

Corpus Linguistics and Corpus-Based Research in Hong Kong: A State-of-Art Review

Qin Xie¹

¹ Department of Linguistics and Modern Language Studies, Hong Kong Institute of Education, Hong Kong

Correspondence: Qin Xie, B4-2f-12, Department of Linguistics and Modern Language Studies, Hong Kong Institute of Education, No. 10 Lo Ping Road, Tai Po, NT, Hong Kong. Tel: 852-2948-8368. E-mail: qxie@ied.edu.hk

Received: June 11, 2013 Accepted: July 12, 2013 Online Published: August 15, 2013

doi:10.5539/ells.v3n3p48 URL: <http://dx.doi.org/10.5539/ells.v3n3p48>

The research is partially supported by Hong Kong Institute of Education via the Start-up Grant to Newly Recruited Lecturers. The author wants to thank Mr. Wong Sze Hon for his assistance in conducting the review. The ideas, views and any remaining inadequacies in this article are all mine.

Abstract

This article reports a state-of-art review of recent development on corpus linguistics and corpus-based research in Hong Kong. A top-down, multi-layer, stratified review identified 29 on-going research projects from the eight research-active universities in Hong Kong. These projects make use of corpus technology to address a wide range of research questions related to issues in language teaching and learning, linguistic studies, cultural studies, information technology, and sign language. Such diversified applications of corpus technology clearly indicate the great potential and opportunities that corpus linguistics can afford for researchers, language educators and learners to explore and collaborate.

Keywords: corpus linguistics, corpus-based research, language teaching and learning, cultural studies, linguistic studies

1. Introduction

The proliferation of corpus systems and techniques has enabled researchers worldwide to conduct research in their own geographical location with minimal hindrance. Over the years, corpus techniques have transformed the landscape of empirical research relating to linguistic studies and language education. It has been increasingly common for researchers to apply corpus techniques in their research as well as compiling their own corpora for specific purposes. This paper reports on a study that systematically reviews how language corpus have been developed and/or used by researchers to address a diverse range of topics in their latest academic endeavours, focusing specifically on research projects conducted in Hong Kong since 2012.

2. Review Scope and Procedures

The focus of the review is on recent corpus and corpus-based research (defined as 2012 onwards) conducted by researchers affiliated with the eight public-sector universities located within the geographic region of Hong Kong. The search was restricted to information available in public domains, such as university websites, individual staff websites, scholarly databases, journals, books, and conference proceedings. A top-down, multi-layer, stratified search was conducted starting from institutes, down to faculties, departments, and individual academic staff. At the individual level, each academic staff's recent journal publications, conference presentations, and on-going projects are reviewed to identify research relating to language corpus. Once individual researchers engaging in corpus-based research is identified, his/her names are used as keywords to search for research outputs in major scholarly databases, such as Google scholar, EBSCOhost, and SSCI. Following this systematic search procedure, the eight tertiary institutes and their relevant faculties and departments were reviewed sequentially. Table 1 lists the departments reviewed in each university. A total of 29 projects (defined broadly referring to projects, studies, and research outputs) are identified and classified. Table 2 and 3 and the reference list provide details of and access to the 29 projects for interested readers.

The 29 projects are published research output (journal papers, books), conference presentations, or work-in-progress from 2012 onwards. Depending on their status at the time of this review (March 2013), available details for each project and presentation may vary. Published research outputs are usually accessible in their entirety, whilst on-going projects often have limited information. Further searches were conducted for those entries with limited information via various search engines and scholarly databases using key words and author names. In the following, the 29 recent and on-going corpus or corpus-based language studies in Hong Kong are briefly described.

Table 1. List of institutions and departments searched

Institutions	Departments
The Hong Kong Institute of Education (HKIED)	Dept. of Linguistics and Modern Language Studies Dept. of Chinese Language Dept. of English Language Education Dept. of Cultural Studies
The Hong Kong Polytechnic University (PolyU)	Dept. of English English Language Center
Hong Kong University of Science and Technology (UST)	Center for Language Education
The Chinese University of Hong Kong (CUHK)	Dept. of Linguistics and Modern Languages Dept. of Chinese language and Literature Dept. of English Dept. of Translation
The University of Hong Kong (HKU)	School of English School of Chinese Center of Applied English Studies Faculty of Education
The City University of Hong Kong (CityU)	Dept. of Chinese, Translation and Linguistics Dept. of English
Hong Kong Baptist University (BU)	Dept. of Chinese Language and Literature Dept. of English Language and Literature Language Center Center for Translation
Lingnan University (LU)	Dept. of English Dept. of Chinese Dept. of Cultural studies Dept. of Translation

