels B1–B2.

Table 2 Number of Texts at Each Level in the Five Textbooks

Place of publication	Textbook title	Author indications of CEFR levels						Total no. of texts
		A1	A2	B1	B2	C1	C2	
Mainland China	Road to Success[6] 成功之路	167	42	41	24	37	53	364
Taiwan	New Modern Chinese[7] 新時代華語	8	12	18	2	0	0	40
Mainland China	Practical Chinese[8] 实用中文	44	43	49	31	0	0	167
France	Le chinois par boules de neige[9] 雪球	0	16	19		0	0	35
Germany	Chinesisch ohne Mühe[10] 漢語	49		56		0	0	105

4.3 Procedure

4.3.1 Selection of Linguistic Features

The CRIE-CFL has developed 30 linguistic features, which can be divided into four categories: lexicon, semantics, syntax, and cohesion (Sung et al., 2015b). This study selects 12 linguistic features using F-score and Trend Analysis F Value to reflect either the key concepts in the CEFR proficiency level or the unique nature of the Chinese language (cf., Table 3) (Sung et al., 2015b). These 12 features represent the most influential aspects of each of the four categories and are used to determine if the difficulty levels of the five Chinese language textbooks are consistent.

First, lexical category is used to measure the complexity of texts and hence text difficulty. The CEFR scale for overall reading comprehension (Council of Europe, 2001)

[6] Editors of the Road to Success sereis (Ed). (2008-2014). *Road to success (成功之路)*. Beijing Language and Culture University Press. China: Beijing.

[7] NTNU Extension School of Continuing Education (Ed). (2012). *New modern Chinese (新時代華語)*. NTNU Extension School of Continuing Education. Taiwan: Taipei.

[8] Chinese Time (Ed). (2009). *Practical Chinese (实用中文)*. East China Normal University Press. China: Shanghai.

[9] Bellassen, J., & Liu, J. L. (2011). *Le chinois par boules de neige (Acces raisonne a la lecture du chinois)（雪球）*. Scérén Cndp-crdp. France: Chasseneuil-du-Poitou.

Bellassen, J., & Liu, J. L. (2012). *Le chinois par boules de neige (Niveau elementaire)（雪球）*. Scérén Cndp-crdp . France: Chasseneuil-du-Poitou.

[10] Kantor P. (2004). *Assimil Pack Chinesisch Ohne Mühe（漢語）*. ASSiMiL GmbH. *Volume 1*. Germany: Köln.

Kantor P. (2006). *Assimil Pack Chinesisch Ohne Mühe（漢語）*. ASSiMiL GmbH. *Volume 2*. Germany: Köln.

states that A1- and A2-level learners can understand short texts, whereas C1-level learners can understand detailed, long texts, and C2-level learners can understand a wide range of long texts. Numerical counts of characters and words are used to measure text length as shown in Table 3.

Table 3 Linguistic Features Selected in this Study

	Feature	Definition
Lexical Category	characters	total number of characters
	high-level words	total number of words listed by the 8,000 Chinese Words[11] as being in the vantage or effective operational proficiency levels
	two-character words	number of two-character words
Semantic Category	content words	number of content words
	sentences with complex semantic categories	number of sentences with a number of semantic categories
	complex semantic categories	number of semantic categories from sentences with complex semantic categories
Syntactic Category	average sentence length	average number of words in a sentence
	simple sentence ratio	the number of simple sentences divided by the total number of sentences
	sentences with a complex structure	the number of sentences containing conjunctions and subordinators
Cohesive Category	conjunctions	number of conjunctions
	positive conjunctions	number of conjunctions with positive meanings
	negative conjunctions	number of conjunctions with negative meanings

As seen in Table 3, in addition to the number of characters in a text, the count of two-character words is also applied as a measure of text length. The main component of Chinese is two-character words (Duanmu, 1999; He & Li, 1987). In order to distinguish a

[11] The 8,000 Chinese Words can be found at https://www.sc-top.org.tw/chinese/download.php.

text's difficulty, this study used the number of *characters* (e.g., *shū* 書 book) and *two-character words* (e.g., *zhī jì* 之際 at the time of) in a text as an indicator of length. Moreover, the CEFR scale for overall reading comprehension states that A2-level learners can understand the *highest frequency vocabulary, high-frequency everyday or job-related language.* B2-level learners can understand *low-frequency idioms*. Therefore, a learner who is at a higher CEFR level can understand harder words, at low frequencies. Accordingly, this study incorporated the number of *high-level words* (e.g. *bǎo cún* 保存 preserve) as an indicator of word difficulty in identifying text difficulty. Note that in this study, all the features in Table 3 are calculated independently. Therefore, some words would be counted more than once. For example, *bǎi tuō* (擺脫, to break away from) was counted both as a high-level word and as a two-character word.

The second measure of text difficulty adopted in this study is semantics. To account for semantics, this study selected three semantic features to examine text complexity: 1) the number of *content words* (e.g., *lán qiú* 籃球 basketball), 2) the number of *sentences with complex semantic categories*, and 3) the number of *complex semantic categories*. Content words are words with independent lexical meanings. More content words within a text represent more concepts in that text and thus higher complexity. According to Hong et al. (2016), semantic categories is defined as the number of meanings in a single word. Words with multiple meanings are more likely to cause semantic ambiguity (e.g., *chī bīng qí lín* 吃冰淇淋, which means either 'eat ice cream' or 'look at an eye candy' when used in Taiwan) at the sentence level. In addition, words with larger numbers of semantic categories usually generate more significant lexical semantic variations (e.g., *dǎ diàn huà* 打電話 call someone/ hit the phone). It has been reported that a higher number of semantic categories also increases sentence difficulty (Cheng, 2005), and therefore was included in this study. Furthermore, polysemous words have more lexical meanings which contribute to lexical ambiguities and increase complexity. More semantic categories also imply more complex lexical meanings.

The third category of measuring text difficulty is syntax. Two crucial components of text complexity are sentence length and sentence structure. For example, simple sentences are semantically independent syntactic units that consist of a subject and a predicate. Complex sentences are formed by combining two or more simple sentences (Hong, Sung, Tseng, Chang, & Chen, 2016). Since the meaning of a complex sentence is broader and more intricate, lower-proficiency learners cannot understand texts with a high number of complex sentences. When lower-proficiency learners read texts with a high number of sentences with complex structures, they experience more difficulties.

The last category of features used in this study to measure text difficulty is cohesion. The three cohesion related indicators that are used to examine text complexity in this study are *conjunction* (e.g., *yīn wèi…suǒ yǐ* 因為…所以 because), *positive conjunction* (e.g., *ér qiě* 而且 and), and *negative conjunction* (e.g., *fǒu zé* 否則 otherwise). Conjunctions are employed within a sentence to indicate that subsequent meanings are systematically connected to preceding meanings (Halliday & Hasan, 1976). Therefore, conjunctions facilitate the establishment of cohesive relationships within texts

(Louwerse & Mitchell, 2003). When texts are longer and more complex, more conjunctions are needed to aid a learner's comprehension.

The aforementioned 12 linguistic features categorized by lexical, semantic, syntactic, and cohesive were selected to calculate text complexity in this study.

4.3.2 Quantitative Feature Analysis of Chinese Textbooks

The CRIE-CFL system was used to determine the 12 linguistic features and then to examine whether appropriate CEFR levels were assigned to each of the selected textbooks within this study. A one-way ANOVA was conducted to identify differences in the linguistic features between different levels within each textbook series, including the CRIE-CFL (i.e. accuracy). The CEFR level served as the independent variable and the value of a linguistic feature was identified as the dependent variable. A significant ANOVA result suggests that the value of the linguistic features of at least one level is significantly higher or lower than that of the other levels; this implies that the linguistic features of different levels of teaching materials are not identical. Alternatively, an insignificant ANOVA result suggests that the values of linguistic features of different levels of teaching material are statistically equivalent; this implies that the linguistic features of different levels of teaching material are the same.

When ANOVA results were significant, a trend analysis was conducted to identify if any special trends were present in the linguistic features of each level or if the changes were simply random. The presence of a significant linear trend would indicate that the linguistic features of different levels do change with CEFR levels, and vice versa. If text difficulty changes with the CEFR level, the value of eleven of the twelve linguistic features (except for *simple sentence ratio*) should be lower in lower-level texts (e.g., A1) than in higher-level texts (e.g., B1).

The second stage of the analysis involved investigating whether the authors of the five selected textbooks assigned CEFR levels consistently; this is indicated by their use of linguistic features within textbooks labeled with the same CEFR level. Since level C is absent from the textbooks used in this study, only the textbooks labeled with A and B levels were compared. Another one-way ANOVA analysis was conducted to test whether there were differences in the linguistic features between the six textbook series (i.e. consistency). In this analysis, the CEFR level was the independent variable while the value of a linguistic feature was the dependent variable. A significant ANOVA result would indicate that the value of the linguistic features of at least one teaching material was significantly higher or lower than that of at least one of the others. That is, the linguistic features of the six teaching materials were not identical. On the other hand, an insignificant ANOVA result would suggest that the linguistic features used across the six textbook series were similar. Across series, textbooks labeled with the same CEFR level were expected to yield similar values in their linguistic features.

5. Results

5.1 Comparing Levels of Text Difficulty within the Same Textbook Series

The mean values of the 12 linguistic features within each level of the five Chinese language textbooks are listed in Table 4. The results show that for all five textbooks there are significant differences and significant linear trends in the following six categories: *characters*, *two-character words*, *sentences with a complex structure*, *content words*, *sentences with complex semantic categories*, and *complex semantic categories*. This means that the values of these six linguistic features either increase or decrease as the CEFR level increases. The values of the linguistic features of four out of the five textbooks (*Road to Success*, *New Modern Chinese*, *Practical Chinese*, and *Chinesisch*) increase with the CEFR level, whereas those of *Boules de neige* decrease; for example, there are fewer *characters* in the B1-level and B2-level texts than in the A2-level text.

This study yielded the following additional observations. All of the textbooks except *Boules de neige* show significant positive linear trends between *high-level words* and level, in that the number of *high-level words* increases as the CEFR level increases. *Practical Chinese* and *Chinesisch* show significant positive linear trends between *average sentence length* and level, in that the *average sentence length* increases as the CEFR level increases. Three textbooks (*New Modern Chinese*, *Practical Chinese*, and *Boules de neige*) show significant negative linear trends between *simple sentence ratio* and level, with the *simple sentence ratio* decreasing as the CEFR level increases. All of the textbooks except *Boules de neige* and *Chinesisch* show significant positive linear trends between *conjunctions* and level, in that the number of *conjunctions* increases as the CEFR level increases. All of the textbooks except *Chinesisch* show significant positive linear trends between *positive conjunctions* and level, in that the number of *positive conjunctions* increases as the CEFR level increases. Finally, two textbooks (*Road to Success* and *Practical Chinese*) show significant positive linear trends between *negative conjunctions* and level, with the number of *negative conjunctions* increasing as the CEFR level increases.

According to the results of this study, the 12 linguistic features (12/12) of *Practical Chinese* change in accordance with the change of the CEFR level. For *Road to Success* and *New Modern Chinese*, the results are similar. Both of the texts have 10 linguistic features (10/12) which change according to the CEFR level, but the remaining two linguistic features do not. The results for *Boules de neige* and *Chinesisch* are similar. Both of these texts have eight linguistic features (8/12) which change based on the CEFR level; however, the remaining four linguistic features of each textbook do not follow this pattern. This indicates that at least four of the five textbook authors did not demonstrate their ability to adopt materials with linguistic features that would appropriately reflect the corresponding difficulty levels of texts within the same textbook series.

Table 4 The Mean Values of the 12 Linguistic Features of the Six Different Levels for CRIE-CFL and the Five Textbooks

	CRIE-CFL						*F*	*Road to Success*						*F*	*New Modern Chinese*				*F*	*Practical Chinese*				*F*	*Boules de neige*		*F*	*Chinesisch*		*F*
Linguistic feature	A1	A2	B1	B2	C1	C2	(η^2)	A1	A2	B1	B2	C1	C2	(η^2)	A1	A2	B1	B2	(η^2)	A1	A2	B1	B2	(η^2)	A2	B1/B2	(η^2)	A1/A2	B1/B2	(η^2)
Characters	64	143	309	481	642	1505	353 (.53)	453	836	1231	1612	1128	1776	72 (.50)	67	162	252	316	84 (.88)	110	190	259	627	66 (.55)	404	236	204 (.86)	90	127	32 (.24)
High-level words	2	9	32	62	102	255	552 (.64)	61	127	195	261	165	295	85 (.54)	5	11	25	39	34 (.74)	4	13	27	82	95 (.64)	38	33	n.s. (.09)	6	9	15 (.12)
Two-character words	12	32	76	127	175	403	412 (.57)	124	225	333	425	301	473	75 (.51)	12	36	66	78	71 (.86)	25	45	70	169	66 (.55)	89	76	9 (.22)	17	26	31 (.23)
Average sentence length	6.07	7.85	8.75	9.19	9.88	10.44	130 (.29)	10.29	9.66	10.15	10.21	9.92	10.32	n.s. (.00)	6.55	8.06	8.13	7.27	9 (.44)	7.38	8.47	8.88	9.39	15 (.21)	8.73	8.82	n.s. (.00)	3.96	4.49	9 (.08)
Simple sentence ratio	.97	.85	.62	.45	.32	.37	310 (.50)	.39	.42	.43	.39	.42	.38	n.s. (.01)	.98	.89	.78	.56	18 (.60)	.86	.71	.53	.56	21 (.28)	.74	.53	16 (.33)	1.00	.99	n.s. (.03)
Sentences with a complex structure	1.37	5.33	12.70	20.59	26.49	63.31	309 (.50)	18.41	34.81	52.83	68.96	50.22	75.32	65 (.48)	1.88	7.42	10.39	12.00	28 (.70)	3.36	7.93	11.24	27.55	61 (.53)	16.13	8.32	93 (.74)	1.98	3.38	12 (.10)
Content words	39	85	175	262	346	816	316 (.50)	245	457	672	872	616	959	65 (.48)	46	104	148	183	73 (.86)	67	110	144	351	61 (.53)	239	116	470 (.93)	57	78	27 (.21)
Sentences with complex semantic categories	5.45	8.99	14.96	20.79	23.62	48.96	152 (.33)	16.66	30.81	43.46	52.54	41.14	58.08	44 (.38)	5.00	9.58	14.06	22.00	30 (.72)	7.16	10.16	11.43	25.32	35 (.39)	19.13	8.63	45 (.58)	13.96	16.98	9 (.08)
Complex semantic categories	2.18	3.26	4.98	6.78	7.54	14.82	107 (.25)	4.96	9.33	13.44	15.85	12.81	17.64	40 (.36)	1.90	3.16	4.78	8.13	18 (.60)	2.69	3.51	3.71	7.98	26 (.32)	6.52	2.66	26 (.44)	6.42	7.91	7 (.07)
Conjunctions	0.35	1.85	5.45	10.27	14.25	32.47	346 (.52)	10.14	16.93	27.02	33.58	24.11	37.58	57 (.44)	0.63	2.08	6.11	3.50	17 (.59)	0.89	2.84	6.73	12.94	54 (.50)	4.81	6.21	n.s. (.05)	0.90	1.32	n.s. (.03)
Positive conjunctions	0.25	1.24	3.84	7.10	9.16	20.33	319 (.50)	6.42	10.86	17.98	20.92	15.22	24.19	58 (.45)	0.00	1.25	3.39	2.00	16 (.57)	0.66	2.00	4.94	8.29	39 (.42)	2.63	4.58	6 (.16)	0.51	0.84	n.s. (.03)
Negative conjunctions	0.13	0.56	1.70	3.00	4.29	10.80	197 (.39)	3.28	5.62	8.07	11.54	7.92	12.45	33 (.31)	0.63	0.92	1.83	0.50	n.s. (.20)	0.18	0.86	1.59	4.45	34 (.38)	1.88	1.53	n.s. (.02)	0.49	0.50	n.s. (.00)

Note. *F* = *F* value in the ANOVA test; CRIE-CFL = Chinese Readability Index Explorer for Chinese as a Foreign Language; n.s. = not significant.

