Using Language Corpora to Enhance Grammatical Proficiency in Chinese

Hsiao-Chien Chen

Brigham Young University - Provo
Using Language Corpora to Enhance Grammatical Proficiency in Chinese

Hsiao-Chien Chen

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Arts

Dana S. Bourgerie, Chair
Matthew B. Christensen
William G. Eggington

Center for Language Studies
Brigham Young University
August 2012

Copyright © 2012 Hsiao-Chien Chen
All Rights Reserved
ABSTRACT

Using Language Corpora to Enhance Grammatical Proficiency in Chinese

Hsiao-Chien Chen
Center for Language Studies, BYU
Master of Arts

School curriculum and pedagogy change over time and are affected by changes in technology. One little used technology in foreign language classrooms is the electronic language corpus. In corpus-based linguistics, language corpora are often used as tools to analyze and observe various language features, including discourse, pragmatics, and syntax. However, language corpora can also act as a tool to assist language teachers by providing greater exposure to features of the language. Using language corpora is especially helpful in exposing learners to so-called authentic language used in a target language culture. Moreover, students can gradually enhance their language proficiency by using a well-developed corpus. In the foreign language-teaching world, most corpora studies focus on using English corpora in ESL settings, and there are only a few studies focused on using foreign language corpora in other foreign language teaching settings. The light use of corpora in other foreign language settings may be partly related to the lack of user-friendly foreign language corpora, or a lack of understanding of how to manipulate different foreign language corpora effectively. This study seeks to demonstrate how language corpora can be used in advanced Chinese classrooms, and how Chinese corpora can help students to enhance their language proficiency. This study’s results show that corpus use in advanced Chinese classrooms can help advanced Chinese learners to improve their understanding of grammar taught in class.

Keywords: Chinese corpora, corpora and language teaching
I would like to thank my thesis committee chair and members, Dr. Dana Bourgerie, Dr. Matthew Christensen, and Dr. William Eggington for their patience and support. I would also like to recognize Dr. Dennis Eggett for helping with statistics and Professor Trever McKay for helping and allowing me to conduct my study in his class. I would also like to express my appreciation for advice and guidance provided me by Drs. Steven Riep and Michael Bush. Lastly I would like to thank AJ Van Wagoner for helping to edit my early drafts.

Special thanks to Dr. Bourgerie for his help in overcoming my language barrier and making everything possible. I would also like to thank the Garner family for their kindness in giving me an opportunity to further my education here, and those who have shown their support and love to me during this process. Without their support, this would not have been possible.
Table of Contents

List of Tables ................................................................................................................................. vi

List of Figures ............................................................................................................................... vii

Chapter One: Introductions ............................................................................................................. 1
  Introduction ................................................................................................................................. 1
  Overview ..................................................................................................................................... 2
  Foreign language instruction and language corpora ................................................................. 3

Chapter Two: Review of Literature ................................................................................................ 8
  Review of the literature ............................................................................................................... 8

Chapter Three: Research Design .................................................................................................. 15
  Participants ................................................................................................................................ 15
  Method ...................................................................................................................................... 15
  Study Design ............................................................................................................................. 17

Chapter Four: Results ................................................................................................................... 20
  Statistical analysis ..................................................................................................................... 20

Chapter Five: Discussion and Conclusion .................................................................................... 24
  Quantitative Analysis ................................................................................................................ 24
  Qualitative Analysis .................................................................................................................. 24
  Conclusions ............................................................................................................................... 28
  Limitations ................................................................................................................................ 30
  Suggestions for further research ............................................................................................... 31

References ..................................................................................................................................... 33

Appendix A: Corpus Survey ......................................................................................................... 37
List of Tables

Table 1: Writing assignment rubric .............................................................................................. 17
Table 2: Mixed Model ANCOVA analysis result......................................................................... 22
List of Figures

Figure 1: Academia Sinica Balanced Corpus of Modern Chinese page ........................................ 6
Figure 2: Academia Sinica Balanced Corpus of Modern Chinese search page ............................ 7
Figure 3: Original diagram from Corpora and Language Teaching, page 40 (Aijmer, Ed., 2009) 7
Figure 4: Academia Sinica Balanced Corpus of Modern Chinese "correlative conjunction" search result ...................................................................................................................................... 13
Figure 5: Academia Sinica Balanced Corpus of Modern Chinese "interjection" search result .... 13
Figure 6: Center for the Chinese Linguistics PKU corpus page .................................................... 18
Figure 7: The Center for Chinese Linguistics PKU corpus "素质" search result .................. 19
Figure 8: Pretest result .................................................................................................................. 20
Figure 9: Posttest result ................................................................................................................. 21
Figure 10: Average grades for each lesson .................................................................................. 22
Figure 11: Center for Chinese Linguistics PKU "以...為主" search ............................................ 30
Chapter One: Introductions

Introduction

When I was an undergraduate student at Brigham Young University, I was introduced to some English language corpora in several of my linguistic classes. As an ESL student, English corpora were helpful for me. I could see authentic language examples from different genres, get word frequency lists, language changes, language variations, and history of the English language. The information I could learn from English corpora is not necessarily found in English dictionaries. Later on, I started to work as a teaching assistant to teach Chinese, and often students would ask me questions such as “Can you give me more examples of X grammar?” “How do native speakers use X word?”, or “Can I use X grammar rule in Y sentence?” Although I am a native speaker of Chinese, sometimes I still had a hard time coming up with enough examples or answers to satisfy some students. Language corpora came into my mind at that time. I can find all of those answers through a well-developed Chinese corpus! But is there one available for people to use? How can I introduce this corpora idea to students when they don’t have high language proficiency? How can I prove language corpora can help students to improve their language proficiency? And how can I encourage students to use corpora in their studies?

When I started searching for Chinese corpora, I found a couple of available online Chinese corpora, one in traditional Chinese and another one in simplified Chinese. The Academia Sinica Balanced Corpus of Modern Chinese (現代漢語平衡語料庫: http://db1x.sinica.edu.tw/kiwi/mkiwi/) was developed in Taiwan in 1990, and the Center for Chinese Linguistics PKU corpus (CCL corpus, 北京大学中国语言学研究中心语料库: }
http://ccl.pku.edu.cn:8080/ccl_corpus/) was developed in Beijing. The Sinica Balanced Corpus contains 5 millions words (data from the Sinica corpus website), and CCL corpus contains 477 million words. There are some other Chinese corpora available, but for the purpose of this particular study, other Chinese corpora will not be discussed here.

