

**Evaluation Report on GCOE-CbLLE's Activities (2007 to 2009),
Tokyo University of Foreign Studies**

Vincent B Y Ooi, PhD
Department of English Language and Literature
National University of Singapore
E-mail: vinceooi@nus.edu.sg

25 March 2009

This evaluation is based on both the GCOE-Interim Report made available to me and the English version of the GCOE-CBLLE website (<http://cblle.tufs.ac.jp/index.php?id=134/>).

1) Project planning

I note that CbLLE (Corpus-based Linguistics and Language Education) is a Global Center of Excellence project from 2007 to 2012, and has the following integrative aims for achieving world-class research:

1. Field linguistics: collection and analysis of naturally occurring language-use data from a wide range of languages through field research
2. Corpus linguistics: compilation and analysis of large-scale corpora of a wide range of language-use data
3. Linguistic informatics: application of corpus-based linguistic analyses to language education and pedagogy

These aims are to be realised by the following outcomes: i) young researchers having broad perspectives of linguistic and cultural diversity, ii) researchers competent in field research, language education, and corpus creation, and iii) the building of an international network of research collaboration and exchange.

I welcome this sustained focus on linguistics as an empirical science, after several decades when the discipline was informed only by various native speakers' intuitions which were non-correlated with large-scale investigations of speech and writing. Empirical linguistics provides not only a 'hands-on' approach to actual linguistic patterning in society, but has also proven to provide new theoretical findings and insights on the nature of language – as evidenced by the established field of corpus linguistics. The corpus enterprise is not concerned with mere collections of data but is associated with sampling data according to the notions of 'representativeness' and 'balance', and is often aided in its analysis by means of the computer. The twin focus on 'field' and 'corpus' linguistics is very much appropriate, and informs the study of language education and pedagogy as a societal imperative.

The aims of the CbLLE project are therefore very much consonant with the overall vision of the Tokyo University of Foreign Studies as a world-class specialist institution on international studies, with language being the central window to culture and norms.

2) Originality

As noted in (1), the combination of field, corpus, and language pedagogy in order to produce

an integrative platform for research is an original idea, and draws strengths from the various expertise that TUFS has.

Notwithstanding minor spelling errors (e.g. Section 7.1. 'Corpra' for 'Corpora', Section 8.3 'Potencial' for 'Potential', and Section 8.8 'Colloquiua' for 'Colloquia'), these strengths are evident in the reports on the original and interesting projects from the three spheres of field linguistics, corpus linguistics, and linguistic informatics, i.e.

1. Compilation of a Word List for Field Research on Khoe Languages
2. Exploration of Interaction between Auditory Phonetics and Phonological Structure
3. Collection of Spontaneous Conversational Data of Swahili
4. Language Description Data
5. Phonetic and lexicological field investigations of selected under-studied speeches of India
6. Development of an electronic dictionary of Russian
7. Development of an electronic dictionary in Thai
8. Corpus compilation of data from medium/minor language groups
9. Collection of spontaneous conversational data in Russian
10. Development of utility manuals for German corpus
11. Compilation of an introductory text book on lexicology based on corpus data
12. Research on corpora for minor Language group in EU countries
13. Research on lexicon/language-use based on Subject-Categorized Corpora
14. Language processing/education technology
15. Part of speech (POS) Search Engine (Spoken French, Spoken Spanish)
16. Basic research on e-learning through Moodle (Japanese Page)
17. Collection of spontaneous conversational data in French/Spanish/Turkish
18. International collaborative project on corpora of learners' language use
19. Corpus of learners' language use
20. Development of Language Tests based on an Error Analysis of a Learner Corpus of English

Incidentally, given that the World Wide Web (the Internet) engenders newer genres and types of linguistic patterning (e.g. text messaging, instant messaging such as Yahoo Messenger, weblogs, social networking utilities such as Facebook, etc), it would be a good idea to explore the possibility of collecting and analysing such data for some future extension of the project. Learners and speakers are increasingly spending their time at the computer, which is giving rise to new electronic textualities.

3) Scientific quality

The scientific quality of the projects mentioned in (2) is measured by the various experiments, collection of data, measurements, software development and methodological analyses afforded by the various researchers in the project. There has been a plurality of methods, but all these have been guided by empirical considerations and the methods used are of high scientific quality. These are noteworthy efforts, and I would urge all the researchers to continue to refine data quality by never losing sight of the twin goals of representativeness and balance.

4) International Contribution :

The international character of the CbLLE project has been evidenced by the local organising of various international workshop/symposia/conferences, collaborative networks and dissemination of results in high quality journals/book volumes, and attendance at international conferences. I am confident that the CbLLE project will continue to raise its international profile by increasing its collaboration with the leading institutions in the areas of field linguistics, corpus linguistics, and language education. Another area which all the 20 or so projects can increase their visibility and contribute to the world community of researchers is to make available their data eventually on the Web and promote the use of web interfaces to access and analyse the data. In this connection, I note that some projects are already doing so (e.g. the electronic dictionary of Russian project has 10000 entries made available over the Internet), and these are steps in the right direction.

5) Feasibility (Results)

It is difficult to have an overall evaluation of the various projects, which are in different states of progress. However, the interim results are very encouraging as the methodologies have been worked out in different stages.

6) Scientific results

Within a short span of 2 years, the CbLLE project now has concrete deliverables: (i) a more experienced team of researchers, ii) a plethora of corpora gathered for the various projects, ii) workshops/symposia/conferences, some of which have been translated into book volumes by international publishers of repute (such as John Benjamins Publishing Company), and iv) exciting new intellectual products, such as the electronic dictionary of Russian and the electronic dictionary of Thai-Japanese/Japanese-Thai. It is highly likely that more dissemination of results from the various projects will follow in the next few years.

In addition, it is very noteworthy that the project has given rise to various modules (or education programs), such as the following:

- i) Field Linguistics: Phonetics and Phonology
- ii) Field Linguistics: Morphosyntax
- iii) Field Linguistics: Local Survey Practical Training
- iv) The Theory and Practice of Corpus Linguistics
- v) Corpus Linguistics: Applied Course
- vi) Quantitative Surveys and Research Methods for Lexicon and Writing Style
- vii) Quantitative Surveys and Research Methods for Grammar and Meaning
- viii) Typology
- ix) Language Informatics

7) Fostering young researchers

It is noteworthy that the CbLLE project has put in place a series of support platforms for pre- and post doctoral fellows, and offers various teaching practice and opportunities. About 12 doctoral degrees have been conferred since the beginning of the project, with another six to

follow soon. There are also another 20 graduate students pursuing PhD study. 38 junior fellows were accepted in 2007, with another 27 junior fellows accepted in 2008. Support has also been given to various research fellows.

I would encourage some of these young researchers to attend international platforms such as the American Association of Corpus Linguistics annual conference, the British Corpus Linguistics Conference (held once in 2 years), and the annual conferences of the International Computer Archive of Modern and Medieval English (ICAME) in order to boost their international experiences.

****End of Evaluation Report****