

GERMANY/HONG KONG JOINT RESEARCH SCHEME
THE PROJECT REPORT
(for Project Completion)

Project Number: G_HK013/11

Title

Textual Revision in Second Language Writing

Particulars

	Hong Kong team				German team	
Name of Project Co-ordinator (with title)	Dr. Lee, John Sie Yuen				Prof. Dr. Lüdeling, Anke	
Name of Co-Investigator (if any)	Prof. Webster, Jonathan					
Institution or Institutional affiliation	<input checked="" type="checkbox"/>	CityU	<input type="checkbox"/>	HKU	<input checked="" type="checkbox"/>	Humboldt-Universität zu Berlin
	<input type="checkbox"/>	CUHK	<input type="checkbox"/>	HKUST	<input type="checkbox"/>	Others: _____
	<input type="checkbox"/>	HKBU	<input type="checkbox"/>	LU		
	<input type="checkbox"/>	HKIEd	<input type="checkbox"/>	PolyU		
Other project team members (if any)					Hagen Hirschmann, Marc Reznicek, Amir Zeldes	

Funding Period

	1 st year	2 nd year (if applicable)
Start Date	01/01/2012	01/01/2013
Completion Date	31/12/2012	31/12/2013

Objective(s) as per original application

1. Characterize the types and nature of textual revision in the writings of learners of a foreign language
2. Analyze quantitatively how textual revision is influenced by comments from language teachers
3. Design and implement query and visualization interfaces to facilitate such quantitative analyses

Details of Report [Please attach relevant document(s)]

i) Outline of proposed research and results obtained

Computer learner corpora have been playing an increasingly important role in research in both Second Language Acquisition and Foreign Language Teaching. Typically, these corpora contain texts written by non-native speakers of the language. Past research has tended to focus on analyzing only the “final version” of these texts; writing, however, is an iterative process with cycles of textual revision. This project aimed to analyze this revision process, with three objectives:

- Characterize the types and nature of textual revision in the writings of learners of a foreign language. *Result: We reported and analysed the distribution of the error types made by learners in our corpus, computed statistics on textual revision, including the amount of editing between drafts, such as the insertion and deletion of sentences, as well as changes in sentence and essay lengths. Detailed analysis is given in Section 2 in the submitted article (see attached).*
- Analyze quantitatively how textual revision is influenced by comments from language teachers. *Result: In a case study on verb tenses, we showed the rate of comment take-up to be mixed, ranging from 76% to 43% for different tenses. We also analysed the range of successful changes made by students based on these comments, which range from 44% to 21% depending on the tense. These differences appear to correlate with students' familiarity with the grammatical phenomenon in question. Detailed analysis is provided in Section 5 in the submitted article (see attached).*
- Design and implement query and visualization interfaces to facilitate such quantitative analyses. *Result: We converted the corpus into TEI and ANNIS, a corpus visualization platform, whose query and visualization interfaces facilitate quantitative analyses on textual revision. The conversion process, example queries and visualizations are provided in Sections 3 and 4 in the submitted article (see attached).*

ii) Significance of research results

A distinguishing feature of our corpus is that it contains not only the final, submitted versions of the essays, but also the students' initial drafts, as well as the language teachers' comments on these drafts which guided the students to produce the final versions. To the best of our knowledge, with almost 8 million words, this is the only large-scale corpus of its kind.

While previous studies relied on much smaller corpora and manual methods, we are the first to use such a large-scale corpus to analyse statistics on the distribution of learner error, and the extent to which learners' revision is influenced by teachers' comments. These statistics are not only more reliable, but can also be easily reproducible by other scholars via the query interface in ANNIS. Understanding the dynamics of textual revision would benefit not only language teachers and second language acquisition research, but also the design of writing assistance tools that provide automatic feedback.

iii) Research output

We have submitted one article to the *Language Resources and Evaluation Journal*: (see attached)

John Lee, Chak Yan Yeung, Amir Zeldes, Marc Reznicek, Anke Ludeling, and Jonathan Webster.
CityU Corpus of Essay Drafts of English Language Learners: A Corpus of Textual Revision in Second Language Writing.

iv) Potential for or impact on further research collaboration

Through this project, we have developed a machine-actionable, web-accessible corpus of textual revision by students of English as a second language. We are continuing our collaboration with the team at Humboldt University on quantitative studies of second-language acquisition, investigating the influence of teachers' comments in other grammatical phenomena. We also seek collaboration at Humboldt and beyond on data-driven development of computer-assisted language learning systems.