2

Learning in Parallel: Using Parallel Corpora to Enhance Written Language Acquisition at the Beginning Level

Brody Bluemel
The Pennsylvania State University

Abstract

This article illustrates the pedagogical value of incorporating parallel corpora in foreign language education. It explores the development of a Chinese/English parallel corpus designed specifically for pedagogical application. The corpus tool was created to aid language learners in reading comprehension and writing development by making foreign language texts more accessible to them. The study follows the application of this parallel corpus in two beginning level high school Chinese classes and describes the experience of both instructor and students in implementing this technology. The positive learning outcomes observed through students improvement in comprehension and composition demonstrate the value of parallel corpora both as a pedagogical tool and innovative technology.

Background

Written language acquisition, as with all aspects of language learning, presents learners with many unique challenges. One of the primary struggles of written language acquisition is the problem of reading comprehension, particularly at the beginning level. Language learners are often limited in the selections of texts that are accessible to them, both physically and intellectually, as they strive to develop their reading and writing skills. Additionally, learners often struggle with sentence composition, especially when learning to use lexical items and grammatical structures that have multiple meanings and/or uses. These challenges can be observed among learners of any foreign language (FL), but are acutely transparent with learners of Chinese. The extensive number and relative complexity of individual Chinese characters, as well as the associated tonal pronunciation, among other features, create several obstacles that learners must face in studying the Chinese orthographic system (Norman, 1988). The current article demonstrates how these challenges can be addressed through the use of parallel corpus technology.

In recent years corpora have evolved into a more accepted and valued tool for both research and FL pedagogy. The continued expansion of corpus research has led to the development of many specialized corpora, and to diverse applications in using this technology. A more recent outgrowth of corpus research is the use of parallel corpora. Although parallel corpora have been used for over two decades in transla-

tion studies and comparative language research, their application in FL pedagogy is a more recent trend (McEnery & Xiao, 2008). Studies suggest the apparent potential pedagogical value of parallel corpora (Fan & Xu, 2002; Johansson, 2009; Wang, 2001). The research presented in this article builds upon these studies by demonstrating how parallel corpora can be incorporated in the classroom, focusing on the beginning level of language learning.

In the present study, the researcher analyzed the application of a specialized parallel corpus in two beginning level high school Chinese classrooms. The corpus developed for this study was designed specifically as a pedagogical tool to help students develop their reading and writing ability in learning Chinese as a FL. The researcher created the Parallel Corpus Teaching Tool with assistance from a third party programmer. After reviewing several pedagogical studies using parallel corpora, the author describes the Parallel Corpus Teaching Tool, accessible at <www.parallelcorpus.com>. A detailed account is then given of how this corpus tool was incorporated into a high school Chinese FL classroom, and accounts for some of the struggles and the successes observed in adapting this parallel corpus in a beginning level FL class. The article explores the potential for pedagogical applications of parallel corpora, and discusses both the advantages and challenges of incorporating this technology in the classroom.

Literature Review

Over the past several decades parallel corpus technology has had a defined presence in translation studies and comparative language research, but its application has been noticeably absent from language pedagogy. Before exploring why this has been the case, it is first necessary to understand what exactly parallel corpora are. Parallel corpora are sometimes referred to as *multilingual corpora*, *bilingual corpora*, *translation corpora*, *comparable corpora*, or *equivalent corpora*. These many terms are occasionally used interchangeably, however, McEnery and Xiao (2008) provide a clear distinction among them by noting that translation corpora is the umbrella term that is then divided into three categories: parallel corpora, comparable corpora, and equivalent corpora. Parallel corpora are composed of source texts and their translations in one or more additional languages. Comparable corpora are two comparable monolingual sub-corpora that are designed using the same sampling frame. In other words, comparable corpora do not include translations but are composed of similar texts in two or more languages. Equivalent corpora are a combination of the first two categories (i.e. parallel and comparable corpora). The terms bilingual or multilingual corpora are simply used to distinguish the number of languages included in a translation corpus. Parallel corpora are the apparent subcategory of translation corpora considered in the current analysis.

Parallel corpora are constructed using either unidirectional, bidirectional, or multidirectional source texts. The unidirectional design includes original source texts from one language, and translations in the other (e.g. ,English source texts / Chinese translations *or* Chinese source texts / English translations). Bidirectional parallel corpora include a balance of source texts from both languages and their translations (e.g., English source texts/ Chinese translations *and* Chinese source / English translations). Multidirectional parallel corpora are multilingual corpora that

include the same piece of writing in numerous languages (e.g., the same article in Chinese, English, and German) (McEnery & Xiao, 2008). Each of these separate designs is best suited for different objectives and functions, and should be taken into account when using parallel corpora.

One final element to highlight is the necessity of textual alignment. The two (or more) textual versions of any text included in a parallel corpus (the original text and translation) are aligned in order to link the corresponding texts together. This alignment can be accomplished using phrasal alignment, sentential alignment, or context-based alignment (Biçici, 2008). Phrasal alignment links set phrases in the corresponding languages together, often at the word level. Sentential alignment links the two translations together in sentence segments. Context-based alignment requires several steps of machine processing in order to determine contextual phrases in the two languages, and then links the corpora together based upon the concept or idea being expressed (Biçici, 2008). Consideration of alignment method becomes important when considering how parallel corpora are used in pedagogy, as the different levels of alignment may require greater negotiation on the part of the parallel corpus user/language learner. After considering all these important characteristics of parallel corpora, a parallel corpus can be concisely defined as a bilingual or multilingual body of aligned corpus texts, consisting of source texts in one language and corresponding translated texts in one or more additional languages.

