

Evaluation of TUFS CbLLE project based on the report on 2007-8

Geert Booij, professor of linguistics, Leiden University, The Netherlands

<http://website.leidenuniv.nl/~booijge/>

### 1) Project planning

The project started in 2007 and is now subject to an interim-evaluation. The range of activities is well-planned and an impressive number of activities has started, at the level of research, research infrastructure, education, and communication. My impression is therefore that a lot of work has been done to give the program a good start, and that the program directors and researchers have achieved a lot in the first 2 years of the project. A very good aspect of the project is the collaboration with relevant experts outside of Japan. Moreover, the idea of training young researchers in a well developed Ph. D. program is important for this project to have a long-standing impact on linguistics in Japan and the world at large.

### 2) Originality

The ambition of this project is excellent because it wants to unify three approaches to linguistic research: corpus-based linguistics, language education, and field research. There is no doubt that modern linguistic research can only be based on proper electronic resources such as corpora, and by making this one of the pillars of the project, this program can achieve excellent results. A second original part of the program is its linking of corpus-based analysis and knowledge to the task of improving language education, by collecting data on actual speech, and also data on errors in second language acquisition. I assume that the focus in this area will be on Japanese learners of foreign languages, because this is the unique material and data that can be collected and used by Japanese researchers. Giving modern language education such a solid empirical base is original indeed.

Computer-assisted language learning is very important but requires a lot of research to guarantee and enhance its quality. Therefore, I consider the design of this research program as original, promising and having a high potential for creating new insights into how to teach foreign languages in Japan, and also abroad. In addition, it will improve our knowledge of how to develop and use linguistic corpora, and how to create databases with language descriptions.

### 3) Scientific quality

The scientific quality of the project must be judged by its present results: books, conference presentations, and above all corpora. As to books, the cooperation with Benjamins, a renowned international publisher is very good. I also noted that conferences are used a lot, and the feedback thus acquired will serve to raise the quality of the work even more. The quality of the corpora may be deemed to be good. I have consulted some of them on the website of the program, and they were very interesting. I also noted with pleasure that for several foreign languages cooperation

with famous institutes such as the IDS in Mannheim, Germany, has been implemented. This is an important aspect, also for the coming years because through cooperation the risk may be reduced that one collects data on foreign languages that are already available. A continuous question to be asked is therefore: which data should be collected in Japan, and which data, relevant for the project, are already available through existing electronic resources abroad.

Further improvement and quality of the results might be obtained by publishing results in peer-reviewed, open, international scientific journals for the relevant fields, such as journals for corpus linguistics and second language acquisition. The problem with publication of results in books like those published by TUFs in cooperation with Benjamins is that they might not be subject to a severe process of peer reviewing, and they are often quite expensive. Also, by publishing more in journals that are usually electronically accessible, the results of this project will be better known and have a larger impact.

Another point that might be considered for the future is that the importance and relevance of corpora depends on asking good theoretical and practical questions in the domains of linguistic theory (architecture of the grammar) and second language acquisition. What kind of information should be stored, and how, in order to answer good scientific questions? In the present program description, the kind of theoretical issues that are broached are not spelled out very extensively. May be they are spelled out in other documents, but in the present document on which I base my evaluation, this kind of information is rather restricted. I would like to suggest that more attention is given to this aspect in the coming years.

A final positive indication of the scientific quality of this project is the field work with respect to minority and endangered languages. UNESCO has indicated that work on endangered languages is very important, and study of these languages will preserve our common global cultural heritage, but also ensure that unique linguistic data will not get lost but kept available for future linguistic research.

#### 4) International Contribution

The potential international contribution is high, The project cooperates already with important relevant institutions abroad that possess and develop corpora.

The corpora and data bases that result from this project should be made accessible as much as possible for the international community of linguists. In this respect, the website of the project could be expanded and provide more access to results. This applies both to databases and to publications. Nowadays, many linguists publish their results on websites, freely downloadable, thus enhancing the impact of their work. I feel that this project could do more in this respect.

As to databases concerning endangered or minority languages, continuation of the cooperation with SOAS as to preservation, accessibility, and formatting is a clear desideratum. Since SOAS is leading in this respect (one might also think of the MPI in Leipzig, Germany) and Australian National University). Linguists of Leiden university such as professor Mous are also very competent in this domain. The consortium of which TUFs is the founder and leader (members: CNRS-Paris, Soas, Singapore, Leiden) can play a positive role here.

I noted with pleasure that outstanding linguists like Austin, Heine, and König, who work in this domain, have been invited to participate by giving lectures and classes.

The project has hosted a number of conferences, also with excellent foreign guest-speakers. The fact that so many linguists of international standing are willing to

participate shows that the project has a high scientific potential and is interesting outside Japan as well.

#### **5) Feasibility (Results)**

TUFS as a university has the potential to make this a very successful program. The results so far are available as corpora, databases, articles, and books. The project has fulfilled its promises for the first years, but lots of subprojects are work in progress, of which the results are not yet readily available.

The policy of focusing on young Japanese talents in developing this program is an excellent idea, but also has the risks that things could have been done better if one also makes use of international expertise. In particular knowledge of recent insights in the domain of language technology might be missed this way. Therefore, I advise to consider for the coming years how it is guaranteed that state-of-the-art-technology in the domain of language engineering is used in the project. One option is that researchers of this program spend more time with cooperating institutions where this knowledge is available.

#### **6) Scientific results**

It is too early to give a proper evaluation of the results since many of the projects are still under way. The books published with Benjamins show that good and publishable scientific results are obtained. A number of articles have appeared in peer-reviewed journals. The presentations at various conferences also indicate scientific quality and potential. However, in the coming years attention to dissemination of the results in various ways may be increased. In particular, it might also be wise to have researchers to present their results at international conferences in the USA, Australia, and Europe. As I said before, it is also very important to link the available data to interesting questions about how language acquisition works, what these data imply for theories about the architecture of the grammars of natural languages, and about the nature of lexical knowledge.

#### **7) Fostering young researchers**

The focus on training young researchers in this program is excellent. The training program is well developed, and the presence of foreign teachers is very good for creating the international atmosphere that is necessary for education top quality young linguists. The program could, as far as I can see, give increase the attention and space given to present day linguistic theories and their relevance for linguistic analysis. I would also like to stress the importance of practical training in abilities like presentation of a paper in a well structured manner and in good English. It is my experience that good presentation abilities are essential for these young linguists having a good career in linguistics, both in Japan and abroad. I appreciate the focus in the program on writing abilities. I trust that students have to present work also orally, and it is also a good idea to give the more mature students the chance to present work at the conferences organized by this project.

A scholarship program is essential for attracting the best students, and it is an excellent feature of the project that such scholarships are provided.

The availability of post doc research grants is an excellent choice.

**In sum, I very much appreciate this part of the program, and continuous attention for learning more about linguistic theory and about presentation of scientific results is a permanent desiderandum.**