

Chinese CALL: Theoretical Foundation, Current Practice and Future Directions

Xujun Tian¹ Chris Shei²

¹China Three Gorges University, China

²Swansea University, UK

Abstract:

This paper discusses computer assisted language learning (CALL) theory and practice for the Chinese language. It examines current advancement in Chinese CALL research and points out the inadequacy of the existing second language acquisition (SLA) theory to serve as the foundation for Chinese CALL research. The paper also explores the current state-of-the-art Chinese CALL systems pointing out their characteristics and inadequacies. The paper hopes to show that there is a gap between the Chinese CALL development and that for the western languages (notably English). However, many technologies for integration into Chinese CALL systems are currently available. This paper introduces a few significant findings along the line of corpus linguistics, web as corpus and construction-based Chinese language teaching prototypes. Our task is to study SLA theory and CALL designing principles carefully and to put the Chinese language technology to good use, taking Chinese linguistics findings into consideration to make the Chinese CALL theory and practice more validated and effective for the future generation.

Key words: CALL, Theoretical Foundation, Current Practice, Future Directions

1. Introduction

Starting from the 1960s, Computer Assisted Language Learning (CALL) has been an interesting area of research and practice for over 50 years. From the simplest programs like flashcards and gap-filling exercises to the more sophisticated applications like Intelligent CALL, virtual world, and Web as corpus, CALL has offered endless possibilities for integrating technology into language learning. However, most CALL-related research so far has focused on learning the language of English. With more and more people in the world wanting to learn Chinese as an additional language, it is imperative that we develop Chinese-oriented CALL research and practice as a matter of urgency.

While there have been relatively few interests and

sparse publications on Chinese CALL in the past, recently a trend is beginning to form in scholarly publication of CALL. On the world-leading *Journal of Computer Assisted Language Learning* alone, Shei and Hsieh (2012) first sets out a principle for teaching Chinese pronunciation, words and phrases in a CALL framework. This publication is immediately followed by Wang (2012) which discusses the role played by e-dictionary in Chinese language reading. This is again closely followed by Qian and McCormick (2012) which discusses the implication of a distance learning functionality to Chinese language learning.

We are now at a junction where a number of issues must be sorted out in order to further develop Chinese CALL. Firstly, we need to go back to the theory and understand whether the language acquisition theory underpinning English CALL is equally applicable to Chinese, and if anything else needs to be added. Secondly, we need to know how much is on offer at present regarding Chinese CALL, including research findings and software and technology available. Finally, we must set out clear guidelines and frameworks for further research on Chinese CALL and development of pedagogical and commercial software. This paper attempts to address some of these important questions.

2. Current state of Chinese CALL

As previously mentioned, most research on CALL so far is based on the learning of English language. English and Chinese are very different languages. As Shei & Hsieh (2012) point out, many problems encountered by English-speaking persons trying to learn the Chinese language involve tone identification, character recognition, and morphological change and syntactic behaviors which are non-existent in the English language. Most particular issues related to Chinese language acquisition have never been addressed by CALL researchers. It is

incorrect to think CALL research outcomes based on English language can be transferred to the Chinese language without some degree of reconceptualisation and modification. It is therefore good to know the current state of Chinese CALL in terms of research and practice and how big the gap is between the Chinese and the English-based CALL.

On the theoretical side, CALL has traditionally drawn from theories of second language acquisition (SLA) for formulation of rationales in designing language learning systems. Are these English-based SLA theories equally applicable to Chinese language learning? We need to examine this fundamental question. For example, Krashen's theory of second language acquisition is widely adopted by CALL researchers and system designers. It consists of five main hypotheses:

- Acquisition-Learning hypothesis,
- Monitor hypothesis,
- Natural Order hypothesis,
- Input hypothesis,
- Affective Filter hypothesis.

As the theory is developed on the basis of English, there will be some differences in aspects of Chinese language acquisition. For example, the Natural Order hypothesis involves the order of acquisition for morphological and grammatical aspects of English language, which are largely absent in Chinese. The *i + 1* input in the Input hypothesis is easier to define in English in terms of vocabulary and grammatical complexity due to the large body of research surrounding English, but it is much less well defined in Chinese language acquisition. Clearly, research on Chinese CALL needs to address this fundamental issue, by clarifying how much of the English-based SLA theory is applicable to Chinese and by building up a particular body of Chinese SLA research in order to serve as the foundation for Chinese CALL theory.

