

Learner Corpus and Academic Writing: Identifying the Error Patterns of Taiwanese EFL Students

Hsiao-I Hou

Department of Applied English

National Kaohsiung University of Hospitality and Tourism, Taiwan, ROC

1 Son He Rd., Kaohsiung, 812, Taiwan, ROC

Tel: 886-8060505 E-mail: Hsiao@mail.nkuht.edu.tw

Received: March 1, 2016 Accepted: March 19, 2016 Published: March 20, 2016

doi:10.5296/jsel.v4i1.9193 URL: <http://dx.doi.org/10.5296/jsel.v4i1.9193>

Abstract

The purpose of this study is to find practical implications for improving academic writing curriculum design by exploring lexical and grammatical errors produced by EFL learners in a vocational institution in Taiwan. To satisfy this purpose, a small learner corpus was compiled in this study. The data were obtained from 58 undergraduate students of a public vocational university from September 2012 to June 2013 in Taiwan. There are 112 essays in the corpus that include 34,426 tokens. Learner errors were annotated based on the error categories found in (Dagneaux et al, 1996). Transfer errors and intralingual errors were analyzed. Lexical Tutor and AntConc software were used to conduct the analyses. Lexical errors, including incorrect word choices, misspellings, and word insertions, deletions or replacements occurred primarily due to students' insufficient mastery of vocabulary. The top three grammatical errors were verb forms, article errors, and preposition errors. Based on the research results, pedagogical implications that focus on teaching EAP writing to vocational university students in Taiwan are discussed.

Keywords: Learner Corpus, Error Analysis, Error Tagging

1. Introduction

This study reports the results of a preliminary analysis that aims to develop supplementary materials for an English for Academic Purposes (EAP) writing class in a hospitality-specialized university in Taiwan. This study was initiated in response to comments from writing instructors regarding the difficulties in correcting students' recurring grammar errors during class activities and assignments. The EAP writing class is designed to develop academic literacy skills and does not focus on the accurate use of grammatical forms, thus the idea was to provide extra input on grammar in the form of supplementary materials that the Taiwanese students could access outside of class. The content of these materials would focus on students' specific grammar problems and would not interfere with the type of language teaching and study skills training conducted during their regular English classes.

To develop materials of this type, it is useful to examine the formal errors this group of students typically makes. A small-scale corpus-based error analysis of the students' academic writing was conducted. As Aston (2000) mentions, there are two ways in which corpora can enhance language teaching: (1) by identifying features to be taught and developing materials to teach them, and (2) by functioning as resources for autonomous language learning. This present study belongs to the first of these. A general description of the study context, the error analysis process, the results, and tentative explanations for the causes of the errors will be presented in the following sections. The teaching implications will be summarized at the end. This study is intended to show that a small learner corpus can help with developing pedagogical materials that are appropriate for particular learners and thereby improve the curriculum design.

2. Literature Review

2.1 Learner Corpora

Corpus linguistics relies on the notion of practical evidence and second language acquisition. Researchers have found corpora—specifically what are known as computer learner corpora—to be particularly useful in objectively investigating learners' learning problems since the early 1990s (Leech, 1998; Tan, 2005). A learner corpus refers to a linguistic system based on the observable output that results from a learner's attempted production of a target language norm (Granger, 2002; 2003). There is an increasing interest in learner corpora both as a pedagogical tool and as a research tool. Learner corpora have been employed mostly to provide information on learners' common errors, such in the Longman Dictionary of Common Errors (Heaton & Turton, 2001) and the more recent Macmillan English Dictionary for Advanced Learners (2007), developed in collaboration with the Centre for English Corpus Linguistics (CECL) at the University Catholique de Louvain and based on the International Corpus of Learner English (ICLE) (Granger et al., 2002). However, although it can be used for pedagogical purposes, ICLE is mainly designed for research purposes along with many other academic and commercial corpora (Gragner, 2002; 2004; Myles, 2005).

As Gragner (2002) mentions, linguistic exploitation of learner corpora may involve one of two methodological approaches: (1) contrastive interlanguage analysis, which involves

quantitative and qualitative comparisons between native and non-native data or different varieties of non-native data from learners with different mother tongues, and (2) computer-aided error analysis, which focuses on errors in interlanguage and the use of computer tools to retrieve them. Computer tools have allowed researchers to handle vast corpora. The number of researchers and practitioners who are collecting their own smaller corpora to cater to the needs of a particular group of learners is still growing (Hidalgo et al., 2007). The present study is an example; by creating a small corpus and identifying and classifying errors in a database, it is possible to design pedagogical materials that are more locally oriented for learners of a particular mother tongue in a particular context.