More detailed information about the standard deviation and F value in trend analysis can be found on http://140.122.96.190/20171107/table4.pdf.

5.2 Comparing Linguistic Features of Texts with the Same CEFR-labels among Textbooks

Table 5 lists the mean values of the 12 linguistic features of the texts labeled as CEFR A-level and B-level for each of the five textbooks and the CRIE-CFL database. The one-way ANOVA results indicate that the 12 linguistic features of the six teaching materials at level A are significantly different (as indicated by the F values in Table 5). The 12 linguistic features of the six teaching materials at level B are also significantly different (see Appendix 2 for the results of post-hoc comparisons). The results show that textbooks labeled with the same CEFR levels yield different values in terms of their linguistic features, such as *high-level words* and *average sentence length*, and should actually be assigned different difficulty levels. A detailed analysis is presented Table 5.

Many discrepancies can be seen when looking at the lexical feature *characters*. Among all A-level teaching materials, *Road to Success* has the highest number of *characters* (mean = 530 characters) while *Chinesisch* has the lowest number (mean = 90 characters), with a difference of 440 characters. Meanwhile, the average number of *characters* is significantly higher in *Road to Success* and *Boules de neige* than in the CRIE-CFL, and significantly lower in *Chinesisch* than in the CRIE-CFL. Among all B-level teaching materials, *Road to Success* has the highest number of *characters* (mean = 1372 characters) while *Chinesisch* has the lowest number (mean = 127 characters), with a difference of 1245 characters. In addition, the average number of *characters* is significantly higher in *Road to Success* than in the CRIE-CFL and significantly lower in *New Modern Chinese*, *Boules de neige*, and *Chinesisch* than in the CRIE-CFL.

The following variances can be seen when analyzing the syntactic feature *average sentence length*. Among A-level teaching materials, *Road to Success* has the longest *average sentence length* (mean = 10.16 words) while *Chinesisch* has the shortest *average sentence length* (mean = 3.96 words), corresponding to a difference of 6.20 words. The *average sentence length* is significantly longer in *Road to Success* than in the CRIE-CFL and significantly shorter in *Chinesisch* than in the CRIE-CFL. Among B-level teaching materials, *Road to Success* has the longest *average sentence length* (mean = 10.17 words) while *Chinesisch* has the shortest *average sentence length* (mean = 4.49 words), corresponding to a difference of 5.68 words. Meanwhile, the *average sentence length* is significantly longer in *Road to Success* than in the CRIE-CFL, and significantly shorter in *New Modern Chinese* and *Chinesisch* than in the CRIE-CFL.

The following observations were made when analyzing the semantic feature *content words*. Among A-level teaching materials, *Road to Success* has the highest number of *content words* (mean = 287 words) while *Chinesisch* has the lowest number (mean = 57 words), corresponding to a difference of 230 words. The number of *content words* is significantly higher in *Road to Success* and *Boules de neige* than in the CRIE-CFL, and significantly lower in *Chinesisch* than in the CRIE-CFL. Among B-level teaching materials, *Road to Success* has the highest number of *content words* (mean = 746 words) while *Chinesisch* has the lowest number (mean = 78 words), corresponding to a difference of 668 words. The number of *content words* is significantly higher in *Road to*

Table 5 The Mean Values of the 12 Linguistic Features in CRIE-CFL and the Five Textbooks at Levels A and B

CEFR	Level A						F	Level B						F
Linguistic feature	CRIE-CFL	*Road to Success*	*New Modern Chinese*	*Practical Chinese*	*Boules de neige*	*Chinesisch*	(η^2)	CRIE-CFL	*Road to Success*	*New Modern Chinese*	*Practical Chinese*	*Boules de neige*	*Chinesisch*	(η^2)
Characters	118	530	124	149	404	90	350 (.67)	381	1372	258	402	236	127	195 (.48)
High-level words	7	75	9	9	38	6	350 (.67)	45	219	27	48	33	9	256 (.55)
Two-character words	26	144	26	35	89	17	410 (.70)	98	367	67	108	76	26	207 (.50)
Average sentence length	7.29	10.16	7.46	7.92	8.73	3.96	30 (.15)	8.94	10.17	8.05	9.08	8.82	4.49	192 (.48)
Simple sentence ratio	.89	.39	.93	.79	.74	1.00	231 (.57)	.55	.41	.76	.54	.53	.99	46 (.18)
Sentences with a complex structure	4.08	21.70	5.20	5.62	16.13	1.98	316 (.65)	16.04	58.78	10.55	17.56	8.32	3.38	187 (.47)
Content words	71	287	81	88	239	57	295 (.63)	212	746	151	224	116	78	177 (.46)
Sentences with complex semantic categories	7.88	19.50	7.75	8.64	19.13	13.96	97 (.36)	17.42	46.82	14.85	16.81	8.63	16.98	65 (.24)
Complex semantic categories	2.92	5.84	2.66	3.10	6.52	6.42	53 (.24)	5.74	14.33	5.12	5.36	2.66	7.91	48 (.19)
Conjunctions	1.38	11.51	1.50	1.85	4.81	0.90	290 (.63)	7.49	29.45	5.85	9.14	6.21	1.32	166 (.44)
Positive conjunctions	0.93	7.31	0.75	1.32	2.63	0.51	211 (.55)	5.22	19.06	3.25	6.24	4.58	0.84	133 (.39)

Negative conjunctions	0.43	3.75	0.80	0.52	1.88	0.49	148 (.46)	2.25	9.35	1.70	2.70	1.53	0.50	103 (.33)

Note. CRIE-CFL = Chinese Readability Index Explorer for Chinese as a Foreign Language; n.s. = not significant.

More detailed information about the standard deviation can be found at http://140.122.96.190/20171107/table5.pdf.

Success than in the CRIE-CFL, and significantly lower in *New Modern Chinese*, *Boules de neige*, and *Chinesisch* than in the CRIE-CFL.

When examining the cohesion feature c*onjunctions* the following disparities can be seen. Among A-level teaching materials, *Road to Success* has the highest number of *conjunctions* (mean = 11.51 conjunctions) while *Chinesisch* has the lowest (mean = 0.90 conjunctions), corresponding to a difference of 10.61 conjunctions. The number of *conjunctions* is significantly higher in *Road to Success* and *Boules de neige* than in the CRIE-CFL. Among B-level teaching materials, *Road to Success* has the highest number of *conjunctions* (mean = 29.45 conjunctions) while *Chinesisch* has the lowest number (mean = 1.32 conjunctions), corresponding to a difference of 28.13 conjunctions. The number of *conjunctions* is significantly higher in *Road to Success* than in the CRIE-CFL, and significantly lower in *Chinesisch* than in the CRIE-CFL.

6. Discussion

This study used the CRIE-CFL system as a tool to calculate 12 linguistic features in five Chinese textbook series in order to examine their CEFR level grading. It compared textbooks within and between series to examine the accuracy and consistency of their CEFR level grading in relation to the actual text difficulty as measured by the CRIE-CFL system.

6.1 Differences within Series

The linguistic features within the same textbook series did not show a meaningful transition between the levels of difficulty. The data produced in this study indicate that most teaching materials do not match their assigned CEFR levels. The trend analysis shows that of the 1578 texts used for CRIE-CFL training, only *Practical Chinese* demonstrated an increasing or decreasing trend corresponding to the assigned CEFR level. That is, linear trends were not found between linguistic features and CEFR levels in *Road to Success*, *New Modern Chinese*, *Boules de neige*, and *Chinesisch*. The linguistic features in these four teaching materials do not vary in accordance with their assigned difficulty level. Although eight of the linguistic features in *Boules de neige* exhibited positive linear tendencies, six of the linguistic features were inversely correlated with the CEFR levels in the remaining five teaching materials. It should also be noted that there are cases in *Boules de neige* where lower-level texts are explained by words from higher-level texts. For example, in Lesson 2 of the A2-level textbook, the word *buzhibujuedi* (unconsciously; 不知不覺地) is explained by the word *zhuyi* (attention; 注意), which is from Lesson 2 of the B-level textbook.

The mismatch between the vocabulary and CEFR levels calls for standardized leveling criteria to select level-appropriate linguistic features for textbooks. This study provides additional evidence to support the observations by Alderson (2007), Hulstijn (2007), and Hulstijn et al. (2010) that the clarity of CEFR's definition of each proficiency level should also be improved for educators.

6.2 Differences among Series

Textbooks with the same CEFR level contained significantly different linguistic features. In this study, textbooks that are assigned the same CEFR level are found to have different difficulty levels. For example, the *Road to Success* has the lowest *simple sentence ratio* while *Chinesisch* has the highest *simple sentence ratio*. In other words, with respect to sentence learning, *Road to Success* is more difficult than *Chinesisch*. Furthermore, this study finds that for these two series, their materials at levels A and B yield different values in the linguistic feature analysis. Even though these textbooks are labeled with the same CEFR levels, they actually have different difficulty levels.

As for *Boules de neige*, its A-level materials have relatively more *characters*, *high-level words*, *two-character words*, *sentences with a complex structure*, *content words*, and *conjunctions*. This indicates that these materials are more difficult than their counterparts in other series. On the other hand, *Boules de neige's* B-level materials are comparatively easier as they contain fewer *sentences with a complex structure*, *content words*, *sentences with complex semantic categories*, and *complex semantic categories*. B-level materials in *New Modern Chinese* are also relatively easy as they contain a higher *simple sentence ratio*, a shorter *average sentence length*, and a smaller number of *sentences with a complex structure* and *content words*. Considering these inconsistencies, our linguistic feature analyses suggest that the compilation of teaching materials cannot be based solely on educators' professional experience. A standardized system is required to determine the difficulty level of L2 teaching materials to ensure accuracy within a series and consistency between series.

The findings of this study provide empirical evidence indicating the inconsistency in the difficulty levels of different Chinese teaching materials. In order to account for these inconsistencies, scholars have independently developed rubrics for evaluating vocabulary, grammar, and reading sections of language textbooks, such as using metrics based on the vocabulary load, vocabulary difficulty, and word frequency (Rahimpour & Hashemi, 2011; Williams, 1983). However, these rubrics take the form of questionnaires, which are still predicated on a subjective evaluation. As Sung et al. (2015b) pointed out, the manual leveling of learning materials presents three problems: high demands on both time and effort, difficulty in reaching a consensus, and ambiguity in the interpretation of leveling criteria. These problems also appeared in the CFL textbooks that were analyzed in this study.

6.3 Types of Linguistic Features that Affect Text Difficulty

According to the CEFR scale for overall reading comprehension (Council of Europe, 2001), learners at levels A1 and A2 should be able to understand high-frequency words, and B2 learners should possess a larger vocabulary than A1. Based on our linguistic feature analyses with CRIE-CFL, A-level materials in *Road to Success* and *Boules de neige* have a considerably higher number of *characters*, *high-level words*, and *two-character words* than other textbook series. Therefore, beginners may find these textbooks difficult. B-level materials in *Road to Success* suffer from the same problem. Textbooks compilers should adjust the difficulty levels of these textbooks by selecting

proficiency-appropriate vocabulary at a given level and then ensure that the vocabulary consistently increases as a learner's proficiency level increases (Rahimpour & Hashemi, 2011).

Regarding syntax, A1 and A2 learners should be able to understand short and simple sentences, while C1 and C2 learners are able to understand long and complex sentences. Our linguistic feature analyses show that the *average sentence length* and the *simple sentence ratio* in *Road to Success* do not change significantly in relation to the CEFR level. Compared to other series, texts at levels A and B in *Road to Success* have a higher *average sentence length* and a higher number of *sentences with a complex structure* but a lower *simple sentence ratio*. Lower-level texts in *Road to Success* tend to be comprised of longer and more complex sentences, which may cause comprehension difficulty for beginners. On the other hand, texts at levels A and B in *Chinesisch* do not show any significant differences in their *simple sentence ratio*, which means that the sentences in the B-level material in *Chinesisch* may be too short and should increase their complexity.

In terms of semantic features, higher-level learners should be able to understand more content words and more semantically complex sentences. For most of the teaching materials, the values for the three semantic features tend to be higher for the higher-level texts than for the lower-level ones; *Boules de neige* is the only exception, in that its three semantic linguistic features move in opposite directions, which suggests that the vocabulary of this textbook needs to be adjusted. *Road to Success* has higher values for the three semantic features than the other textbooks. This indicates that its texts are more difficult and include a larger number of concepts that require more time to process.

The linguistic features adopted by the CRIE-CFL correspond to those in the CEFR reading comprehension grading standards, and also include an additional three linguistic features: *conjunctions*, *positive conjunctions*, and *negative conjunctions.* These three features were added because conjunctions help learners to establish cohesion when reading a text (Louwerse & Mitchell, 2003). Cohesion is an important component of reading comprehension (Benjamin, 2012; Graesser, McNamara, & Kulikowich, 2011; Graesser et al., 2004; McNamara, Louwerse, McCarthy, & Graesser, 2010). The combination of cohesive sentences, consistent text, and cohesive semantics contribute to the creation of texts that are more readable to learners (Gernsbacher, 1990; McNamara & Kintsch, 1996; McNamara, Kintsch, Songer, & Kintsch, 1996).

Our analyses of the five textbooks and the texts in the CRIE-CFL system have shown that the number of *negative conjunctions* in the CRIE-CFL training data set, *Road to Success*, and *Practical Chinese* increases with the CEFR level. These results suggest that higher-level texts contain more transitions, as more cognitive resources are required to process the complex relationships between sentences in the texts. However, no significant differences in the three cohesive features were found between different level textbooks in *Chinesisch.* This suggests that *Chinesisch* did not take into account the effect of conjunctions on reading comprehension.

6.4 Pedagogical Implications

Findings of this study have several pedagogical implications. First, developers of language materials who are looking to incorporate the CEFR scale should carefully consult the statements regarding the horizontal dimension of the CEFR scales in order to obtain an in-depth understanding of various topics, the scope of language use, and language proficiency. These statements define the level of proficiency in listening, speaking, reading, writing, and translation. Additional training led by experienced experts in CEFR and language learning material grading may help users to better understand the criteria in the lexical, syntactic, and semantic aspects of language proficiency. Language-specific proficiency standards should also be developed to make the description of each standard more objective and precise.

