

Learner corpus building and a ‘living’ university foreign language curriculum

Kevin Mark

Meiji University

1. An educational slant on learner corpus research

In this paper I would like to outline an educational approach to learner corpus research that can be contrasted with what can be called an applied linguistic approach. Briefly, I see the prime aim of an applied linguistic approach to learner corpus work as this: to be able to provide descriptions of learner language in relation to the target language that would not otherwise easily be accessible, if at all. An educational approach, on the other hand, sees a concern with quality of linguistic description as being subordinate to the quality of learner experience: particularly in terms of self-confidence, self-esteem, enthusiasm and engagement. I am deeply interested in the linguistic insights that can be gained from learner corpus work, and by the potential for learner corpus research to enhance materials and test development, but I am a teacher before I am a researcher: I feel comfortable saying, as some anonymous teacher is credited with having said, “I don’t teach English. I teach people.” What I would like to briefly discuss here is how this attitude is finding expression in the new phase of corpus building that I am about to embark on.

2. The TRIO learner corpus project

For some eight years or so years now I have been working on the TRIO project, a parallel corpus of Japanese university learner English, which now consists of roughly 160,000 words of learner English. It is a 3-way parallel corpus, because the learners’ English has been reformulated by me, and a Japanese version. has been subsequently added. Sentences or—in some cases—sub-sentential word groups are aligned directly with each other within the Filemaker database designed for this purpose, and it is possible to toggle from an individual line to seeing it within the passage from which it was taken. In the process of exploring or analyzing the data, a typical search will involve searching for all lines in which a given item is present in the reformulated version but not in the learner version. I have written a number of papers on the methodology and long-term potential of the project, as well as some fairly detailed—if not rigorous or exhaustive—descriptions of various contrastive aspects of the native English and learner versions. Some of these papers are listed in the references (Mark 1997, 1998, 2000, 2001, 2002, 2003).

3. An educational current in corpus work: data-driven learning

Although it is not conventional to frame a discussion of corpus research—or even learner corpus

research—in terms of general educational principles, there is a clear educational current in what we call "data-driven learning"(DDL), now well-established as a way for learners to become engaged with the language by exploring it as they find it in a corpus. Aston (2001, p19) summarizes the processes of data-driven learning as a form of 'discovery learning', a term derived from general education rather than language teaching, and which is associated with the work of Jerome Bruner. Bruner's depiction below of the notion of 'discovery' (in relation to ideas) could easily be modified so that it relates to language rather than ideas. It clearly resonates with the benefits claimed for DDL: that giving language learners opportunities to explore the language through a corpus can encourage them to formulate questions and hypotheses that they can test them for themselves; and that being able to do so can also be powerfully self-affirming.

Just as a physicist has certain attitudes about the ultimate orderliness of nature and a conviction that order can be discovered, so a young physics student needs some working version of these attitudes if he is to organize his learning in such a way as to make what he learns usable and meaningful in his thinking. To instill such attitudes by teaching requires something more than the mere presentation of fundamental ideas. Just what it takes to bring off such teaching is something on which a great deal of research is needed, but it would seem that an important ingredient is a sense of excitement about discovery—discovery of regularities of previously unrecognized relations and similarities between ideas, with a resulting sense of self-confidence in one's abilities. (Bruner 1963, p20)

4. From a long period of gestation and experimentation: a new kind of course

My department, the School of Political Science and Economics, is about to open a new course known as the Advanced Communicative English (ACE) program. Before I go on, I should say that the word "Advanced" is a little pretentious for what we are actually doing: this appellation has been selected, to some extent, for diplomatic and promotional reasons rather than for the reality of the English with which we are dealing. In fact we have a wide band of abilities within the class; and we also go out of our way to encourage students to enter the course if they feel drawn towards the activities it offers, even if we sense that their English is fairly weak. We do this providing they seem to be sufficiently motivated to persevere. Since there is a strong self-access element, it is possible to cater for different levels, and the meaning of "advanced" should therefore be taken to be this: by the time a student completes the 2-year program his or her English should, if all has gone well, have significantly advanced.