The main purpose of this study is to demonstrate how language corpora can be used in a foreign language classroom, how can it assist class instruction, and how can it benefit language learners. The hypothesis of this study is that the use of a Chinese corpus can enhance advanced Chinese learners’ understanding of grammar principles taught in class.

**Overview**

The main question in this study is: Can using language corpora improve advanced Chinese learners’ language proficiency? More specifically, can using corpora enhance students’ understanding of the Chinese grammar taught in class?

The first challenge in designing this study was deciding how to incorporate a Chinese corpus into normal classroom instruction. An advanced Chinese instructor provided a possible way to incorporate this corpus “treatment” into his classroom: There was one control group and one experimental group in this study. Both groups had the same instructor, the same teaching assistant, and used the same class materials. The control group received regular grammar instructions in class from the instructor, and participants in this group had to write a grammar assignment for each chapter; the experimental group received detailed instruction in corpus use in addition to their regular class instruction, and participants in this group were also required to use a Chinese corpus when writing their grammar assignments. In order to get more information from participants, they were each asked to take a Chinese reading test twice. The first time they
took it as a pretest, the second time as a posttest. All of the participants’ grammar assignment scores, pretest scores, and posttest scores were recorded for further analysis.

**Foreign language instruction and language corpora**

Languages change in different ways for different reasons, whether for political, social, or cultural reasons. In the history of foreign language instruction, scholars have emphasized different approaches and methods in different periods of time. If language corpora had been available during the end of the eighteenth century, how might teachers have incorporated different types of language corpora into the approaches of the time? And how could language corpora have benefited those approaches and methodologies? One good early example is Grammar Translation. Grammar Translation was long used in teaching Latin and Greek. Students focused on both translating printed texts in order to learn grammar rules and also on memorizing bilingual word lists to aid translation. If there had been some well-developed Latin and Greek language corpora available during that time, then the corpora could have assisted this methodology. For example, let’s assume there was a bilingual corpus available in Latin and English. Students could then get a bilingual vocabulary list sorted by word frequency. Moreover, some corpora allow users to access word frequency lists, so language corpora could definitely benefit the Grammar Translation approach. Corpora can be used to make this approach more effective because they allow users to get a word frequency list just by a simple search.

Other second language acquisition approaches—such as Stephen Krashen’s Second Language Acquisition Hypotheses—could also incorporate language corpora. One of these, the Input hypothesis, explained that “i+1” level comprehensible input can stimulate second language learning (Krashen, 1985). Language corpora are definitely one good source of “i+1” level input. Although learners might not have enough linguistic competence to understand examples in a
corpus, according to Krashen, regular language exposure can help a person to learn more efficiently. Language corpora provide authentic language examples for language teachers and learners (Krashen, 1985).

Using corpora in foreign language classrooms can be seen as both bottom-up and top-down approaches, depending on which one is the focus. In *Blackwell Handbook of Linguistics*, Treiman explains the differences between bottom-up and top-down processes in reading: “bottom-up processes are those that take in stimuli from the outside world -- letters and words, for reading -- and deal with that information with little recourse to higher-level knowledge.” For top-down processes, she explains it is “the uptake of information is guided by individual’s prior knowledge and expectations.” (Aronoff & Rees-Miller (Eds.), 2001). The application of bottom-up processes in a corpus-based approach to learning allows teachers to search for a specific word or usage in a corpus and see how the language is correctly used. On the other hand, for a top-down approach, teachers can search for content from a certain genre in a corpus, then ask students to find a certain word, grammar structure, or usage in that content.

Language corpora can also assist in other ways of looking at language, such as Analogical Modeling (AM). This model may be less familiar than Grammar Translation, Stephen Krashen’s Second Language Acquisition Hypothesis, or bottom-up and top-down approaches, so the following is a basic introduction to Analogical modeling. “Analogical Modeling (AM) is an exemplar-based approach designed to predict linguistic behavior on the basis of stored memory tokens” (Skousen, 1995). Analogical Modeling theory, as applied to language acquisition, suggests that a language learner (either L1 or L2) uses comparison and categorization to organize new language concepts as they are encountered. This means essentially that the acquisition of new elements of language, as well as the ways they are used, will combine to enhance the
learner’s ability to make predictions about the language. That is, language learners are able to use their linguistic episodic memory or a problem-specific dataset to predict language behavior. For example, if a learner of English as second language knows the word “run”, then he or she can predict the meaning of the word “runner” because “runner” is associated with the word “run.” In the classroom, a teacher can first introduce a new sound in the target language, and then help students to make connections with the new sound. The final stage of the acquisition process suggested by this model would then be to help learners store more linguistic episodic events in the learners’ individual dataset. Eddington and Lonsdale (2007) have pointed out that most implementations of Analogical Modeling “involve feature vectors that specify relatively low-level linguistic features such as sound segments, phonetic or orthographic environments, syllable structure, and word boundaries.” Once a language learner has established a certain language dataset, then this language learner can use the existent to predict new or relevant language behaviors. Moreover, Eggington points out “if we acquire language patterns through surface level processing, then it is likely that we will sub-consciously model frequent and prestigious language patterns” (Eggington, 2009).

Language teachers of almost any language should be able to use tools that implement pedagogically sound principles that have been derived from AM. The end result would be a way to develop class materials that would help students increases their language proficiency. AMould help learners establish connections between and among various elements of their individual linguistic knowledge. Such a process would require time, however, and it may benefit advanced language learners more than novice learners. The reason for this limitation is that novice learners might not have enough “data” stored in their minds to make connections with. Once a language teacher helps learners establish this “dataset,” then learners will be able to increase their ability to learn the target language. This increased ability will be the result of learners combining their
old and new datasets to help them to predict more language phenomena in their future learning experiences. For example, to help learners establish a vocabulary dataset, a language teacher can teach homonyms in the target language that will enable learners to compare and make connections between words they know. Another extended application for AM is using AM during the development of language corpora materials. A corpus itself is a large and structured set of naturally occurring text, and the use of this type of text can help learners to observe and compare more language phenomena than would otherwise have been possible, and this in order to gain more linguistic knowledge. Language corpus data is more structured and controlled for a number of reasons, but the data they contain is also easier for both language teachers and learners to manipulate. Corpus users can see many more language examples of a certain grammar usage or vocabulary than by using a dictionary or a textbook. For example, if corpus users of the Academia Sinica Balanced corpus (see Figure 1) want to look up a word’s usage, then they first go to the Sinica corpus website: http://db1x.sinica.edu.tw/kiwi/mkiwi/).

