

# DDL in L2 Writing in China—A Retrospect and Prospect<sup>\*</sup>

AN Xue-hua, GAO Gui-zhen, CHENG Li-xia

Dalian University of Technology, Dalian, China

This paper reviews papers, in the field of DDL (Data Driven Learning) in L2 (second language) writing, published in international journals and in Chinese journals. Specifically, it compares the journals published in China with journals published internationally, aiming to inform DDL in L2 writing in the context of China. Differences are found in research methodology, research perspective, and research realm. It is proposed that DDL in L2 writing needs to be localized taking Chinese learners' characteristics into consideration, to inform L2 writing research and teaching in China. Furthermore, empirical research needs to be conducted to discuss DDL in writing teaching. Therefore, future research may seek to localize DDL teaching and research in China.

*Keywords:* DDL (Data Driven Learning), L2 (second language) writing, corpora, China

## Introduction

The advent of computers has provided a new and powerful research and teaching methodology for language related studies. Corpus linguistics, based on computer technology, can be utilized to analyze large data efficiently, which is a new research paradigm (L. WANG, 2008). Corpus linguistics has greatly impacted L2 (second language) writing teaching and research, widened L2 research methodology (Lee & Swales, 2006), and enriched teaching approaches (Flowerdew, 2010; Hyland, 2006). From the 1980s, corpus linguistics is one of the most promising methodologies that can inform L2 writing and research (Hyland, 2003; Bloch, 2007; Conrad, 2000, 2008; Granath, 2009; Stapleton & Radia, 2009). Corpus-based L2 writing will have an attractive and promising prospect in the future, and data-driven writing teaching should be emphasized in China (L. WANG, 2008).

## Literature Review

In corpus linguistics, corpus refers to large collection of naturally occurring texts for linguistic study (Biber, Conrad, & Reppen, 1998). With the aid of software, researchers can conduct relevant language theory and application research. Concordancer is the list of retrieved words and sentences with given linguistic items in the middle. The KWIC (Key Word in Context) is the common format with key words in the middle of the concordance display. With the rapid development of computer technology, up to date there have been several corpora with abundant language data in the world. The corpora is updated regularly. BNC (British National Corpus), is made up of 4,124 texts of different genres, with 10% of written texts and 90% of spoken texts. COCA

---

<sup>\*</sup> This paper is supported by teaching reform fund of Dalian University of Technology (Project No. MS201312 and No. JGXM201254).

AN Xue-hua, lecturer, School of Foreign Languages, Dalian University of Technology.

GAO Gui-zhen, professor, School of Foreign Languages, Dalian University of Technology.

CHENG Li-xia, associate professor, School of Foreign Languages, Dalian University of Technology.

(Corpus of Contemporary American English), created by Mark Davies of Brigham Young University, is the largest freely-available corpus of English, and a large and balanced COCA. The corpus contains more than 450 million words of text and is equally divided among spoken, fiction, popular magazines, newspapers, and academic texts. The interface allows you to search for exact words or phrases, wildcards, lemmas, part of speech, or any combinations of these. Besides, there are corpora for specific purpose, for example, the BAWE (British Academic Written English Corpus), covering Arts and Humanities, Social Sciences, Life Sciences, and Physical Sciences. The MICASE (Michigan Corpus of Academic Spoken English) focuses on academic discourse, including settings in lectures, experiment, office, and library services in American universities. In China, the last decade has witnessed the growth of corpora and related studies. CLEC (Chinese Learner English Corpus) is a large collection of Chinese English learners' English compositions, from different levels of proficiency, ranging from senior high school to university learners. The corpus is error-tagged according to an error marking scheme of 61 types of error, including various lexical, grammatical, semantic, and sentence level errors. CLEC is developed to study English learners' errors in English learning and provide feedback to English teaching practice (H. YANG, GUI, & D. YANG, 2005). SWECCCL (Spoken and Written English Corpus of Chinese Learners) includes both spoken and written language data (WEN & L. WANG, 2005).

DDL (Data Driven Learning) is rooted in computer science and was first coined by Johns in 1991 (Johns, 1986, 1991). DDL can improve learners' inductive learning ability through the analysis of large data. Based on traditional language learning 3P process, *presentation*, *practice*, and *production*, Johns (1997) advocated a new trilogy: *research*, *practice*, and *improvisation*. The previous studies have shown that DDL is beneficial to language learning in several ways. Firstly, KWIC will provide learners the most common language formats and structures. Secondly, learners can experience "real language" in authentic contexts, but not artificial language in grammar books and dictionaries. Thus learners' language awareness could be raised. Thirdly, learners could improve inductive learning capability in analyzing large quantities of language data.