With a clear understanding of what parallel corpora are, the value of their use in translation and comparative language studies is evident. The application of parallel corpora in pedagogy has not always been so intuitive, but with advancements in technology and user-friendly corpus interfaces, the untapped potential of parallel corpora has recently begun to rise to the forefront. The limited number of studies that have been conducted provide foundational evidence of the value of parallel corpora in pedagogy, but also highlight the need for additional classroom applications and continued research (Fan & Xu, 2002; Johansson, 2009; Laviosa, 2002; Wang, 2001).

Previous studies have shown the potential for applying parallel corpora to FL classrooms. For example, Laviosa (2002) reported that navigation of a parallel corpus "reveals precisely the information that the learner needs to acquire in order to establish mental links between first language (L1) and second language (L2) schemas and create new L2 schemas when there is not reciprocity between the two language" (p.110). Laviosa's statement is substantiated by Tsai and Choi's (2005) study of lexical development among English L1 Chinese language learners. Their study analyzed the lexical acquisition and retention of American learners of Chinese using parallel corpus concordances to learn new vocabulary items in comparison with a control group who were presented with the same material in a traditional format (i.e., textbook, dictionary). The corpus-based group had a greater observed level of acquisition and retention of the tested lexical items in analyzing pre- and posttest results. More notably, though, they demonstrated a functional understanding of lexical terms with complex and/or multiple meanings as assessed by their ability to use new terms correctly in multiple contexts in which the meaning and form of the lexical item varied.

Frankenberg-Garcia (2005) provides further insight to Tsai and Choi's findings in the specific context of reading comprehension. "When reading in a foreign language, L2-L1 parallel concordances can help learners to understand foreign words,

meanings and grammar that they are unfamiliar with [… and] boost language comprehension" (ibid, p. 194). In support of this claim she details the experience of a Portuguese learner of English in understanding the *no matter how* meaning of the word *however* in the sentence *No programme, however good, can replace the role of the teacher.* The student had difficulty comprehending this usage of *however,* but the parallel corpus enabled him to conceptualize the meaning in his L1. These implications are also observed in Fan and Xu's (2002) study that had students use a Chinese/ English parallel corpus of legal documents to answer comprehension questions. Their study not only supports Frankenberg-Garcia's assertions, but also reported that the students actually preferred using the parallel corpus as opposed to other means. These studies demonstrate the unique role parallel corpora can play as a tool for learners grappling with complex structures, terms, and concepts by helping them to conceptualize these complex items in their native language.

An additional study by Xu and Kawecki (2005) provides interesting insight into the diverse potential for pedagogical applications of parallel corpora. These researchers used an aligned Chinese/English/French trilingual parallel corpus in a French FL class. Their study took place in a Hong Kong University with a classroom composed of Chinese L1 learners with high proficiency in L2 English. The learners used the corpus to derive meaning and use of new lexical items in French by comparing the French terms with the English and Chinese counterparts and contrasting their form and function. Findings showed that the use of the trilingual corpus enabled students to draw upon their knowledge of both their L1 and L2 when learning a third language. The authors concluded that the trilingual parallel corpus aided students' comprehension of linguistic concepts that are often pragmatically and semantically challenging.

Several other studies have also shown similar successful applications of parallel corpora in the FL classroom (Fan & Xu, 2002; Frankenberg-Garcia, 2000; 2005; St. John, 2001). This line of research has clearly established the pedagogical value of parallel corpora for FL learning. Though parallel corpora evidently have their place in the classroom, the primary limitation that currently impedes their application is the underdevelopment of this resource. There are relatively few parallel corpora available for use, and of those that are available, many have been developed for linguistics research and not necessarily as pedagogical tools accessible to students. For example, a review of the available corpora indicates that there are two Chinese/English parallel corpora accessible without download and software: *The Babel English-Chinese Parallel Corpus* and *E-C Concord.* While both corpora are valuable resources, they require some metalinguistic knowledge in order to best maneuver their interfaces. This limitation emphasizes the need for further application and development of parallel corpora, but likewise implies the importance of careful corpus and research design in developing new corpora and in creative planning in pedagogical applications of extant corpora.

Parallel Corpora in the Classroom

The research presented in this article answers calls for the development of new parallel corpus tools and research on their application(s) in the FL classroom. For this project a new Chinese/English parallel corpus was designed and constructed

specifically for the purpose of language learning. The primary aim in creating this parallel corpus tool was to develop a resource that would make written texts more accessible to language learners, and improve upon current approaches to written language instruction. As previously mentioned, L2 learners face two challenging tasks in written language acquisition: namely, reading comprehension and sentence composition and construction. These tasks are particularly difficult to address at beginning levels of language learning because learners may not have developed adequate vocabulary or functional knowledge of the language. The parallel corpus tool designed for this project addresses these and other tasks of language development by making written language more accessible to language learners at all levels.