Figure 1: Academia Sinica Balanced Corpus of Modern Chinese page
Users can select a certain desired genre, style, or topic to manipulate the search. If users do not have any preference, then they select “all” for a general search. This corpus also allows users to set more filters (See Figure 2), such as different features and patterns (e.g. reduplications):

![Academia Sinica Balanced Corpus of Modern Chinese search page](image)

**Figure 2: Academia Sinica Balanced Corpus of Modern Chinese search page**

Moreover, Johansson discussed many uses of corpus. In his diagram, he use a corpus as a central point to show different uses of corpus:

![Diagram from Corpora and Language Teaching](image)

**Figure 3: Original diagram from Corpora and Language Teaching, page 40 (Aijmer, Ed., 2009)**
Chapter Two: Review of Literature

Review of the literature

Compared to a decade ago, more and more language learners rely on Computer-Assisted Language Learning (CALL), Technology-Enhanced Language Learning (TELL), Computer-Mediated Communication (CMC), and other technologically based multimedia to learn new languages more efficiently and effectively. These technologies enable new ways of language learning, but language learners are required to learn how to use them first before they can be effective. An electronic language corpus is a good example of one of these new technologies. In Trust the Text, Sinclair discusses some of the earlier corpora works and the new “corpus-driven linguistics” which gave an overview of corpora used (Sinclair, 2004). Cheng, Warren & Xu, (2003) argue that the use of corpus in the classroom provides an opportunity for learners to become researchers, and also for native speakers to examine their intuition and introspection in their own language. Corpus data is also the most reliable source of target language descriptions. As Aston (2001) has argues corpora can be seen as “sources of knowledge about the language.”

Corpus linguistics is a relatively new field of applied linguistics, and language corpora are not often used in classroom settings because of the lack of knowledge and training about their use. The basic definition of a corpus, according to Huston (2002), is "a collection of naturally occurring examples of language, consisting of anything from a few sentences to a set of written texts or tape recordings which have been collected for linguistic study." A corpus usually does not contain new information about a language, but it is rather a large number of collected texts that are produced by native speakers. A text in a corpus can be considered authentic text and also natural language occurrence because they are from native speakers’ natural utterances or writings (with the exception of learner corpora). Several researchers have taken advantage of “authentic
text” in corpora to investigate the lexicon, semantics, pragmatics, morphology, syntax and other parts of languages. One such example is Hong, Chen and Chiu’s (2009) research “Automatic extraction of new words based on Google News corpora for supporting lexicon-based Chinese word segmentation systems.” The Hong et al study demonstrates how Google can act as a corpus for language studies. More specifically, Google’s search engine can serve effectively as a “corpus” because it contains a large amount of data in various forms and from a variety of registers. However, compared to most other language corpora, Google cannot filter out certain language features that users may not need.

Using corpora in a language classroom can be effective and helpful for language learners in many different ways. With today’s highly developed technology, language books and dictionaries are not the only source of reference tools for second language learners. There are now a large number of corpora available for use in research, language learning, and teaching. A typical dictionary usually contains information on part of speech, pronunciation, examples, usage, etc. Examples in a dictionary have often been carefully selected and prescribed to represent a certain usage. On the other hand, corpus examples provide more practical and realistic examples. Moreover a corpus allows learners to see both formal and informal usages. Authentic texts are more meaningful and helpful for learners in a language learning setting. Gavioli and Aston’s article asked a question, “can corpora capture reality?” (Gavioli & Aston, 2001). Most language learners take the opportunity to apply what they’ve learned in real life, and corpora provide a way to them to see real life, native usage, which dictionaries do not provide.

A language corpus can also provide word lists for language beginners to learn. Most language learners want first to know the most frequently used words in the target language. A corpus can provide this. Furthermore, using corpora also allows users to access more examples in
different dialects and registers. Scholar Yukio Tono discussed how to use corpora effectively in English teaching in Japan (Frankenburg-Garcia, Flowerdew & Aston, 2011).

Different types of corpora can be beneficial in a number of professional fields. Language translators can use comparable corpora to study usage and patterns of language in translations. Language teaching is another field in which language corpora have been infrequently used, and which could benefit from their use. Using existing corpora, classrooms could help learners achieve higher language proficiency levels and help teachers gain more confidence in discovering facts about the target language. Mark Davies, a corpus linguist at Brigham Young University has noted: “A goal of many language learners is to more fully understand the range of syntactic variation in the second language and thus move beyond the simplistic rules that are presented in many textbooks” (Davies, 2004). Evidence shows data-driven learning can help language learners to become more successful at examining and understanding the target language.

Learner corpus research has only existed since the late 1980’s (Granger, 2002). Using corpora can provide better descriptions of the target language and can also provide new ways to teach a foreign language. Biber (1988) shows that using corpus-based techniques in the study of language variation can help reveal the distinctive patterns of distribution of each variety. Granger also notes that “corpus-based studies conducted over the last twenty or so years have led to much better descriptions of many of the different registers” (Granger, 2002). Training language learners to become language researchers can lead them to become more successful in their studies. Researchers Cheng, Warren, and Xu state, “The teacher becomes a facilitator of language study instead of being seen as the language expert responsible for both teaching and research.” Students in this type of study “acquire a new role as language investigator in addition
to that of language learner” (Cheng, and et al, 2003). They begin looking for patterns in the target language rather than looking for the correct answers. According to Cheng et al’s research, most students found this type of data-driven approach helpful in learning a language, also DDL (Data-driven learning) “can lead a strong sense of achievement to the serendipitous linguistic discovery experienced by some learners” (Cheng, and et al, 2003). Charles’ (2007) research combines discourse analysis with corpus investigation approaches to show how top-down and bottom-up approaches can be reconciled in English for academic purpose writing. Charles asks the question of whether or not the combination of discourse analysis with corpus investigation can provide the enriched input necessary for students to make the connection between general rhetorical purposes and specific lexico-grammatical choices. This is another example of data-driving learning approach use in language classrooms.

Wang (2002) discusses recent corpora in China. One of the examples in this article is the Modern Chinese Language Corpus (MCLC), which Wang argues could help in five areas of language learning:

1. Information processing in the Chinese language
2. The unification and standardization of the Chinese language
3. Academic research
4. Language education
5. Application of the Chinese language

(Wang, 2002)

These specific functions of language corpora can benefit both Chinese teachers and learners.
Corpora allow users to see language changes over time, and also allow users to see the differences within different registers. Corpora also provide more details of how certain words used in various contexts. A language dictionary typically can only provide small number of examples, but corpora can provide many examples in a single search. More importantly, using corpora in language classrooms can help language learners to develop linguistic knowledge.

