Company Name:

NLP lab, National Tsing Hua University

Productions Being Shown (Name and Version):

TransArts Prototype 1

Company URL:

nlp.cs.nthu.edu

Company information:

Natural Language Processing team at National Tsing Hua University

Representative Names:

Hung-ting Hsieh, [*]Chung-chi Huang (corresponding author), Mei-hua Chen, [*]Ming-hong Bai, Jason S. Chang, and [*]Keh-jiann Chen

[*]Institute of Information Science, Academia Sinica, Taipei, Taiwan

Representative Email:

{vincent732,u901571,chen.meihua,mhbai.tw,jason.jschang}@gmail.com;
kchen@iis.sinica.edu.tw

Product Description:

TransArts is a web-based in-page computer-assisted translation assistant for specific domain of arts and exhibitions. It aims to help translators understand and translate source documents and provides functionalities such as term sense disambiguation and translation based on online Web data or off-line crawled corpora and Chinese segmentation while translating. The in-page translation and sense disambiguation make the tool more interactive and promising.

Language Supported:

Chinese to English

Unique Features:

A web-based computer-assisted translation tool incorporates modules of term disambiguation and translation via online or offline data sets, segmentation while translation, document translation, each of which constitutes a research topic and software application.

An In-Page Computer-Assisted Translation Tool for Specific Domain

Many translators submit texts to machine translation (MT) systems on the Web in the hope of accelerating their translation process. However, most online MT systems such as *Google Translate* are not for specific domain such as arts and so on.

Consider the text of arts "北宋徽宗政和六年的仿商宋鼎" (*The tripods in the Northern Song period mimicking those in the Shang and Song Dynasty*). The best translation for this sentence is probably not "North Song governance and six years of imitation, Song Ding" returned by *Google Translate* trained on general domain. A good response, on the other hand, might contain sense disambiguation for words like "鼎" and translation or transliteration for terms like "北宋", and "商".

We present a Chinese-to-English computer-assisted translation system, *TransArts*, that learns to translate terms or texts in the specific domains of arts and exhibitions (i.e., term or text translation for art displays of National Palace Museum in Taiwan). *TransArts* provides document translation as well as in-page term disambiguation and translation and learns translation from the off-line parallel sentences or online mixed-code sentences. We will now take a closer look at *TransArts*' functionalities, each of which is a research issue and is currently based on well-known solution. The working prototype is at http://nlp-service1.cs.nthu.edu.tw:8080/~nlplab/vincent/assistedTranslation/.

Term Disambiguation. *TransArts* disambiguates the senses of the user highlighted terms using Lesk algorithm which compares the word overlapping ratios between sense definitions and the input article. For instance, if a user highlights "鼎" in the above text, *TransArts* disambiguates it as "ancient word use for metal container" for assistance in monolingual understanding.

Online Term Translation. *TransArts* finds translations for the highlighted terms on the Web, if indicated. In this module, we submit the source terms, which may be accompanied by translations of partial source terms or relevancy feedback as query expansion, to search engines, collect the returned snippets, and extract possible in-page translations via patterns. Take a source term "資料探勘" and its translation "data mining" for example. A pattern "(" lies between them and there are more. In implementation, our patterns are learned by submitting bilingual phrase pairs and analyzing their snippets. Such online term translation is designed to accommodate new words/senses.

Seg and Trans. *TransArts* translates at character level to bypass the word boundary issue in Chinese-to-English MT. Specifically, *TransArts* generates top-n in-page term translation while segmenting. We segment and translate the highlighted Chinese terms or sentences using greedy and dynamic programming algorithm based on bilingual translation probabilities.

Document Translation. *TransArts* integrates Moses as MT server for sentence translation. In the user interface, we allow for editing and reordering the translation to speed up the translation process and alleviate translators' burden.

To translate texts in the specific domains of arts and exhibitions, we crawl and sentence align the bilingual articles on National Palace Museum websites. Apart from the parallel corpora such as Hong Kong Parallel Corpus, FBIS and Wikipedia's language links, *TransArts* further trains on domain-related snippets containing term translation pairs in Bilingual WordNet and multi-lingual Art & Architecture Thesaurus. That is, we submit phrase pairs to search engines to collect snippets in the domain of interest for better translation variety and coverage. The functionality of **Translation Memory** in *TransArts* returns translations by referencing aforementioned data collections.

In summary, we have introduced a method for learning to assist translators in understanding and translating source documents in a specific domain. The method incorporates functionalities, each of which constitutes a research topic, such as term disambiguation, online and off-line term translation, translation while segmenting, and document translation. We have implemented the method as applied to computer-assisted translation and we look forward to its deployment among translator community and its extensive evaluation.