Secondly, language educators need to increase their awareness of the influence of linguistic features, such as *characters and words*, *semantics*, *syntax*, and *cohesion*, on text difficulty and reading comprehension. The awareness of linguistic features enhances an educator's ability to select proficiency-appropriate materials for leaners. Such awareness also facilitates the process of selecting CEFR-graded teaching materials, comparing textbooks published by different publishers or in different regions, and complying with the language proficiency standards in the CEFR.

Lastly, analytics tools, such as the CRIE-CFL, can be useful for quantifying linguistic features to ensure that textbook contents are consistent with both the vertical and the horizontal dimension statements of the CEFR scales. The automatic analysis functions of CRIE-CFL can also help educators efficiently develop parameters that reflect a text's level of difficulty and therefore enhance the objective evaluation of textbook levels.

7. Conclusion

Despite the importance of grading language textbooks for teaching and learning, few studies have addressed the issues of consistency, accuracy, and efficiency in the grading of texts. Based on the CEFR framework and the analytic tool CRIE-CFL (Sung et al., 2016), this study examined the accuracy and consistency of text grading within and between textbook series. Based on our linguistic feature analyses, we found that most of the textbooks we examined did not use the linguistic features reflective of their corresponding proficiency levels. The language used in these textbooks does not always increase in difficulty as the level increases. Our analyses also show that even textbooks labeled with the same CEFR level yielded different values in terms of their use of linguistic features, therefore indicating a varying level of difficulty. The results of this study call for a standardized system for educators to use in determining the difficultly level of teaching materials as manual text grading is no longer effective or reliable.

Finally, there are four major Chinese proficiency standards adopted across continents: ACTFL (US), CEFR (EU), HSK (China), and TOCFL(Taiwan). As conversions among these proficiency standards are, in fact, rather straightforward (e.g.

CEFR A2 would be equivalent to ACTFL Intermediate or HSK 4[12]), future research will extend the CRIE-CFL model to other proficiency standards to further validate the findings of this study. Thus, with CRIE-CFL, educators as well as textbook developers will be able to make use of the tool when they select and compile texts suitable for students at various proficiency levels.

Acknowledgments: This work was supported by National Taiwan Normal University's Chinese Language and Technology Center. The center is funded by Taiwan's Ministry of Education (MOE), as part of the Featured Areas Research Center Program, under the Higher Education Sprout Project.

References

Alderson, J. C. (2007). The CEFR and the need for more research. *The Modern Language Journal*, *91*(4), 659–663. doi: 10.1111/j.1540-4781.2007.00627_4.x

Benjamin, R. G. (2012). Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty. *Educational Psychology Review*, *24*(1), 63–88. doi: 10.1007/s10648-011-9181-8

Bonnet, G. (2007). The CEFR and education policies in Europe. *The Modern Language Journal*, *91*(4), 669–672. doi: 10.1111/j.1540-4781.2007.00627_7.x

Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1721-1730), Sydney, Australia.

Chang, T. H., & Sung, Y. T. (2019). Automated Chinese essay scoring based on multi-level linguistic features. In X. Lu, & B. Chen (Eds.), *Computational and Corpus Approaches to Chinese Language Learning* (pp. 258-274). Singapore: Springer.

Chang, Y. W., & Lin, C. J. (2008, December). Feature ranking using linear SVM. In *Proceedings of the Workshop on the Causation and Prediction Challenge at WCCI 2008* (pp. 53–64), Hong Kong, China,.

Chen, B., & Hsu, Y. C. (2019). Mandarin Chinese mispronunciation detection and diagnosis leveraging deep neural network based acoustic modeling and training techniques. In X. Lu, & B. Chen (Eds.), *Computational and Corpus Approaches to Chinese Language Learning* (pp. 219–237). Singapore: Springer.

Chen Y. W., & Lin, C. J. (2006) Combining SVMs with various feature selection strategies. In I. Guyon, M. Nikravesh, S. Gunn, & L. A. Zadeh (Eds), *Feature extraction. Studies in fuzziness and soft computing* (Vol. 207, pp. 315-324). Springer, Berlin, Heidelberg.

Cheng, C. C. (2005). Computing the degree of difficulty in lexical semantics and sentence reading. In *The 6th Chinese lexical semantics workshop* (pp. 261-265), Fujian: Xiamen University. 261–265. [鄭錦全. (2005) 詞彙語意與句子閱讀難易度計量, 261-265.]

Collins-Thompson, K. (2014). Computational assessment of text readability: A survey of current and future research. *ITL-International Journal of Applied Linguistics*,

[12] https://www.actfl.org/sites/default/files/reports/Assigning_CEFR_Ratings_To_ACTFL_Assessments.pdf

165(2), 97-135.
Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment.* Cambridge, UK: Cambridge University Press.
Crossley, S. A., Louwerse, M. M., McCarthy, P. M., & McNamara, D. S. (2007). A linguistic analysis of simplified and authentic texts. *Modern Language Journal*, *91*(2), 15–30. doi: 10.1111/j.1540-4781.2007.00507.x
Crossley, S. & McNamara, D. (2008). Assessing L2 reading texts at the intermediate level: An approximate replication of Crossley, Louwerse, McCarthy & McNamara (2007). *Language Teaching*, *41*(3), 409–429. doi: 10.1017/S0261444808005077
Dale, E., & Chall, J. S. (1948). A formula for predicting readability: Instructions. *Educational research bulletin*, *27*(2), 37–54.
Ding, S. (2009, November). Feature selection based F-score and ACO algorithm in support vector machine. In C. Zhao, Y. Wu, J. Wang, & Q. Liu. (Eds), *2009 Second International Symposium on Knowledge Acquisition and Modeling* (Vol. 1, pp. 19-23). IEEE Computer Society 10662 Los Vaqueros Cir, Los Alamitos, CA 90720-1314.
Duanmu, S. (1999). Stress and the development of disyllabic words in Chinese. *Diachronica*, *16*(1), 1–35. doi:10.1075/dia.16.1.03dua
Faison, E. W. (1951). Readability of children's textbooks. *Journal of Educational Psychology*, *42*(1), 43. doi: 10.1037/h0060329
Figueras, N. (2012). The impact of the CEFR. *ELT Journal*, *66*(4), 477–485. doi: 10.1093/elt/ccs037
Flesch, R. F. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(3), 221-233.
Fulcher, G. (2004). Deluded by artifices? The common European framework and harmonization. *Language Assessment Quarterly: An International Journal*, *1*(4), 253–266. doi:10.1207/s15434311laq0104_4
Fulcher, G. (2007, December). The reification of the Common European Framework of Reference (CEFR) and effect-driven testing. Paper presented at the 14th International Conference of Applied Linguistics, Thessaloniki, Greece.
Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Earlbaum.
Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multi-level analyses of text characteristics. *Educational Researcher*, *40*(5), 223–234. doi: 10.3102/0013189X11413260
Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 193–202. doi:10.3758/BF03195564
Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
He, K. & Li, D. (1987). *Three thousand most commonly used words in modern Chinese*. Beijing, China: Beijing Normal University Press. [何克抗, & 李大魁. (1987). *现代汉语三千常用词.* 中国北京: 北京师范大学出版社.]
Hong, J. F., Sung, Y. T., Tseng, H. C., Chang, K. E., & Chen, J. L. (2016). A multilevel analysis of the linguistic features affecting chinese text readability. *Taiwan Journal of Chinese as a Second language*, *13*, 95–126.
Hsu, F. Y., Lee, H. M., Chang, T. H., & Sung, Y. T. (2018). Automated estimation of item

difficulty for multiple-choice tests: An application of word embedding techniques. *Information Processing & Management*, *54*(6), 969-984.

Hulstijn, J. H. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language Proficiency. *The Modern Language Journal*, *91*(4), 663–667. doi:10.1111/j.1540-4781.2007.00627_5.x

Hulstijn, J. H., Alderson, J. C., & Schoonen, R. (2010). Developmental stages in second-language acquisition and levels of second-language proficiency: Are there links between them?. In I. Bartning, , M. Maisa, & I. Vedder (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 11–20). Amsterdam, Netherlands: Eurosla.

Klare, G. R. (1984). Readability. In P. D. Pearson, R. Barr, M. L. Kamil, & P. B. Mosenthal (Eds.), *Handbook of reading research*. New York: Longman

Lee, L. H., Chang, L. P., & Tseng, Y. H. (2016). Developing learner corpus annotation for Chinese grammatical errors. In Y. Lu (Ed), *Proceedings of the 20th international conference on Asian language processing* (pp. 254–257). Tainan, Taiwan.

Li, Y., Wen, X. H., & Xie, T. W. (2014). CLTA 2012 Survey of College-Level Chinese Language Programs in North America. *Journal of the Chinese Language Teachers Association*, *49*(1), 1-49.

Lin, S. Y., Chen, H. C., Chang, T. H., Lee, W. E., & Sung. Y. T. (2019). CLAD: A corpus-derived Chinese lexical association database. *Behavior Research Methods*, *51*(5), 2310-2336. doi:10.3758/s13428-019-01208-2

Little, D. (2006). The Common European Framework of Reference for Languages: Content, purpose, origin, reception and impact. *Language Teaching*, *39*(3), 167–190. doi: 10.1017/S0261444806003557

Little, D. (2007). The Common European Framework of Reference for Languages: Perspectives on the making of supranational language education policy. *Modern Language Journal*, *91*(4), 645–655. doi: 10.1111/j.1540-4781.2007.00627_2.x

Louwerse, M. M., & Mitchell, H. H. (2003). Toward a taxonomy of a set of discourse markers in dialog: A theoretical and computational linguistic account. *Discourse Processes*, *35*(3), 199–239. doi: 10.1207/S15326950DP3503_1

Lu, X., & Chen, B. (2019). Computational and corpus approaches to Chinese language learning: an introduction. In X. Lu, , & B. Chen (Eds.), *Computational and corpus approaches to Chinese language learning* (pp. 6–14). Singapore: Springer.

McNamara, D. S. & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247–287. doi:10.1080/01638539609544975

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1–43. doi:10.1207/s1532690xci1401_1

McNamara, D. S., Louwerse, M. M., & Graesser, A. C. (2002). *Coh-Metrix: Automated cohesion and coherence scores to predict text readability and facilitate comprehension*. Technical report, Institute for Intelligent Systems, University of Memphis, Memphis, TN.

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010).

Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes*, *47*(4), 292–330. doi:10.1080/01638530902959943

Rahimpour, M., & Hashemi, R. (2011). Textbook selection and evaluation in EFL context. *World Journal of Education*, *1*(2), 62–68. doi:10.5430/wje.v1n2p62

Sung, Y. T., Chang, T. H., Lin, W. C., Hsieh, K. S., & Chang, K. E. (2016). CRIE: An automated analyzer for Chinese texts. *Behavior Research Methods*, *48*(4), 1238–1251. doi:10.3758/s13428-015-0649-1

Sung, Y. T., Chen, J. L., Cha, J. H., Tseng, H. C., Chang, T. H., & Chang, K. E. (2015a). Constructing and validating readability models: The method of integrating multilevel linguistic features with machine learning. *Behavior Research Methods*, *47*(2), 340–354. doi: 10.3758/s13428-014-0459-x

Sung, Y. T., Lin, W. C., Dyson, S. B., Chang, K. E., & Chen, Y. C. (2015b). Leveling L2 Texts Through Readability: Combining Multilevel Linguistic Features with the CEFR. *Modern Language Journal*, *99*(2), 371–391. doi:10.1111/modl.12213

Sung, Y. T., Chen, J. L., Lee, Y. S., Lee, Y. S., Cha, J. H., Tseng, H. C., Lin, W. C., Chang, T. H., & Chang, K. E. (2013). Investigating Chinese text readability: Linguistic features, modeling, and validation. *Chinese Journal of Psychology*, *55*(1), 75–106. doi: 10.6129/CJP.20120621. [宋曜廷, 陳茹玲, 李宜憲, 查日穌, 曾厚強, 林維駿, 張道行, & 張國恩. (2013). 中文文本可讀性探討：指標選取、模型建立與效度驗證. *中華心理學刊, 55*(1), 75-106.]

Tseng, H. C., Chen, B., Chang, T. H., & Sung, Y. T. (2019). Integrating LSA-based hierarchical conceptual space and machine learning methods for leveling the readability of domain-specific texts. *Natural Language Engineering*, *25*(3), 331-361.

Vapnik, V. (1995). *The nature of statistical learning theory*. New York, NY: Springer Verlag.

Vinther, J. (2013). CEFR - in a critical light. In J. Colpaert, M. Simons, A. Aerts, & M. Oberhofer (Eds.), *Language testing in Europe: Time for a new framework?* (pp. 242-247), University of Antwerp, Belgium.

Williams, D. (1983). Developing criteria for textbook evaluation. *ELT Journal*, *37*(3), 251–255. doi:10.1093/elt/37.3.251

Appendix 1

Textbooks used in this study

Bellassen, J., & Liu, J. L. (2011). *Le chinois par boules de neige (Acces raisonne a la lecture du chinois)* （雪球）. Chasseneuil-du-Poitou, France: Scérén Cndp-crdp.

Bellassen, J., & Liu, J. L. (2012). *Le chinois par boules de neige (Niveau elementaire)* （雪球）. Chasseneuil-du-Poitou, France: Scérén Cndp-crdp.

Chinese Time (Ed). (2009). *Practical Chinese (实用中文)*. Shanghai: East China Normal University Press.

Editors of the Road to Success sereis (Ed). (2008-2014). *Road to Success(成功之路)*. Beijing: Beijing Language and Culture University Press.

Kantor P. (2004). *Assimil Pack Chinesisch Ohne Mühe (漢語) Volume 1.* Köln, Germany: ASSiMiL GmbH.

Kantor P. (2006). *Assimil Pack Chinesisch Ohne Mühe (漢語) Volume 2.* Köln, Germany: ASSiMiL GmbH.

Liu, X. (2007). *New Practical Chinese Reader*（新实用汉语課本）. Beijing : Beijing Language and Culture University Press.

National Taiwan Normal University (Ed). (2008). *Practical Audio-Visual Chinese* (2nd Edition) *(新版視聽華語).* Taipei: Cheng Chung Bookstore.

NTNU Extension School of Continuing Education (Ed). (2012). *New Modern Chinese (新時代華語)*. Taiwan: NTNU Extension School of Continuing Education.

Yeh, T. M. (Ed). 2008. *Far East Everyday Chinese (遠東生活華語).* Taipei: Far East Book Company.

Appendix 2 The Results of the Post-hoc Comparisons Between the Five Textbooks and the CRIE-CFL

The values across the horizontal rows and down the vertical columns are for the linguistic features in level-A and level-B textbooks, respectively. The upper right-hand and lower left-hand corners of the table provide the post-hoc comparisons of level-A and level-B textbooks, respectively.