3. Results

Spanning across a wide range of topics, the 29 studies are classified according to their primary research focuses or target readers. Five major focuses are identified; Table 2 presents the number of studies identified for each focus. Twelve studies conducted to inform language teaching and learning are classified under “Teaching and learning”. Eleven research studies with their primary focus on linguistics and languages are put under “Linguistic research”. Three studies with their primary focus on cultural issues, such as inter-cultural understanding, Chinese classics, and Cantonese movies are put under “Cultural studies”. Two studies are classified under “Information technology”, these two studies were conducted to inform and improve the efficiency of search engines and remote access to information. While 28 of the 29 studies primarily focus on Chinese or English languages, there is only one project focusing on sign language, and this project is singled out as one distinct category – “Sign language”. It should be noted that some studies have several focuses and thus can be put under two or three categories. For ease of presentation and understanding of the research landscape, we only consider the primary focus of each study during the classification.

Table 2. Distribution of projects according to primary focuses

Teaching and learning	Linguistic research	Cultural studies	Information technology	Sign language
12	11	3	2	1

The following section reports on research under the five key focuses. Given the limited space for this report, not all research can be reported in full detail. Instead, relevant research is cited as examples to demonstrate how corpora are used to address respective research purposes. Interested readers can refer to the reference list and Table 3 for further details.

3.1 Teaching and Learning

A total of 12 projects have made use of corpora to answer research questions relating to language learning and teaching, including one edited book (Hyland & Wong, 2013). Studies under this category represent the largest group, which amounts to almost half of the 29 studies. These 12 studies commonly analysed the target corpus or corpora, compared and annotated upon it in order to enhance learners’ awareness of specific language items. For instance, the research titled “A corpus of textual revisions in second language writing” (Lee & Webster, 2012) made use of a corpus of L2 learners’ writing from the researcher’s own institute. This corpus includes original drafts, tutor’s feedback, and subsequent corrections of the learners’ writing. Errors made in L2 learners’ writing are calculated for their occurrence and classified at word, sentence, and paragraph levels. The study sheds light on the process of writing and informs the design of writing assistant tools.

Similarly, errors made by L2 users of academic English are also of great interest for cognitive linguists and second language acquisition researchers in order to study how a second language is learned or acquired. Error analysis has become a potent means of finding out how L2 learners conceptualize their newly learnt language under the influence of their native language. The research titled “Wrong to right: A cognitive linguistic approach to the analysis of second language errors” (Holme, 2012), investigated the phenomenon of language transfer amongst Chinese L1 users of academic English, whereby errors attributable to the interference of their native language were studied. This project used two corpora, the first one being the one million word corpus of academic writing in English from Chinese L1 users, the other being the *International Corpus of Learner English* (ICLE). First, consistent errors from the Chinese learner corpus are identified and classified according to their lexical-grammatical features. These selected errors were then translated into Chinese and searched for their frequency in the ICLE. The comparison helped to determine whether the errors are specific to the interference of L1 (Chinese). At the end, an empirically verified inventory of constructions that Chinese L1 users find difficult will be compiled. Error analysis also found its application in another project (Yan, 2012b) that aimed at creating an *Error Annotated Learner Translation Corpus*. The corpus is composed of translation work samples from 150 students from the Translation Department of the researcher’s institute. Translation errors were identified, analysed, and logged so as to provide students with a more comprehensive view on translation techniques.

Several studies focus on a particular lexical and grammatical phenomenon in L2 users’ language production. For instance, one project focuses on L2 users’ use of five high frequency abstract nouns generated from a news article corpus. The participants were given a list of concordance for the five abstract nouns to study on their own

before writing an essay using those nouns. Meanwhile, the control group was given the same writing task, but relied on dictionaries to understand the same five abstract nouns. The study found that participants using the concordance tools demonstrated significantly higher variety of collocational and colligational patterns than the control group, thus affirming the hypothesis that concordance can improve abstract noun use in L2 writing. Another project focused on the use of the common verb *make* by L2 users and its complementation patterns. The project employed two corpora: the British component of the *International Corpus of English* and the *Chinese English Learner Corpus*. The *Chinese English Learner Corpus* is further divided into a beginner and advanced level for comparison. Five types of complement patterns were studied for the verb *make*, they were then counted according to their occurrence in each category. L2 learners were found to have a tendency to overuse the complex transitive and intransitive patterns and to rely on pronouns as the direct object or the object complement. In comparison, native speakers make use of a wider range of complements for *make* including, pronouns, noun phrases, and *that* clauses. The findings of this project suggest that teaching materials need to be enriched as well as be cleared of “language skeletons” where certain complementary patterns are taught as dogmas.