Learner corpora may also be used for classroom methodology by exploiting the corpus in class for inductive learning (Aston, 2000). In addition, learner corpora can be used as a basis for curriculum design (Hutson, 2002). An advantage of using learner data for these purposes is that they are authentic data—although as Granger mentions, the notion of authenticity in reference to learner data are somewhat problematic. Sinclair's (1996) definition of authentic data are those gathered in real communication situations, with people going about their normal business. In that sense, learner data are very rarely authentic, and learner corpora are often experimental. Granger (2002) defines learner data as authentic when the data are generated from authentic classroom activity, as in the case of the written essays collected for this study. It is believed that by having a good understanding of learners' difficulties, teachers and teaching materials can help students become better error and contrastive analysts by fostering language awareness, with the ultimate purpose of promoting proficiency. It is in this sense that a learner corpus can be used as a powerful pedagogical tool.

2.2 Error Analysis

The analysis of learner corpora is key in determining the items of the language code that are worth teaching the most in English for Specific Purposes (ESP) courses. In the 1950s and 1960s, contrastive analysis (CA) was the favored paradigm for second and foreign language teaching and learning (Mendikoetxea, A., Murcia, S. & Rollinson, P., 2010). The general idea behind CA was that difficulties in learning were associated with differences in structure between the mother tongue and the target language. It proceeded by describing and comparing features of the mother tongue with those of the target language with the goal of formulating predictions about areas that could cause interference and error. However, CA was discredited by the early 1970s—not only as a result of its association with structuralism and behaviorism but also because its predictions were thought to be unreliable (James, 1998). Error analysis (EA) was the paradigm adopted to replace CA. It proceeded by first describing the learner's interlanguage and the target language and then comparing the two by searching for mismatches without referring to the mother tongue. In the 1980s, EA also went out of fashion because of the advent of the communicative approach. However, James (1998) still suggests that “EA has never been abandoned, but has rather lain in the doldrums perhaps awaiting the signal to ply the main” (p.19).

EA was criticized due to the data used: the material explored was broadly associated with lists of errors gathered from elicited practice sessions where little attention was paid to task,

learner or language variables. Therefore, the results usually materialized as unnatural and heterogeneous collections of data (Ellis, 1994). A further weakness attributed to EA is that errors were explained according to inadequate taxonomies often grounded on unobservable, subjective characteristics and that included overlapping categories. To attempt to overcome the inadequacies of error classifications, EA researchers have reformulated error taxonomies (Lenko-Szymanska, 2003). Currently it is generally claimed that taxonomies should be grounded on descriptions of observable data and include well-defined linguistic categories to minimize subjectivity in the process of error diagnosis and categorization (Dagneaux et al., 1998). According to Granger (2002; 2004), for an error annotation system to be fully effective, it should be consistent, informative, flexible and reusable. To make every manual error tagging operation consistent and informative, clear tagging guidelines that include detailed information is necessary. Flexibility in error tagging means that the tagset should provide the information that one wishes to search for. Tagset reusability means that the tagsets must be suitable for use in learner corpora for a variety of languages. Different types of errors must be taken into consideration while developing the tagging system. In error taxonomies, spelling, grammar and lexis are the most common linguistic features to be covered. Phonetic, pragmatic or discourse errors do not seem to always be present in the error tagging system.

Current EA is said to be in need of further work to standardize error typologies. Unlike other areas such as learner corpus design, where more standardization might exist, corpus researchers have yet to agree on a general scheme of error annotation (Tono, 2003). In general, research groups often appear to design their own error tagging systems and explore different tagging models and error typologies. Indeed, the diversity of error tagging systems seems to be evidence of the constant questions about emerging approaches to error annotation and also of the need for a benchmark for analyzing computerized learner errors. Although language instructors may be aware of areas of difficulty that their students have, the only way of uncovering error frequencies and, in general, of becoming more aware of their performance is through the analysis of error-tagged materials (Granger, 2002). Moreover, as has been claimed, the analysis of errors provides actual evidence of the areas that learners still need to master and therefore disclosing their pedagogical needs (Izumi, 2004).