The course, which will be starting in April this year, represents an expansion of an existing program that has been going for twelve years. This course has been small enough in terms of student numbers to allow ideas and projects to germinate and grow. As a prototype, it is now sufficiently tested as to be 'safe' with a larger group of students. One of the most pleasing aspects of this course is

that in recent years the level of student enthusiasm has been steadily gathering momentum. The positive atmosphere seems to derive primarily from the students' understanding that the course does not represent a static body of material which they have to master, but rather that it is something dynamic and living, and that their input is crucial to its success as a whole. It is, in other words, a 'living' curriculum. With this positive student energy as a base, we are about to enter a new phase of corpus building.

5. Corpus building within a 'living' curriculum: a diary writing project

The curriculum that began to crystallized last year, that will form the basis for curriculum development in ACE, can be described as self-renewing, 'living' and 'owned' by students in the sense that they themselves generate, through the activities in which they take part, the content of the materials which they, their peers and their juniors use for study. It is thus possible for students within a particular group to study with materials that have emerged from their own or their peers' activities, or for students in a group to work with materials produced by their seniors.

Figure 1 shows both how this process of materials generation is also combined with corpus building. In order to give a more concrete sense of how this 'living curriculum' works, I will supplement the diagram with the description of a diary writing project that went on over the course of the April to July semester in 2002. The success of this experience was one of the main factors that led to the planning of the much more ambitious and comprehensive teaching / research project in the 2003-4 academic year (which is about to begin this April).

The class last year had 15 students, and over a period of roughly 8 weeks, two students each week were asked to send to me by e-mail at least 100 words about their daily experiences. Although nominally 'diary entries, the writing tended to go well beyond simple narratives, and included a good deal of reflection, soul-searching and humour. Many were much longer than the minimum requirement. Once I received these diaries, they were rewritten by me so as to sound native-like, almost invariably within a day or two. I usually gave no comment about the language itself, but would send back to the student a personal comment of some kind.

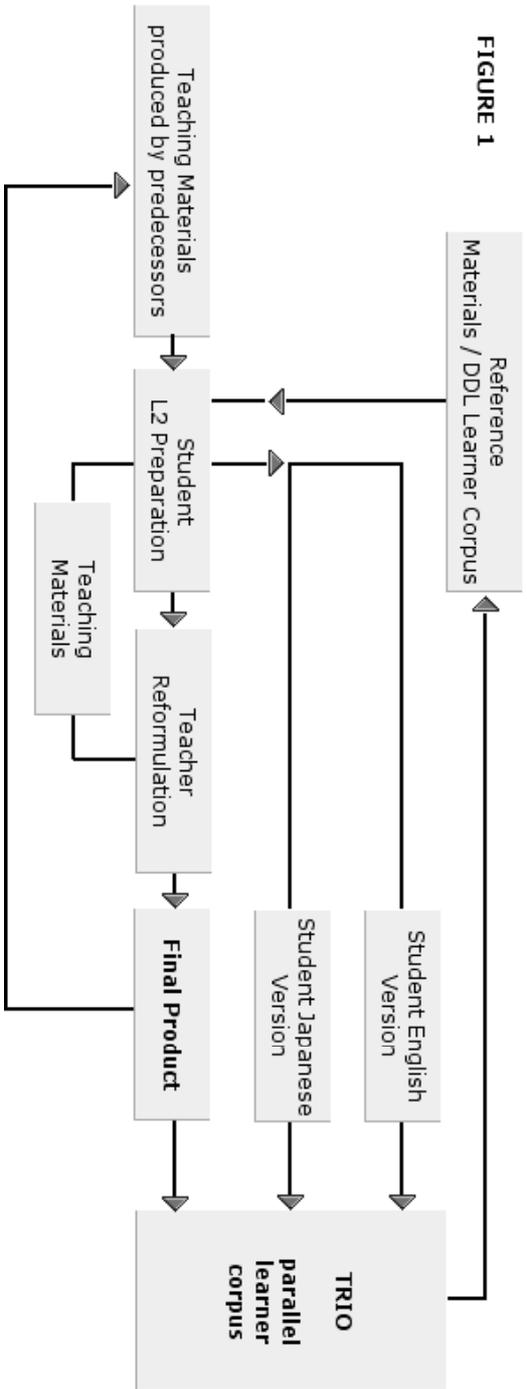


FIGURE 1

The very positive response from the students to this activity seemed to be primarily due to these factors:

- a) They knew that their writing would be read with interest, both by their peers and by myself.
- b) The rewriting was done very promptly.
- c) The students got to know each other much better through the activity.
- d) The activity was reasonably challenging without being over-demanding.
- e) Since the writing which each person produced was read by fellow students only in a teacher-corrected version, there was no embarrassment or pride associated with the quality of the English itself.

When this process was completed, I asked each student to look at both my version and at their own, and to produce as natural-sounding a Japanese version as possible. Once this was done, I was able to enter the three versions of the writing into TRIO, thereby increasing the size of the corpus by about 10,000 learner words. In the process of producing the Japanese version, several students commented that they noticed many intriguing differences in the native and learner versions that they had not noticed when they originally read my rewritten version.

While the rewriting was going on, I had been developing some cloze software with which the students could review and practice with my version of their own and each others' writing. This software allows for quite a sophisticated range of over 20 types of cloze and other exercises, including multi-word items of up to five words, as well as for dictation practice. I was able to give the students several hours of practice with it in class time. At the time, although it worked tolerably well in the classroom situation, I judged that taking it home might result in glitches that could be time-consuming to deal with. The bugs are now ironed out, and the software is ready to be used by students at home during the new academic year. It incorporates a further dimension of corpus building, in that it allows for the students' input to be sent on to the teacher by e-mail, in an attached database file. This means that special parallel corpora of cloze items, as well as a parallel corpus of entire dictation sentences, will in future be generated.

One of the pleasing aspects of this experience has been that several if not all of the students not only enjoyed the process, but also developed an understanding of the methodology behind the course. Knowing what a course is setting out to achieve, how the processes fit together and understanding that one is responsible as a student for the success of one's own course is a powerful educational experience in itself, independent of language learning.

6. Sub-corpora based on other types of activity.

Similar cycles, with different activities, were started last year and although rewriting is not yet complete, they were carried out with sufficient success to make it seem worthwhile to have them incorporated into the new phase of development. It is therefore planned that there will be several sub-corpora incorporated into the TRIO project, each of which will be directly applicable to instruction that prepares students for specific activities.

6.1 Simulation language

One of these activities is a 45-minute management simulation. Last year one such session was carried out, with relatively little student preparation. Three groups of five students each took part, separately, in the same simulation. Two groups spoke in English, while one group spoke only in Japanese. A contrived native English version of this spoken language, which mirrors each of these spoken language mini corpora, is being prepared. New groups of students will perform the same simulation next year, but will first have the opportunity to prepare for it by using worksheets and computer materials designed to focus their attention on language missing from English of the two original groups, and on English, dictation and cloze practice based on the mini TRIO corpus generated by this activity. It is anticipated that a number of simulations will be treated in this way, allowing for the building up of a corpus-based 'simulation course'.

6.2 A 'questionnaire corpus'

A limited amount of work has so far been done to develop a 'questionnaire corpus'. Experience has shown that students tend to find it interesting to devise questionnaires, conducting a survey and reporting back on their findings. The question writing for these can be seen as a kind of pseudo-spoken language, since many of the questions which students ask are ones which they would want to be able to use in spoken conversation.

6.3 Video production

One of the most impressive achievements of this particular group of students was a 30-minute video story, an entirely student-produced murder mystery, with an accompanying live trial scene on stage, in which the audience participates by listening to the defence and prosecution and then voting as a jury. At the time of writing key scenes from this production are being prepared for a computer-based multimedia study unit that will be used with first-year students as a way of introducing them both to the use of computers for self-paced study and to the creative possibilities of the course.