Characters

		CRIE-CFL	Road to Success	New Modern Chinese	Practical Chinese	Boules de neige	Chinesisch 000
		118	530	124	149	404	90
CRIE-CFL	381	—	***			***	**
Road to Success	1372	***	—	***	***	***	***
New Modern Chinese	258	***	***	—		***	
Practical Chinese	402		***	***	—	***	***
Boules de neige	236	***	***		***	—	***
Chinesisch	127	***	***	***	***	***	—

High-level words

		CRIE-CFL	Road to Success	New Modern Chinese	Practical Chinese	Boules de neige	Chinesisch 000
		7	75	9	9	38	6
CRIE-CFL	45	—	***			***	
Road to Success	219	***	—	***	***	***	***
New Modern Chinese	27	***	***	—		***	
Practical Chinese	48		***	***	—	***	*
Boules de neige	33	***	***		*	—	***
Chinesisch	9	***	***	***	***	***	—

Two-character words

		CRIE-CFL	Road to Success	New Modern Chinese	Practical Chinese	Boules de neige	Chinesisch 000
		26	144	26	35	89	17
CRIE-CFL	98	—	***		*	***	***
Road to Success	367	***	—	***	***	***	***
New Modern Chinese	67	***	***	—		***	
Practical Chinese	108		***	***	—	***	***
Boules de neige	76	***	***		*	—	***
Chinesisch	26	***	***	***	***	***	—

Average sentence length

		CRIE-CFL	Road to Success	New Modern Chinese	Practical Chinese	Boules de neige	Chinesisch 000
		7.29	10.16	7.46	7.92	8.73	3.96
CRIE-CFL	8.94	—	***				***
Road to Success	10.17	***	—		***		***
New Modern Chinese	8.05	***	***	—			*
Practical Chinese	9.08		***	***	—		***
Boules de neige	8.82		***	*		—	**
Chinesisch	4.49	***	***	***	***	***	—

Note. CRIE-CFL = Chinese Readability Index Explorer for Chinese as a Foreign Language.

* $p < .05$, ** $p < .01$, *** $p < .001$.

(continued)

Simple sentence ratio

		CRIE-CFL	Road to Success	New Modern Chinese	Practical Chinese	Boules de neige	Chinesisch 000
		0.89	0.39	0.93	0.79	0.74	1
CRIE-CFL	0.55	—	***		**	**	***
Road to Success	0.41	***	—	***	***	***	***
New Modern Chinese	0.76	***	***	—	**	**	
Practical Chinese	0.54		**	***	—		***
Boules de neige	0.53			**		—	***
Chinesisch	0.99	***	***	***	***	***	—

Sentences with a complex structure

		CRIE-CFL	Road to Success	New Modern Chinese	Practical Chinese	Boules de neige	Chinesisch
		4.08	21.7	5.2	5.62	16.13	1.98
CRIE-CFL	16.04	—	***			***	***
Road to Success	58.78	***	—	***	***	***	***
New Modern Chinese	10.55	***	***	—		***	**
Practical Chinese	17.56		***	***	—	***	***
Boules de neige	8.32	***	***	.052	***	—	***
Chinesisch	3.38	***	***	***	***	***	—

Content words

		CRIE-CFL	Road to Success	New Modern Chinese	Practical Chinese	Boules de neige	Chinesisch 000
		71	287	81	88	239	57
CRIE-CFL	212	—	***			***	*
Road to Success	746	***	—	***	***	***	***
New Modern Chinese	151	***	***	—		***	
Practical Chinese	224		***	**	—	***	**
Boules de neige	116	***	***	***	***	—	***
Chinesisch	78	***	***	***	***	***	—

Sentences with complex semantic categories

		CRIE-CFL	Road to Success	New Modern Chinese	Practical Chinese	Boules de neige	Chinesisch
		7.88	19.5	7.75	8.64	19.13	13.96
CRIE-CFL	17.42	—	***			***	***
Road to Success	46.82	***	—	***	***		***
New Modern Chinese	14.85		***	—		***	***
Practical Chinese	16.81		***		—	***	***
Boules de neige	8.63	***	***	***	***	—	
Chinesisch	16.98		***			***	—

Note. CRIE-CFL = Chinese Readability Index Explorer for Chinese as a Foreign Language.

* $p < .05$， ** $p < .01$， *** $p < .001$.

(continued)

Complex semantic categories

		CRIE-CFL	Road to Success	New Modern Chinese	Practical Chinese	Boules de neige	Chinesisch 000
		2.92	5.84	2.66	3.1	6.52	6.42
CRIE-CFL	5.74	—	***			**	***
Road to Success	14.33	***	—	***	***		
New Modern Chinese	5.12	***	***	—		**	***
Practical Chinese	5.36		***		—	**	***
Boules de neige	2.66	***	***	***	***	—	
Chinesisch	7.91	***	***	***	***	***	—

Conjunctions

		CRIE-CFL	Road to Success	New Modern Chinese	Practical Chinese	Boules de neige	Chinesisch
		1.38	11.51	1.5	1.85	4.81	0.9
CRIE-CFL	7.49	—	***			**	
Road to Success	29.45	***	—	***	***	***	***
New Modern Chinese	5.85		***	—		**	
Practical Chinese	9.14		***	*	—	*	**
Boules de neige	6.21		***			—	**
Chinesisch	1.32	***	***	***	***	***	—

Positive conjunctions

		CRIE-CFL	Road to Success	New Modern Chinese	Practical Chinese	Boules de neige	Chinesisch 000
		0.93	7.31	0.75	1.32	2.63	0.51
CRIE-CFL	5.22	—	***			*	*
Road to Success	19.06	***	—	***	***	***	***
New Modern Chinese	3.25	**	***	—		*	
Practical Chinese	6.24		***	***	—		**
Boules de neige	4.58		***			—	**
Chinesisch	0.84	***	***	***	***	***	—

Negative conjunctions

		CRIE-CFL	Road to Success	New Modern Chinese	Practical Chinese	Boules de neige	Chinesisch
		0.43	3.75	0.80	0.52	1.88	0.49
CRIE-CFL	2.25	—	***			*	
Road to Success	9.35	***	—	***	***	**	***
New Modern Chinese	1.70		***	—			
Practical Chinese	2.70		***		—		
Boules de neige	1.53		***			—	.055
Chinesisch	0.50	***	***	**	***	*	—

Note. CRIE-CFL = Chinese Readability Index Explorer for Chinese as a Foreign Language.

* $p < .05$, ** $p < .01$, *** $p < .001$.

华语电影在高年级中文教学上的应用：以共享互动式教学模式的设计为例 (The Application of Films in Advanced Chinese Language Courses: A Cooperative-Collaborative Learning Model)

蔡晶晶
(Cai, Jingjing)
和理大学
(College of the Holy Cross)
jcai@holycross.edu

陈素宜
(Chen, Su-I)
克莱姆森大学
(Clemson University)
schen@clemson.edu

摘要：本文从电影和语言课程的结合出发，延伸探讨适合以电影为教材的教学平台以及学习模式。笔者以电影《我不是药神》和微电影《感染者为 0 的城市—南京》为例，展示电影在四年级中文课程中的应用，并以 Google Docs, Google Sheets 和豆瓣作为共享互动平台，设计合作式学习和协作式学习任务，举例阐释两种互动交流模式在教学应用上的优缺点，并分享在高年级中文教学中以电影作为语言教材的一些心得及方法。

Abstract: This article focuses on the use of films, teaching platforms, and teaching models in a fourth-year Chinese language class. Using the film *Dying to Survive* and the documentary short film *A City with Zero New Cases of COVID-19* as examples, the authors demonstrate the effectiveness of using films as major course materials, and using Google Docs, Google Sheets and Douban as sharing and interactive platforms in language courses. The article also highlights the advantages and disadvantages of cooperative and collaborative learning models, shedding light on the versatility of teaching methods in advanced Chinese language courses.

关键词: 电影、高年级中文教学、 共享互动式学习、合作式学习模式、协作式学习模式

Keywords: Films, advanced Chinese language learning, sharing and interactive learning, cooperative learning, collaborative learning

1. 前言

在高年级中文课堂运用电影教学已经司空见惯，但是该如何平衡电影内容教学与语言教学，还是众说纷纭。因为电影内容的丰富性，学生可以从电影中学到多元

的语言和文化知识。学生通过参与讨论和生生互动不仅能增进对各种不同观点的理解，而且能锻炼他们在不同语境下使用语言的能力，从而推动内容及语言学习。本文将介绍如何以共享互动模式，用电影作为教材进行四年级中文语言课的教学设计，以此提供一个结合电影内容与语言教学的案例， 以供同侪批评参考。

2. 文献综述

2.1 以电影为语言教材的相关文献综述

在语言课上使用电影能有效地提升学生的学习动机与兴趣是学界已有的普遍看法。多位学者的研究指出，外语电影可以让学生更轻松地学习第二语言和文化习俗，并增强学生对他们学习语言能力的信心（Kitajima & Lyman-Hager, 1998; Kramsch, 2004; Sundquist, 2010; Wood, 1995）。Garn（2012）指出，“以内容为导向的高级语言课程 （advanced ‘content’ courses in the language)，特别是以电影和语言结合的课程……为了解另一种文化提供了独特的窗口，这将极大地帮助学生提高学习动机以及知识和语言水平”（p.40）。Harrison（2009）认为在语言课上利用电影教学可以帮助学生发展文化认知，也可激发学生参加海外学习项目的欲望，甚至可能引起对目标语（target language）的进一步学习的兴趣。学生对目标语的兴趣正是学生参与课堂互动的原动力。Kabooha（2016）的研究表明，相较于使用教科书或其他学习材料，使用电影作为教材时，老师发现学生在课堂中的参与和互动明显地增加了，而且学生也认为使用电影有助于发展他们的互动技巧。

相比传统的语言教材，在语言课上使用电影拓宽了语言课教学的文化范畴。Sundquist (2010) 认为使用电影授课的外语课实际上可能使语言学习者“超越其教科书所涵盖的主题，从而洞悉文化之间的异同”（p.130）。众所周知，语言学习和文化学习是不可分割的（Chen，2009）。很多学者在研究中指出电影是一种增强语言和文化认知的有效办法（Bueno, 2009; Hughes, 2019；Ning, 2009; Sturm, 2012; Zhang & Yu, 2008）。Luo （2018）也指出运用电影来加强中文语言及文化学习的优点，即中文电影丰富的文化背景可以展现中国固有的思维方式、风俗习惯、价值观、社会结构和地理特征的多样性等方面。

因为电影本身的特质，学者们对电影在文化和语言教学上的桥梁作用充满信心。一方面，电影是视觉的艺术。Wood（1995）认为电影的关键特征是它们在视觉和言语上描绘了现实。Ning（2009）认为中文和英文之间语言上的鸿沟可以通过电影来填补，这是因为影像的呈现能帮助观众理解电影的内容。 Sturm （2012）认为电影提供了与目标语和文化的独特链接，因为它是视觉的、真实的、容易获得的，而且对于数字时代的学生来说，他们已经习惯了多媒体环境，因此电影非常有吸引力。另一方面，电影中使用的语言具有真实性。Zhang（2011）指出电影里口语表达的真实性以及电影里丰富的视觉和文化元素都是学习文化观点不可或缺的渠道，所以为了实现语言和文化相结合的教学目标，在语言课上使用电影是必不可少的。但是，如果直接采用为教学而专门拍摄的视频也有不足之处。比如，Yu（2009）的研究

表明，参加教学实验的学生普遍反映，相较于传统的教科书以及专门为教学设计的视频，电影中那些在真实情境中自然使用的语言更适合作为语言教材。

从语言教学角度上来看，有学者认为影片中的真实情境所提供的线索可以更全面地促进学生语言技能的学习，比如帮助学生理解主题，识别生词的含义，进而让他们能够在各种不同语境中使用适当的语言（Chen, 2009；Kitajima & Lyman-Hager，1998）。在 Ghajar et al.（2018）的著作中，很明显地将电影当作促进词汇教学的有效工具。他们认为随着逐渐熟悉不同的电影类型，学生不但学习了新词汇而且在写作中应用了这些词汇。Kabooha（2016） 则进一步指出学生还从电影中学习了传统教科书中不常有的俚语、非正式短语和非语言沟通。电影除了可以促进词汇教学以外，Hayati & Mohmed （2011）的研究结果还表明，在放映电影时使用字幕可以帮助学生提高听力水平，而在听的同时阅读字幕也可以促进阅读能力的提升。至于字幕是选择目标语还是母语，他们认为这取决于学生的语言程度。低年级适合使用母语字幕，中年级则推荐采用目标语字幕，而听力水平已经达到极高水平的高年级学生，则不推荐使用字幕。

综上所述，在语言课中以电影为教材，一方面可以帮助学生更精确地掌握语言的使用，实现语言学习的目标，另一方面也可以通过电影对文化丰富的视觉呈现，促进学生对目标语文化的了解和兴趣，进而提升学生对语言及文化学习的信心。

2.2 共享互动式学习：合作式学习与协作式学习

笔者在使用电影教学的四年级中文课程上采用了共享互动式学习法。顾名思义，共享互动式学习指的是通过共享学习资料，师生或生生在互动中相互学习，相互借鉴，共同完成教学目标的教学方式。本文强调的共享互动式学习法主要指的是生生之间的互动，它包含两种模式，一是合作式学习模式（cooperative learning model），二是协作式学习模式（collaborative learning model）。

根据 Oxford（1997）的定义，合作式学习模式指的是在课堂上通过培养学生之间的相互依赖来促进学生认知和社交能力发展的学习技巧，而协作式学习模式则是以“社会建构主义”（social constructivist theory）为理论基础，把学习看作一种在社会框架下的知识建构，因此鼓励个人通过涵化（acculturation）融入学习群体当中的个人哲学。合作式学习模式在指导学生完成群组活动时比协作式更直接且结构更清楚。Johnson et al.（2007）强调学生应遵循五个要素来一起完成共同目标。这五个要素指的是正面的互相依赖（positive interdependency），个人承担自己应负的责任（individual accountability），互相鼓励式的互动（promotive interaction），适当地使用社交技巧（appropriate use of social skills），和全组协同向目标前进（group processing）。与之不同的是，根据 Andreu-Andrés（2016）的描述，协作式学习模式并没有合作式学习模式那样有明确的任务解决办法，它只有开放式的复杂任务，而且老师只扮演协助者和推动者的角色，不再是课堂上的权威。

Panitz（1999）也指出合作式学习模式是通过学习者之间的分工来完成任务，从而实现学习目的的一种教学方法。学生在老师的指导下以小组形式活动，从而实现共同的学习目标。合作式学习模式被认为比协作式学习模式更有条理，关于课堂活动方面老师的规定更多，对学生如何在小组中一起活动的指导建议也更多。与合作式学习模式相比，多位学者都指出协作式学习模式指的是具有相同学习目标的学习者共同参与，努力来构建知识和解决问题（Dillenbourg, 1999；Roschelle & Teasley, 1995）。协作式学习模式给了学生更多的自主性和自由度，因此老师的参与度也相对较低（Kao, 2012）。