Finally, a couple of the studies focused on developing learning systems to enhance learning on the basis of relevant corpora. One such learning system is the pronunciation learning system for Hong Kong/Mainland university students (Chen, 2012), which is based on a learner pronunciation corpus. The goals of the system are to identify learners’ recurrent pronunciation difficulties, and to offer remedies to students and teachers to improve students’ English pronunciation. Another corpus-based learning system was developed to enhance the learning of lexical-grammatical items for English majors (Ma, 2012). This system was based on three corpora including an English for Academic Purposes (EAP) learner corpus, an EAP corpus from professional academic authors, and a Chinese for Academic Purposes corpus.

3.2 Linguistic Research

Under the umbrella term of linguistic research, there are 11 research outputs, including three edited books on corpus linguistics and corpus-based research in language studies (Cheng, 2012a; Hyland, 2012; Hyland & Wong, 2013). Six research projects investigated various linguistic phenomena or particular features of the languages under study. One study made use of the switchboard corpus of telephone conversations to study dialogue acts in the naturally occurring spoken language (Fang, Cao, Liu, & Bunt, 2012). Telephone dialogues and conversations were recorded and annotated according to their communicative functions. For example, a dialogue segment could be tagged as yes/no questions or statement/opinion. The study identified 15 high-frequency dialogue acts, and then analysed their lexical dispersions as well as similarities. The study found a particular clustering pattern across the dialogue acts. Furthermore, dialogue acts are found to be highly distinctive from one another with regards to the lexical/word types, some word types were found to be used by fewer than four dialogue acts. This finding implies that it is possible to predict the occurrence of certain dialogue acts by specific lexical cues. Another study (Cheung, 2012) made use of two specialised corpora (Hong Kong Financial Services Corpus and Hong Kong Engineering Corpus) to study and describe examples of phraseologies that are discipline-specific or more frequently found in the financial services or engineering corpus.

Gaskell (2012) investigated the use of eight abstract nouns as epistemic stance markers within the *Michigan Corpus of Academic Spoken English* (MICASE). The eight most common shell nouns are: *sense, hypothesis, argument, feeling, thought, expectation, concept, and impression*. Of the eight nouns, *argument* was found to be the most interesting and thus chosen to be studied in greater depth. The study found that *argument* appears most prominently in the field of social science and least in that of physical science. Also there are frequent collocational uses of the word *make* before *argument* as well as *be* and *about* after it. Finally, the use of *argument* as anaphoric and cataphoric reference was investigated. *Argument* was found to be used for both anaphoric (in front) and cataphoric (back) references and the frequencies of the two uses were roughly equal.

Three studies have a focus on particular language arts. One research (Kwong, 2012) aimed at understanding how meanings and morals are realized in fables through various linguistic means and configurations. The research involved the compilation of a bilingual (Chinese and English) corpus of the *Aesop’s Fables*, including a short and a long version. The study annotated these fables at structural, discourse, and semantic levels. Ten well-known stories in both languages and both versions are selected for further analysis. Surface linguistic features, narrative structures, and discourse relationship are analysed to see the linguistic similarity and differences across different versions. The other research focused on classical Chinese poems from the Tang dynasty (Lee, 2012a). The study involves the tagging of the parts of speech (POS) in these poems. One particular difficulty for this task is related to the ambiguity of Chinese word boundaries. To solve this difficulty, the study developed a special scheme for tagging POS at two levels. The first level is described as “strings without internal structures” where the POS labels are only assigned to the word level with the number of multi-character words

kept at the minimum. The other level is where all compound words are considered to be “strings with internal structure” in which they are tagged according to various possible boundaries. For instance, the Chinese word *Yellow River* functions as a proper noun in denoting a geographical location, *Yellow* could also be tagged as a modifier of the noun-*River*, thus resulting in a nested POS where multiple tags are given. In a similar vein, Lai (2012) investigated Chinese parsing in an attempt to compile a syntactically parsed Chinese language corpus.