### Review of International Studies

This paper reviews papers in the field of linguistics, published in top journals, like *TESOL Quarterly*, *Applied Linguistics*, and *ELT Journal*. Also journals in the field of EAP (English for Academic Purposes) and L2 writing journals, like *Journal of Second Language Writing*, *Journal of English for Academic Purpose*, are also covered. Furthermore, papers in journals focusing mainly on the application of computers in language studies, for example, *CALL*, *Language Learning & Technology*, and *System*, are reviewed. The empirical research on the application of corpus in L2 writing can be, for discussion purpose, generally divided into writing process research and writing product research.

**Writing process research.** (1) L2 Writing teaching research including DDL, corpus reference model in practice, writing design, etc. (Cresswell, 2007; Lee & Swales, 2006). Lee and Swales (2006) investigated four non-English speaking doctoral students and explored how the students made use of corpora. One of the tasks is to compile a corpus of their own writing, and compile an expert corpus, and make comparison between the two. It is revealed that students found it useful to make the comparison between the two corpora, and compared with dictionary, they preferred to refer to corpus for language help. The researchers also argued that the doctoral students could make better use of corpora due to their high motivation and motivation is a significant factor.

(2) Learners' acquisition studies, including learning strategies, learning process, learning characteristics, etc. (Todd, 2001; Gaskell & Cobb, 2004; H. Yoon, 2008; Kennedy & Miceli, 2001, 2010). Kennedy and Miceli (2010) developed and evaluated "apprenticeship" in learners' corpus consultation. Taking into the learners' language level, the researchers trained three learners the skills of pattern-hunting and pattern-defining tools. The result shows that learners' attitude and computer literacy influence the effect of DDL. Gaskell and Cobb (2004) conducted a research on 20 low-intermediate English learners in Canada, and investigated their ability to make revision with the help the concordancer. The result showed that concordancer reference may not be the "miracle cure for sentence errors" (Gaskell & Cobb, 2004, p. 315).

(3) Learners' individual differences, including learning motivation, characteristics, learning attitude, computer literacy, and their impact on DDL (Turnbull & Burstons, 1998; H. Yoon & Hirvela, 2004; Chambers & O'Sullivan, 2004; O'Sullivan & Chambers, 2006; Park, 2012). Chambers and O'Sullivan (2004, 2006) conducted a two-phase corpus training project, involving eight graduate students and 14 undergraduate students respectively. The researchers concluded that DDL was particularly useful to the acquisition of lexical grammatical patterns. The two-phase study revealed that postgraduate students are more active than the undergraduates. However, H. Yoon and Hirvela's (2004) study involved intermediate and advanced learners, and the result showed that intermediate learners benefit from DDL more than the advanced learners.

(4) Writing teachers' training research. O'Keeffe and Farr (2003) offered examples of corpus-based tasks for initial language teachers.

**Writing product research.** (1) Corpus-based text analysis, including the use of words, collocates, word frequency, etc.. Hyland (2002) investigated the use of directives in the published expert corpora and learners' corpora of around 2.5 million words. The result showed that in academic writing, the use of directives differ in different genres and disciplines.

(2) Learners' corpora analysis, including learners' language use, interlanguage, and error analysis based on corpora analysis. Rogers (2012) compiled a Japanese English learner corpora and found that culture and metalanguage may lead to errors.

### Review of Studies in China

This paper reviews papers published in the top journals in the field of linguistics and applied linguistics. The papers are also divided into two types, according to writing process and writing product.

**Writing process research.** (1) The teaching paradigm of DDL in teaching writing. LIU, GE, and G. LI (2012) conducted an experiment, comparing DDL and traditional writing teaching modes. The result is that DDL is more helpful to learners than traditional writing teaching methodology.

(2) Corpus and genre analysis teaching paradigm. Y. ZHANG and XU (2012) combined genre analysis and corpus approach in language teaching, and argued that the paradigm worked better.

(3) Theoretical discussion. J. WANG (2012) analyzed the motivation of corpus-based English vocabulary teaching from the perspective of Schema Theory, Lexical Approach, and Lexical Grammar, and explored the application of corpora use in vocabulary teaching.

**Writing product research.** Written product research includes research on error analysis, word acquisition, lexical chunks, and text analysis, based on self-compiled learners' corpora. S. ZHANG (2002) investigated the

English majors' use of the verb "make". DENG and XIAO (2005) studied grammatical verbs in university learners' argumentative writing. L. WANG and Y. ZHANG (2006) analyzed university students' use of lexical chunks in argumentative writing. LIANG (2006) revealed that text cohesion and the overall writing capability were correlated. HE (2009) investigated word misuse in 290 essays written by Chinese learners. CHU and ZHAO (2011) examined the use of adversative conjunctions in English argumentative essays by engineering majors in a Chinese university, based on self-compiled learners' corpora.