To establish a reliable high-quality corpus takes great effort and it’s also a complex process. In 中文語料庫構建及管理系統設計 (Chinese Corpora Construction and Management System Design) article, authors described the process of building a Corpus. First one needs to collect large amount of texts in a corpus database, and those texts need to be able to represent certain genres or styles, i.e., those texts needed to cover different topics and modes. After having texts from a variety of sources, one needs to sort texts into relevant categories, and provide information about each text (i.e. author, publisher, and type etc.). Tagging is important in the corpus building process and almost all of tagging process can be done with computer software, but it often requires human post editors to verify the data in a corpus to increase reliability. Different corpora use different ways of tagging, but every corpus usually has a list of their tagging symbols for users to use as reference (Ma, and et al, 2010). For example, when using Academia Sinica Balanced Corpus of Modern Chinese, if users want to search for the most common correlative conjunction in this corpus data without a searching within a specific mode (e.g. written or spoken text), domain (e.g. newspaper, academic journal etc.) or register, and users only need to put in “Cbb (stands for correlative conjunction)” in the “List of POS” column. Parts of the results are shown in Figure 4. If users are looking for the most common interjection in the corpora, then the search stream needed to be changed to “I (stands for interjection)”, and
Figure 5 shows a small part of the results. A corpus can be considered as a “filter” that screens out certain parts of the target language and keeps the part that we are looking for.

Another search that can be done in language corpora is a search by words. In this function, corpora users can investigate a specific word through a myriad of examples. By seeing more examples of how a specified word is used in context can help language learners to have a better understanding of a particular word.

Although there is great interest in corpora for language teaching in China, most of the work done in China centers on use of corpora for EFL. Only a small amount of work has been done to advance Chinese as a foreign language. Simon Smith (2011) addresses the topic in “Corpus-based tasks for learning Chinese: a data-driven approach.” In this paper, he demonstrates how to use corpus-based tasks for intermediate-level Mandarin learners. In
addition, Tao who is a member of Center for Advanced Language Proficiency Education and Research at the Pennsylvania State University (CALPER) established some of the earlier work related to corpora use. Tao and Meyer state “There is considerable evidence that a corpus can enrich our understanding of language and, in many cases, provide linguistics with examples they would have never considered had they relied only on data obtained through introspection” (Meyer & Tao, 2005).
Chapter Three: Research Design

Participants

The sample size of this study is not ideal due to some limitations in the study design and classes offered during Winter semester, 2012. In this study, we tried to control for different variables to improve validity and reliability study, but this also leads to a small sample size problem.

The subjects were 36 students that enrolled in the advanced Mandarin course at Brigham Young University. There were two sections of the same class with the same instructor and teaching assistant. One section had 17 students with three females and 14 males, and all students in this section met at eight o’clock in the morning. The other section had 19 students with four females and 15 males. All students in this section met at twelve o’clock. Each section of the class met for fifty minutes each day, from Monday to Friday. Students who enrolled in this class have at least two-year of previous Chinese studies. Most students in both sections were also exposed to Chinese speaking environments for one to two years.

Method

The experimental group received detailed instruction in corpus use on top of regular class instruction, and the control group only received the regular classroom instruction. The purpose of this study was to see if exposure to corpus data helps students’ overall reading ability and grammatical accuracy. Participants were first asked to take a reading test (Computer Adaptive Test for Reading Chinese (CATRC) at the beginning and end of semester to see if they would show any global improvement. The CATRC is automatically scored on a continuous scale (0-900), and this test is currently used as a placement test at Brigham Young University. This test will be aligned with ACTFL guidelines in the near future.
All participants were asked to turn in a writing assignment each week during the fourteen-week semester. This writing assignment required students to form ten sentences according to the grammar patterns they learned from each lesson. Participants in one section, the twelve o’clock section, were treated as the control group and did not receive the corpus treatment; the other section was the experimental group and received corpus treatments for the writing assignments each week. The experimental group had to use the assigned corpus to copy five more sentences from the corpus for each grammar principle they’ve learned in addition to the ten sentences they had to create for each lesson.

The treatment group received detailed corpora training in class. The researcher conducted the corpus training, and the experimental group participants were introduced to language corpora to learn how to manipulate an online corpus tool. The corpus tool used in this study was developed by the Peking University Center for Chinese Linguistics. The class instructor was asked to use the corpus tool in class when participants needed to see more examples related to the grammar principles, phrase, or vocabulary they learned in class. Participants in the treatment group were asked to find sentences from the provided corpus (http://ccl.pku.edu.cn:8080/ccl_corpus/index.jsp?dir=xiandai), and write down three more sentences with the featured grammar patterns in addition to the ten sentences they had to write. Participants in the control group only needed to write ten sentences total for each lesson, but the participants in the experimental group had to copy five more sentences from the assigned corpus as part of their assignment. Participants’ grades from two sections were recorded for further comparisons. A rubric designed by the instructor was used to grade students’ sentences to avoid possible objectivity problems.
The rubric used for this assignment was based on 100 point total, with 10 points for each sentence. 0.2 points were deducted or each wrong character as in table 1 below.

**Table 1: Writing assignment rubric**

<table>
<thead>
<tr>
<th>Used grammar correctly, fluent sentence</th>
<th>10 points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Used grammar correctly, but awkward sentence</td>
<td>7 points</td>
</tr>
<tr>
<td>Used partial grammar correctly (half correct)</td>
<td>5 points</td>
</tr>
<tr>
<td>Didn’t use grammar, or used grammar incorrectly</td>
<td>0 point</td>
</tr>
</tbody>
</table>

The grader was a trained teaching assistant and a native speaker of Chinese. All graded assignments were under proper supervision of the instructor. At the end of semester, participants in the experimental group were asked to complete a survey about their personal opinions toward the corpus used in class.

**Study Design**

The study design used a triangular method to examine how language corpora could assist students to enhance their understanding of Chinese grammar. In order to provide evidence for this study, the researcher recorded participants’ pretest, posttest, and assignment scores for comparison purposes. A survey was also used in this study to allow participants to self-evaluate and also to provide an opportunity for the researcher to examine participants’ cognitive processes. To encourage participants in the experimental group to use the assigned corpus, the class instructor assigned students to write three more sentences of each given grammar patterns for each lesson. Furthermore, the class instructor was asked to use the assigned corpus simultaneously with his grammar instructions in the experimental group.