The present study adopted the coding scheme from Dagneaux et al. (1998) and Chuang and Nesi (2006). Their error tagging system is hierarchical: error tags consist of one major category code and a series of sub-codes. There are seven major category codes: formal, grammatical, lexical-grammatical, lexical, register, word redundant/word missing/word order and style (Dagneaux et al., 1998). In the study by Dagneaux et al., which analyzed 75,000 words of essays written by French-speaking university students from the ICLE database, the most common error types are article usage (28%), vocabulary error (25%), prepositions (18%), and verb-subject agreement (15%). In Chuang and Nesi's (2006) study, their corpus contained fifty essays written by intermediate and upper-intermediate Chinese learners. Their research results show that grammatical errors were considerably more common than lexical and lexico-grammatical errors, suggesting that Chinese students had more problems with morpho-syntactic features than with lexis. In addition, the top ten most frequent types of error

are shown in Table 1. The first, second, third and ninth error types all concern the article system. Several factors contribute to the high frequency of article errors in Chinese students' writing. First, articles are used extremely frequently in writing. Second, the English article system is complex, making it difficult for non-native English speakers to master. Third, the use of articles are closely related to the features of nouns. The Chinese language does not distinguish between count and non-count nouns and does not distinguish between singular and plural nouns. Some other common errors also include preposition problems, lexical problems and tense problems (Chuang & Nesi, 2006).

Table 1. Summary of Error Analysis Results of Chuang and Nesi's Study (2006)

Error type	% of all errors
1. Missing definite article	10.1
2. Bare count noun for plural	8.8
3. Redundant definite article	8.5
4. Mis-selection of preposition	6.1
5. Lexical misconception	5.8
6. Incorrect tense and aspect	3.8
7. S-V non-agreement	2.4
8. Incorrect collocation	2.1
9. Missing "a/an"	2.0
10. Comma splice	2.0

Both studies have produced comprehensive error tagging systems and provided useful teaching implications to ELT. EA is a powerful technique that will help ELT materials designers produce a new, more learner-aware pedagogical tool.

3. Method

3.1 The Learner Corpus

To satisfy the purpose of this research effort, a preliminary corpus-based error analysis of the Taiwanese students' academic writing was initiated by working with a small pilot corpus of twenty essays. The essays were collected from Taiwanese freshmen students who are majoring in English at a hospitality-specialized university in the south of Taiwan. All the students were between 18 and 20 years of age, and Mandarin was their mother tongue. Before enrolling at this university, they all had completed their high school educations in Taiwan.

The students enrolled in the English program had come through different admissions channels, thus their English proficiencies varied; however, in most of the students' classes, English is the medium of instruction and evaluation.

3.2 The Coding Scheme

Based on the twenty essays, a tagging system was developed by a native and a non-native speaker of English based on an integration of the previous studies of Dagneaux et al. (1998) and Chuang and Nesi (2006). The error system used in the present study focuses primarily on grammatical errors. For reliability, as Dagneaux et al. (1998) suggested, each error category should be described in terms of linguistics categories to distinguish it from other categories. Table 2 summarizes the lexical and grammatical coding scheme of this study. The errors listed in Table 2 are the main focus of this research.

Table 2. The Coding Scheme of the Present Study

Error Type	Code	Description
Article error	A	Any error involving an article
Preposition Error	P	Any error involving a preposition
Noun Error	N	Error involving plural/singular confusion of a noun
Verb Form	V	Errors in verb tense and verb inflections
Word Form	W	Correct lexeme, but wrong suffix
Spelling	S	Error in spelling
Punctuation	Pun	
Word misuse, insertion, deletion, or replacement	O	Other corrections that do not fall into any of the above categories

After the error tagging system was developed, the essays were collected between September 2012 and June 2013. The corpus contained a total of 112 essays written by 58 students. These essays originated as writing assignments for an English composition III class. In this writing class, students learned to write different types of short essays with an emphasis on essay structures and development. The essays collected in this corpus are narrative and process analysis essays. Students chose their own topics for the writing assignments and they used word processors to write the assignments. All essays were first drafts.

After all the essays were collected, the errors were manually tagged by two raters: a senior

student and native English speaker who had formerly worked as a TA for a English composition class and an English instructor with a specialty in teaching languages and who had prior experience in language annotation. Based on the research goal—to understand the formal errors that Taiwanese learners typically make—an error tagging scheme was developed. The two raters had an agreement rate of 82% on identifying and categorizing error features on 8 (37.5%) randomly selected essays.