两种学习模式也存在着一些相似之处。无论是合作式还是协作式，都是在群组中进行的活动。Johnson et al.（2007）指出，群组活动的精髓在于成员之间基于共同目标的相互依赖。这种相互依赖导致整组变成一个“动态的整体”（dynamic whole），即每个成员的变动都会影响全组整体的变动，而小组成员之间的内在张力则能促进全组共同目标的完成。有多位学者指出两种学习模式都通过小组活动来最大限度地提高学习者自己及同组学员的学习能力, 都强调学习者之间直接的正面关系和相互支持，而不是他们之间的竞争关系（Johnson et al., 2007；Kirschner, 2001; Vuopala et al., 2016）。研究也表明，学生珍视在一起学习的机会，因为合作与协作活动可以激发和增进他们对内容的理解，增强学习者的参与度，从而实现深度学习（deep learning）的教学目的（Baeten et al.，2010；Cavanagh，2011; Herrmann, 2013；Johnson et al., 2007；Vuopala et al., 2016）。

Liang et al.（1998）指出，合作式和协作式学习模式都能增加学生语言输入和产出的机会，从而极大地促进学生的第二语言学习。但是 Liang 也发现了这种共享互动式学习模式的缺陷，他认为学生在小组活动的互动环节，往往倾向于使用简单的口语，而忽略对复杂、高级语言的使用。Johnson et al.（2007）则指出这两种学习模式的优点在于小组活动能够让学生在互相交流中观察并发现较优秀的组员，以他们为榜样，从而激励其他组员向榜样看齐，以期有更好的表现。Cohen（1986）也发现了群组活动中榜样的重要性，他主要聚焦有双语背景的学生在采用这两种模式的语言课中的正面作用。比如在英语为第二语言的课上，有西班牙语和英语双语背景的学生可以起到桥梁作用，帮助只会说西班牙语的学生学英语。Wong-Fillmore et al.（1985）也肯定了双语背景的学生在合作互动中不可或缺的作用，他们认为这些学生可以用两种语言给其他学生提供任务说明和信息补充。

3. 以电影作为语言教材：选片考量

笔者在长期观看华语电影以及参加电影小组的定期讨论后，积累了一个小有规模的华语电影库[1]。根据个人对用电影教中文的经验，笔者选出 27 部适合语言教学

[1] 此处感谢陈雅芬教授组织的 CLTA 电影兴趣小组全员。这个电影小组是由 11 位在美国大学及高中任教的中文老师组成，成员的专业背景分别为中文语言教学，语言学，文学和电影等方面。这些老师都有用电影教中文语言和文化的经验。笔者从定期的华语电影讨论中获取了很多的电影相关知识。

的影片（参见附录 1）。在笔者教的四年级中文课中选用了其中的 3 部故事片，再加上应时事需要选用的 1 部微电影纪录片（请参见表 1）。

表 1 电影列表

电影名称	电影年份	电影主题
青春派	2017	高考/早恋/师生关系/家长和考生
我不是药神	2018	罪案/社会/现实/人性
找到你	2018	女性/职场/婚姻/家庭
感染者为 0 的城市—南京（微电影）	2020	新冠肺炎/传染病防治/衣食住行

笔者之所以选用这几部片有以下几个考量：

第一，选用能生动呈现现实语境的电影。电影能呈现两种层面的真实，一是事实层面的真实，比如纪录片中报导的真人真事。《感染者为 0 的城市—南京》这部纪录片反映了当时新冠疫情在全球肆虐的状况，真实地呈现了南京抗疫措施的方方面面，给学生提供了客观的讯息。另一种是建构出来的模拟真实，如一般电影中呈现的生活化的场景给人的真实感。透过电影，学生可以身临其境地体验到电影中情景。比如《我不是药神》是由真实事件改编的故事片，它有效地结合了两种层面的真实，既有事实层面的，也有建构出来的。而影片如《找到你》和《青春派》虽然是完全虚构的故事，但都有现实基础，即女性的现实生存困境和中国高考考生面临的压力，学生也能有代入感或产生文化比较的心理。

第二，选用善于利用声光和剪辑手法来烘托气氛或交代剧情的电影。比方说，在《我不是药神》中有一个片段是程勇被法庭判刑后，从法庭到监狱的路上遇到了无数的病友来夹道送别（参见图 1）。因为这个电影场景有了恰如其分的背景音乐、灯光、脸部特写等，所以特别令人感动。电影中甚至用了剪辑手法让已经去世的吕受益和彭浩梦幻般地出现在欢送的病友中。这让程勇感动到落泪，觉得自己当初不惜犯法来救人的选择是值得的。借由这个片段，学生能为程勇和病友们的深厚情谊所感动，在课堂讨论能畅所欲言, 提高参与度。再比如在《找到你》中有一个在医院的片段（参见图 1），利用蒙太奇手法呈现了两个女主角生命中的第一次交集，通过几个高效率的闪回镜头，为之后的剧情展开做了充分的说明，让学生更能与人物产生共情。

图 1《我不是药神》（左）和《找到你》（右）影片截图

第三，选用场景中包含丰富文化讯息的电影。比方说，在《我不是药神》中有一个程勇出场的片段，这一片段伴随着一段典型的印度音乐缓缓展开，通过一系列的特写镜头描绘了程勇的保健用品店内场景（参见图 2）。观众可以看到店里摆满了各种文化物件，有来自异域的文化物件，也有中国本土的文化物件。程勇的桌上就摆放着一尊来自印度的女神像，他办公桌的边上杂乱地摆放着各种印着各国性感女郎的成人用品，而角落里则供奉着一尊中国财神爷的塑像。这一看似不经意的场景其实给学生提供了丰富的文化信息，也预示了接下来程勇将在中印两国之间扮演一个售卖仿制药的掮客角色。再比如在《青春派》中经常出现的教室场景，中国学生被桌上成堆的书籍包围，直观体现了高考的压力（参见图 2）。美国学生看到这些场景后，能自然地跟自己申请进大学的经历相比，从而在文化讨论中言之有物。

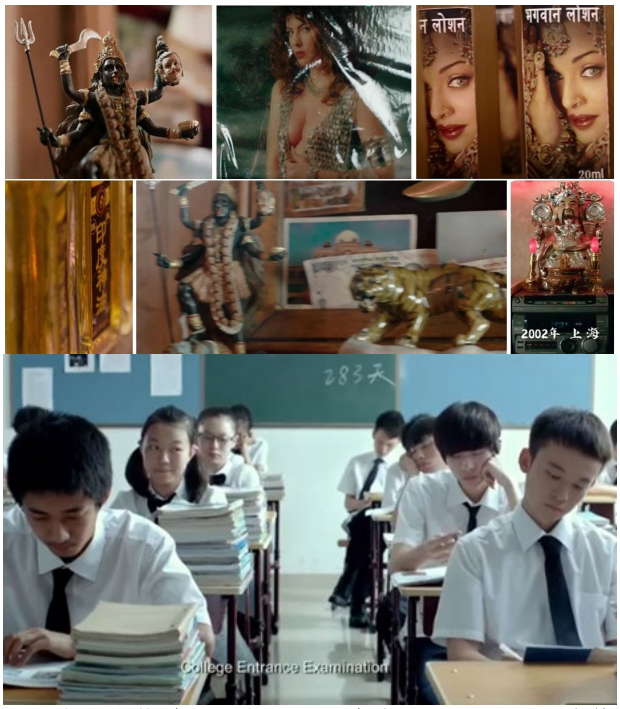

图 2《我不是药神》（左）和《青春派》（右）影片截图

第四，选用包含多种语体的电影。比方说，《我不是药神》这部电影既有非常市井的俚语，也有在正式场合如法庭使用的正式语（参见表 2）。表 2 中吕受益和程勇的对话还原了生活化场景中的语言。因为片中吕受益和程勇在组队卖印度仿制药的过程中，建立了深厚的情谊，所以他们的对白非常诚恳、亲切，用的语言也是很地道的日常生活表述，非常适合让学生听纯正的非正式口语，继而模仿并在适当的语境中使用这些词汇。此外，这一段话也说明了语言应用与人际关系之间的联系，学生能了解到在一般情况下，只有跟关系很熟的平辈或晚辈才能使用俚语和非正式的表述。比方说，吕受益跟程勇对话的时候，他用了“搞不好我能当爷爷啊”这样口语化的表述。表 2 中程勇律师和审判长的发言则展现了用电影来教正式语的范例。这一段是关于法庭对程勇案的判决，程勇的辩护律师先为程勇申辩，再由审判长来宣判法庭对程勇的最终判决。由于这是一个非常正式甚至有些严肃的场合，而且是由两个专业性非常强的人物所说的话，所以这两段话里包含了很多比较正式的表述，比如“而非、陈述完毕、倾家荡产、给予” 等。

表 2《我不是药神》非正式语和正式语的比较

非正式语：吕受益和程勇的对话	
吕受益	我刚查出病的时候啊，他妈已经怀他五个月了，那个时候啊天天想死，结果他一出生，我第一眼看到他就不想死了，就想听他叫声爸爸，但是现在好了，有药了，也有钱了，他要是早点结婚， 搞不好我能当爷爷啊。
程勇	那肯定的啊！
正式语：程勇律师和审判长的发言	

程勇律师	我的当事人虽然触犯了法律法规，但是一年多来有近千名的慢粒白血病人，是通过他代购的药保住了生命，格列宁在全球的定价，如此高昂，多少病人倾家荡产也负担不起，试问他们这样定价，真的合理吗？我们必须要清楚，程勇的主观意愿是救人，而非盈利，我的陈述完毕。
审判长	本庭宣判，被告人程勇，犯走私罪，销售假药罪，犯罪证据充足，犯罪事实成立。同时，对程勇帮助病人，购买违禁药物的行为，给予一定程度的理解。综上，判处被告人程勇，有期徒刑五年。

第五，选用发音相对纯正的电影。笔者选用的 4 部电影均以普通话发音，方言相对较少，电影台词可以直接作为听说材料。学生听到的是在现实生活中词汇和语法的使用状况，包括最切合情境的发音、语调、语速和口气，为学生在现实生活中与母语者的沟通做了绝佳的临场准备。比如在《我不是药神》中有很多性格、生活遭遇完全不同的人物（参见表 3），有卖假药的张长林用讽刺的语气陈述穷人看不起病的惨况，有患有白血病的老太太用恳切的语气哀求警察对程勇卖印度仿制药的犯罪事实网开一面，还有程勇在法庭上自陈其贩卖仿制药的心路历程。学生可以通过认真去听片中人物说话的方式，在口语练习中模仿他们的语调、语气和语速。

表 3 《我不是药神》中不同人物的语言表达方式

 领导，求你个事啊，我，我就是想求求你，别再追查印度药了，行吗？我病了三年，四万块钱一瓶的正版药，我吃了三年，房子吃没了，家人被我吃垮了，现在好不容易有了便宜药，你们非说他是假药，那药假不假，我们能不知道吗？那药才卖五百块钱一瓶，药贩子根本没赚钱，谁家能不遇上个病人，你就能保证你这一辈子不生病吗？你们把他抓走了，我们都得等死，我不想死，我想活着，行吗？	 听说你这次不挣钱，挺仗义的，不过哥得劝你几句啊，我卖药这么多年，发现这世上只有一种病，穷病，这种病你没法治啊，你也治不过来，算了吧。  我犯了法，该怎么判，我都没话讲，但是，看着这些病人，我心里难过，他们吃不起进口的天价药，他们就只能等死，甚至是自杀，不过，我相信今后会越来越好的，希望这一天，能早一点到吧。

第六，选用有中英双语字幕的电影。在阅读训练方面，这 4 部电影提供了阅读字幕与凝听语音材料同步进行的方式，学生在看电影时，不仅在听觉上有了语音输入，同时在视觉上也有文字输入。在文字输入方面，这 4 部电影既有中文字幕也有英文字幕，提供了新的词汇及语义配合。因此用这些电影教词汇可以把词汇的字形、字义和字音在适当的语境中同时呈现给学生，既有效率，又能加深学生对词汇的初步印象和理解，比如，在《感染者为 0 的城市—南京》里有一个片段出现了“口罩”这个生词（参见图 3），学生通过看电影里的“口罩”实物，阅读“口罩”这个词的中英文字幕，听“口罩”的发音，以及观察地铁上每个人都戴口罩的情境，可以帮助学习“口罩”这个词。

图 3 “口罩”一词在《感染者为 0 的城市—南京》中的呈现方式

4. 教学设计与分析

以下笔者设计的教学内容所针对的对象为在美国大学选修四年级中文课的学生，总人数有 13 个。选修该课程的学生大部分是双专业的，其中必有一个专业是中文，因此他们学中文的动机普遍比较强，对各种中国社会、文化议题都非常感兴趣。全班有 2 位华裔学生，非华裔学生里面有 5 位曾在中国大陆留学一个学期或一个学年，剩下 6 位学生则是从一年级开始都在笔者所在的学校学习中文。因此学生的语言程度跨度较大，平均在中级中（Intermediate-mid）到高级中（Advanced-mid）之间。针对这个班学生的语言水平，以及与学生交流中了解到的他们感兴趣的电影题材，笔者选出了两部社会事件主题、一部青春校园主题的剧情片，以及一部新冠疫情相关的微电影纪录片作为教学材料（参见表 1）。

笔者采用共享互动式教学法的两个模式，即合作式和协作式学习模式来教这门课。至于为什么要采用这两种学习模式来教以电影为教材的中文课有以下考量：

其一，电影文本本身没有词汇表，通过共享平台，例如 Google Sheets，让学生在线同步合作完成词汇表的创建，不仅能为本课提供词汇表，而且能提高学生自学能力。此外，到了高年级，学生的语言能力参差不齐，而传统的词汇表有统一的标准和限制，很难满足学生多元的需求。如果让学生看完电影后依照自己的背景、语言能力和兴趣来挑选想学的词汇，这可以弥补由老师、课本制定生词表，学生学习意愿不足的缺陷。

其二，电影作为真实语料（authentic material），涵盖了诸多可供讨论的议题。每位学生从一部电影中所学的语言、文化知识和观点各有不同。通过生生互动，在合作/协作完成任务的过程中，学生不仅可以互相鼓励、互相学习，还可以通过分工把任务化整为零，在短时间内高效地完成任务。

单部电影的共享互动式教学实例，笔者以“4.1 以合作式学习为主导的教学法：以电影《我不是药神》为例”和 “4.2 兼具合作式与协作式学习的教学法：以微电影

《感染者为 0 的城市—南京》为例”展示如下。

4.1 以合作式学习为主导的教学法：以电影《我不是药神》为例

《我不是药神》是 2018 年在中国上映的一部现实主义剧情片。该片讲述了神油店老板程勇从一个自顾不暇的利己主义者，一跃成为万千白血病患者的救星的故事。这部电影的主题涉及当下中国老百姓最关心的民生问题之一：“看病难，看病贵”问题，具有一定的现实性，因此给学生提供了一个了解当代中国的绝佳视角。此外，该片的语言运用相当多元，既有书面的、正式的用法，也有口语化、生活化的表达，方便让学生学习不同语境下不同语体的运用。下面笔者结合这部电影的主题以及四年级语言课的要求，谈一谈具体的教学目标和教学设计。