On the practical side, there is relatively little Chinese CALL software on the market with advanced features like natural language processing and speech generation, intelligent tutoring or user modeling. Most Chinese language learning materials published so far are still based on text and audio recordings of the text (e.g. Chapin and Zhang 2007, Kantor 2006, Huang et. al. 2002). Some software employs relatively basic elements of technology to help with Chinese vocabulary learning. For example, Gengo WordPower: Learn Chinese Vocabulary

offers recording facility for learners to make basic comparison with native speaker pronunciation. Another basic feature of CALL – flashcards testing, is also included. Learn Chinese 2008, on the other hand, integrates vocabulary into reading and offers the functionality of wordlist compilation. Some slightly more

advanced features are present in some software. For example, Wenlin Software for Learning Chinese offers very useful stroke-by-stroke animations and editable dictionary. Rosetta Stone Chinese (Mandarin) uses speech recognition technology, language-enhancing games, live online lessons, and mobile apps.

Interestingly, a major brand name TeLL me More Chinese boasts “cutting-edge speech recognition software”, “automatic detection of pronunciation errors” and so on, allowing learners to “practice speaking in an interactive manner and to perfect [their] pronunciation”. Surprisingly, however, the Amazon user review of this software averages only 1.5 stars from 10 customers, with 7 users giving only one star. A reviewer says: “This software is basically a series of exercises” which means there are few elements of instruction. And “the fatal flaw though remains the lack of instruction” which means although the technology is present, it is not well harnessed or suitably integrated into the teaching model. This echoes the aforementioned point of the lack of research foundations in Chinese SLA and CALL theory and applications in general.

3. New development and future directions

A relatively recent development in CALL is using corpora as a resource for language learning and teaching. Corpora can be directly accessed by the learner with some kind of human or machine mediation to help with the process of language learning (Breyer 2009 offers a good summary and case study). Many CALL programs also use corpora to extract language patterns such as collocation or vocabulary as basis of learning (e.g. Shei and Pain 2000; Shei 2001; Chang et. al. 2008). Corpora can also be used to identify learner needs by analysing the gap between learner vocabulary as observed from learner corpora in comparison with teaching resources (Dodigovic 2005). Most of the existing literature in this CALL branch, again, is based on English or other western languages such as German or French.

While there has been relatively little work on corpus-related CALL targeting the Chinese lan-

guage, Shei (2005) makes a comparison between English and Chinese in terms of fixed expressions in both languages as observed in corpora on a specific genre. The research outcome is implicational for designing Chinese-specific pedagogical models. In addition, Shei (2008) offers a useful tool for identifying frequently used Chinese structures and expressions via the

web search engine. The methodology works on the basis of Google hits as the frequency of a word string. The more frequency a word string receives, the more established the expression is as a native-speaker usage. Moreover, the relative robustness of a frequently used string is established by the relatively small frequency gaps between its substrings. For example,

Table 1: Robustness of an established Chinese expression as evidenced by smaller frequency gaps

Chinese expression	儿童肥胖问题不容忽视		儿童肥胖造成智力低下	
n Google hits	儿童肥胖问题不容忽视	25,600	儿童肥胖造成智力低下	2
n-1 Google hits	儿童肥胖问题不容	31,500	儿童肥胖造成智力	2
n-2 Google hits	儿童肥胖问题	919,000	儿童肥胖造成	158,00
n-3 Google hits	儿童肥胖	994,000	儿 童 肥 胖	994,00
n-4 Google hits	儿童	81,300,000	儿童	81,300,00

By comparing Google hits in Table 1, we can establish that 肥胖问题 (‘obesity problem’) is a frequently used collocation in contemporary Chinese, at least when talking about children. This is because the little frequency gap between 儿童肥胖(994,000) and 儿童肥胖问题 (919,000) tells us that when the expression 儿童肥胖 is used, it is most frequently followed by the word 问题, establishing the strong bond between 肥胖 and 问题 in the given domain. Similarly, 不容忽视 is a strong collocation because the frequency drop accountable between 不容 (31,500) and 不容忽视 (25,600) is negligible, meaning that when the word 不容 is used (in the given context), the expression is most likely to end with 忽视. By contrast, there is no such small gap between substrings for the expression 儿童肥胖造成智力低下 (except the two 2 hits which is too astro-

nomically small to explain anything).

Shei and Hsieh (2012) further proposes a connectionist model to teach Chinese syllables, words and phrases in an integrated fashion, using the Shei (2008) machinery to identify the idiomatic value of a given expression. In addition, Shei and Hung (2012) uses Construction Grammar to integrate relevant aspects of a construction into a pedagogical framework for teaching the Chinese language. Each construction frame is based on numerous examples of authentic language use searchable from the web. Figure 1 shows an example construction.