3.3 Cultural Studies

Three projects have a cultural focus. In one study by Cheng and Lam (2013), a textual discourse analysis was conducted to trace Western perceptions of Hong Kong for ten years after its handover. Western and Chinese media texts taken from the *Western Media Corpus* and *Chinese Media Corpus* (English language) were analysed for their lexical and semantic patterns. The study found certain collocation patterns such as *political/Hong Kong* frequently featured across the corpora. Employing concordance tools such as *ConcGram* and *Wmatrix*, the occurrence of certain collocates were statistically analysed and grouped according to their semantic preference and prosody. Take the collocate of *political/Hong Kong* as an example; a number of texts may have a positive opinion about political reform in Hong Kong, whilst others may have a negative opinion of it.

Yan (2012a) focuses on Western views of Chinese culture. This project conducted a cross-cultural analysis of Chinese cultural terms found in the English translation of Chinese classic works to describe how certain Chinese classics are understood in the West. A number of Chinese cultural terms are selected for analysis; the translated terms are then compared with the classics such as the *Analects*, giving some insight into how Chinese cultural terms are interpreted in Western culture. Apart from the two studies focusing on cross-cultural understanding, one study (F. K. Lee, 2012) applies corpus techniques to explore and understand the dynamics within Chinese culture defined broadly with a particular focus on gender representation in Hong Kong ESL textbooks and their changes overtime.

3.4 Information Technology

Whilst information technology has played an important role in facilitating the creation and analysis of language corpus, it can also benefit from corpus linguistics research. Two studies have employed corpus linguistic research to enhance the efficiency and effectiveness of information technology systems. One study (Liu & Hao, 2012) focused on improving an automated question recommendation system, where a corpus containing 216,563 questions (spread into 30 categories) from Yahoo! was used to match questions raised on the Internet to a list of potential answerers who are most willing and best suited to answer them. The other research (Webster, 2012) on a remote access translation system is a multi-lingual, corpus-based system designed to provide an automatic translation service, as well as offering a platform for corpus-based linguistic investigation. The system is designed for Internet and mobile communication use, whereby an automatic translation service could be enhanced and provided in the most convenient ways.

3.5 Sign Language

One project focuses on the compilation of a child sign language corpus (Tang, 2012) for the international community. A sizeable track record of various studies on sign language are also found for the same researcher and others attached to this particular institute (CUHK) spanning over a decade. In line with international standards, the team has been using sign language corpora to analyse classifier constructions, verb types and agreement, negation, modals, perfective aspect, and handshape phonology in child Hong Kong sign language.

4. Summary and Overview

This systematic review provides a snapshot of the most recent and on-going research conducted within the geographical location of Hong Kong. The diversity of research topics, the number of corpora used and languages under study give us a strong impression that corpus and corpus-based research is a vibrant field of research in Hong Kong. Although we are not in a position to speculate on possible research directions, general patterns were observed amongst the 29 studies.

Firstly, in terms of languages studied, English language corpora are used for the majority of the studies (19 out of 29), whilst some projects use Chinese language corpora (four out of 29) exclusively in the study of Chinese culture, language arts, and linguistics. Meanwhile, corpora of both languages (four) are used in some of the studies such as error analysis or translation studies. The remaining two consist of a multilingual corpus and a corpus of sign language. Language used in the research involves both the spoken and written form and covers a wide range of genres such as academic writing, naturally occurring dialogue, media texts, and literary works.

Secondly, a significant number of identifiable corpora have been employed and or developed by the researchers for their particular research. There are a total of 11 established corpora employed by the researchers. On top of

that, 13 corpora were compiled by the researchers themselves as part of their projects. Although some of the established corpora may bear resemblances with each other, there is no clear evidence of corpora use amongst different projects.

Whilst attempting to classify the collection of studies according to their key focuses, we found that some research are not as clear cut as the others in that they seem to carry attributes of multiple focuses. Take a particular work on the media discourse (Cheng & Lam, 2013) for example; whilst it employed a meticulous method in discourse analysis commonly adopted by linguistics research, the goal of the research is to examine Western perceptions of Hong Kong. In other words, the ends of cross-cultural analysis are achieved by the means of linguistic research. In Chin's project (2012a, b) on the Cantonese spoken corpus, 21 Cantonese movies produced in the mid-twentieth century (between 1950 and the 1970s) are transcribed, annotated, and incorporated. Whilst this corpus of Cantonese represents a distinctive local culture for a particular historical period, it also provides a platform for further linguistic research on lexical and grammatical variation of the Cantonese dialect over time. The potential of some research to inform different areas suggests the inter-disciplinary tendency of corpus-based research, whereby the collaboration of researchers with expertise from diverse areas is not only desirable but also necessary. Related to this, ten research projects are identified as teamwork (defined as projects with more than one investigator or publications with more than one author). Out of the ten team research projects, six involve collaboration of researchers from the same department or institute, and four involve research collaborators from other institutes.