《我不是药神》教学目标

总目标：在完成本课学习后,学生可以

1. 就豆瓣影评写一篇 250-300 字的书面回应，在文中使用至少 10 个生词以及 3 个回应常用的结构
2. 采访母语者对于看病和医疗保险的看法，在课堂上做 3 分钟的口述采访总结
3. 为 XX 病患者做一个 3 分钟的募捐演讲，使用至少 10 个生词和 2 个新学的成语或俗语

子目标

1. 掌握讨论医药议题所需的 40 个基本词汇
2. 掌握 5 个语法和句型，包括：即便是……也，verb 不 complement, 更何况，作为……，一旦……
3. 掌握 9 个成语及俗语的用法，包括：劫富济贫，救人一命胜造七级浮屠，我不入地狱谁入地狱，人命关天，谋财害命，自投罗网，无理取闹，伤天害理，义不容辞
4. 就医药相关问题写一个 250-300 字的豆瓣影评书面回应
5. 就医药相关问题进行采访，再做一个 3 分钟的口述采访总结
6. 就医药相关问题做一个 3 分钟的募捐演讲

为了实现以上的教学目标，笔者用 10 个课时来完成每部长电影的教学任务（参见附录 2）。每个课时五十分钟，每个星期上三次课，外加两个由助教上的词汇和语法练习的小班课（practicum）。为了能够善用课堂时间，让学生充分互动、交流、练习，笔者采用了翻转课堂（flip classroom）的教学理念，给学生设计每一课时的预习单，让学生在进入课堂前已经有了一定的心理和语言上的准备。每个预习单大概预期学生在 1 个小时内完成，至于学生合作制作词汇表的预习功课，大概预期每个学生在 15 分钟左右完成自己的部分。课前预习任务共占总成绩的百分之三十。预习单的设计理念是要充分考虑内容教学和语言教学的平衡，所以在设计练习题的时候，笔者针对每个电影片段（参见附录 3）设计了阅读理解题、词汇填空题、句型应用题和综合练习题（参见附录 4）。

接下来笔者将以表 4 展示课程的流程，以及如何在各个课时结合 Google Docs, Google Sheets 和豆瓣来完成合作式教学设计：

表 4《我不是药神》课程流程表

一、分享：豆瓣短评 （第一课时） 学生课前准备：看完整部电影，依据笔者准备的思考题在豆瓣上分享自己的短评（参见附录 5）。 课堂上：围绕思考问题和学生写的豆瓣短评展开课堂讨论。
⇩
二、合作分享：生词表、对电影不同主题的理解和看法 （第二-五课时） 学生课前准备：学生根据自己的意愿选出想学的生词，合作共享在 Google Sheets 上。完成预习单上的作业。 课堂上：在笔者讲解语法后，学生应用本课的语法和生词回答预习单里关于电影内容的问题。学生先分组讨论，再按笔者要求回答问题。
⇩
三、分享：写作、采访及口试 （第六-八课时） 从第六课时开始到第八课时，课堂教学以任务导向型（task-based）为主。学生在课前完成写作、采访及口头报告的准备，在课堂上与同学分享自己任务的成果。

第一课时的目的就是让学生在整体上理解电影的人物和剧情，为接下来第二到五课时的主题讨论做好铺垫。刚看完电影，学生的课堂讨论往往非常热烈，基本上每个人都能根据自己写的豆瓣短评，发表自己的看法。此处笔者想说明的是，豆瓣短评不同于长影评，后者通常要求写作者具有一定的电影专业知识，并以此为基础进行评论，而豆瓣短评只相当于简单的个人观后感，非常主观且字数限制在 150 字以内。因此写豆瓣短评的练习只是为了让学生写出看完电影后的感想，为课堂讨论作准备，至于内容完全由学生决定。在课堂讨论的时候，鉴于有些学生不主动发言，笔者会让他们先分组讨论，练习怎么组织语言，然后再面向全班发表看法。笔者在学生发言之后，会负责词汇和语法上的纠错，也会根据实际课堂互动提出进一步探讨的问题。

第二到五课时是笔者根据电影中选出来的四个主题设计的。以《我不是药神》这部电影为例，笔者一共选出了以下四个主题：程勇五人小组群像（人物介绍），程勇的心路历程（角色发展），孰对孰错：警察、药厂和病人的不同立场（观点分析），情与法的天平：如何判决程勇的罪？（观点讨论）。根据这些主题，笔者要求学生观看指定的电影片段，一般是每个主题 2-3 个片段，每个片段 1-2 分钟。学

生观看的电影片段都含有中英双语字幕。笔者预期学生能理解片段中大部分的意思，即使对具体的词汇和语法使用还存在疑问，也能完成预习单上的练习。学生看完电影后的课前练习包括三个方面，一是选出 3 个自己认为重要的生词，分享在班级共享的 Google Sheets 上，选中的生词不得与他人的重复。除了列出生词以外，还要列出拼音、英文释义和影片中的中英文例句。每人 3 个生词，笔者再根据教学目标从学生选词中精选出每部电影 40 个生词。二是在课前学习同学分享的生词，为课堂讨论提供必要的词汇准备。三是依据大家合作完成的生词表，完成预习单的作业。

至于课前练习的评量，生词部分共 1 分，为了鼓励学生参与这样的合作式任务，他们只要完成即可得 1 分。稍有错误，笔者帮忙修改后不扣分。预习单共 9 分，根据学生的表现优劣给分。所有的课前练习占学生总成绩的百分之三十。

第六到第八课时，笔者给学生设计了一系列的重点任务。第六课时的重点在于训练学生的写作能力（writing skill）。笔者要求学生看一篇名为“药化科研工作者千思万虑的两小时”的豆瓣影评后写一篇 250-300 字的回应（参见附录 6）。这篇影评是笔者根据学生的语言程度在豆瓣原文的基础上进行修改后的版本。笔者要求学生在回应文章中至少使用 10 个本课的生词以及 3 个回应常用的结构。

第七课时的重点在于加强学生与母语者的面对面交流的口语能力（interpersonal skill）。笔者要求每个学生采访 1 个中文母语者，了解采访对象对中美两国看病和医疗保险的看法，预习单中笔者提供给学生一些跟主题相关的问题作为参考，也要求学生自己想出 2 个采访问题。采访的过程全程录音，作为学生的口语作业。课堂上每个学生在 3 分钟内依次总结自己的采访内容，笔者根据学生的采访总结再问学生 1-2 个问题，并根据学生表现评分。

第八课时的重点在于培养学生用正式语言演讲的能力（presentational skill）。笔者要求学生以“为 XX 病患者募捐的演讲”为题在课堂上作 3 分钟的口头报告。笔者给学生提供了一个演讲稿的范文，供学生作格式上的参照。正式报告前一天的小班课，学生会先跟助教练习。正式演讲的课堂上每个学生作完报告后，笔者现场给出语言运用上的反馈。

因为每个课时之间都相隔一天甚至跨周末，而且在两个课时之间都有小班课可以给学生做练习和准备，所以学生都能按时完成任务。在评量方面，写作文和采访母语者都是课前预习作业的一部分，如前所述，所有的课前预习占总成绩的百分之三十，而整个学期的四次口头报告占学生总成绩的百分之十。为了鼓励学生做好准备，学生在准备时，可以事先写一个演讲稿，交给笔者修改，修改的过程不计分。在所有学生讲演完后，全班投票选出 2 个最佳演讲者，获得本次口头报告满分奖励。因为演讲的过程全程录音，课后笔者根据录音，再给学生一份详细的书面反馈并给学生评分。笔者随时在预习单作业的批改和课堂教学上观察学生的表现并给予及时的提醒和纠正。除此以外，笔者会定期组织集体纠错环节，把学生的常见错误总结后在课堂上说明。

4.2 兼具合作式与协作式学习的教学法：以微电影《感染者为 0 的城市—南京》为例

微电影《感染者为 0 的城市—南京》是由日本导演竹内亮（Takeuchi Ryō）执导的，拍摄于 2020 年新冠疫情在中国大陆得到初步遏制之后，以南京这座过去一个月感染者为零的城市为例，探讨中国大陆广泛采用的抗疫措施及其有效性的一部微电影纪录片。该片片长约 12 分钟，分为六个章节。每个章节的主题分别是外食、交通、户外、公司、网课和隔离生活。以下为本课的教学目标。

《感染者为 0 的城市—南京》教学目标

总目标：学生在完成本课学习后，可以对新冠肺炎和传染病相关议题进行全组 6 分钟的口头陈述和 250-300 字的书面报告。

子目标

1. 掌握讨论新冠肺炎和传染病所必须的 20 个词汇
2. 针对微电影内容与同学两两配对，在共享平台上完成 4 次书面问答，使用至少 5 个生词
3. 针对波士顿新冠疫情状况，学生 4 人一组，在课堂上完成 6 分钟的口头报告，包含三个主题：户外，外食，交通
4. 针对波士顿新冠疫情状况，学生 4 人一组，在共享平台上完成 250-300 字的书面报告，包含三个主题：在家隔离，核心行业（超市、快递、药店）继续办公，在家上网课

这部电影一共六个主题，以两个课时完成，每个课时讲三个主题，分三部分循序进行，具体步骤请参考表 5：

表 5《感染者为 0 的城市—南京》课程流程表

步骤一、合作学习之制作生词表 课前准备：笔者根据学生程度分三个混合能力小组，每个主题由一组学生负责。每组共同完成每个主题的生词表。
⇩
步骤二、合作学习之两两问答 课前准备：延续步骤一的分组，学生在预习单上跨组两两交换提问及回答。
⇩
步骤三、协作学习之小组活动完成短文 课堂活动：笔者完成词汇和语法讲解，学生延续在步骤一的分组协作完成指定的写作任务。

从上表可以看出，每个课时都分三部分进行。第一部分和第二部分都是学生的课前预习，具体来说，第一部分是学生分三组在 Google Sheets 上合作共享生词，每位学生负责 3 个生词，笔者再根据教学目标从中精选出 20 个生词供全班学习。笔者会在上课前一天晚上 9 点修改学生合作共享完成的词汇表，帮学生纠错，学生再根据笔者修改过的词汇表来完成第二部分的问答环节。第二部分是学生在笔者的作业指示下完成 Google Docs 上的两个互动问答：每人针对别组的内容提一个问题，每人再回答一个别人的问题（参见表 6）。笔者是根据学生程度来预先给学生分好组的，每组都有不同程度的学生，且笔者已经将他们两两配对，明确列出每个问题的提问者和回答者。2 个华裔学生和 5 个出国留学归来的学生，因为程度高于其他学生，所以笔者把他们分散在不同的组里面，希望他们能在各自组中起带头作用。表 6 中为了保护学生隐私，用“学生 1，学生 2”等替代学生的真实姓名，且实际配对在此省略。

表 6《感染者为 0 的城市—南京》课前互动交流之两两问答

时长	主题	提问者	回答者
1:55 - 3:25	外食	学生 1，学生 2，学生 3，学生 4，学生 5	学生 6，学生 7，学生 8，学生 9
3:26 - 4:50	交通	学生 6，学生 7，学生 8，学生 9	学生 10，学生 11，学生 12，学生 13
4:51 - 7:20	户外	学生 10，学生 11，学生 12，学生 13	学生 1，学生 2，学生 3，学生 4，学生 5

第三部分是在课堂上进行的，学生在 Google Docs 上自主协作完成短文写作（请参考表 7）。学生共同讨论，写出句子，再讨论句子之间的衔接，最后完成短文写作任务。

表 7《感染者为 0 的城市—南京》课堂小组协作活动

小组活动要求：假设明天你就要回波士顿，但是目前波士顿的新冠疫情还很严重，请你写一个注意事项列表，随时提醒自己在户外、外食、交通方面要注意些什么。这是一个小组活动（Group Project）, 请组员们互相商量，用 Zoom 沟通以后，写出你们这个组的注意事项，每组不少于 6 个注意事项，每个主题至少 2 个注意事项。

因为这个单元的第二课时和第一课时形式相同，只是讨论的电影主题改为公司、网课和隔离生活，所以此处不再重复列出分组细节。因为疫情的影响，这一个单元已改成上网课，所有的活动改在 Zoom 上完成。

这里要指出的是，学生在第三个部分第一次进入到协作式写作环节，因为是初次尝试，所以课堂上笔者利用 Zoom 的 breakout room 的功能，让学生分组一边讨论，一边完成协作式写作练习。笔者也依次进入每组的 breakout room，一边帮助学生答疑解惑，一边在 Google Docs 上观察学生完成短文的进度，随时提供指导和纠

错。在这个过程中，笔者要求学生用中文进行讨论和向笔者提问，笔者在每个breakout room里待一段时间，观察学生的互动写作情况，如发现有不太主动参与的学生，即刻鼓励那位学生说出自己的看法。

至于评量方面，由于第一和第二部分都是学生在课前要完成的预习工作，所以都算作一次课前的作业，生词 1 分，问答 9 分，整个学期的课前作业占总成绩的百分之三十。而第三部分算学生的课堂参与和表现的分数，整个学期的课堂表现占总成绩的百分之十。

微电影《感染者为 0 的城市—南京》的教学设计也为期末项目（Final Project）做好了铺垫。学生通过这一课的学习，初步了解了跟新冠疫情相关的生词以及从合作到协作完成任务的模式。接下来的期末项目，笔者在微电影的教学设计的基础上，分三个课时帮助学生在课内和课外交流协作完成期末项目“新冠疫情与 X 国”[2]。具体步骤请参考表 8：

表 8 期末项目：“新冠疫情与 X 国”流程表

步骤一：课前自学可能会用到的生词 由笔者提供补充生词表，学生课前先自学所有生词，以备在课堂上小组共同协作时使用
⇩
步骤二：课上协作找出与主题相关的资讯 （第一-二 课时） 学生 2-3 人一组，自行分组，找出与 X 国有关疫情现状及措施的资料，讨论后加以整理，选出可以用于报告撰写的资料。
⇩
步骤三：课后协作完成期末报告的撰写 课后，由原组学生用 Google Docs 完成期末报告的撰写，分工方式由每组自行决定。
⇩
步骤四：课上合作口头报告 （第三课时） 在课堂上，组员轮流向全班分享他们合作的成果。

以“新冠疫情与 X 国”为期末项目的主题是因为这个题目很具时效性，在学生完成期末报告的这个时段里，全球所有国家都还处在与新冠疫情抗争的关键期。这个

[2] 此处鸣谢 STARTALK team at the University of Virginia 在疫情开始之初给 CLTA 成员提供的线上教学培训，梁新欣教授在培训中的教学实践给笔者提供了期末报告选题的灵感。