After all the texts were tagged, two corpora were created. One is intended as an advanced learner corpus and the other is intended as an intermediate learner corpus. Students who have a TOEIC score higher than 750 are considered advanced learners (a TOEIC 750 is equivalent to a CEF B2 level). *Lexical Tutor* and *AntConc* were used to analyze the data.

3.3 Error Tagging Sample

Below is an example of an essay tagged by the error raters of this research.

The summer **//S//vocation** was coming, but I **//V//have** not found a job yet. When I was really depressed and anxious, my phone rang, and the **//O//captain** finally responded to me. Afterwards, I was really shocked because I **//V//found** my first job, and the location was the department store.

I worked as a **//O//bartender** at **//A//the** delicious **//O//food street**. In the beginning, the staff **//O//directed** me how to use **//A//** cash **//O//** desk and introduced the products **//O//of** the shop **//V//offering**. I felt really **//S//pressed** because I was afraid that my carelessness **//V//may** cause the company to lose money. However, the staff was really friendly to me, and seldom blamed me. And this **//O//work** usually **//O//met** Chinese or **//A//the** HK people because they thought Taiwan was famous for tea; therefore, our shop **//O//earned** lots of money from these customers. When the summer **//S//vocation** **//O//finished**, this **//O//work** became boring, but usually busy **//P//at** weekends.

From this experience, I thought I was really a very fortunate person because I found **//A//the** nice **//O//work** and met nice staff **//Pun//**, **//V//learning** more about **//O//knowledge of** tea. Moreover, the most important thing I **//V//learn** **//V//is** **//V//deal** with customers and how to get along with staff **//Pun//**, I thought it **//V//is** very important to students who study **//A//the** hospitality

To sum up, I really **//V//enjoy** this work, and I **//V//become** more outgoing than before. I **//V//believed** this experience will help me for the internship.

4. Results

Table 3 summaries the profile of the corpora. There are 19,393 tokens in the advanced learner corpus and 15,033 tokens in the intermediate learner corpus. These data show that advanced learners tend to write longer essays compared with the intermediate learners. In addition, advanced learners also used more of the “1000 most frequently used English words” in their writing than their intermediate counterparts. However, the intermediate learners tend to use more English words from their academic word lists than their advanced counterparts. In the advanced learner corpus, there are 1,151 errors and in the intermediate learner corpus, there are 1,583 errors. The intermediate learners made more errors in their writing compared with

the advanced learners. On average, each essay in the advanced group contained 52 errors, while the average error count for intermediate group essays was 72.

Table 3. Profile of the Corpora for the Present Study

	Advanced Learners	Intermediate Learners
Tokens	19,393	15,033
Lexical Density	0.53	0.48
K1 words	15,879 (81.88%)	12,171 (80.96%)
K2 words	1,507 (7.77%)	1,385 (9.21%)
AWL words	448 (2.31%)	442 (2.94%)
Off-list words	1,559 (8.04%)	1,055 (6.98%)
Errors	1,151	1,583

Table 4 shows a breakdown of the errors. While examining the types of errors that Taiwanese learners usually made, the most frequent are the lexical errors, which include incorrect word choices or word insertions or deletions. The second most frequent errors are verb forms, which include verb tense, subject-verb agreement and incorrectly substituting a verb for a gerund. The third most frequent type of error is article usage, which includes missing definite articles, redundant articles, missed a/an, and confusing the and a.

Table 4. Errors Analysis Results of the Present Study

Error Type	Advanced Learner	Intermediate Learners
Errors per 100 words	6.5	10.23
Article Error	18.58%	15.15%
Preposition Error	11.63%	13.67%
Noun Error	9.34%	8.56%
Verb Form	20.71%	18.73%
Word Form	7.41%	4.62%
Spelling	1.32%	1.63%
Punctuation	6.48%	8.71%
Word insertion, deletion, or replacement	24.53%	28.93%

Here are some examples of incorrect word choices, word insertions, deletions and replacement errors that were extracted from the corpora:

- (1) In the beginning, the staff *directed* me how to use cash *desk* (intermediate learner # 6).
- (2) I *promised* a proposal and I asked them to vote (intermediate learner # 13).
- (3) From this work experience, I also *reaped* something (advanced learner #20).
- (4) When you are done, you can also *dress* your macaroons with small chocolate chips (advanced learner #20).