Table 3. Tabulation of the 29 studies by focus and the primary investigator's affiliation

Universities	Learning and teaching	Linguistic research	Cultural studies	Information technology	Sign language
CityU		Fang (2012a)			
	J. Lee (2012b)	Fang et al. (2012)			
	J. Lee (2012c)	Fang (2012b)		Liu & Hao (2012)	
	J. Lee & Webster (2012)	Lai (2012)	Yan & Cheung (2012a)	Webster (2012)	
	Pan et al. (2012) Yan (2012b)	J. Lee (2012a) Kwong (2012)			
HKIEd	Chen & Wang (2012)				
	Holme, et al. (2012)	Chin (2012a, b)	F. K. Lee (2012)		
	Ma, et al. (2012)				
CUHK	Huang (2012)				Tang (2012)
	Xu, et al. (2012)				
Others		Cheng (2012a)			
	Cheung (2012)	Cheng (2012b)			
	Hyland & Wong (2013)	Gaskell (2012)	Cheng & Lam (2013)		
		Hyland (2012)			

Finally, there seems to be some clustering of researchers in terms of the researchers' affiliated institutions (Table 3). Specifically, we found 14 projects from the Department of Chinese, Translation and Linguistics, at the CityU; five projects from the Department of Linguistics and Modern Languages, at HKIEd, and three from the Department of Linguistics and Modern Languages, at CUHK. The rest of the projects are scattered across the other five universities in Hong Kong. Whilst this finding may suggest that corpus and corpus-based research may have taken on a lower priority in the other five tertiary institutes, this could also be due to the limitations of our search scope, which is restricted to publically available information. Table 3 provides a breakdown of the distribution according to the first author or primary investigator's affiliations. Judging from our findings, the projects and output from CityU seem to suggest interests both in teaching and learning and information

technology, whilst much of the research from HKIED seems to have direct or indirect implications for language teaching and learning. CUHK is the only institute with a focus on sign language (further information can be accessed via the Department's webpage).