### **Critical Appraisal of Literature**

There has been extensive research on the application of DDL and corpus-based language research published. However, the applicability is more talked about than tested with empirical research (C. Yoon, 2011). Comparing the literature reviewed, there are differences between research published in international journals and research in China.

The international studies are mainly in-depth qualitative studies and case studies (Cresswell, 2007; Lee & Swales, 2006; Chambers & O'Sullivan, 2004; O'Sullivan & Chambers, 2006; Kennedy & Miceli, 2010), addressing concrete and specific questions. Therefore, the generalization of the research results should be exercised cautiously (C. Yoon, 2011), and some of the results are contradictory with each other. For example, the two-phase study of Chambers and O'Sullivan (2004, 2006) showed that postgraduates are more motivated than the undergraduates, whereas H. Yoon and Hirvela's (2004) study showed the intermediate learners benefited from DDL more than the advanced learners. The studies in China are generally large-scale quantitative research and theoretical discussion (LIU, GE, & G. LI, 2012; J. WANG, 2012). The research questions are more general and broad. The results would have proved to be more useful to future teaching and research.

The international studies focus more on writing process, including teaching design, corpus as language reference model, writing design, learning strategy, learner differences, etc.. The studies in China focus more on the written product analysis. Based on large quantity of learner corpora, language use is much discussed.

The studies in international articles mainly deal with English for academic purpose, while the studies in China study the behavior of English for general purpose (S. ZHANG, 2002; W. LI, 2003; DENG & XIAO, 2005; L. WANG & Y. ZHANG, 2006). However, there is an urgent need for academic reading and writing teaching in high school in China (LUO & CHEN, 2012). CAI (2010, 2012) has been appealing to redirect English teaching from test-oriented EGP (English for General Purposes) to research-oriented EAP in Chinese high school. In published articles, the theoretical discussions outnumber empirical studies.

### **Future Research in the Context of China**

Writing, as well as academic writing, is much influenced by social, cultural norm, and political factors. Therefore, localized writing pedagogy, taking local characteristics into consideration, is very important. DDL in L2 writing in the context of China should be more informed and tested by research.

Based on DDL, writing teaching resources should be developed to meet the needs of Chinese English learners. So far, DDL learning has not yet fundamentally impacted foreign language teaching, especially teaching materials (Romer, 2009, as cited in ZHEN & H. WANG, 2010). Therefore, teaching resources based on DDL can be a promising prospect for L2 writing teachers.

With the establishment of English as a lingua franca, and the improvement of basic English education, there is an urgent need of English publishing skills among postgraduate students. At present, a lot of universities would

require the graduates to published in SCI journals in order to get a diploma (Y. LI & Flowerdew, 2007). LUO and CHEN (2012) conducted a need analysis of academic English among the universities of science and technology. The result showed the postgraduate learners need academic reading and academic writing more than listening and speaking. Furthermore, in recent college English reform program, CAI (2010, 2012) has always advocated the establishment of EAP courses.

### Conclusions

This paper reviewed papers, in the field of DDL in L2 writing, published in international journals and in Chinese journals. A further comparison of the journals shows that there are some differences, from three aspects: research methodology, research perspective, and research realm. Based on the differences, future research in the context of China should seek more varied research methods. Furthermore, EAP teaching should be more emphasized in Chinese universities' English education, and more teaching resources should be developed to meet the demand of teaching and learning.

### References

- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge, UK: Cambridge University Press.
- Bloch, J. (2009). The design of an online concordancing program for teaching about reporting verbs. *Language Learning & Technology*, 13(1), 59-78.
- CAI, J. (2010). On the reorientation of Chinese college English education. *Foreign Language Teaching and Research*, 42(4), 306-308.
- CAI, J. (2012). On college English perishing and trend analysis. *Foreign Language Research*, 3, 46-52.
- Chambers, A., & O'Sullivan, I. (2004). Corpus consultation and advanced learners' writing skills in French. *ReCALL*, 16(1), 158-172.
- CHU, L., & ZHAO, C. (2011). A corpus of analysis of Chinese engineering majors' use of adversative conjunctions in English argumentative essays. *Foreign Language World*, 5, 9-15.
- Conrad, S. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century?. *TESOL Quarterly*, 34(3), 548-560.
- Conrad, S. (2008). Myth 6: Corpus-based research is too complicated to be useful for writing teachers. In J. M. Reid (Ed.), *Writing myths: Applying second language research to classroom teaching* (pp. 115-139). Ann Arbor, M.I.: The University of Michigan Press.
- Cresswell, A. (2007). Getting to "know" connectors? Evaluating data-driven learning in a writing skills course. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the foreign language classroom* (pp. 267-287). Amsterdam: Rodopi.
- DENG, Y., & XIAO, D. (2005). A study of college learners' use of collocates of grammatical verbs. *Foreign Language Teaching and Research*, 7, 7-10.
- Flowerdew, L. (2010). Using corpora for written instruction. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics*. London, UK: Routledge.
- Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors?. *System*, 32, 301-319.
- Granath, S. (2009). Who benefits from learning how to use corpora?. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 47-65). Amsterdam, Netherlands: John Benjamins.
- HE, H. (2009). A corpus-based error analysis of non-English majors' essays. *Foreign Language World*, 3, 2-9.
- Hyland, K. (2002). Directives: Argument and engagement in academic writing. *Applied Linguistics*, 23(2), 215-239.
- Hyland, K. (2003). *Second language writing*. Cambridge, UK: Cambridge University Press.
- Hyland, K. (2006). *English for academic purposes: An advanced resource book*. London, UK: Routledge.
- Johns, T. (1986). Micro-concord: A language learner's research tool. *System*, 14(2), 151-162.
- Johns, T. (1991). Should you be persuaded: Two example of data-driven learning. In T. Johns & P. King (Eds.), *Classroom concordancing* (pp.1-16). *English Language Research Journal*, 4.

