- Contents

    general/: contains geneeral feedback comment data (UTF-8)

    prep/: contains feedback comment data for preposition use (UTF-8)

    doc/guideline_feedback_comment.en.pdf: guideline for feedback comment annotation

    j2e/: contains dictionaries (UTF-8) for translating Japanese feedback comments into English. Note that not all the feedback comments are included in them.

- Base corpus

The provided data consit of the essays sampled from The International Corpus Network of Asian Learners of English (ICNALE). Each essay is stored in a file, of which name specifies the original essay in ICNALE (e.g., CHN_PTJ_001_B1_1).

Each essay is split into sentences (sentence per line) and then tokonized by the Stanford Parser. Tokons are separated by blank.

- Feedback comment Annotation

Each essay is annotated with two types of comments. One concerns general writing techniques (hereafter, general feedback comments). The other is feedback comments on preposition use. All comments were manually annotated by using the commenting function of the WORD software. This means that annotated essays have a file extension ending with '*.docx'. For accessibility, each essay in the WORD format was converted to the TSV format with the UTF-8 encoding by using the tool called Wex, which we also developed. In this format, learner sentences are located in the first column. If a learner sentence has a feedback comment, the comment is located in the second column with its offset in the third column. An offset consists of two integers separated by ':' denoting to which words the corresponding comment apply; the range starts from zero. If there are more than one feedback comments in a learner sentence, the second and third column are repeated an arbitrary number of times (depending on the number of feedback comments. An example is:

From this reason , I agree it . [TAB] Use another <preposition> is often used with the <noun> <<reason>> to indicate a relation. 0:0 [TAB] Since the <verb> <<agree>> is an <intransitive verb>, the <object> requires a

<preposition>. [TAB] Look up the use of the <verb> <<agree>> in the dictionary. [TAB] 5:6

   For general feedback comments, 12 annotators were hired. They were either item writers/editors/ex-editors of English learning materials, or raters for English proficiency tests. Each essay was annotated with general feedback comments by two out of the 12 annotators, meaning that each essay has two versions of general feedback comments, which are distinguished by the directory names (annotation01 or annotation02).

   For preposition feedback comments, two professional annotators, who had a good command of English, were hired. Both of them had experience in English syntactic annotation for more than ten years. Unlike general feedback comments, each essay has only one version of feedback comments for preposition use.

The basic annotation procedures for general feedback comments are as follows:
   1. Read the entire essay first before annotation
   2. Determine what is relevant to the writer most
   3. Annotate the given essay with about five to ten feedback comments based on 2 (Note that the target essay is assumed to consist of 200 to 300 tokens.)
   4. After annotation, double-check the results
   5. Revise the results (if necessary)

For preposition feedback comments, the second procedure is replaced by ``Determine where preposition errors exist and other places that require feedback comments.'' because all preposition errors are the target of the annotation.

   This is the big picture of our corpus design and annotation guidelines. The complete details are in the guidelines accompanying the dataset.

- Citation
 If you make a publication using this dataset, please kindly refer to the following two papers.

Shin'ichiro Ishikawa, The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. In S.Ishikawa (Ed.), Learner corpus studies in Asia and the world, 1, pp.91-118, 2013.

Ryo Nagata, Toward a task of feedback comment generation for writing learning, Proc. of 2019 EMNLP.
(alternatively, 永田亮, 石川慎一郎, 乾健太郎, ``解説文生成研究のためのライティング技術解説付き学習者コーパス,'' 第 25 回言語処理学会年次大会論文集, pp.308-3011, 2019.)