

## The Feasibility of Trilingual Education Based on Corpus-linguistics

Huang Xin

Khampa Development Research Center  
Sichuan University for Nationalities,  
Kangding, China  
huangxin2009@foxmail.com

**Abstract**—This electronic document discusses the feasibilities of trilingual education (TE) based on corpus-linguistics, which includes the following: providing TE with teaching basis and abundant teaching resources, constructing the mode of autonomous learning, cultivating the empirical capability of teaching and research, etc. Also some methods on how to construct a multilingual corpus on trilingual education in Tibetan areas of China (TETAC) are illustrated with some examples, too.

**Keywords**—trilingual education; corpus linguistics; feasibility

### I. INTRODUCTION

Corpus means a collection of texts, of the written or spoken words, which is stored and processed on computers for the purposes of linguistic research; or a collection of naturally occurring language text chosen to characterize a state or variety of a language. (Liang Maocheng & Li Wenzhong, 2010:4) And depending on the purpose of research, corpus can be classified into different types in order to characterize the different state or variety of a language. The usual types are monolingual corpus, parallel/bilingual corpus, multilingual corpus; or general corpus and specialized corpus. (Huang Xin, 2012:88-93) Due to the trilingual feature of TEATC, i.e. the trilingual corpus collecting natural and authentic materials on Tibetan, Chinese and English, so it may be called multilingual corpus, too.

Hunston, a professor from University of Birmingham has ever commented on it like this: Corpus linguistics has made great influence on language teaching; even those who have never heard of it may have used the corpus-products frequently. (Hunston, 2002) In recent years, the theoretical and practical researches on corpus linguistics have played an active role in language teaching, for example, vocabulary syllabuses and Tibetan-Chinese-English Dictionary and others may be completed with corpus. In addition, the achievements in foreign language teaching and research based on corpus linguistics have made it possible to apply the new approach based on multilingual corpus to TETAC. Here are the three reasons: 1. providing TETAC with teaching basis and abundant teaching resources; 2. constructing the mode of autonomous learning; 3. cultivating the empirical capability of teaching and research.

### II. PROVIDING TETAC WITH TEACHING BASIS AND ABUNDANT TEACHING RESOURCES

In recent years, more and more scholars advocate the task-based approach, which requires the students “learning by doing”. This approach essentially emphasizes the authenticity and contextualization in language. Similarly, multilingual corpus for TETAC also provides a lot of real and natural corpus or language texts. Another, a dispute on “what to teach?” and “how to teach?” has been disputed for a long time among trilingual teachers and researches. As for the question of “what to teach”, we believe in that the multilingual corpus for TETAC can answer it in a good way because in the process of trilingual teaching, the multilingual corpus for TETAC can provide TETAC with scientific foundation and abundant teaching resources.

According to our investigation, not all knowledge in the target languages is acquired by the pupils from Tibetan areas. Actually, what they really want to learn is those that the native speakers use frequently or possibly, such as grammaticalized lexis and conventionalized phrases. This idea is similar to that of Tomlinson, a professor from England, who has ever commented on language learning like this: we must from the beginning help the pupils realize that the way of speaking a foreign language is just like that of their mother tongue—not all the sentences must be in accord with the grammatical rules. (Tomlinson, 1998:88) Therefore, if the criteria to the teaching materials for TETAC are just for the completeness in grammatical system not for the practicality, TETAC are doomed to be of blindness. To this problem, multilingual corpus for TETAC with its numerous utterances, examples can allow the students after their observation and comparison to summarize the language points, evaluate and reflect themselves; and then, understand and master the language points.

With the characteristics of specialization and relevance, the multilingual corpus will facilitate both the lesson-preparation for teachers and consultation from students, and will overcome the shortcomings like dullness and irreality in teaching exemplars from traditional approach. Also, it will promote the effectiveness in TETAC through providing the students with the different means to express the same or similar language function. In addition, due to the essential feature: electric-corpus, i.e., collection of texts, the classroom-teaching requires electrification, too, i.e.,

adopting the modern educational technology to present the teaching contents objectively.