In order to train students to use corpora better, the following instructions were given to demonstrate how to use the Center for Chinese Linguistics PKU corpus for their assignments:
Step 1: Go to the Center for Chinese Linguistics PKU corpus website (See Figure 6)

Step 2: Look at the sentence patterns you learned in class that day.

Step 3: Identify a searchable term from each sentence pattern.

Step 4: Type in your search term, then select “查询”

Step 5: Pick FIVE examples that match your textbook descriptions for each sentence pattern you learned in class. (If a grammar pattern has two parts, search for one part first, then use “command + F” and type in the second part in the search bar.)

Step 6: Copy these 5 examples and then create 3 sentences of your own.

For example, type in “素质” in the search bar, and this is what you see:
Figure 7: The Center for Chinese Linguistics PKU corpus "素质" search result
Chapter Four: Results

Statistical analysis

There were a total of 36 participants, 17 participants in the experimental group, and 19 participants in the control group. The maximum possible score on this test was 900, and the control group’s average grade was 529, the experimental group’s average grade was 580. Figure 8 shows the pretest results, and it is evident that all participants have varying language ability. The experimental group had somewhat higher reading scores than the control group in the pretest and might suggest a higher beginning aptitude for that group. However, 51 point difference is relatively small on the 900 scale and likely does not indicate a significant difference.
At the end of the semester, a posttest was administered to both groups. Note that six of the 36 (two from the experimental group, and four from the control group) did not participate in this second administration of the CATRC (the reading proficiency test). Therefore the numbers are not perfectly comparable with the pretest, but one can see the general pattern. The results are as follows:

![Posttest result](image)

Figure 9: Posttest result

With respect to the in-class work, each student submitted a total of seven individual assignments, which included grammar and usage. The following figure shows the average grades of participants:
Figure 10: Average grades for each lesson

Figure 10 shows that the experimental group’s average grades were higher than the control group’s. For further analysis of this study, both qualitative and quantitative analyses were used in this study. An ANCOVA was used for the quantitative part, which used the pretest score as a covariate, and the post-test score as the dependent variable and treatment as independent variables. Results can be found in the following table:

Table 2: Mixed Model ANCOVA analysis result

<table>
<thead>
<tr>
<th>Effect</th>
<th>Number</th>
<th>Den</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect</td>
<td>DF</td>
<td>DF</td>
<td>0.0204</td>
</tr>
<tr>
<td>CATRC: PreTest</td>
<td>1</td>
<td>181</td>
<td>0.0030*</td>
</tr>
<tr>
<td>Lesson</td>
<td>6</td>
<td>181</td>
<td>0.0007*</td>
</tr>
<tr>
<td>Treatment</td>
<td>1</td>
<td>33</td>
<td>0.0007*</td>
</tr>
<tr>
<td>Treatment lesson</td>
<td>6</td>
<td>181</td>
<td>0.0522</td>
</tr>
</tbody>
</table>

*Significant at the .05

Mixed models are used for repeated measurement, or multiple sources of variation, that include “fixed effects (where all levels of an effect of interest are assumed to be known and represented) and random effects (where levels of an effect are assumed to be random draws from
a population of effect values)” (Bell & Grunwald, 2004). This is a mixed models ACNOVA blocking on individual students. Each student has multiple scores from the various assignments. Thus the observations are not independent (which is an assumption of ACNOVA), thus we use mixed models ACNOVA that accounts for the lack of independence in the observations. The score from each assignment is the dependent variable. The independent variables are pretest (covariate), lesson numbers, treatment (corpus), and the lesson numbers by treatment interaction.

The survey (see Appendix A) used in this study is considered as a qualitative analysis, which allows the researcher to examine participants’ cognitive process, and also to receive some feedback on corpus use in their classrooms. The details of this survey’s results can be found in Appendix B. All participants are numbered from 1 to 14 in order to maintain anonymity.
Chapter Five: Discussion and Conclusion

Quantitative Analysis

According to the pretest results (see figure 8), participants in both groups had different reading abilities to start with. However, a person’s reading ability often can not be determined by a single reading examination, and instead often requires multiple tests and retests in order to analyze their ability from every possible angle. Due to the fact that this study is considered as a small pilot experiment, and lacked the time to conduct more tests on participants, the CATRC was the best choice. By having participants take this test, we could get a rough idea of their reading proficiency level.

Figure 10 shows the average grades of the experimental group are slightly higher than the control group. Furthermore, the experimental group’s average grades on their homework increased from 90.6 (Lesson 6) to 96 (Lesson 10), but decreased after lesson 10. We can see the experimental group’s average grades improved almost 6 points (from 90.6 to 96) after using the assigned corpus for about one month. The experimental group’s average grades on the last two assignments are higher than the control group’s.

ACNOVA was used to get more reliable statistics, and the result (see Table 2) shows that p-value of the corpus treatment is 0.0007, which is smaller than 0.01, so the result is statistically significant.

Qualitative Analysis

A survey was used for qualitative analysis. The language corpora idea was fairly new for participants, and the researcher wanted to know if any participants in the experimental group actually used the assigned corpus besides doing their grammar assignments. Most participants only used the assigned corpus when they needed to do
their assignments, but three out of fourteen participants actually used it to look up some new vocabulary usages.

The second survey question allows participants to examine whether language corpus actually helped them to understand the language. Only one participant said that the corpus wasn’t effective, but a lot of other participants thought that the corpus helped them to understand the usage better. The following are some of the other responses to this survey question:

Participant number 1: “It helped me learn how to use grammar better, but for the most part it provided further examples so I could see how it is used.”

Participant number 3: “It helped me become more confident in writing my sentences and using the grammar patterns in spoken Chinese.”

Participant number 6: “It was very helpful in learning the grammar to improve my speaking and writing skills.”

Participant number 7: “Put the vocab in context, and gave me new ideas how to use the grammar.”

Participant number 13: “It helped me to see many different examples using the specific grammar principle I was learning, I felt that after looking at the corpus, I understand the grammar better.”

Using a corpus to learn a language helps learners to put words in context, and this makes language learning more meaningful because corpora provide a way for language learners to see how language is used, and more importantly, how native speakers use the language in their daily lives. Corpora also make learning a language is not just learning
grammar and memorizing new vocabulary words, but it creates an environment for language learners to become language researchers by providing a tool for them to create a self-learning environment.

The third survey question asked participants if the corpus helped them write better sentences, and this question allowed participants to do a self-evaluation of their own progress. Most of participants felt the corpus helped them to create better sentences, but there were three participants who thought it didn’t help very much.