References

- Chen, H. C., & Wang, L. (2012). A corpus-based online pronunciation learning system: The pedagogical applications of spoken corpus in improving Hong Kong/Mainland university students' English pronunciation. Retrieved February 28, 2013, from https://oraas0.ied.edu.hk/rich/web/search_prj_result.jsp?dept_code=LML
- Cheung, L. (2012). Enhancing a corpus-based approach to academic literacy and language learning in the fourth-year English-in-the-Discipline course in the new dental curriculum. Hong Kong. Retrieved February 28, 2013, from http://facdent.hku.hk/learning/cide_projects_008.html
- Cheng, W. (2012a). *Exploring corpus linguistics: Language in action*. London: Routledge.
- Cheng, W. (2012b). Exploring the limits of phraseology. *Asia Pacific Corpus Linguistic Conference*. Auckland: University of Auckland.
- Cheng, W., & Lam, P. (2013). Western perceptions of Hong Kong ten years on: A corpus-driven critical discourse study. *Applied Linguistics*, 34(2), 173-190. <http://dx.doi.org/10.1093/applin/ams038>
- Chin, A. C. (2012a). Spoken corpus construction and linguistic analysis of mid-20th century Cantonese. Retrieved February 28, 2013, from https://oraas0.ied.edu.hk/rich/web/search_project.jsp
- Chin, A. C. (2012b). New resources for Cantonese language studies: A linguistic corpus of Mid-20th century Hong Kong Cantonese. *Newsletter of Chinese Language*, 92(1), 7-16.
- Fang, A. C. (2012a). A corpus-based empirical study of paratactic constructions in contemporary English. Retrieved February 28, 2013, from <http://iris.cityu.edu.hk/>
- Fang, A. C. (2012b). A corpus-based investigation of paratactic and hypotactic chains in China English. Retrieved February 28, 2013, from <http://iris.cityu.edu.hk/>
- Fang, A. C., Cao, J., Liu, X., & Bunt, H. (2012). Lexical characteristics of dialogue acts in the switchboard corpus of telephone conversations. *Foreign Language Teaching and Research*, 44(3), 346-358.
- Gaskell, D. (2012). Evaluation in spoken discourse: An exploratory investigation of stance in the MICASE. *Asian Pacific Corpus Linguistics Conference*. Auckland: University of Auckland.
- Holme, R. H., Ma, Q., & Wang, L. (2012). Wrong to right: A cognitive linguistic approach to the analysis of second language errors. Retrieved February 28, 2013, from https://oraas0.ied.edu.hk/rich/web/search_prj_result.jsp?dept_code=LML
- Huang, Z. D. (2012). Can concordance improve the lexico-grammatical use of abstract nouns in L2 writing? *Asia Pacific Corpus Linguistic Conference*. Auckland: University of Auckland.
- Hyland, K. (2012). *Corpus applications in applied linguistics*. London: Continuum.
- Hyland, K., & Wong, L. L. (2013). *Corpora and innovation in the language classroom*. London: Routledge.
- Kwong, O. Y. (2012). Analysing story structure with a bilingual corpus of Aesop's Fables. *Asia Pacific Linguistics Conference*. Auckland: University of Auckland.
- Lai, B. Y. (2012). Dependency-based parsing of Chinese—Development of a parsed corpus and a supporting lexicon. Retrieved February 28, 2013, from <http://iris.cityu.edu.hk/>
- Lee, F. K. (2012). A corpus-based study of Hong Kong English language textbooks—Do they create and sustain a sexist culture? Retrieved February 28, 2013, from https://oraas0.ied.edu.hk/rich/web/search_prj_result.jsp?dept_code=LML
- Lee, J. (2012a). A classical Chinese corpus with nested part-of-speech tags. *The 6th EACL Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities* (pp. 75-84). Avignon, France: Association for Computational Linguistics.
- Lee, J. (2012b). An annotated corpus of Hong Kong ESL learners. Retrieved February 28, 2013, from <http://iris.cityu.edu.hk/>
- Lee, J. (2012c). Manual and automatic creation of learner corpora. Retrieved February 28, 2013, from <http://iris.cityu.edu.hk/>

- Lee, J., & Webster, J. (2012). A corpus of textual revisions in second language writing. *The 50th Annual Meeting of the Association for Computational Linguistics* (pp. 248-252). Jeju Island, Republic of Korea: Association for Computational Linguistics.
- Liu, G., & Hao, T. (2012). User-based question recommendation for question answering system. *International Journal of Information and Education Technology*, 2(3), 243-246.
- Ma, Q., Wang, L., He, A. E., & Decoursey, M. W. (2012). A corpus-based online learning system: Improving undergraduates use of lexico-grammatical items. Retrieved February 28, 2013, from https://oraas0.ied.edu.hk/rich/web/search_project.jsp
- Pan, H. H., Wong, K. F., & Zhou, M. (2012). Corpus-based linguistic knowledge acquisition for Chinese noun phrase. Retrieved February 28, 2013, from <http://iris.cityu.edu.hk/>
- Tang, W. L. (2012). Child sign language documentation. Hong Kong. Retrieved February 28, 2013, from http://www.cuhk.edu.hk/lin/new/en_research_project.html#7
- Webster, J. (2012). A remote-access multilingual corpus-based system for linguistic applications (RAMCORP). Retrieved February 28, 2013, from <http://iris.cityu.edu.hk/>
- Xu, Q., Ying, M., & Mak, W. F. (2012). What can MAKE tell us about second language learning? A corpus-based study of the verb complementation patterns for MAKE. *Asia Pacific Corpus Linguistic Conference*. Auckland: University of Auckland.
- Yan, X. J., & Cheung, P. K. (2012a). A corpus-based approach to the translation of Chinese cultural terms. Retrieved February 28, 2013, from <http://iris.cityu.edu.hk/>
- Yan, X. J. (2012b). Error annotated learner translation corpus: An interactive and constructive evaluation of student translations. Retrieved February 28, 2013, from <http://iris.cityu.edu.hk/>

Note

Note 1. For the web address <http://iris.cityu.edu.hk/>, it is the web address given to all projects found in the CityU website. Readers once onboard the iris system, can click *general public*, then *search for researchers* or *search for projects*, to find details relating to each project.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).