In a word, the multilingual corpus will not only provide TETAC with teaching basis and abundant teaching resources, but also increase the opportunities for the students to experience the multi-cultures, and then perceive them unconsciously.

### III. CONSTRUCTING THE MODE OF AUTONOMOUS LEARNING

In the early 1990s, Tim Kohns and Chris, professors from University of Birmingham have ever applied corpus linguistics to the language teaching and put forward to the concept of data-driven learning, whose central idea is also similar to those of the concepts: student-centered and autonomous learning. They all emphasize the new learning style which is independent, exploring and cooperative. The mode of autonomous learning involves the dynamic integration of vocabulary, grammatical structure, concordance, etc., provided by the use of corpora and corpus tools, view of text (some with audio streaming or video streaming), and cyber-interaction to provide a bottom-up learning environment in which learners are encouraged and expected to learn a foreign language independently and/or collaboratively. In short, it highlights the students' sensitivity to cultural information embodied in the language, so it requires a high occurrence rate and distinctive features in the language input.

Corpus with its inherent function of concordance does not only provide trilingual, bilingual or monolingual examples by search term centered, but also highlight the contextual features in the right or left of the search term. In this way, corpus will become a focus point to catch the students' attention. In the matter of TETAC, teachers may ask the students to draw the conclusions on the language points after their observation, hypothesis and deduction, etc. Here is an example in detail:

- A. *Teaching aim: cultivating the intercultural communication competence of students from the Tibetan areas*
- B. *Teaching aid: sub-corpus for Tibetan, Chinese and English cultures*
- C. *Teaching procedures: with the help of the teacher and sub-corpus, the students:*
  - a. In the sub-corpus read or observe the same or similar cultural topic such as Tibetan New Year, Han's New Year, and Christmas Day;
  - b. Compare the three festivals and understand the similarities and dissimilarities between the three--this step does not only enrich their knowledge, but also cultivate their sensitivity of cultural difference);
  - c. Discuss with each other and then draw some conclusions on how to understand and deal with the differences—this step help the students make a transform in emotional attitude and shape the proper

cultural values, i.e. facing in the Tibetan New Year, Han's New Year, and Christmas Day with a rational and tolerant attitude;

d. (According to the idea of Life-long Education) in their future life deal with the cultural differences properly and flexibly.

Finally, corpus, with electronic text as its main component, is characterized by autonomy and openness. Thus, a good interaction of man-machine will be formed when the Tibetan students collect and process the trilingual corpus from their own lives. Naturally, the process of enriching the multilingual corpus on TETAC will in turn accelerate the construction of the mode of autonomous learning.

### IV. CULTIVATING THE EMPIRICAL CAPABILITY OF TEACHING AND RESEARCH

Only under the joint efforts from Chinese education departments, Tibetan local governments, and elementary and secondary schools in Tibetan areas of China, can TETAC be carried out successfully. Comparatively, the schools in Tibetan areas are more important for this program, which will need both the teaching links and teaching staff. The teaching links are the main process to carry out the trilingual education and cultivate the tri-culture talents while the teaching staff is the fundamental guarantee for the talent training. TETAC based on multilingual corpus will cultivate the teachers' empirical capability of teaching and research. Meanwhile, it has the characteristics of low cost, short period and high efficiency, etc.

Firstly, compared with the traditional trilingual teaching mode without richness and authenticity in teaching resources, the teachers affected by that mode have to explain the monolingual or bilingual knowledge with their own inspirations, without deep analyses on or comparison with Tibetan, Chinese and English...so the traditional trilingual teaching mode has influenced TETAC a lot. With the help of multilingual corpus, both the teachers and students can find the teaching conclusions satisfactory to each other; or confirm their hypotheses through the rich and natural utterances in the corpus; or do more researches on vocabularies, topics, language points and others through "sub-corpus for teaching materials" co-constructed by schools or teachers from different areas; or do some studies on the students' diction or pragmatic features in their works just like the assignments, writings, and all kinds reports, etc.; even on their acquisitions of Chinese and English. Then, the trilingual teachers improve their teaching and research capabilities.