The next survey question asked participants if the regular use of a corpus helped them to improve their language proficiency, and asked them to specify why or why not? Only two participants felt that using a corpus regularly would help to increase their language proficiency, but most of them felt that if they could take full advantage of the corpus and use it more frequently, then it would help them to improve, but stated that they didn’t have time to do it or they thought some corpus examples are too complicated and technical. The followings are some of participants’ responses:

Participant number 3: “I think it helps improve language proficiency generally because it helps the vocab/grammar patterns become less strange.”

Participant number 4: “I think it helps writing/speaking, because it gives you more ideas of when to use it. But only if used over a long period of time would it actually improve overall proficiency.”

Participants number 7: “Yes, the corpus sentences were a little too complex but I believe it helped me most with understanding how to personally use the grammar patterns.”
Participant number 9: “Yes, it does. It helps with both reading and just in general because it gives you examples of everyday uses of Chinese.”

Almost all languages have distinctions between written and spoken language or between formal and less formal language. A language learner needs to understand those different forms or styles of a language to become successful in the target language culture and society. Some specialized corpora allow their users to see these different language forms, some regular corpora don’t provide this distinction. Although the Center for Chinese Linguistics PKU corpus doesn’t provide full information for its corpus examples, but third-year advanced Chinese learners should be able to have the ability to distinct the different forms or styles of a language. The fifth survey question asked participants if the corpus helped them better understand various styles of the language, and most of participants didn’t think it helped. The reason for this might be that participants didn’t pay attention to the language style, and both the researcher and the class instructor did not emphasize it in this study.

Participants were encouraged to use corpora, but most participants are still not very familiar with the corpora used in their Chinese studies, so the next survey question asked if participants would continue to use language corpora in the future to help them learn Chinese. Almost all participants said they would use it in the future.

The last question asked participants if they would encourage other Chinese learners to use corpora in their Chinese studies, and all of them said they would. Although not every participant had a positive experience using the corpus to assist them to do their grammar assignments, they definitely perceived the value of corpora.
Conclusions

Using language corpora in foreign language classrooms is not a theory or a new pedagogy, but corpora can be seen as a reference tool to assist language learning and teaching, and corpora can also help learners to enhance their understanding of a language. Moreover, pedagogically speaking, this study shows a significant impact of a corpus used in Chinese language classrooms. We found that participants who used the assigned corpus to assist them in writing their grammar assignments created more accurate sentences because they had better understanding of the grammar they learned in class.

The following are several benefits of using corpora:

For language learners:

1. Language learners can improve their understanding of a target language by using well-developed corpora that can represent the target language to increase their “actual language” exposure.

2. Language learners can become language researchers. Corpora allow learners to investigate specific elements of a language.

3. Many learners often ask the question: how can I continue to learn a language when I am out of school? Corpora provide a source for lifelong learning and self-learning. People who are not in school can consult corpora to continue learning a language, but this is for those who already have higher language abilities.

For language teachers:

1. Language teachers can use language corpora to develop class materials or tests. For example, teachers can find an article in a certain genre to make a fill-in-the-blank test.
2. Language teachers can also use corpora to help students to understand language beyond the basic word or grammar level.

3. When students need to see how a target language is used, teachers can use corpora to provide authentic examples for students. Sometimes a language teacher or even native speakers of a language can only come up with so many language examples on the spot, but corpora can show more examples with just a simple search.

Some other researchers also suggest using small corpora to teach a specific genre, using Corpus to teach and compare cultures, and also using corpus to teach writing (Henry & Roseberry, 2001). Bennett mentions, “All language skills can be taught using corpora” (Bennett, 2010), and Beaugrande stated the importance of corpora in English learning: “The use of corpora are surely most urgent for non-native speakers who have not had extensive exposure to fluent English” (Henry & Roseberry, 2001). Although Beaugrande’s statement was for English learning, the same statement can apply to any other language. It is important to create a learning environment that provides sufficiently natural language so language learners have enough language exposure, and corpora can be one of the best resources in assisting the foreign language learning process. It is not necessary that every language teacher or learner be a corpus expert, but having a basic understanding of how to refine searches can help teachers to establish a better language learning environment.

The challenge now is how to teach students to use corpora to assist their language learning? If language teachers can incorporate corpora in their teaching to assist students in learning a foreign language—as this study illustrates—not only will students’ reading abilities be enhanced, but their general understanding of a language will also increase.
Limitations

One of the main concerns in the present study is that participants had limited corpus training, so some participants might have limited Chinese knowledge to interpret corpus data. For example, when participants looked up one of the grammar patterns from their textbook — 以...为主 (meaning to focus primarily on something) — in the assigned Chinese corpus, these are some examples they might see from this corpus search:

他领导了国民党统治区波澜壮阔的第二条战线。这条战线，以学生为主，有广大工人、农民、教员、了全国的爱国民主运动。在中国共产党的号召下，一个以学生运动为主的反内战运动一时席卷了整个国家。没有工业领导，农业就无法发展。”我们的国家建设，是以国内力量为主，即自力更生为主，生产建设上、阜新和鹤岗的煤炭工业，沈阳的机械制造工业和吉林的电力工业为主要内容的东北工业基地的建设；

Figure 11: Center for Chinese Linguistics PKU "以...为主" search

The first sentence example has the pattern 以...为主 ‘takes x as the main point’ or ‘main point’ — which is the usage that participants had learned. However 主体 (meaning “main part”), in the first sentence above, “主” and “体” can not be separated which also means this first example is not the grammar pattern searched for: “以...为主”. Native speakers of Chinese can filter out this example quickly when they are looking for “以...为主” in a corpus, but L2 learners might not have enough general language proficiency to do. This problem occurred frequently throughout the study, which suggested participants might have not been ideally suited to the task.

Corpora provide large quantities of authentic language examples, but many texts it is required that users have a relatively high level of general proficiency to interpret corpora data effectively. Some students may feel overwhelmed or inadequate due to the lack of language proficiency. Language teachers need to guide or assist students to interpret corpora data when necessary.
For most students, taking a test is usually a serious matter because the test result might affect their grades, performances, or even future occupation choices. However, for a test that will not affect any of the above, test takers usually may not take it as seriously as a graded assessment. We were not able to make the CATRC a part of participants’ semester grade, but we gave out extra credit to encourage them to take it seriously. Therefore, it was still difficult to make sure all participants took the test seriously.

