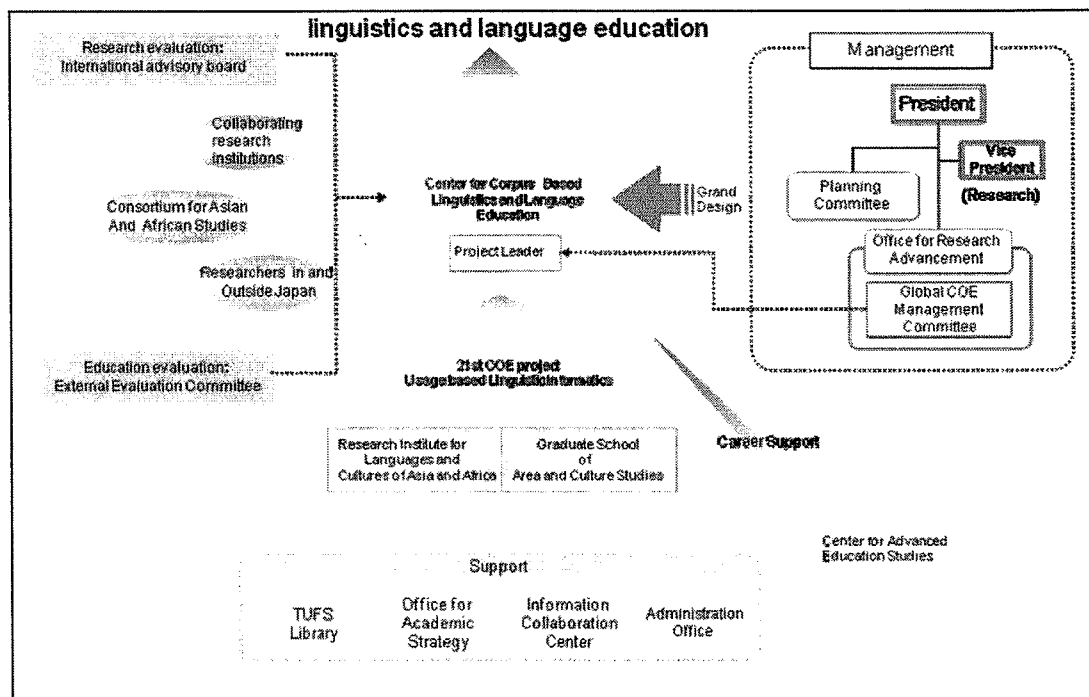


## Assessment of *Corpus-based Linguistics and Language Education* (CbLLE) Project at Tokyo University of Foreign Studies (March 2012)

### 1) Project planning

This project (from 2007 to March 2012) builds upon the accomplishments of the 21<sup>st</sup> COE program 'Usage-Based Linguistic Informatics' (2002-2006). Under the present structure, diagrammatically represented by the project as follows,



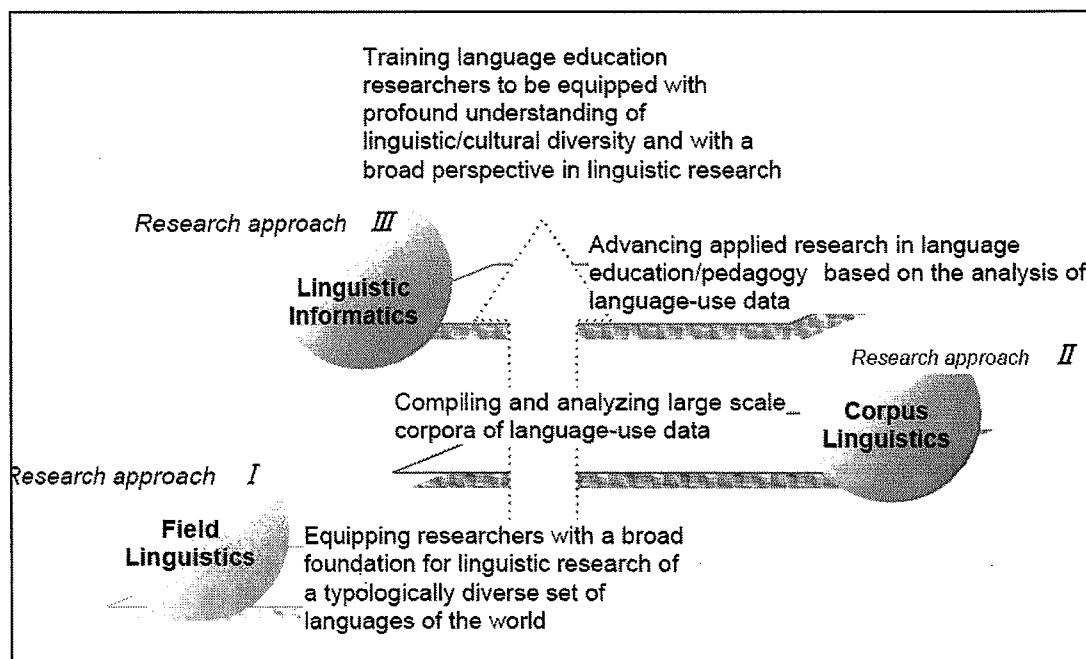
the project planning ensures a synergy between the administrative, research, teaching, and international collaborative agencies, with the Project Leader at the centre to liaise between these various agencies.