While reading comprehension and writing development in any FL can be challenging, the unique and complex structure of written Chinese presents learners with additional obstacles. The Chinese orthographic system is composed of thousands of individual characters, with estimates suggesting that an individual must know 3,000–4,000 characters in order to read general texts such as newspapers (Norman, 1988). The large amount of characters that one needs to acquire makes learning written Chinese especially challenging for beginners since the Chinese writing system is not based on an alphabet, and students cannot simply sound out words or phrases. In alphabetically based languages, learners can sound out words and read through entire documents after mastering the alphabet associated with the language. This provides learners with the benefit of aural recognition as well as context to derive the meaning of unknown terms or phrases and to decipher textual meaning. Though being able to read through a document and understand it are two different things, it is important to remember that learners of Chinese face additional obstacles before even being able to read through a text. Consequently, Chinese language learners have less access to contextual information and/or the ability to draw on aural recognition when encountering new texts than do learners of alphabetic languages.

Another unique feature of Chinese is the composition of words. A Chinese word can be composed of one, two, or even more characters, and many characters can have several meanings. For example the character 会 (*huì*) generally means *to be able to*, but has additional uses. It can also be combined with other characters to form two-character words such as 社会 (*shèhuì*), which means *society*. Though the same character 会 (*huì*) appears in both words, the two words have very different meanings and function differently within sentences. So an individual who had learned the word 会 (*huì*) but not 社会 (*shèhuì*) would be easily confused and likely completely misunderstand a text that contained the latter term. This polysemic feature of many Chinese characters presents learners with several challenges. At one level, simply confusing word order within a sentence could result in an alternate meaning being expressed in the text. At another level, remaining unaware of multiple meanings and functions of terms greatly limits a learner's ability to effectively function and communicate within a FL.

These inherent challenges to learning Chinese orthography are specifically addressed and made more accessible through the use of parallel corpus technology. These specific benefits of parallel corpus technology add to the already established claim that using parallel corpora in language teaching enable students to conceptualize the target language through schemas in their L1. The corpus tool designed for

this study improves upon this inherent characteristic of parallel corpora by including added features that enable learners to read through Chinese texts fluidly, and not be limited by the characters they may not know yet. This allows individuals to not only benefit from aural recognition and context clues, but it also makes more advanced texts accessible to learners. Additionally, parallel corpora also provide learners with an efficient method for addressing the challenge of polysemy and multiple functions of characters. Corpora, in general, work well in addressing polysemy, but parallel corpora assist learners in more readily comprehending concepts through their first language. Again, it is because learners are able to explicitly link polysemous terms and complex constructs in the target language with the conceptual meaning expressed in the aligned L1 text.

When designing the parallel corpus tool for learners, addressing the aforementioned challenges of acquiring written Chinese needs to be the primary focus, with the ultimate goals of enhancing current methods for approaching Chinese reading and writing and making written texts more accessible to learners. Two key research questions that address pedagogical implementation and learner experience guided this study:

1. How do the students and instructor use the tool?

2. How effective is the parallel corpus tool in aiding students' acquisition of written Chinese?

The Development of the Parallel Corpus Tool Design

The Parallel Corpus Teaching Tool (Bluemel, 2013) used in this study was designed specifically for pedagogical application in the FL classroom setting. The creation of a new corpus tool, instead of adapting an extant one, was pursued for several reasons. First, the need for more and better-developed parallel corpora has already been established (Fan & Xu, 2002; Johansson, 2009; Laviosa, 2002; Wang, 2001). Second, developing a new corpus tool allowed for innovative functions that more precisely address the challenges of Chinese orthography. Third, research has demonstrated that designing a parallel corpus specifically for pedagogy improves students learning experience by allowing for an interface that is more accessible and easier to integrate into the curriculum (Lavid, Hita, & Zamorano-Mansilla, 2010). As previously mentioned, the majority of parallel corpora currently available have been created for linguistics research, without considering the possibility of pedagogical application. The design and features of the Chinese/English parallel corpus created for this study were greatly influenced by the intent for the pedagogical application of the tool.

Xu and Kawecki's (2005) study using a trilingual English/Chinese/French parallel corpus suggests the value of presenting parallel corpora in more than just the standard bilingual format. Though the parallel corpus tool created for this study is bilingual (Chinese/English), it includes texts in four language formats: Chinese characters, Chinese characters + tone marks, pinyin, and English. Just as the students in Xu and Kawecki's (2005) research were able to use both English and Chinese in learning concepts in French, the design of the Parallel Corpus Teaching Tool enables

students to use tone marks, pinyin, and English as aids in learning Chinese characters. In order to realize the significance of this design structure, it is necessary to first consider some basic elements of Chinese.

Though Chinese is written using characters, a corresponding system known as pinyin has become the standard writing system for transliterating Chinese characters by using the Roman alphabet. It is used both by FL learners of Chinese as well as native speakers of the language. The pinyin system allows for the alphabetic representation of characters, which can aid in reading, understanding, and typing Chinese. Another pertinent feature of Chinese orthography is that, generally, each character corresponds to one syllable, and every syllable/character has a tone mark. Chinese is a tonal language, and the tone associated with each syllable/character functions to indicate the meaning. Mandarin Chinese has four tones, plus a fifth neutral tone, making it possible for a syllabic utterance such as *ma,* to have five different possible meanings based upon the tone. It is therefore imperative that utterances in Chinese are spoken with the correct tone, and that students learn the correct tones associated with characters and meanings of words.

To aid in the mastery of tones, tone marks are used as part of the pinyin system to clearly demarcate tone. Tone marks appear in two forms, either as numbers following a syllable or, more typically, as diacritics written above the syllable, as illustrated in Table 1. In studying Chinese, learners typically begin by first learning the pinyin system and tone marks before moving onto characters.