Another challenge was time. Corpora contain a plethora of data, and those data can be manipulated in different ways. To see a significant impact through corpora use in language classrooms, then treatments over a longer period of time, and with a larger sample size would be more effective. The ideal situation would be to find a group of participants with similar language abilities, and after using language corpora for a certain period of time, measure those participants’ progress. It is difficult to achieve this goal because of the many variables needed to control for an ideal experiment, such as each participant’s language background, anxiety, learning motivation, and other language exposure.

Lack of data is also a limitation. Due to some “natural human behavior” of university students, some students like to procrastinate or do not want to turn in their assignments. This phenomena leads to a lower participation.

Suggestions for further research

Corpora can be very useful and beneficial for language learning and teaching. However, due to lack of information on using corpora in teaching foreign languages and other limitations, foreign language corpora are not widely used in the classroom. Language teachers and learners need to develop a new corpus approach in teaching and learning foreign languages, and also understand how language corpora can be beneficial in classroom settings. More research needs to
be done in using corpora in non-ESL language classrooms. Furthermore, more constructive language corpora need to be created in other languages, and more instruction needs to be given to language teachers and learners. Corpora are of limited value by themselves, but once people learn and know how to use them effectively, they can truly benefit not only linguists in research, but also language teachers and learners, who can become more successful in teaching and learning a foreign language.
References


Appendix A: Corpus Survey

1. How often do you use the CCL (Center for Chinese Linguistics PKU) corpus outside of your class? (E.g. I only use it when I need to write my class assignments. / Once or twice a week in addition to the assigned homework.)

2. How did the CCL corpus help you to better understand various parts of the language that you learned in class? (For example, grammar, vocabulary usage, etc.)?

3. Did the corpus help you to better create your own sentences?

4. Do you think using corpus regularly helps you improve your language proficiency? (Reading or generally?) Why or why not?

5. Did using the corpus help you better understand various styles of the language. For example, more or less formal (书面语 / 口头语)?

6. Do you think you will continue to use language corpora in the future to help you learn Chinese? Why or why not?

7. Would you encourage other Chinese learners to use Chinese corpora in their Chinese studies? Why or why not?
### Appendix B: Corpus Survey Results

1. How often do you use the CCL (Center for Chinese Linguistics PKU) corpus outside of your class? (E.g. I only use it when I need to write my class assignments. / Once or twice a week in addition to the assigned homework.)

<table>
<thead>
<tr>
<th></th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Only use it for my class assignments</td>
</tr>
<tr>
<td>2</td>
<td>Once a week</td>
</tr>
<tr>
<td>3</td>
<td>I only use it when I need to write my class assignments because it takes a lot of time to look up the words in the sentence that I didn’t know.</td>
</tr>
<tr>
<td>4</td>
<td>I use it for the homework, and a few other times during the week with new vocab.</td>
</tr>
<tr>
<td>5</td>
<td>When writing homework assignment.</td>
</tr>
<tr>
<td>6</td>
<td>I only use it when I need to write my assignment.</td>
</tr>
<tr>
<td>7</td>
<td>About once more in addition to the assigned homework.</td>
</tr>
<tr>
<td>8</td>
<td>Generally I only use for assignments, but I actually used it a number of times independently. Probably 2+ a week.</td>
</tr>
<tr>
<td>9</td>
<td>I usually only use it for homework or once or twice a week besides that if I don’t understand the usage of a word.</td>
</tr>
<tr>
<td>10</td>
<td>Only when I need + for assignment.</td>
</tr>
<tr>
<td>11</td>
<td>I used it when I had class assignment.</td>
</tr>
</tbody>
</table>
I only use it when I need to write my assignment.

I only use it when I need to write my class assignment.

Once or twice a week in addition to assigned homework.

2. How did the CCL corpus help you to better understand various parts of the language that you learned in class? (For example, grammar, vocabulary usage, etc.)?

1. It helped me learn how to use grammar better, but for the most part it provided further examples so I could see how it is used.

2. It wasn’t that effective

3. It helped me become more confident in writing my sentences and using the grammar patterns in spoken Chinese.

4. It primarily just helped me understand usage better.

5. Vocabulary usage mostly

6. It was very helpful in learning the grammar to improve my speaking and writing skills.

7. Put the vocab in context, and gave me new ideas how to use the grammar.

8. It helped me understand the vocab word’s usage within a sentence.

9. Gave me lots of good grammar examples and showed me the correct usage of confusing vocab words.
When I did use it, it helped me understand how words and phrases were used and put together.

It helped me better understand how to use the grammar correctly because I saw examples.

Vocabulary usage

It helped me to see many different examples using the specific grammar principle I was learning, I felt that after looking at the corpus, I understand the grammar better.

Most helpful was seeing how a native used the words we learned.

<table>
<thead>
<tr>
<th>3. Did the corpus help you to better create your own sentences?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
</tbody>
</table>
Yes, I got a better concept of how to use the grammar so making my own sentences was better.

I sure hope. Yes?

Um, I still had a rough time with the correct grammar sometimes.

I believe so, yes.

Yes, it did.

Yes.

Yes.

A little.

4. Do you think using corpus regularly helps you improve your language proficiency? (Reading or generally?) Why or why not?

I think it could but I feel I didn’t take full advantage of it. I would quickly glance over sentences and write the shorter ones. Time was an issue.

No, I just did it for the assignment

I think it helps improve language proficiency generally because it helps the vocab grammar patterns become less strange.

I think it helps writing/speaking, because it gives you more ideas of when to use it. But
only if used over a long period of time would if actually improve overall proficiency.

5 It helps in some times. But sometimes it the examples are lousy.

6 It helps in speaking and writing, but not much in reading because there are still too many sentences I didn’t recognize.

7 Yes, the corpus sentences were a little too complex but I believe it helped me most with understanding how to personally use the grammar patterns.

8 I think it improves how to use words in general. This specific corpus often gave examples that were a little too technical.

9 Yes, it does. It helps with both reading and just in general because it gives you examples of everyday uses of Chinese.

10 I think it would-even though I didn’t do that-because we get more examples to draw from.

11 I think it definitely would if I had used it more diligently, because it gives tons and tons of examples to help you see how to better use the things you are learning.

12 Yes, because it teaches you new vocabulary.

13 I think regular use would help improve reading skills and comprehension skills.

14 I think it is helpful but not so much in the context in which we used it in class. For most of our assignments it didn’t help understanding and it just felt like busy work, because the words and phrases we used that corpus for were arbitrarily assigned and I already
understand the usage. I did find the corpus helpful when I used it on my own though because I only used it when I really wasn’t sure of usage.