- Johns, T. (1997). Contexts: The background, development and trialling of a concordance-based CALL program. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 100-115). Harlow: Addison Wesley Longman.
- Kennedy, C., & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology*, 5(3), 77-90.
- Kennedy, C., & Miceli, T. (2010). Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning & Technology*, 14(1), 28-44.
- Lee, D., & Swales, J. (2006). A corpus-based eap course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25(1), 56-75.
- LI, W. (2003). A corpus-based study of key words by English learners. *Modern Foreign Language*, 3, 284-293.
- LI, Y., & Flowerdew, J. (2007). Shaping Chinese novice scientists' manuscripts for publication. *Journal of Second Language Writing*, 16(2), 100-117.
- LIANG, M. (2006). Learners' written discourse cohesion study. *Modern Foreign Language*, 3, 284-330.
- LIU, B., GE, L., & LI, G. (2012). An empirical research on corpus-based college English writing teaching mode. *Foreign Language and Literature*, 28(4), 131-135.
- LUO, N., & CHEN, C. (2012). A need analysis of EAP among engineering postgraduates. *Contemporary Foreign Languages Studies*, 5, 38-42.
- O'Keeffe, A., & Farr, F. (2003). Using language corpora in initial teacher education: Pedagogic issues and practical application. *TESOL Quarterly*, 37(3), 389-418.
- O'Sullivan, I., & Chambers, A. (2006). Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing*, 15(1), 49-68.
- Park, K. (2012). Learner-corpus interaction: A locus of microgenesis in corpus-assisted L2 writing. *Applied Linguistics*, 33(4), 361-385.
- Stapleton, P., & Radia, P. (2009). Tech-era L2 writing: Towards a new kind of process. *ELT journal*, 64(2), 175-183.
- Todd, R. W. (2001). Induction from self-selected concordances and self-correction. *System*, 29, 91-102.
- Turnbull, J., & Burston, J. (1998). Towards independent concordance work for students: Lessons from a case study. *On-Call*, 12(2), 10-21.
- WANG, J. (2012). A corpus-based vocabulary teaching: Theory and application. *Foreign Language Research*, 4, 127-130.
- WANG, L. (2008). Corpus—A new perspective of English writing and teaching research in China. *The 4th International Conference on Teaching and Researching in the Chinese Context*. Beijing.
- WANG, L., & ZHANG, Y. (2006). A corpus-based analysis of lexical chunks in college learners' essays. *Computer-Assisted Foreign Language Education*, 110, 36-41.
- WEN, Q., & WANG, L. (2005). *Spoken and written English corpus of Chinese learners*. Beijing: Foreign Language Teaching and Research Press.
- YANG, H., GUI, S., & YANG, D. (2005). *A CLEC based analysis of Chinese English learners*. Shanghai: Shanghai Foreign Language Education Press.
- Yoon, C. (2011). Concordancing in L2 writing class: An overview of research and issues. *Journal of English for Academic Purposes*, 10(3), 130-139.
- Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning & Technology*, 12(2), 31-48.
- Yoon, H., & Hirvela, A. (2004). Esl student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13(4), 257-283.
- ZHANG, S. (2002). An investigation of Chinese English majors' use of "make". *Journal of PLA University of Foreign Languages*, 25(4), 58-63.
- ZHANG, Y., & XU, S. (2011). A corpus-based genre analysis—An empirical study of English majors' writing research. *Foreign Language World*, 6, 49-55.
- ZHEN, F., & WANG, H. (2010). The application of learner corpora in foreign language teaching. *Foreign Language World*, 6, 72-78.