Table 1

Tone Marks in Mandarin Chinese.

Tone	Tone indicated by Number	Tone indicated by Diacritic
1st Tone	ma1	mā
2nd Tone	ma2	má
3rd Tone	ma3	mǎ
4th Tone	ma4	mà
Neutral Tone	ma	ma

As the parallel corpus tool was designed, the function of pinyin and tone marks in learning written Chinese was integrated into the tool by aligning texts in four language formats: Chinese characters, Chinese characters + tone marks, pinyin, and English. Figure 1 depicts these four different formats. The Chinese character and English formats lie at opposite ends of the spectrum, and require little discussion as they simply represent the two languages. The character + tone marks and pinyin formats function as a form of *learner language* to assist learners in the acquisition of characters. While the pinyin representation of any character is available in most Chinese language technologies, the Character + tone mark format is a novel approach, which, to the researcher's knowledge has not been previously implemented elsewhere. These two additional formats specifically address the challenges learners face in being able to read complete texts, and are illustrated below.

Figure 1

Chinese Character	Chinese Character + Tone Mark	Pinyin	English
马	马̌	mǎ	horse

Though the texts are aligned in all four of the language formats, students are able to select which formats are visible. The goal is for students to use the least amount of mediation necessary to read and understand texts. Ideally, they would be able to read the Chinese characters, but if additional information is needed, they can then add tone marks to help with pronunciation. Thus the tool aids learners in written language acquisition by enabling them to comprehend texts by using additional formats as necessary, but also places an emphasis on character acquisition by presenting learners first with just the characters. This contrasts with a majority of beginning level learning materials that always juxtapose characters with pinyin, enabling students to completely ignore characters. In using the parallel corpus, learners can click on a character to add pinyin, but the primary focus in on character acquisition, only making pinyin visible as needed. Last, the English translation is also available for words or characters they are not able derive the meaning of using the other three textual formats.

The aim of this aspect of the design was to provide a learning environment in which students focus on character acquisition, and tone marks and pinyin are used as a form of interlanguage to assist them in the ultimate goal of learning characters. Systematically, the corpus did not have a tool that forced learners to use the tool in this manner, but classroom instruction on proper usage as well as screen-recordings of students' use of the tool ensured its application for the intended purpose. The corpus enabled allowed users to view entire texts, or they could also query a specific term or phrase using Chinese characters, pinyin, and/or English. A character input system was necessary to search for Chinese characters, whereas the tool included a feature that not only enabled, but also required, students to input the correct tone mark in order to search for the pinyin. When a student types a pinyin letter that could possibly contain a diacritic, the possible options appear on the screen and require the student to select the correct one.

The approach to compiling texts into the digital parallel corpus tool is also somewhat unconventional, and again intended to maximize the pedagogical value of the tool. The corpus is composed of bidirectional texts (Chinese L1 to English L2 translations *and* English L1 to Chinese L2 translations) that are functionally accessible to students, meaning the students are familiar with the content addressed in the included texts. The texts included in the corpora were selected according to two criteria: 1) the content of the texts related to the content covered in the course curriculum, and 2) the texts were challenging in that students did not necessarily know all words or characters, but appropriate for students' language level. The basic set of texts included the students' textbook, *Learn Chinese with Me* (Chen, 2009) along with several supplementary articles and books, which were selected by the in-

structor. The supplementary texts used in the corpus included bilingual short stories (Hou, 2006; White, 2004), bilingual textbook articles (The Overseas Chinese Affairs Office, 2007a; 2007b; 2007c), bilingual published speeches (Xu, 2011), and bilingual online articles (National Foreign Language Center, 2010).

While only a small subset of texts was initially made accessible through the corpus tool, the course instructor was able to add, change, or alter source texts throughout the semester. This means that the corpus size was continually expanding, and by the end of the research period the corpus contained 45 texts aligned in Chinese and English, of which approximately 18 were from the textbook and the remaining were from the selected supplementary articles and books. These 45 texts contained 26,563 words [6,512 Chinese words or tokens (11,039 Chinese characters) - 6,512 Chinese characters + tone mark tokens, 6,512 pinyin tokens, and 7,057 English tokens). By enabling the instructor to control the data that students are able to access, the goal is not to limit their exposure, but rather to insure that students are able to access material that is consistent with their learning level and with the content being covered in class. By doing so, the instructor can add more challenging texts as more vocabulary is learned throughout the semester, and hopefully provide texts that continue to challenge students in their development. This included adding texts that contained new lexical items as well as gradually adding texts of greater length.

All texts used in the corpus were aligned at both the word and the sentence level. These two levels of alignment allow for different functions within the tool. First, the word alignment allows students to search for words, and have the corresponding translation equivalents highlighted in the corresponding text formats (See Figure 2). The figure depicts the search term 马 with its corresponding pinyin *mǎ* and English *horse* all highlighted in green in the text. Additionally, an interactive feature of the tool allows the student to view the text in the character with tone mark format after clicking on the Chinese character. Next, the sentence level alignment was chosen for two reasons. First, the fluidity of text sources required a more standardized system of alignment. Second, the sentence alignment forces students to analyze the structure of the entire sentence, and observe how the two languages differ grammatically. By having both word-level and sentence-level alignment students are also able to identify specific terms and then compare how those terms function within the two different languages.