Construction framework like this can be used in CALL environment as described by Schulze and Penner (2008).

<p>谢谢 + NP</p> <p>Discourse: polite, positive Pragmatic: “Expressive” speech act (revealing speaker’s psychological state) Semantic: speaker expresses gratitude towards NP Syntactic: 谢谢 NP (2nd person pronoun, kinship terms, professional titles, organizations, and so on) Phonetic and prosodic: normally spoken with grateful tone and facial expression with possible gestures</p>
--

Figure 1: An example construction (谢谢你, 谢谢妈妈, 谢谢老师, 谢谢政府...)

This line of research has brought the field of Chinese as an additional language closer to existing English-based linguistics and CALL research findings. Adopting corpora for language teaching and learning, especially using the web as corpus, offers the large amount of language data necessary to activate Krashen's 'i+1' Input machinery and provides the best possibility for extracting usage patterns to serve as pedagogical models. This is clearly one of the best directions where the Chinese SLA and CALL research should be heading.

4. Conclusion

In this paper, we have briefly described the current state of Chinese CALL in terms of both theoretical and practical aspects. As CALL is a versatile field with many possible types of designs and applications, in this short article, we have focused solely on the use of authentic text as basis for language learning and teaching. In other words, we discussed somewhat more extensively how the web can serve as a large corpus for patterns of usage to be extracted and applied to language learning and teaching. For lack of space, we have been unable to touch upon other interesting aspects of CALL relevant to Chinese language learning such as intelligent tutoring systems, advanced speech technology, natural language processing, dialogue systems, virtual world, mobile applications, and so on. However, the issues we brought up in this article, namely, the transferability of English-based SLA and CALL research to the Chinese language, will be the most important issue to be addressed by all researchers and practitioners of Chinese CALL. This paper has shown an example of how to develop Chinese CALL in line with English-based research findings. Future researchers interested in any of the CALL departments not addressed in this paper can follow the example and help establish CALL paradigms particularly useful to the Chinese language.

References

1. Breyer, Y. "Learning and teaching with corpora: reflections by student teachers". *Computer Assisted Language Learning*. 22.2 (2009.): 153-172.
2. Chang, Y, J. Chang, H. Chen and H. Liou. "An automatic collocation writing assistant for Taiwanese EFL learners: A case of corpus-based NLP technology". *Computer Assisted Language Learning*. 21.3 (2008): 283-299.
3. Chapin, A. and J. Zhang. *iSpeak Chinese Man-*

darin: The Ultimate Audio + Visual Phrasebook for Your iPod. New York: McGraw-Hill Professional. 2007.

4. Dodigovic, M. "Vocabulary profiling with electronic corpora: A case study in computer assisted needs analysis". *Computer Assisted Language Learning*. 18.5 (2005): 443-455.
5. Huang, Z. and Beijing Language Institute Staff. *Chinese for Today*. Hong Kong: The Commercial Press Ltd. 2002.
6. Kantor, P. *Chinese with Ease*. Noida, India: Gopsons Papers Ltd. 2006.
7. Qian, K. and R. McCormick. "Building course cohesion: the use of online forums in distance Chinese language learning". *Computer Assisted Language Learning*. iFirst article (2012): 1-26.
8. Schulze, M. and N. Penner. "Construction Grammar in ICALL". *Computer Assisted Language Learning*. 14.2 (2001.): 117-121
9. Shei, C. "Follow You!: An Automatic Language Lesson Generation System". *Computer Assisted Language Learning*. 14.2 (2001.): 129-144.
10. Shei, C. "Fixedness in genre-specific language and intercultural differences: Comparing English and Chinese fire news corpora". *International Journal of Corpus Linguistics*. 10.2 (2005): 199-225.
11. Shei, C. "Discovering the hidden treasure on the Internet: Using Google to uncover the veil of phraseology". *Computer Assisted Language Learning*. 21.1 (2008.): 67-85.
12. Shei, C. and S. Hung. "Construction Approach to the Teaching of Chinese as a Second Language". 2nd International Conference on Chinese as a Second Language Research. NTNU, Taipei, 17-19 August, 2012.
13. Shei, C. and H-P. Hsieh. "Linkit: A CALL System for Learning Chinese Characters, Words and Phrases". *Computer Assisted Language Learning*. 25.4 (2012.): 319-338.
14. Shei, C. and H. Pain. "An ESL Writer's Collocational Aid". *Computer Assisted Language Learning*. 13.2 (2000.): 167-182.
15. Wang, J. "The use of e-dictionary to read e-text by intermediate and advanced learners of Chinese". *Computer Assisted Language Learning*. 25.5 (2012): 475-487.