Secondly, in the whole process to construct or enrich the corpus the trilingual teachers will promote them because the teachers must free themselves from the hurdles from the traditional teaching mode, and extend the trilingual teaching outside of the class. Maybe, many years later, the trilingual teachers will have a better command of both three languages and have a systematic understanding to the teaching materials. Then they will have the ability to evaluate the teaching materials objectively, use the teaching materials

effectively, and finally learn to compile some school-based textbooks. (Huang Xin, 2012:88-93)

Of course, the teaching mode and the related researches mentioned above are centered around the multilingual corpus on TETAC. In addition, if the teaching and scientific researches are dealt with perfectly, TETAC is doomed to be successful in the recent future.

## V. THE CONSTRUCTION OF MULTILINGUAL CORPUS ON TETAC

Like most of other corpus construction, the multilingual corpus on TETAC can be completed through four steps, namely, preparatory phase, design phase, execution phase and completion phase, which respectively refers to confirming the design principles, designing the criteria to text classification and annotation, collecting and processing the texts, and application and development. In consideration of the topic, this essay will make a brief description on how to construct the multilingual corpus on TETAC, especially the “design principles” and “the criteria to text classification”.

It is well known that “the idea and principle to design a corpus are subordinated to the constructing goals or using purposes.”(Tomlinson, 1998: 91)The constructing goals or using purposes of this multilingual corpus aim at trilingual teaching and researching on TETAC and cultivating a lot of talents with the ability to cross-cultural communication for the Tibetan areas. In detail, the using goals of this multilingual corpus are: providing both trilingual teachers and students with the rich and authentic teaching materials, and after their experiencing to the target languages helping them summarize the linguistic characteristics of Tibetan, Chinese and English, and helping them reflect and evaluate on their teaching and learning respectively...etc. Therefore, we before the constructing should take into consideration the proportionality, capacity, source, etc., and take samples according to these.

Firstly, according to the domain-criterion we may classify the collected texts into: politics, economy, education, culture, literature, religion, medicine, tourism, and law, etc. Or, according to the topic-criterion we classify the collected texts into: personal information, family & friends, environment, daily activities, school life, hobbies & interests, emotion, interpersonal relationship, plan & wish, transportation, health & diet, etc. The different texts may be collected into different sub-corpus according to the classification criteria above and they may be aligned in trilingual or bilingual form, even in monolingual form. In order to make it more convenient for the trilingual teaching and research, especially for the contrastive analyses in the future, we advocate the trilingual or bilingual form.

Secondly, we should endeavor to balance the collection proportion. All the texts should be copied from publications, public documents, white papers, e-books, teaching materials, handouts, and others from internets, etc., and the optimization is based on characteristics—highlighting regional, ethnic, practical, adaptable and multicultural features. In addition, the capacity of each sub-corpus should be more than ten thousand entries or words, and all the texts

should be published from 1990 to 2013; written or spoken, American or British languages, Khampa, Amdo or Central Tibet dialects may be collected into the corpus, but they should have a special annotation.

Thirdly, only the metadata-markup will be discussed in this essay because the others to process the collected texts just like clearing up the nonstandard format, symbols, empty lines, unreadable codes, as well as the tokenization and POS tagging, semantic annotation, etc., can be done with some softwares like Microsoft Word, Power Grep, Tokenizes, Segmenter, SDL Trados, D<sup>é</sup> à Vu and CLAWS. Due to the constructing goals of multilingual corpus on TETAC, the collected texts will be annotated with XML. And the metadata-markup includes two types: external information and structural information. The external information contains the medium, text-head, author/translator, domain, period, ID number, etc., and the structural information has chapter-title, paragraph, sentences, etc.

