

## **Towards an Accuracy-oriented Alignment and Web-based Concordances of Multilingual Parallel Texts**

Hong Huaqing, Zheng Jianzhen  
National Institute of Education, Nanyang Technological University

Well-aligned parallel texts are an excellent source of information for contrastive linguistic and translation studies. There already exist quite a few alignment tools and parallel text concordancers, but the aligners found in the literature mostly work towards an automated alignment of words or sentences in a restrictive or ideal domain and context (Barlow 1995; Eberling 1998; Véronis 2000; Caseli and Nune 2003; to name just a few), and most parallel concordancers in general provide source language and target language keyword search facilities, and present the search results in the form of traditional KWIC concordances. These tools either fail to extend their functionality to the real life situation or end up with fuzzy match of the parallel equivalents. Our understanding of the reasons underlying their success or failure is still at the initial stage, but the present study has no attempt to probe the issue further, rather, along with our on-going projects, contributes to this area through our effort to build an accuracy-oriented alignment toolkit and an easy-managed online query package. The objective is to identify design considerations and strategies for an open-sourced handy tool package which can efficiently publish parallel texts online for translation study and contrastive analysis.

In this paper, we first elucidate the increasing demands and defining attributes of an accuracy-oriented aligner and some key factors enabling an easy-managed online parallel text concordancer. On the basis of that, we explore the motivating and inhibiting factors for developing such a toolkit, together with the challenges encountered and the solutions proposed. Lastly, we demonstrate how the toolkit handles a variety of formats of raw data. In particular, the experience we obtained and points we paid special attentions are also highlighted in the demonstration. In so doing, we believe that the result of this study can shed light on research in parallel texts processing, and that the design and application of the toolkit can facilitate our future work and other similar endeavors in the community.

### **References:**

- Barlow, M. (1995). ParaConc: A Concordancer for Parallel Texts. *Computers & Texts*, Vol. 10, (CTI Textual Studies).
- Caseli, H. and Nunes, M. (2003). Sentence Alignment of Brazilian Portuguese and English Parallel Texts. Paper presented in Simposio Argentino de Inteligencia Artificial 2003, Buenos Aires, Argentina, 1-5, Sept, 2003. Retrieved on April 15, 2008 at [http://www.frcu.utn.edu.ar/deptos/depto\\_3/32JAIIO/asai/asai\\_21.pdf](http://www.frcu.utn.edu.ar/deptos/depto_3/32JAIIO/asai/asai_21.pdf).
- Eberling, Jarle (1998). The Translation Corpus Explorer: A Browser for Parallel Texts. In: S. Johansson and S. Oksefjell (eds). *Corpora and Cross-linguistic Research. Theory, Method, and Case Studies*. Amsterdam: Rodopi. 101–112.
- Véronis, J. (2000). *Parallel Text Processing: Alignment and Use of Translation*. Springer.