6.4 Web site

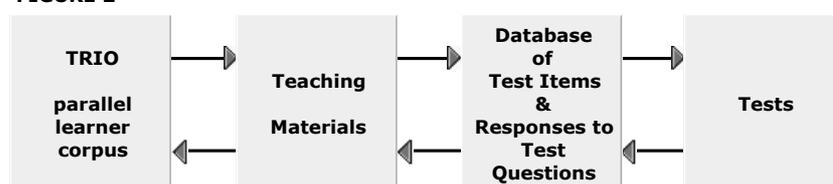
Students have created a bilingual web site including a wide range of information related to the course, opportunities for study abroad, travel and also short reports on pleasure reading and video watching. All of the different types of material that are found in the web site are valid input for the main TRIO corpus, but there is

one genre that seems to be worthy of being given the status of a TRIO sub-corpus: that of writing a report on a book or film that one has enjoyed.

6.5 Building a parallel corpus of test items

In addition to the 3-way sub-corpora outlined above, the corpus-building / teaching system outlined in this paper also allows (Figure 2) for the rapid development of a database (generated by the cloze and dictation practice) of items correctly or wrongly matched by students. It is anticipated that, as the ACE programme develops, the corpus and the teaching materials will feed into a large pool of potential test items which will form the basis for adding a further dimension to the project.

FIGURE 2



7. Envoi

This has been a very brief sketch, and there are inevitably many aspects of the TRIO project that are left untouched. The reader is likely to feel there are many unanswered questions. In particular, the concept of reformulating large amounts of learner written and spoken language, the rights and wrongs of so doing and the practical problems it entails have not been touched upon at all. It can be said, however, as a final note, that students do not seem to have found this ‘hypercorrection’ to be intrusive, but rather seem to find it very valid in the context of the activities in which it is used. When one wishes to make a presentation or give a performance, one naturally wants to put as much care into it as one can.

Let me refer in conclusion to another classic of education which captures one aspect of the ACE curriculum, although I risk sounding patronizing of students since the quotation refers to children. I believe it captures well the spirit of the way the ACE programme combines corpus building and teaching:

All children like to express themselves through the medium of form and color. If you simply indulge this interest by letting the child go on indefinitely, there is no growth that is more than accidental. But let the child first express his impulse, and then through criticism, question, and suggestion bring him to consciousness of what he has done, and what he needs to do, and the result is quite different. (Dewey 1956)

References

Aston G, 2001 Learning with corpora: an overview. In Aston, G (ed), *Learning with corpora*. Houston,

Athelstan, pp 7-45.

Bruner J, 1963 *The process of education*. New York, Vintage.

Dewey, J, 1956 *The School and Society*. Chicago, The University of Chicago Press

Mark, K, 1997 Curriculum development issues and the need for a learner corpus. In Berry, V, Adamson, B., Littlewood, W.(eds), *Applying Linguistics—Selected Proceedings of the 1996 International Conference of the Institute of Language in Education* The English Centre, University of Hong Kong pp 139-151

Mark, K. 1998, The significance of learner corpus data in relation to the problems of language teaching, *Meiji University Bulletin of Arts and Sciences* No. 312, September pp. 77-90

Mark, K. 2000, Analysis of a corpus of Japanese learner English: applying insights to the development of 'frequent error' reference materials, *Meiji University Bulletin of the Institute for Research in the Humanities* No. 47, pp 233-254

Mark, K. 2001 A Parallel Learner Corpus: Using Computers in a Humanistic Approach to Language Teaching and Research, *The Japanese Journal of Language in Society*, Vol.4 No. 1 pp 5-16

Mark, K. 2002 The TRIO Project: an Integrative Approach to teaching and Learner Corpus Building, *Meiji University Bulletin of Arts and Sciences* No. 360 (September) pp 69-89

Mark, K. 2003 A Learner Corpus-Based Approach to Lexical Syllabus Specification and Curriculum Design: Examples with the Verb BE, *Meiji University Bulletin of Arts and Sciences* No. 361 (January)