Such a structure ably underpins the trilateral research objectives of the project to bring together the traditionally distinct areas of Field Linguistics, Corpus Linguistics, and Linguistic Informatics in order to see what research and collaborative synergies such an experiment would bring about.

### 2) Originality

Until more recently, some main theories of linguistics did not pay enough attention to how ordinary people used language in everyday life. Linguistic evidence was taken principally from one's personal intuitions or casual observations (and somewhat derogatorily referred to as "armchair linguistics"). On the other hand, the current project appropriately focuses on the rigour of systematic empirical investigations into how language really functions in society.

From this perspective, then, I congratulate the project planners of the CbLLE project for envisaging the scaffolding of layers of empirical investigation, beginning with Field Linguistics and Corpus Linguistics, and then investigating their relevance to language education and pedagogy (as diagrammatically represented below):



This project is therefore highly original in combining typological research and corpus-based research, with a pedagogical leaning in order to benefit members of the wider society at large.

### 3) Scientific quality

Each of the three principal components of the project has come up with a number of group-research projects.

Under the rubric of Field Linguistics, there are 5 projects that collectively involve the following: Khoe languages, auditory phonetics, phonological structure, Swahili, Tankhul Naga dialects, lexicology, and field investigations.

Under the rubric of Corpus Linguistics, there are altogether 8 group-research projects involving the following (sub) disciplines: computational lexicography (Russian, Thai), corpus data from medium/minor language groups such as Hmong, Santali, Lhaovo, and Toda, conversational data from Russian, utility manuals for German, textbooks for Japanese vocabulary education, and corpus-based semantics of Japanese, corpus-based and language policy studies of Romansch, Latvia and Swiss German. There is also a study of German motion verbs and adjectives

Under the rubric of Linguistic Informatics, there are projects which build upon the basic foundational research in both Field Linguistics and Corpus Linguistics in order to advance

language pedagogy. Technical advances include the development of search engines, tagging of computerized texts, part-of-speech tagging of spoken French and spoken Spanish, the development of a multi-lingual learning environment based on 17 languages, a spoken corpus of French, Spanish, and Turkish, a learner corpus of French, and an interlanguage corpus of English essays from several countries, and a learner corpus of Japanese.

All these projects show the wide-ranging engagement with the world's major international languages, and not just with Japanese and English. All the experiments are empirical and have high scientific quality.

#### 4) International Contribution

The project is certainly international, in at least three major ways:

- i) The international contributors that make up the three John Benjamins volumes of *Corpus Analysis and Variation in Linguistics*, *Geographical Typology and Linguistic Areas*, and *Corpus-based Analysis and Diachronic Linguistics*
- ii) The international students in the PhD programme and
- iii) The international collaborators to the various group-research projects, workshops and conferences

#### 5) Feasibility (Results)

I have a positive view of the overall feasibility of the programme. The various projects have proceeded at varying speeds and I note that quite a lot of corpora have been/are being collected. When ready, such corpora can of course form the basis for the production of more language materials (beyond what has been stated in the report) for learners and wider dissemination, including school textbooks, dictionaries and grammar books for the various languages under study.

#### 6) Scientific results

In the span of 5 years, there have been seven volumes of the TUFs produced *Working Papers in Corpus-based Linguistics and Language Education* series.

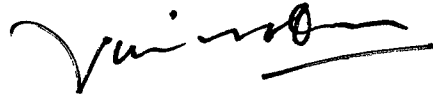
In particular, the three John Benjamins volumes -- *Corpus Analysis and Variation in Linguistics* (2009), *Geographical Typology and Linguistic Areas* (2011), and *Corpus-based Analysis and Diachronic Linguistics* (2011) -- have certainly brought about extra prominence to the Tokyo University of Foreign Studies in the international arena. Together, these volumes are making a significant impact in terms of their contribution to our collective understanding of the wider relationship between language, society and education.

I also note the various published URLs (albeit in varying content) and accomplishments associated with the various group projects, e.g. <https://sites.google.com/site/bhaperi/> regarding the Toda language. In addition, I would encourage those projects which have solid commercial and knowledge enterprise potential, e.g. the project involving the creation of a Thai electronic dictionary, to explore the possibility of being put not only on the Web but also penetrate the smartphone market, i.e. as Apps on the Iphone and Android platforms.

## 7) Fostering young researchers

In particular, Sections 3, 5.4, 7.2, 8, 8.14 and 9.4, 9.6, and 11 of the CbLLE report indicate how young researchers are ably supported in terms of their training for the various principal components of the project: workshops, fieldwork, information technology, thesis writing and PhD training, international conferences, academic exchange and teaching opportunities as post-doctoral research fellows etc. Section 9.4 gives an excellent indication of how the project is contributing to Japan's present and future intellectual training in wide-ranging areas pertaining to linguistics in various languages.

Overall, in assessing Sections (1) to (7), the project may be said to be highly successful, and I warmly congratulate the Project Leader and his team on their great efforts to ensure the success of the project.



Dr Vincent B Y Ooi, PhD (Lancaster, United Kingdom)  
Associate Professor, Department of English Language and Literature  
concurrently: Assistant Dean (External Relations and Student Life), Faculty of Arts and Social Sciences,  
National University of Singapore  
E-mail: vinceooi@nus.edu.sg  
URL: <http://courses.nus.edu.sg/course/ellooiby>

Date: 10 March 2012