

Instructions for January 2003 Dev Set Human Translations

Important:

Please read the original translation guidelines whether or not you participated in the 2001 MT translation project. The document title is “Guidelines for the Linguistic Data Consortium’s Project to Support Automatic Evaluation of Language Translation Technology (Chinese-to-English)”. You should have received a copy if you’re reading this.

Source Documents

There are 100 documents to be translated, among which 50 are from Xinhua News Agency of mainland China and 50 from AFP. They are all very recent hard news.

Document Format

There’s some minor changes to the file format. The new format is as follows:

```
<DOC docid=XIN20020318.0154 sysid= >  
<headline>
```

```
--Segment 1--  
Chinese Text to Be translated
```

```
<p>  
--Segment 2—
```

The docid corresponds to the file name. Please put your sysid in the space provided.

Segments, which correspond to a Chinese sentence, are sequentially numbered. It is very likely that some Chinese sentences may be translated into two or more English sentences. Please *DO NOT* insert additional Segment tags, nor should you break multiple sentences into different lines. The completed translation should have the exact same number of Segment tags as the source file.

The <headline> tag is followed by the news headline. The translation should also be in headline style. You do not have to be concerned with paragraph tags <p>. In short, do not remove or alter any of these tags.

To ensure that there’s a one-to-one match between the source and translated documents (other than the possibility that one Chinese sentence is translated into multiple English sentences), we suggest that you make a copy of the source file and perform translation directly in the copy. Remember, the output must be plain text files. That is, they should NOT contain any double-byte characters,

including numbers, punctuations, and white spaces. If you use a word processor such as MS Word, make sure you save it as a text file.

Work Flow

Once you're done with the first 5 documents, please send them to us for quality checking. This is to ensure that the translation guidelines have been interpreted correctly and the files are in correct format. You should continue with your translation whether or not you hear any comments from us. We may also check with you on the progress once or twice.

If you have any questions, please contact us directly or your agency.

Xiaoyi Ma

xma@ldc.upenn.edu

3600 Market Street, Suite 810

LDC, University of Pennsylvania

Philadelphia, PA 19104

Phone: (215) 573-5491

Fax: (215) 573-2175