5. Did using the corpus help you better understand various styles of the language. For example, more or less formal (书面语 / 口头语)?

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-A little bit. Didn’t focus so much on.</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>no</td>
</tr>
<tr>
<td>4</td>
<td>Yes, but only when the vocab, etc. was identified as 书面语 or 口头语 in class.</td>
</tr>
<tr>
<td>5</td>
<td>Not particularly because it doesn’t tell you what is what when reading it.</td>
</tr>
<tr>
<td>6</td>
<td>The only way I know it was whether or spoken was if the teacher told us in class</td>
</tr>
<tr>
<td>7</td>
<td>No not really, about the same. I still couldn’t tell which was which.</td>
</tr>
<tr>
<td>8</td>
<td>Mostly only 书面语</td>
</tr>
<tr>
<td>9</td>
<td>Eh, it didn’t help one more than the other. I often didn’t know if it was 书面语 or 口头语.</td>
</tr>
<tr>
<td>10</td>
<td>I wasn’t look for it, so, no. If I would have been looking for it, it probably could have.</td>
</tr>
<tr>
<td>11</td>
<td>Not entirely, I didn’t really learn much in this aspect from the corpus.</td>
</tr>
</tbody>
</table>
6. Do you think you will continue to use language corpora in the future to help you learn Chinese? Why or why not?

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I will use it in the future when I need examples of usage.</td>
</tr>
<tr>
<td>2</td>
<td>Yes, if I really want to learn a grammar structure on my own, I feel it is a good tool.</td>
</tr>
<tr>
<td>3</td>
<td>No, because most of the sentences have so much vocab that I don’t know so I might use it when I have more time.</td>
</tr>
<tr>
<td>4</td>
<td>Yes, I will, because it gives a broader understanding of usage.</td>
</tr>
<tr>
<td>5</td>
<td>Yes. I find it helpful for understanding how to use unfamiliar vocab.</td>
</tr>
<tr>
<td>6</td>
<td>Probably, because of its usefulness in learning grammar.</td>
</tr>
<tr>
<td>7</td>
<td>Yes, a good resource to see a variety of examples.</td>
</tr>
<tr>
<td>8</td>
<td>Yes, because it’s good to know how it is placed in a sentence.</td>
</tr>
<tr>
<td>9</td>
<td>Yeah, for sure! It is very helpful in clarifying things I am confused about.</td>
</tr>
<tr>
<td>10</td>
<td>I know I should, but I am probably a little lazy.</td>
</tr>
<tr>
<td>11</td>
<td>Yeah I think I will because it is a great asset that allows you to know how to use</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>grammar correctly.</td>
</tr>
<tr>
<td>12</td>
<td>Yes, because it has helped with my assignment.</td>
</tr>
<tr>
<td>13</td>
<td>Yes, when there is a phrase, grammar, or vocabulary that I don’t know.</td>
</tr>
<tr>
<td>14</td>
<td>Yes, but not as busy work</td>
</tr>
</tbody>
</table>

7. Would you encourage other Chinese learners to use Chinese corpora in their Chinese studies? Why or why not?

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-Yes. Using the internet to find reliable Chinese study help is a mystery to me. This helped find a good source.</td>
</tr>
<tr>
<td>2</td>
<td>Yes, it could be a useful resource if it wasn’t forced.</td>
</tr>
<tr>
<td>3</td>
<td>Yes, just to see where it can help them.</td>
</tr>
<tr>
<td>4</td>
<td>Yes, reason noted above.</td>
</tr>
<tr>
<td>5</td>
<td>Yes, because of the reason mentioned above.</td>
</tr>
<tr>
<td>6</td>
<td>Yes, because it requires that you spend more time in Chinese, which with help improve your Chinese.</td>
</tr>
<tr>
<td>7</td>
<td>Yes, can help people be exposed to real examples of Chinese grammar.</td>
</tr>
<tr>
<td>8</td>
<td>Definitely, because it is helpful in putting in context.</td>
</tr>
<tr>
<td>9</td>
<td>Yup! Same reason as above.</td>
</tr>
<tr>
<td></td>
<td>Yes, it can help for sure.</td>
</tr>
<tr>
<td>---</td>
<td>---------------------------</td>
</tr>
<tr>
<td>11</td>
<td>Yes I would because it helps to see more examples.</td>
</tr>
<tr>
<td>12</td>
<td>Yes, because it will help them generate better sentences.</td>
</tr>
<tr>
<td>13</td>
<td>Yes, it is helpful to see examples of the specific principle you are focusing on</td>
</tr>
<tr>
<td>14</td>
<td>Yes, because it is helpful to see native usage.</td>
</tr>
</tbody>
</table>
Appendix C: Consent to be a Research Subject

Introduction
This research study is being conducted by a graduate student, Rita Chen at Brigham Young University, Center for Language Studies to determine if Chinese corpus tool used in Chinese classroom can enhance students’ overall reading and writing abilities. You were invited to participate because you are enrolled in Chinese 301, Advance Mandarin class.

Procedures
If you agree to participate in this research study, the following will occur:

- You will need to take a reading test twice during this semester
- You will learn how to manipulate a Chinese corpus tool
- You will need to use the corpus tool to complete your class assignments
- You will need to fill out a survey in the end of this semester
- The researcher may contact you later to ask you to fill out a survey
- Total time commitment will be 240 minutes

Risks/Discomforts
There are minimal risks for participation in this study. You may, however, feel some discomfort when answering questions on the survey. If you feel embarrassed about answering a particular question, you may choose to decline or excuse yourself from the study.

Benefits
There will be no direct benefits to you. It is hoped, however, that through your participation researchers may learn about how effectual corpus is to your Chinese studies and may benefit your future Chinese studies.

Confidentiality
The research data will be kept on a password-protected computer and only the researcher will have access to the data. At the conclusion of the study, all identifying information will be removed and the data will be deleted.

Compensation
No compensation will be given for participation.

Participation
Participation in this research study is voluntary. You have the right to withdraw at any time or refuse to participate entirely without jeopardy to your class status, grade, or standing with the university.

Questions about the Research
If you have questions regarding this study, you may contact [researcher’s name] at [contact information] for further information.

Questions about Your Rights as Research Participants
If you have questions regarding your rights as a research participant contact IRB Administrator at (801) 422-1461; A-285 ASB, Brigham Young University, Provo, UT 84602; irb@byu.edu.

Statement of Consent
I have read, understood, and received a copy of the above consent and desire of my own free will to participate in this study.
Name (Printed): ____________________  Signature ____________________  Date: ____________