Figure 2

Methods

Participants

The study of the application of the parallel corpus took place in a Chinese FL course in an American high school. The participants (*N* = 15) in the study were all beginning level high school students (grades 9-12) enrolled in a second-semester Chinese class. They were between 15 and 18 years of age. The participants came from two different classes and did not include heritage learners. Students had three hours of classroom interaction each week as well as having a one-hour Chinese culture class each week. The first class had four students and the second eleven. All students participated in the study and the same instructor taught both classes.

The Study

This research took place over a three-month period in which all participants in both classes received the same treatment. During the first month of the study, the class continued unaltered, but with participant data (assignments, exams, etc.) being collected. Assignments were completed every other class period (once or twice per week), and there was an exam administered approximately every two weeks. The parallel corpus tool was introduced at the beginning of the second month of the study. At this time, participants were taught how to use the tool and encouraged, but not required, to use it in their study and class work as a resource for looking up new lexical items as well as examples of sentence construction. By the beginning of the third month of the study, the tool was completely integrated into the classroom as a regular part of participant activity used during all exams and for all in-class assignments. Participants had continual access to the parallel corpus tool during class time, and were encouraged to use it as needed even if the focus of the class did not specifically require its use.

Several varied data sources were collected to analyze the pedagogical implementation of the tool and the learners' experience. First, all course assignments and projects were gathered during the duration of the three-month study. Participants' performance on the assigned tasks and projects were used to assess and analyze performance in reading, writing, and lexical acquisition. Additionally, the participants' use of the corpus tool in completing projects and exams in the classroom was recorded using screen-capture software in order to examine exactly how participants applied the tool. All subjects were also required to keep learner logs as part of the course curriculum. These learner logs were completed with each assignment as well as with all exams and other projects. The logs provided details on how the tool was used for each individual assignment, and gave insight into the individual learner's experience in using the tool. Last, the instructor maintained an autoethnography throughout the semester, which details how he implemented and used the tool. At the end of the research period, participants also completed a brief questionnaire in which they evaluated the tool and gave written feedback on their experience.

As stated, learners were assessed through a series of assigned tasks and unit projects. The participants were asked to complete several assignments throughout the semester. These tasks were given both as in-class assignments and as homework assignments. The assignments included lexical acquisition tasks, reading compre-

hension tasks, as well as writing tasks. For example, after learning Chinese kinship terms, the participants used the parallel corpus tool to read a letter in which the author, Xiūmíng, described the members of his family. The letter included many of the new vocabulary terms recently covered in class and additional unfamiliar terms. In addition to reading and explaining the meaning of the letter through a written prompt, participants were asked to identify new lexical items and then search those items in the parallel corpus in order to find additional examples of their use and then describe the meaning of the terms and how they were used. Last, as homework, participants completed a separate writing task in which they had to use the new lexical items they had individually identified in writing a response letter to Xiūmíng by describing their own family. Participants in both classes were given the same assignments. During the first month of the study they completed these tasks without the aid of the parallel corpus tool. In the second month, they had the option of using the tool. In the third month three they were required to use the parallel corpus in completing their in-class assignments and exams.

Screen recordings were also taken of the participants while they completed in-class assignments and exams. *QuickTime* screen-capture software was used to record participants' screens during the unit assessments in order to document exactly how each participant chose to implement the corpus tool in completing their work. These recordings were taken to provide insight into what features of the tool the participants used most frequently, and which features the participants seemed to ignore.

The parallel corpus tool was created as a web-based tool. The website was designed for both computer as well as tablet access. One of the classes used school-issued *iPads*. Both groups had open access to computers, *iPads*, or both during class time, but also had full access to the tool from home. Thus, from the beginning of the second month of the study participants had open access to the parallel corpus tool.

Findings

With regard to the first research question about how the participants and instructor used the tool, the integration and application of the parallel corpus in the classroom was not without problem. However, the overall experience of both the instructor and the participants resulted in a very positive learning outcome. Taking the time to allow participants and the instructor to become familiar with the tool during the second month of the study allowed for a fluid adaptation that seemed to encourage both instructor and participants in using the technology. Participants all expressed an overall enthusiasm for the tool as expressed in their evaluations and learner logs, and their learning outcomes suggest reason for the instructor and researchers to share in their enthusiasm.

First, while the pedagogical implementation of the corpus tool proceeded quite smoothly, there were a couple of issues that arose. The instructor's autoethnography detailed how adapting the parallel corpus tool did require a couple additional hours of work each week in both preparing and designing lesson plans as well as in finding and preparing text sources for the parallel corpus tool. Specifically, lesson plans had to be restructured in order to incorporate the content and activities that were added using the parallel corpus. The greatest amount of preparation time was spent

in finding and preparing text sources to be added to the parallel corpus tool, as this required additional effort beyond the already established lesson plans. On average, each added text took approximately two hours to format and upload to the corpus database. Additionally, two different class periods had to be dedicated to teaching the participants to use the corpus tool. The instructor, though, reported that this extra effort and time was worth the effort due to the positive participant response and learning outcomes.

One of the challenges that arose in implementing the tool derived from its web-based interface. On two separate occasions the school's Internet service went down during the class. On both occasions the instructor delayed the planned lesson material – which required the use of the corpus tool – and did alternative interactive speaking/listening activities. This was particularly challenging the second time this happened as it was during an exam time, in which the participants were required to use the parallel corpus in completing an essay. Though the effect of these incidences was relatively minor, it does suggest a limitation of adapting technology based upon the quality of both the hardware and technology systems available.