Fourthly, the construction of multilingual corpus on TETAC is a complex and systematic project. So its application and development will need the joint efforts from governments, local schools, and trilingual teachers, etc. The multilingual corpus on TETAC featured by open-source, transportability should support distributed operation for scalability and availability, and have two application systems: retrieval system and maintain system. The retrieval system provides language retrieval for the trilingual teachers in Tibetan areas, researchers on TETAC, and students from the Tibetan areas, while the maintain system provides administrators with authority to backup, maintain, and add new texts, etc. In addition, the multilingual corpus should be managed by someone designated.

## VI. THE EXAMPLE FOR THE TRILINGUAL TEACHING BASED ON MULTILINGUAL CORPUS ON TETAC

The preceding paragraphs have stated the feasibilities of new approach based on the multilingual corpus. Actually, the feasibility of the approach is tested by many trilingual teachers, here is an exemplar again:

A. *Teaching aim: helping the students to understand the grammatical position of adjective used as attributive*

B. *Teaching steps:*

a. Classing the students into several groups, and asking each group underline the adjectives from a passage;

China is a unified multi-ethnic country. Tibet is an inseparable part of China, and the Tibetan ethnic group is an important member of the big family of the Chinese nation. The Tibetan ethnic group has a long history and a splendid culture. Tibetan culture is a lustrous pearl of Chinese culture as well as a precious part of world culture.

b. Using the multilingual corpus to ask each group to search some adjectives such as important, big, and long, etc. from the selected;

c. Guiding the students to write down the Left Sort and Right Sort of the Key Word,;

KEYWORD: \_\_\_\_\_ (adjective)

LEFT SORT: Are there any words (or word classes) that appear frequently before the keyword? Write any that you find below:

RIGHT SORT: Are there any words (or word classes) that appear frequently after the keyword? Write any that you find below:

SUMMARY: Are there any grammar rule/s that you think might work for this word? Write them in below and then decide how you can test your idea by studying further examples.

d. Asking the students exchange their retrieved result and summarize the grammatical position of adjective used as attributive;

e. Then the trilingual teacher commenting on their summaries.

## VII. CONCLUSION

TETAC aims at cultivating the Tibetan students' intercultural communication competence, which means: "one's understanding, accepting to some culture, and his change and self-construction in order to adapt to the culture." (Lu Ziwen & Wang Duqin, 2006:85; Moran, 2004:116) Also, TETAC plays some indispensable roles in promoting the national reunification, ethnic unity and social stability, developing the ethnic culture, accelerating the ethnic civilization, and realizing ethnic modernization. But the realization of the aim requires the trilingual teachers in

Tibetan areas to do like the following: making full use of the theories on the ethnic foreign language teaching in China and corpus linguistics, constructing the multilingual corpus on TETAC, carrying out the trilingual teaching based on the multilingual corpus, and highlighting the contrastive analysis during the trilingual teaching.

## ACKNOWLEDGMENT

This essay is sponsored by the projects from Chinese Ministry of Education (Project No. 12YJC740037, and Project No. GPA115077).

## REFERENCES

- [1] Huang Xin. Special Purpose Corpus : A New Perspective of Competence Training for International Publicity Translation Talents in Tibetan Area [J]. Journal of Sichuan University for Nationalities, No.05, 2012.
- [2] Lu Ziwen & Wang Duqin. New Pedagogical Theory on English [M].Shanghai: East China Normal University Press, 2006.
- [3] Hunston, S. Corpora in Applied Linguistic [M]. Cambridge: Cambridge University Press, 2002.
- [4] Liang Maocheng & Li Wenzhong. Using Corpora: A Practical Course-book [M]. Beijing: Foreign language Teaching and Research Press, 2010.
- [5] Moran, P.P. Teaching Culture: Perspective in Practice [M].Beijing: Foreign language Teaching and Research Press, 2004.
- [6] Tomlinson, B. Material Development in Language Teaching [M]. United Kingdom: Cambridge University Press, 1998.