主题不仅与学生的生活息息相关，而且学生也能较容易地找到大量的研究资料和最前沿的讯息，这能极大地激励学生的研究兴趣和表达欲望。

整个期末项目分为四个步骤，第一个步骤是由老师提供新冠疫情相关词汇，帮助学生打好词汇的基础，其实学生在前面微电影那个单元已经练习过大部分生词，但老师补充了更多疫情讨论相关的生词，以备期末报告作更深入的研究之用；第二个步骤是学生自行结组 2-3 人一组，同组的组员在课堂上合作找出与期末报告主题相关的资讯；第三个步骤是学生协作共同完成期末报告的撰写，可以每人写几句话，也可以共同撰写。写完后笔者帮忙完成词汇和语法运用以及行文逻辑上的纠错与润饰。因为这是课后完成的，所以完全由学生自主分工，老师并没有全程观察以便随时介入，学生的自由度很高；第四个步骤是学生根据已完成的期末报告，练习口语的表达，在课堂上轮流向全班分享合作研究和写作的成果，报告 X 国的疫情现状及未来展望，并预测疫情过后的影响。

期末项目的评量分两部分，一是底稿的撰写，全组写完后提交一份。二是口头报告，由个人单独认领底稿的一部分完成。这两部分各占总分的百分之五十。因为底稿是由学生协作完成的，笔者的目的是鼓励全组学生以积极的态度参与到底稿的撰写当中，在写作的过程中向其他组员学习或成为全组的榜样，从而促进底稿的顺利完成，所以最后全组同分，而口头报告的部分则是由个人根据小组分工完成的，所以由笔者根据个人表现给分。期末项目占总学期成绩的百分之十。

《感染者为 0 的城市—南京》是兼具合作式和协作式学习模式的，而期末项目则是以协作式为主导的。之所以这样设计，是因为笔者觉得学生经过了一整个学期的合作式学习模式的实践，学生之间的互动交流已经非常自然，也非常熟练了，而且学生的语言能力经过一学期的训练也有了一定的提升。因此笔者觉得在期末的最后一个单元，可以尝试给学生更多的自由度和自主权，比方说让学生自由分组找队友，让学生从他们熟悉的合作式学习模式开始，逐渐过渡到对他们来说相对陌生且自由度更大的协作式学习模式。这对老师和学生来说，都是一次有意义的教学实验。在实验的过程中，双方都可以探索学生协作的效果，同时也为下一个阶段的教学设计铺路。

5. 教学启示与反思

5.1 应用华语电影的启示和反思

从教学效果来看，大部分学生在教学评鉴中提到他们对以电影为教材的教学方式持正面的看法。在他们看来，首先电影不仅有文字，还有影像和声音，这能提高他们的注意力，让他们在学习过程中更加专注；其次电影里的人物使用的语言很自然，这增加了他们在自然语境中沉浸式学习的机会；第三点也是最重要的一点，电影有故事性，一个个引人入胜的故事情节极大地提升了他们对中文学习的兴趣。

至于电影本身，学生对于电影的片长和类型并不太挑剔。比如他们觉得像《我不是药神》这样 2 个小时左右的故事片和像《感染者为 0 的城市—南京》这样 12 分钟左右的微电影纪录片都有学习的价值。学生对它们的学习兴趣并无高下之分。对他们来说，电影的主题是否吸引人最重要。这班学生对于电影的主题有强烈的偏好性，他们普遍偏好社会议题的电影，比如《我不是药神》和《找到你》。因此笔者建议老师选片要更多元，更符合学生兴趣。如果条件允许的话，最好在教学大纲完成之前先在学生中做一个选片调查，选出学生最感兴趣的电影主题。

至于电影的字幕问题，学生反映他们看电影的时候比较喜欢有字幕，而且最好是中英双语字幕。因为这个班的学生程度跨度较大，所以本课的所有电影片源都包含中英双语字幕，确保学生能根据自己的语言程度，各取所需。建议老师在选电影的时候根据学生语言程度来决定采用什么语言的字幕。

5.2 采用合作式与协作式学习模式的启示和反思

采用合作式学习模式的最大优点是学生合作完成老师布置的任务，老师可以把一个看起来复杂的任务化整为零，因此分配给每个学生的任务并不多，学生反馈说他们比较能够轻松地完成自己的部分。同时，因为时间上的充裕，每个学生都能达到个人的较佳表现，而整体的任务也能得到较好的完成，请参见前文 4.1 提到的合作制作生词表的任务。

采用协作式学习模式的最大优点是让学生自由决定任务完成的方向及他们预计的成果。老师只提供大方向上的指引，并没有指定具体的任务细节。同组的学生可以自主选择他们感兴趣的主题，然后小组自己决定如何分工找相关资料来完成任务。相比合作式学习模式来说，协作式学习模式让学生的自主度更高，更可以激励学生的创造力和思辨能力，请参见前文 4.2 提到的期末项目。

从期末项目的完成结果来看，有的组完全采用协作式模式，有的组则采用协作式和合作式并用的方式。对于由程度相对较低的学生组成的小组来说，他们最后并没有完全采用协作式模式来详细讨论内容以及一起完成期末报告的撰写，因为在期末的时候各科作业都很多，时间很难协调，所以他们采用的是合作式与协作式并用的模式。他们先通过组内共同讨论决定主题和需要的细节，然后分工找资料，最后分工完成报告的撰写。因为把一个大项目分成了几个小任务，每个学生的研究都可以做得比较详尽，所以报告的整体水平还不错。但是，由语言程度比较高的学生组成的小组则反馈说他们完全采用了协作式的模式，先对主题进行了详细的讨论，然后在写作的过程中也没有分工，而是一边现场讨论，一边撰写报告。对他们来说，这样的协作式完成期末报告的方式既有效率，又能产出比他们个人独立完成的报告更好的成果。由此可以看出，协作式模式可能更适合语言程度整体较高的学生群体。总之，全班都对期末报告可自主决定学习模式的评价相当正面。

至于合作式和协作式学习模式的难点和不足之处：其一，相比完全由老师主导，以师生互动为主的教学模式，合作式与协作式模式更加鼓励生生互动。但是，把主

导权交由学生以后，师生互动相对减少了，这其实对老师的要求也更高了。建议老师随时根据生生互动的具体情况来调整相应的教学方向和细节，灵活把握老师对生生互动介入的程度。

其二，合作式和协作式学习模式都要求生生互动能够顺利有效地进行，但现实是生生之间也需要一定的时间去磨合，学期初可能会遇到一些困难，诸如生生之间沟通不畅或者意见不同导致的问题，比如，如果小组内有不合作的同学，会导致组内任务分配不均，最终影响任务完成的质量。建议在学期之初就增加学生小组讨论的机会，让学生之间通过分组增加熟悉度，也让学生尽早适应并培养对共享互动学习模式的认可度。

其三，由于高级中文课的学生组成语言程度差异较大，有华裔和非华裔的差别，也有具备留学经验和不具备留学经验的差别，因此在协作过程中，程度相对较低的学生可能会感到挫折，而程度较高的学生可能会觉得协作浪费时间，不如自己独立完成任务效率更高。建议老师根据实际情况多多鼓励学生并及早引导，让学生形成在共享互动中每个人都能受益的共识。

其四，从老师评量的角度来看，合作式和协作式学习模式在老师评量学生方面挑战较大。由于无法确定每个组员对任务的具体贡献，老师可能在公平、准确地评量每一位学生的表现方面会感到困扰。这也是合作式和协作式学习模式最大的不足之处。在此，笔者想抛砖引玉，提出一个未来针对评量的调整方案。以学生的期末报告为例，建议在布置期末报告的时候，要求学生在交报告时附上一份简单的贡献总结。总结可以用英文写成，简要说明自己在完成报告的过程中做了什么贡献。这个总结不但能够让学生提升参与的自觉性，而且对老师最后的评量也有一定的参考价值。如果时间允许，老师也可以在共享互动结束之后，跟学生面对面约谈，了解学生参与共享互动的反馈及修改建议，这将对未来活动设计及评量都有参考价值。

5.3 共享互动平台分析与启示

笔者在这个课程里有意识地使用了两个共享互动平台，一个是Google Drive, 另一个是豆瓣网。选择学生熟悉的 Google Drive 可以节约学生适应新的共享平台的时间。使用豆瓣是因为豆瓣是目前中国最大的影评共享平台，笔者想通过让学生和中文母语者共同使用一个讨论电影的平台，在现实语境中提升他们的阅读和写作技能。

针对这些平台的使用情况，笔者通过跟学生面对面讨论，还有看学生写的教学评鉴，发现他们普遍喜欢使用Google Drive（包括Google Docs 和 Google Sheets），因为他们对这个平台都很熟悉，觉得用起来容易且便利，因此对 Google Drive 大多持正面看法。但是他们对使用豆瓣持保留的态度。学生认为写豆瓣短评，以及写长评的回应对中文学习有一些帮助。比方说，有的学生提到由于豆瓣短评是有字数限制的，所以这强迫他们锻炼用精炼的语言来写概括（summary）和评论（comment）的能力。但是大部分学生觉得豆瓣的网页设计杂乱无章，对非母语者来说用起来比较困难，他们很难找到同侪的短评去阅读和评论。其次，对于一些语言程度较低的

学生来说，即便是笔者已经给他们指定相对切合他们语言程度的豆瓣影评来阅读，他们还是觉得这些影评难度较高。再次，有的学生反应豆瓣上没有人对他们的长评回应有及时的评论，这让他们觉得很难取得及时的反馈意见，所以对语言学习的帮助不大。

根据学生的反馈意见，笔者后期调整了“短评”和“回应”的共享平台，把在豆瓣上分享改为在 Google Docs 上分享，把教学目标也从让学生跟母语者通过写影评和回应进行交流改为同学之间的互相评论与交流，而且笔者也对豆瓣上的一些影评进行了编辑和修改，再把修改后的影评通过 Google Docs 分享给学生看。这样的调整虽然有违笔者原先的教学目标，但有两个明显的好处：一是学生可以直接看到同学写的短评，不需耗费时间去浩如烟海的豆瓣平台上搜寻他们能看得懂的短评；二是由老师出面去挑选适合学生阅读水平的长评，在修改后作为阅读练习，这样避免了学生直接阅读母语者写的长评的挫折感。总而言之，虽然开放豆瓣平台让学生直接去写“短评”和“回应”的尝试并不成功，但是笔者还是建议以后在用电影教中文的课上适当地使用豆瓣影评。因为虽然这个平台本身不太适合直接用于教学，但是平台上的内容如果善加利用，还是能给课堂补充很多真实语料。

6. 总结

总而言之，本文肯定了电影在高年级语言教学上应用的优势，证实了有效采用电影进行教学对中文学习的正面影响。其次，本文探索了共享互动模式对运用电影为教材的语言课的贡献及其重要性，并探讨了共享互动模式的优势和不足之处，给相关的研究提供了切实的实践经验和范例。第三，本文评估了包括 Google Docs, Google Sheets 和豆瓣在内的共享平台在高年级语言教学应用上的有效性，给现有的共享平台的优缺点做了一个总结。

未来笔者想继续探究采用电影作为中低年级语言课教材的可能性，同时继续扩大对其他共享互动平台的评估。此外，笔者还希望能进一步开发现有的共享互动模式的潜能，以及探索这种模式在其他年级（比如中低年级）和课程类型（比如不是以电影为教材的语言课）上应用的潜力，策划更多元的合作式和协作式任务，进一步激发学生的兴趣，提高学习效果。

参考文献

Andreu-Andrés, M. Á. (2016). Cooperative or collaborative learning: Is there a difference in university students' perceptions? *Revista Complutense de Educación*, *27*(3), 1041-1060.

Baeten, M., Kyndt, E., Struyven, K., & Dochy, F. (2010). Using student-centered learning environments to stimulate deep approaches to learning: factors encouraging or discouraging their effectiveness. *Educational Research Review*, *5*(3), 243-260.

Bueno, K. A. (2009). Got film? Is it a readily accessible window to the target language and culture for your students? *Foreign Language Annals*, *42*(2), 318-339.

Cavanagh, M. (2011). Students' experiences of active engagement through cooperative learning activities in lectures. *Active Learning in Higher Education*, *12*(1), 23-33.

Chen, A. M. (2009). Developing and studying the effectiveness of EFR annotations for Chinese language learners (Unpublished master's thesis). Brigham Young University, Utah.

Cohen, E. (1986). *Designing group work: Strategies for the heterogeneous classroom*. New York: Teachers College Press.

Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), *Collaborative learning: Cognitive and computational approaches* (pp.1-19). Oxford: Elsevier.

Garn, R. (2012). Teaching the five Cs with cinema. *Journal of the National Council of Less Commonly Taught Languages*, *12*, 37-71.

Ghajar, G, S., Hekmati, N., & Navidinia, H. (2018). Movie-generated EFL writing: Discovering the act of writing through visual literacy practices. *International Journal of Language Studies*, *12*(2), 51-64.

Harrison, L. (2009). Foreign films in the classroom: gateway to language and culture. *Journal of College Teaching and Learning*, *6*(8), 89-93.

Hayati, A., & Mohmedi, F. (2011). The effect of films with and without subtitles on listening comprehension of EFL learners. *British Journal of Educational Technology*, *42*(1), 181-192.

Herrmann, K. J. (2013). The impact of cooperative learning on student engagement: Results from an intervention. *Active Learning in Higher Education*, *14*(3), 175-187.

Hughes, H. Z. (2019). China through the lens: teaching Chinese language and culture through film. *Dimension*, 42-56. Retrieved from https://files.eric.ed.gov/fulltext/EJ1214286.pdf.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (2007). The state of cooperative learning in postsecondary and professional settings. *Educational Psychology Review*, *19*(1), 15-29.

Kabooha, R. H. (2016). Using movies in EFL classrooms: A study conducted at the English Language Institute (ELI), King Abdul-Aziz University. *English Language Teaching*, *9*(3), 248-257.

Kao, S. (2012). Examining the mediating effect of knowledge sharing willingness on the relationships between learning motivation, learning interaction, supportive learning platform, and learning satisfaction. *Commerce & Management Quarterly*, *13*(1), 75-98. [高淑珍. (2012). 以知識分享為中介變數探討學習動機、學習互動以及學習平臺對協同學習滿意度的影響. *商管科技季刊, 13*(1), 75-98.]

Kirschner, P. A. (2001). Using integrated electronic environments for collaborative teaching/learning. *Learning and Instruction*, *10*(1), 1–9.

Kitajima, R., & Lyman-Hager, M. A. (1998). Theory-driven use of digital video in foreign language instruction. *CALICO*, *16*(1), 37-47.

Kramsch, C. (2004). *Context and culture in language teaching* (Vol. 6). Oxford, Great Britain: Oxford University Press.

Liang, X., Mohan, B. A., & Early, M. (1998). Issues of Cooperative Learning in ESL Classes: A Literature Review. *TESL Canada Journal*, *15*(2), 13-23.

Luo, L. (2018). Applying film in teaching Chinese to enhance culture and language learning. *Journal of Chinese Teaching and Research in the U.S*. 124-132. Retrieved from https://www.clta-gny.org/journal/journal18_contents.pdf.

Ning, C. (2009). Engaging a "truly foreign" language and culture: China through Chinese film. *New Media for Higher Education*, *42*(1 & 2), 29-35.