Despite these challenges, the instructor was motivated to continue using the tool based on the effects he was observing among participants as they used it. Initially, the participants showed enthusiasm simply by the ability to use technology in the class-room. All of the participants were very familiar and comfortable with using technology, and were motivated by the ability to use it in studying Chinese. During the second month of the study, after the parallel corpus was introduced, nine of the 11 participants' homework assignments had notably improved. The writing responses were longer and better developed than in previous assignments, and included sentences with more complex grammatical structures. These outcomes will be discussed in greater detail shortly after first considering all data sources collected during the study.

Learner logs

In reviewing participants' learner logs there was an apparent consistency in participants reporting on their enthusiasm for the corpus tool, and reporting more time spent on their homework using the tool. While the increased enthusiasm cannot be attributed inherently to the parallel corpus, it can be attributed, and was explicitly so by participants, to the adaptation of technology. The use of technology appeared to boost their motivation for Chinese learning, and participant motivation remained relatively constant throughout the remainder of the study as detailed in their explicit responses in their learner logs.

Screen recordings

While this general observation of the participants' implementation of the tool was evident through their learning logs, greater insight into how participants chose to use the source is evidenced through the screen-recordings taken during in-class assignments and exams. Figure 3 depicts a screenshot of one participant screen recording taken during an exam, which was very typical of how participants' used the tool during exam periods. In this instance the subject was using the parallel corpus to search for the term 看到 (*kàndào*) which means *to see*. The parallel corpus returns a list of concordances of the term that the participant can use to both observe the meaning of the term and how it is used. Two of the five listings returned are

also shown in Figure 3. Participants were observed using the corpus tool to write answers to exam questions as well as to construct essays. By using the tool in such a manner, they then looked to the corpus texts as models for constructing their own writing. The instructor observed that this led participants away from giving standard textbook-style answers to questions and essays, and articulating more dynamic sentences that were more grammatically complex and expressive.

Figure 3

Qualitative Review Task

With regard to the second research question about the effectiveness of the parallel corpus tool in aiding students' acquisition of written Chinese, a qualitative analysis of an in-class review task completed immediately after fully implementing the parallel corpus revealed how the corpus tool functioned in aiding the participants' written language acquisition. Before this particular review task, participants had learned the vocabulary for several different animals, and then read a story about the twelve Chinese zodiac animals using the parallel corpus tool. The review assignment required them to identify five new lexical items, describe the meaning and use of these items, and then compose sentences with them. Additionally, participants were asked to identify the 12 zodiac animals and write a brief summary of the story. Finally, they were presented with five images of different animals and then prompted to write a sentence about each image.

The example of one participant, Erin (pseudonym), illustrates the general findings observed in both classes. In first evaluating Erin's review assignment, the ef-

fect of the parallel corpus tool on her development was not explicitly evident. She completed the assignment receiving full marks, demonstrating a clear understanding of both the lexical items being evaluated, as well as evidencing competence of the grammatical structures used in composing both her sentences, and overall summary. While Erin's improvement on this task indicated a stark improvement over comparable previous assignments (in which she averaged 84%), it was not possible to conclude from the assignment score alone that the parallel corpus tool was the variable that aided in her development. However, after then analyzing the screen recording of her use of the tool in completing this task, and considering her self-evaluation in the learner log, it became apparent that the parallel corpus tool was the element that led to her performance improvement. For instance, Erin identified the term 第 (dì) as a new lexical item, and she described the term as meaning "something similar to *the*, but only seems to come before numbers." As observed in the screen recording, Erin derived her meaning of the term by searching the parallel corpus and then evaluating the 12 tokens of use presented in the corpus. After reviewing the tokens of use of the term, she was able to not only provide a definition that showed an understanding of the concept and its use, but also to use it correctly in writing her sentence 第一个动物是老鼠 (*dì yígè dòngwù shì lǎoshǔ*) which translates as *the first animal was the rat*.

　　A similar observation was made when examining how she composed five sentences describing the images of animals. For the image of the tiger she wrote the following two sentences: 老虎很大。虎听音乐。(*Lǎohǔ hěn dà. Hǔ tīng yīnyuè*), which translates as *The tiger is very big. The tiger listens to music*. While the sentences are simple, and abstract in meaning as she describes the tiger listening to music, this example illustrates two aspects in which the parallel corpus tool aided Erin's conceptual development and understanding. First, she used the term 老虎 (*Lǎohǔ*) to refer to the tiger in the first sentence, and the term 虎 (*Hǔ*) in the second sentence. Both of these terms correctly indicated tiger, one was simply the longer two-character form, and the second was the one character form. While many nouns in Chinese follow this pattern and can be represented by either one or two characters, Erin and the other participants had not yet encountered or been instructed on this point. Erin's screen recording, however, showed her recognition of this distinction in the corpus text by identifying both characters corresponding to the English *tiger*, and then applying this knowledge correctly in her basic sentence construction. Also, it can be noted in this example was her use of the verb-object 听音乐 (*tīng yīnyuè*). This was a new verb phrase that Erin identified separately in the parallel corpus. Erin was observed in her screen recording identifying this phrase in the parallel corpus and then looking up other tokens of its use to derive the meaning. She identified the term and then compared it to the aligned English texts. Then, she used it grammatically correct in composing her own sentence. While the sentence, *the tiger listens to music*, was not a typical sentence one would expect to see when describing a tiger, it was correct. Furthermore, the analysis of the learning episode revealed how Erin was able to use the parallel corpus tool in order to learn a new term and then was able to apply it in creating her own sentence. Erin's example showed how the corpus tool was used to aid participants in learning not only the material covered in the class, but also to acquire additional linguistic knowledge in the process.