From Table 4, it is apparent that lexical errors or vocabulary errors are the most frequent errors in the corpora. The learners tended to use the 1000 most frequent English words in their writing; therefore, it can be concluded that the learners have insufficient mastery of vocabulary. They made many mistakes in word choice. This phenomenon was mentioned by Flowerdew (2003), who stated that learners' problems are not so much *bad* English or *incorrect* English, but rather *insufficient* English.

For verb form errors (the second most frequent errors in the corpora), the following is an example that was extracted from the corpora.

- (5) To sum up, I really *enjoy* this work, and I *become* more outgoing than before. I *believed* this experience will help me for the internship (intermediate learner #14).

One possible explanation for this phenomena is that because half of the collected texts are narrative essays and in narrative essay writing, time is a primary factor in telling a story. As a result, the learners are often confused about which verb tense to use, especially past tense, present perfect and past perfect. Compared with other research, for example Chung and Nesi's (2006) study, verb tense errors occurred less frequently. This occurred because their learners' texts were mainly argumentative essays. In argumentative essays, time is not a primary concern. Therefore, different types of texts collected in the corpus will show different results in EA.

Table 5 presents the results for article errors. It shows that redundant use of *the* is the most common article error, followed by omission of *the* and *a*. This result is similar to Chuang and Nesi's (2006) study. Similar findings have also been reported in other studies. For example, Milton (2000) examined Hong Kong University students' writing and found four types of article errors among the top ten most frequent errors. Papp (2004) also found many errors concerning the article system in a 200,000-word corpus of writing by Chinese ESL university students. It appears that the complex concepts expressed within the English article system make it particularly difficult for learners of English to master. This is especially true for those whose first languages have no articles. The first language of the participants for our corpus was Mandarin Chinese, which does not have an article system.

Table 5. Article Errors from the Present Study

Total Errors	Missing <i>the</i>	Missing <i>a</i>	Extra <i>the</i>	Extra <i>a</i>	Confusion	Other
454	20.22%	11.24%	48.31%	6.74%	10.11%	3.38%

5. Conclusion

This preliminary study reveals some salient error features of the Taiwanese students' EAP writing. Some of the results are the same as found in other studies (Chuang & Nesi, 2006; Milton, 2000; Papp, 2004). Although lexical errors (wrong word choices, word insertion, deletion, or replacement and misspelling) are important due to students' lack of sufficient vocabulary, grammatical errors made up 74.15% of the total for the advanced group of learners and 69.44% of the total for the intermediate group of learners. The top three grammatical errors are verb forms, article errors and preposition errors. Both verb forms and article errors are influenced by the participants' mother tongue because their mother tongue has neither tense nor article systems. The preposition errors occurred in higher frequency due to the complicated preposition system in English. Compared with English, the Chinese preposition system is not as rigid and complicated.

As Chuang and Nesi (2006) suggest, Chinese students should be able to correct many of their mechanical errors when editing their own writing. All the above mentioned errors are identified as mechanical errors. Therefore, supplemental materials should focus on these particular areas and build activities that raise students' consciousness to make them aware of these problematic features and help them understand the correct forms.

Two small corpora were constructed during this study, but the corpora should be expanded to make this database a useful resources that can be used as the basis for new pedagogical materials for the learners. By using the database, different tasks for each of the grammatical error categories in this study can be designed as learning resources. In addition, these tasks can be classified into different levels for different learners depending on their degree of difficulty. It is suggested that while designing the task, the starting point should always be an error made by a learner that can be taken from the learner corpora. Each error can then be followed by the corrected version. Next, grammatical explanations and comments should be provided. The purpose of providing the comments is to assist interlanguage development and to help learners distinguish between correct or appropriate use of the target language structure and incorrect or inappropriate use. After the errors and comments, a variety of exercise or activity types such as error correction, cloze, rewriting and discovery learning may be delivered.

This study has shown how small learner corpora can assist in understanding learners'

difficulties and how they may lead to some possible ways to design learning activities to overcome those difficulties by focusing on errors. Although the analyses could be time-consuming, the process provides valuable experience in exploring and understanding our won learners' English writing problems.

Acknowledgement

This study had been presented in the 2015 North East Regional (NEAR) Language Education Conference in Niigata Japan. The travel fees for attending the presentation was sponsored by the Office of International Affairs at National Kaohsiung University of Hospitality and Tourism, Taiwan.