Oxford, R. (1997). Cooperative learning, collaborative learning, and interaction: three communicative strands in the language classroom. *The Modern Language Journal*, *81*, 443-456.

Panitz, T. (1999). Collaborative versus cooperative learning: a comparison of the two concepts which will help us understand the underlying nature of interactive learning. (Report No. CS 217 306). Lanham, Maryland. (ERIC Document Reproduction Service No. ED 448 443)

Roschelle, J., & Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.) *Computer supported collaborative learning* (pp.69-97). Berlin: Springer-Verlag.

Sturm, J. (2012). Using film in the L2 classroom: a graduate course in film pedagogy. *Foreign Language Annals*, *45*(2), 246-259.

Sundquist, J. (2010). The long and the short of it: the use of short films in the German classroom. *Die Unterrichtspraxis/Teaching German*, *43*(2), 123-132.

Vuopala, E., Hyvönen, P., & Järvelä, S. (2016). Interaction forms in successful collaborative learning in virtual learning environments. *Active Learning in Higher Education*, *17*(1), 25-38.

Wong-Fillmore, L., Ammon, P., McLaughlin, B., & Ammon, M. (1985). *Learning English through bilingual instruction.* Final Report to National Institute of Education. (Report No. FL 015 141). (ERIC Document Reproduction Service No. ED 259 579)

Wood, D. (1995). *Film communication theory and practice in teaching English as a foreign language*. New York, NY: Edwin Mellen Press.

Yu, K. F. (2009). Learning English through films: a case study of a Hong Kong class (Unpublished master's thesis). University of Hong Kong, Pokfulam, Hong Kong SAR. Retrieved from http://dx.doi.org/10.5353/th_b4324124

Zhang, D., & Yu, Y. (2008). Context of learning and requesting in Chinese as a second language: An exploratory study of students learning Chinese in study abroad and at home contexts. *Journal of the Chinese Language Teachers Association*, *43*(1), 73-92.

Zhang, L. (2011). Teaching Chinese cultural perspectives through film. *L2 Journal*, *3*, 201-231.

附录 1 推荐电影目录

关于列表中推荐电影的详细讨论，请参见下面的 CLTA 中文电影教学兴趣小组脸书网页及群组的网站:
脸书网页：https://www.facebook.com/groups/317635671923482/?ref=bookmarks
群组的网站：https://sites.google.com/site/clfedex/film-list

电影主题	电影名称（年份）	关键词
青春校园	请为我投票 （2007）	班长、选举、候选人、候选人家长、民主
	青春派 （2013）	高考、早恋、师生关系、家长和考生
	少年的你 （2019）	校园霸凌、受害者与加害者、沉默的大众、老师、警察
社会议题	大腕 (2001)	葬礼、生死的观念、广告、商业化
	钢的琴 (2011)	下岗、离婚、父女之情、兄弟情
	十二公民 (2014)	司法公正、陪审团制度、公民义务、社会阶层、推理
	狼图腾 (2015)	人与动物、大草原、汉民族与蒙古族关系
	少年小赵 (2015)	民族主义、青少年成长、高考、支教、拆迁
	我不是药神 (2018)	社会正义、医药制度、走私、假药、法庭、病友、警察、医药代表
	找到你 (2018)	女性困境、保姆、社会阶层、法庭、抚养权
	西虹市首富 （2018)	投资理念、一夜暴富、继承权、足球比赛、减肥、人命与金钱的选择
	虚你人生 (2018)	网络直播、社会阶层、网红、屌丝、土豪、农民工
	流浪地球 (2019)	科幻片、未来社会、人与自然、亲情、希望、人工智能

婚恋爱情	失恋三十三天 (2011)	失恋、劈腿、蓝颜知己、婚庆公司、闪电约会、相亲、婚恋观
	心花路放（2014）	公路片、寻找自我、离婚、杀马特文化、兄弟情
	超时空同居 (2018)	时空穿越、时代变迁、人生选择、命运
	半个喜剧 (2019)	婚恋观、社会阶层、富二代、凤凰男、望子成龙、剩女
女性 / 家庭	重返 20 岁（2015）	返老还童、人生选择、家庭关系、校园乐队、老人生活
	七月与安生（2016）	青少年成长、姐妹情、三角恋、叛逆少女、模范人生
	相爱相亲（2017）	三代女性的婚恋观、贞节牌坊、安土重迁的观念、母女关系、媒体角色
	别告诉她（2019）	善意的谎言、绝症、个人与家庭的关系、两代人、中美生活习俗的差异、婚丧嫁娶
	阳光普照（2019）	家庭关系、青少年犯罪、心理健康、奉子成婚、刑满释放人员重返社会
历史议题	活着（1994）	历史变迁、内战、大跃进、文革、家庭与社会、皮影戏、塞翁失马焉知非福
	归来（2014）	文革、创伤、背叛、失忆、夫妻之情
	明月几时有（2017）	香港、殖民、抗日、文学、共产党、母女之情
	幸福路上（2018）	动画片、台湾、历史变迁、出国、跨国婚姻、多元文化
	哪吒之魔童降世（2019）	动画片、宿命论、人性善恶、神话、友情、师徒、亲情

附录 2《我不是药神》的课时安排

课时	主题	教学目标	技能训练
一	课前看电影、写豆瓣短評 课堂讨论短评内容	1. 看完整部电影后，学生能用自己的语言写出豆瓣短评 2. 课堂上，针对学生写的范文讨论用词、语法以及行文逻辑，一起完成纠错任务	阅读：电影梗概 写作：豆瓣短评 听说：课堂讨论短评
二	电影主题讨论一：程勇五人小组群像	1. 通过自选生词，学习新的词汇 2.通过语法练习，掌握语法的运用 3.透过预习作业和课堂讨论，理解电影文本	听力：看电影相关片段 阅读：阅读相关片段的电影对白 说话：课堂讨论电影主题一
三	电影主题讨论二：程勇的心理路程	1. 通过自选生词，学习新的词汇 2.通过语法练习，掌握语法的运用 3.透过预习作业和课堂讨论，理解电影文本	听力：看电影相关片段 阅读：阅读相关片段的电影对白 说话：课堂讨论电影主题二
四	电影主题讨论三：孰对孰错？执法者（警察）、药厂、病患的不同立场	1. 通过自选生词，学习新的词汇 2.通过语法练习，掌握语法的运用 3.透过预习作业和课堂讨论，理解电影文本	听力：看电影相关片段 阅读：阅读相关片段的电影对白 说话：课堂讨论电影主题三
五	电影主题讨论四：情与法的天平：如何判决程勇的罪？	1. 通过自选生词，学习新的词汇 2.通过语法练习，掌握语法的运用 3.通过预习作业和课堂讨论，理解电影文本	听力：看电影相关片段 阅读：阅读相关片段的电影对白 说话：课堂讨论电影主题四
六	写作练习： 看豆瓣影评后回应	1. 通过阅读豆瓣影评加深对电影主题的了解，了解专业影评的行文逻辑及术语的运用	阅读：阅读豆瓣影评 写作：根据阅读的豆瓣影评，写回应文章 听说：课堂讨论豆瓣影

		2.通过写回应文章，掌握议论文的写法 3.通过课堂讨论，分享不同的观点，进一步了解电影涉及的社会议题	评及学生的回应
七	采访练习：采访中文母语者对看病和医疗保险的看法	1.针对电影涉及的社会议题，向母语者提问 2.听懂母语者的回应，根据母语者的回应给出恰当的反馈 3.总结母语者的陈述，在课堂上完成口头报告 4.完成报告后，能够回答老师和同学的后续提问	听说：与母语者自然交流（interpersonal）； 课堂上完成口头报告（presentational）； 回答问题 (interpersonal)
八	口头报告：为抑郁症患者募捐的演讲	1.根据指定话题，写出符合格式的倡议书 2.根据倡议书的目的，运用演说技巧说服听众	写作：完成口头报告的讲稿 演讲：根据演讲稿，完成指定话题的演讲
九	考前复习	1.复习生词、句型以及电影主题 2.准备综合考试	阅读：复习生词、句型以及对主题内容的理解 写作：以生词和句型为主的单句写作练习及翻译练习
十	考试	在语言上和内容上评量对整部电影的学习效果	阅读：了解涵盖生词以及句型的单句以及短文 写作：完成以生词和句型为主的单句写作、短文写作

附录 3《我不是药神》预习单（四）片段三文本

01:20:44 01:24:26	警察四处搜查、曹斌以法说之，老太太以情恳求
曹斌（警察）	我听说你们大家不配合，我们办案的讲证据，从你们每一个人身上，都搜出那种假药，你们配合还是不配合，案子都一样办，包庇犯罪也是犯罪，我只提醒你们一件事情，你们这么做，不是在帮他，是在害他。
老太太（病人）	领导，求你个事啊，我，我就是想求求你，别再追查印度药了， 行吗？我病了三年，四万块钱一瓶的正版药，我吃了三年，房子吃没了，家人被我吃垮了，现在好不容易有了便宜药，你们非说他是假药，那药假不假，我们能不知道吗？那药才卖五百块钱一瓶，药贩子根本没赚钱，谁家能不遇上个病人，你就能保证你这一辈子不生病吗？你们把他抓走了，我们都得等死，我不想死，我想活着，行吗？

附录 4《我不是药神》预习单（四）片段三练习题

1.（阅读理解题 1） 根据曹斌对病友的说法，下面的陈述，哪一个是错的？(Highlight the incorrect answer)

a. 这些病友不说出卖药的人是谁，算是包庇犯罪，那他们也算犯了罪
b. 警察办案一定要有证据
c. 如果这些病友不配合，警察就没有办法继续办案了
d. 这些病友不说，其实是在害卖药的人，而不是在帮他

2. （阅读理解题 2）根据老太太的说法，下面的陈述，哪一个是错的？(Highlight the incorrect statement)

a. 老太太吃了三年原药厂的药，房子没了，家人钱也没了
b. 老太太吃不出来便宜药是真的还是假的
c. 老太太知道五百块钱一瓶的药，卖药的人根本不赚钱
d. 如果警察把卖药的人抓走了，病人就没有药吃，他们就是在等死

3. （词汇填空题）用『配合、证据、包庇、提醒、追查、正版、吃垮、保证』完成下面的句子。

a. 今天老师向我说明了这次处事上的一些问题，我也向老师_____以后不会再犯同样的错误了。
b. 我想_____大家，这次的校外活动要到山上天气变化比较大的地方，希望大家做好保暖的准备。
c. 这次警察查毒行动迅速(fast)又严厉(strict)，如有_____罪犯的行为，绝对不轻放。
d. 任何的法律诉讼都要有事实的 _____ ，才能在法庭上赢得辩论。
e. 最近学生考试作弊(cheat)的事，学校一定会谨慎地(carefully)_____清楚， 以防以后再有类似的事情发生。
f. 药物使用一定要特别谨慎小心，宁可(rather)多花一些钱买_____的药，也不要为了省钱买便宜的假药。
g. 这次的作业是小组项目，需要所有的组员_____才能完成。
h. 这么多人要吃饭，却只有一个人在工作，再这样下去，我们这家很快就要被_____了。

4.（句型使用题）请用 “非.....不可” 填空或者翻译下面的句子。

1. 我不让他喝那么多，可是_____。
2. 这次足球赛他不想参加，可是他的教练说他_____。
3. It’s so cold today, but my daughter is determined to wear a skirt.
4. My girlfriend made me give her roses on Valentine’s Day.

5.（综合练习题）请把下面的句子翻译成英文：

“我病了三年，四万块钱一瓶的正版药，我吃了三年。房子吃没了，家人被我吃垮了，现在好不容易有了便宜药，你们非说他是假药，那药假不假，我们能不知道吗？”

附录 5《我不是药神》预习单（一）针对豆瓣短评的思考问题

1. 引用电影里你印象最深刻的一段话，说说你对这段话的看法。
2. 选一个片中的角色，说明你为什么最喜欢 / 最不喜欢 / 觉得印象最深刻 / 觉得演得最好 / 让你最感动
3. 选一个片段，或是剧中人做的事，说明为什么这一段很重要

附录 6《我不是药神》 预习单（六）写作练习： 看豆瓣影评后回应

药化科研工作者千思万虑的两小时（剪辑版 Edited）

影评人：Andreja

https://movie.douban.com/review/9498742/

影片中的药企是最大的“反派”，但在现实生活中他们却是有苦说不出的“受害者。”药企为了弥补早期的所消耗的大量经费，售卖高价的药，它们是为了生存，穷苦的患者为了重见明日的曙光，购买低价的仿制药，他们也是为了生存，前者是“止损盈利”的生存，后者是“生与死”的生存，一个关乎利益，一个关乎道德，天平毫无疑问会倾向于道德一方，如此一来，“受害者”便成为了“加害者”。

片名为《我不是药神》，“神”却在全片中只出现过一次，徐峥金盆洗手一年后重出江湖，只身前往印度，他在药店门口拿着格列宁，路人捂着鼻子嘴巴行色匆匆的与他擦肩而过，戴着防毒面具的人，面无表情的喷洒着消毒剂，一辆推车拉着迦梨女神的雕像向他驶来。影片给了迦梨女神和徐峥面部表情的特写，这段是全片的点睛之笔，也是徐峥从“人”到“神”的过渡。影片开头，徐峥对黄毛说，“你还挺仗义的。”影片后半段时，假院士对徐峥说，“你还挺仗义的。”第一次徐峥为了钱卖“假”药，第二次徐峥为了“义”卖“假”药。

迦梨女神集毁灭和创造于一身，徐峥也是如此，他为了私利不顾病人的安危，将他们推入了绝望的深渊，他给了病人们希望又亲手毁灭了他们，他与迦梨女神四目相对时，他突然明白，既然是他摧毁了病人的生之欲望，那么也得有他来创造出病人的生之欲望，这一次，他不仅仅是“创造”，他也是在给自己的过往“赎罪”。

看了 Andreja 写的这篇文章，请到豆瓣《我不是药神》以下链接处去添加你的回应（参考 syllabus Appendix 2 如何使用豆瓣的“短评”、“评论” 跟“回应”）：

https://movie.douban.com/review/9498742/

要求：

回应字数：	250-300 字。（必须使用至少 10 个本课的生词，请把使用的生词划下划线）
回应内容：	针对 Andreja 的观点，可以回应他的某一个观点或者某几个观点。你可以赞同或者反对他的观点，最好能提出影片中的具体内容作为论据。
回应角度：	现实生活中药企是不是反派？程勇是药神吗？程勇为什么会突然转变？等等。

回应常见结构：	使用至少 3 个结构： ……是全片的点睛之笔 ……让人误解 ……有失公允 作为……，可以理解 虽……，但…… 我觉得……、……很奇怪 既然……，为什么……？ 即使是……，也不够有说服力 电影本身与……相差实在太远 所有……的描述都是.....的。因此不免有……的嫌疑、 这就使影片除了……外，缺少.....的意义。 影片开头，（一句话），影片后半段时，(同一句话再次出现），第一次……为了……，第二次……为了……。

URL (网址): http://www.tclt.us/journal
Email (电子邮件): editor@tclt.us