Discussion

The purpose of this research was to discuss the creation of a new digital corpora tool and examine its effectiveness in a beginning level high school Chinese course. Data from this study suggest that the Parallel Corpus Tool is a worthwhile and effective pedagogical application.

The observation of how the parallel corpus text influenced participant responses may be attributed in part to the nature of the corpus texts themselves. As parallel corpora are composed of source texts plus their translations into another language, translated language undoubtedly has an effect on the participants' learning. Frankenberg-Garcia (2004) points out that "it is well documented in the literature that the language of translation is not the same as language which is not constrained by source texts from another language" (p. 225). The *language of translation* should always be understood as a representation of the meaning expressed in the source language and not as a direct equivalent. While this observation is a reality of translation and therefore an inherent characteristic of parallel corpora, it is undoubtedly a strength that can aid learners in developing their conceptual knowledge of a language, as was observed in the participants' application of the parallel corpus in constructing their written responses.

Translations do not simply expose learners to two linguistic variations of a text, they also provide a written example of how a language expert (translator) chose to represent the meaning of the source text in his/her translation. As Aijmer (2008) observes, "translation is one of the very few cases where speakers evaluate meaning relations between expression not as part of some kind of metalinguistic, philosophical or theoretical reflection, but as a normal kind of linguistic activity" (p. 98). The cognitive decisions made by translators in order to best represent the meaning of a source text in a grammatically appropriate context in a second language provide an ideal model for participants to learn from. Thus, translation texts present learners with a model of a translator's conceptual knowledge of a language, and parallel corpora enable learners to take advantage of this knowledge in developing their own language skills and conceptual understanding. Participants in the current study demonstrated this by composing responses modeled after the parallel corpus texts that used more complex structures and diction than previously witnessed in their work. Though this observation does not prove that participants mastered these concepts, it does demonstrate how parallel corpora can be used to aid in developing these concepts and encouraging participants to become more independent language learners.

Corpora, in general, are important in language learning as they bring students in contact with the types of actual language structure and vocabulary that is encountered in authentic texts and communication. Though the parallel corpus used in this study included the students' textbook, it juxtaposed it with authentic texts, which allowed students to compare *textbook language* with *authentic language.* In other words, corpora expose learners to conceptual knowledge, or conceptual understanding of meaning, from the beginning stages of language learning. Parallel corpora go one step further by then enabling learners to link this conceptual knowledge with established schemas in their first language.

One of the participants in the study articulated this exact sentiment in a learn-

ing log toward the end of the study, noting that after using the corpus tool, she felt like she could actually understand things in Chinese instead of just learning new words. The participant was excited by her realization that, with the aid of the parallel corpus, she could understand concepts and ideas in Chinese. In effect Chinese became a functional language for her for the first time. In this instance, and as observed overall in this study, participants demonstrated an improved level of language comprehension through using the parallel corpus, and as previously discussed the corpus tool evidently assisted them in developing more articulate and meaningful writing in Chinese.

Future Research Directions

As a field of research, using parallel corpora as a pedagogical tool remains a blossoming area that currently has more questions than answers. As observed in the current study, parallel corpus technology has the capacity to aid in the learning of specific challenging aspects of the Chinese orthographic system. Each language comes with a unique set of characteristics and anomalies, and a parallel corpus could undoubtedly be similarly adapted to address the issues associated with numerous other languages. There is a great need not only for further research using parallel corpora, but also in the development of additional parallel corpora that are designed specifically with pedagogy in mind.

The pedagogical research applying parallel corpora is also evidently lacking in more specific case studies. It has been demonstrated that parallel corpora can function as a great tool for language learning, but exactly *how* to best apply this technology remains relatively underexplored. Are there specific concepts or constructs that would be made particularly salient to learners through parallel corpora? More content and concept specific research and applications of parallel corpora would greatly inform the research community, but more importantly it would benefit the academic community interested in pursuing this technology in teaching FLs.

Conclusions

The primary goal of this article was to investigate the potential value of parallel corpora as a pedagogical tool in FL education, and to discuss the strengths and issues of its application. Prior research has affirmed the practicality and viability of parallel corpora in pedagogy (Johansson, 2009; Wang, 2001; Fan & Xu, 2002), but there remain many unexplored questions in this area of research. The study presented here was one approach to addressing some of the unexplored issues.

The corpus-learning tool described in this study was created specifically for use as a pedagogical tool in teaching Chinese as a FL to English L1 students. The design of this tool sought to address the general areas of reading comprehension and writing development by making Chinese texts more accessible to language learners. Specifically, the challenge of reading and writing associated with character acquisition and correct tonal pronunciation were targeted. As was then observed, this corpus tool aided participants in more efficiently acquiring written Chinese.

The parallel corpus tool was adapted into a beginning level high school Chi-

nese classroom, and effectively aided participants in improving their reading and writing. As noted by one participant, the corpus enabled her to understand concepts *through* Chinese for the first time. This and other observations suggest an overall improvement in participants' written language skill and understanding of texts, leading to the conclusion the parallel corpus was effective in aiding participants' experience in acquiring Chinese. Additionally, consideration of both the instructor and the participants' experiences in implementing the tool into the classroom provides insight into how to effectively incorporate this technology, and what type of challenges can be anticipated. Overall, the parallel corpus tool was observed to be a very effective language-learning tool in addressing challenges of written language.