References

- Aston, G. (2000). Corpora and language teaching. In Burnard, L. and McEnery, T. (eds), *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt am Main: Peter Lang, 7-18.
- Chuang, F.-Y., & Nesi, H. (2006). An analysis of formal errors in a corpus of Chinese student writing. *Corpora*, 1(2), 251–271. <http://dx.doi.org/10.3366/cor.2006.1.2.251>
- Dagneaux, E., Denness, S., & Granger, S. (1998). Computer-aided error analysis. *System: An International Journal of Educational Technology and Applied Linguistics*, 26(2), 163–174. [http://dx.doi.org/10.1016/S0346-251X\(98\)00001-3](http://dx.doi.org/10.1016/S0346-251X(98)00001-3)
- Flowerdew, L. (2003). A combined corpus and systemic-functional analysis of the problem-solution pattern in a student and professional corpus of technical writing. *TESOL Quarterly*, 37(3), 489–512. <http://dx.doi.org/10.2307/3588401>
- Granger, S. (2004). Computer learner corpus research: Current status and future prospects, in Connor, U. & Upton, T. (eds), *Applied Corpus Linguistics. A Multidimensional Perspective*. Amsterdam: Rodopi, 123-145.
- Granger, S. (2003). Error-tagged learner corpora and CALL: A promising synergy. *CALICO Journal*, 20(3), 465-480.
- Granger, S. (2002). A bird's view of learner corpus research, in Granger, S., Hung, J. and Petch-Tyson, S. (eds), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins, 3-33. <http://dx.doi.org/10.1075/llt.6.04gra>
- Granger, S., Dagneux, E., & Meunier, F. (2002). *International Corpus of Learner English*. Louvain: Presses Universitaires de Louvain.
- Heaton, J., & Turton, N. (2001). *Longman Dictionary of Common Errors*. London: Longman.
- Hidalgo, E., Quereda, L., & Santana, J. (eds, 2007). *Corpora in the Foreign Language Classroom*. Amsterdam: Rodopi.
- Hutson, S. (2002). *Corpora in Applied Linguistics*. Cambridge: CUP.

Izumi, E., Uchimoto, K., & Isahara, H. (2004a). SST speech corpus of Japanese learners' English and automatic detection of learners' errors. *The International Computer Archive of Modern and Medieval English (ICAME) Journal*, 28, 31–48.

James, Carl. (1998). *Errors in Language Learning and Use: Exploring Error Analysis*. London: Longman.

Leech, G. (1998). Learner Corpora: What they are and what can be done with them, in Granger, S. (ed), *Learner English on Computer*. London: Longman, xiv-xx.

Lenko-Szymanska, A. (2003). Lexical problem areas in the advanced learner corpus of written data. In B. Lewandowska Tomaszczyk (Ed.), *Practical Applications in Language Corpora (PALC 2001)*, Papers from the International Conference at the University of Lodz, pp. 505-520, Frankfurt am Main: Peter Lang.

Mendikoetxea, A., Murcia, S., & Rollison, P. (2010). Focus on errors: learner corpora as pedagogical tools. In M. C. Campoy-Cubillo, B. Bellés-Fortuño, & M. L. Gea-Valor (Eds.), *Corpus-based Approaches to English Language Teaching*. London; New-York: Continuum.

Milton, J. (2000). Elements of a written interlanguage: a computational and corpus-based study of institutional influences on the acquisition of English by Hong Kong Chinese students. Hong Kong. Retrieved from <http://repository.ust.hk/dspace/bitstream/1783.1/1055/1/>

Myles, F. (2005). Interlanguage corpora and SLA research. *Second Language Research*, 21(4), 373-391. <http://dx.doi.org/10.1191/0267658305sr252oa>

Papp, S. (2004). The use of learner and reference corpora to foster inductive learning and self-correction in Chinese learners of English. Paper presented at the 'Meeting the Needs of the Chinese Learner in Higher Education' Conference, University of Portsmouth, 17-18 July.

Tan, M. (2005). Authentic language or language errors? Lessons from a learner corpus. *ELT Journal*, 59(2), 126-134. <http://dx.doi.org/10.1093/eltj/cci026>

Tono, Y. (2003). Learner corpora: design, development and applications. In D. Archer, P. Rayson, A. Wilson, & T. McEnery (Eds.), (Vol. 16, pp. 800–809). Presented at the Corpus Linguistics 2003 Conference (CL 2003), Lancaster (UK): Lancaster University: University Centre for Computer Corpus Research on Language.

Copyright Disclaimer

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/>).