References

Aijmer, K. (2008). Translating discourse particles: A case of complex translation. In G. Anderman & M. Rogers (Eds.), *Incorporating corpora: The linguist and the translator* (pp. 95-116). New York: Multilingual Matters Ltd.

Biçici, E. (2008). Context-based sentence alignment in parallel corpora. In A. Gelbukh (Ed.), *CICLing, LNCS: Vol. 4919.* (pp. 434-444) Heidelberg: Springer-Verlag.

Bluemel, B. (2013). *Parallel corpus teaching tool.* <http://www.parallelcorpus.com>.

Chen, F. (2009). *Learn Chinese with me (*跟我学韩语*)*. Beijing, PRC: People's Education Press.

Fan, M., & Xu, X. (2002). An evaluation of an online bilingual corpus for self-learning of legal English. *System*, *30,* 47-63.

Frankenberg-Garcia, A. (2000). Using a translation corpus to teach English to native speakers of Portuguese. *Special issue on translation of Op. Cit, Journal of Anglo-American Studies, 3,* 65-73.

Frankenberg-Garcia, A. (2004). Lost in parallel concordances. In G. Aston, S. Bernardini, & D. Stewart (Eds.), *Corpora and language learners* (pp. 213-229). Amsterdam/ Philadelphia: John Benjamins Publishing Company.

Frankenberg-Garcia, A. (2005). Pedagogical uses of monolingual and parallel concordances. *ELT Journal, 59*(3), 189-198.

Hou, G. B. (侯冠滨). (2006). *Ant and cricket (*蚂蚁和蟋蟀*)*. (G.P Cai蔡关平, Trans.). Beijing, PRC: Hăitún Chūbănshè (海豚出版社).

Johansson, S. (2009). Some thoughts on corpora and second-language acquisition. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 33-46). Amsterdam/ Philadelphia: John Benjamins Publishing Company.

Lavid, J., Hita, J. A., & Zamorano-Mansilla, J. R. (2010). Designing and exploiting a small online English-Spanish parallel textual database for language teaching purposes. In M. C. Campoy-Cubillo, B. Belles-Fortuno, & M. L. Gea-Valor (Eds.), *Corpus-based aproaches to English language teaching* (pp. 138-148). New York: Continuum International Publishing Group.

Laviosa, S. (2002). *Corpus-based translation studies: theory, findings, applications.* Amsterdam & New York: Rodopi.

McEnery, T., & Xiao, R. (2008). Parallel and comparable corpora: What is happening? In G. Anderman & M. Rogers (Eds.), *Incorporating corpora: The linguist and the translator* (pp. 18-31). New York: Multilingual Matters Ltd.

National Foreign Language Center. (2010). *Read Chinese!* Available from, http://readchinese.nflc.org/

Norman, J. (1988). *Chinese*. Cambridge, UK: Cambridge University Press.

St. John, E. (2001). A case for using a parallel corpus and concordance for beginners of a foreign language. *Language Learning & Technology, 5*(3), 185-203.

The Overseas Chinese Affairs Office of the State Council, The Office of Chinese Language Council International (国務院侨务办公室, 国家汉语国际推广领导小组办公室). (2007a). *Common knowledge about Chinese culture (*中国文化常识*)*. Beijing, PRC: Gāoděng Jiàoyù Chūbǎnshè (高等教育出版社).

The Overseas Chinese Affairs Office of the State Council, The Office of Chinese Language Council International (国務院侨务办公室, 国家汉语国际推广领导小组办公室). (2007b). *Common knowledge about Chinese geography (*中国地理常识*)*. Beijing, PRC: Gāoděng Jiàoyù Chūbǎnshè (高等教育出版社).

The Overseas Chinese Affairs Office of the State Council, The Office of Chinese Language Council International (国務院侨务办公室, 国家汉语国际推广领导小组办公室). (2007c). *Common knowledge about Chinese history (*中国历史常识*)*. Beijing, PRC: Gāoděng Jiàoyù Chūbǎnshè (高等教育出版社).

Tsai, C., & Choi, H. (2005). *Parallel corpus and lexical acquisition in Chinese learning.* Proceedings of Fourth International Conference on Internet Chinese Education. Taipei, Taiwan.

Wang, L. (2001). Exploring parallel concordancing in English and Chinese. *Language Learning & Technology, 5*(3),174-184.

White, E. B. (2004). *Stuart Little (*精灵鼠小弟*)*. (R.R. Ren 任溶溶, Trans.). Shanghai, PRC: Yìwén Chūbǎnshè (译文出版社). (Original Work Published 1945)

Xu, X., & Kawecki, R. (2005). Trilingual corpus and its use for the teaching of reading comprehension in French. In G. Varnbrook, P. Danielsson, & M. Mahlberg (Eds.), *Meaningful texts: The extraction of semantic information from monolingual and multilingual corpora* (pp. 222-228)*.* New York: Continuum.

Xu, X. J. (徐晓健), (Ed.). (2011). *The world's famous speeches (*世界著名演说*)*. Qingdao, PRC: Qīngdǎo Chūbǎnshè (青岛出版社).