A Category-oriented Approach to Paraphrase Corpus Construction

**Background**
- Towards intelligent treatment of text data on ubiquitous computing network → Processing contents written in natural language is important
- Handling paraphrases benefits various applications
- Paraphrase generation → Reading assistance, Machine translation, etc.
- Paraphrase recognition → Question answering, Multi-document summarization, etc.

**Goal**
- To create a large-scale paraphrase corpus (collection of examples)
  - To analyze how we paraphrase
  - To improve paraphrase generation and recognition systems
  - by inducing paraphrase patterns from corpus
  - by identifying useful linguistic knowledge for paraphrase generation
  - As a standard test-set for evaluating paraphrase generation systems

**Method**
1. **Manual description**
   - Step 1: Manual description
   - Paraphrasing rules and dictionaries
2. **Manual evaluation and correction**
   - Step 2: Manual correction
   - Paraphrase corpus DB
3. **Automatic generation**
   - Step 3: Automatic generation

**Features**
- Category-oriented
  - Use rules and dictionaries for particular category of paraphrase phenomena
- Semi-automated
  - Use an existing paraphrasing system to collect all possible paraphrases, reducing human labor and bias

**Advantages and future work**
- Advantages of the proposed method
  - It allows amateur to engage in judge
  - It leads to rapid corpus construction (three-times faster than Shirai et al., 2001)
  - It collects incorrect examples (for analysis)
  - We will create corpora for other categories
  - Discussion on required expertise
  - To collect reliable and diverse data

**Product**
- Paraphrase corpus
- Collection of paraphrase examples

**Trials**
- To confirm the feasibility of the proposed approach through creating several types of corpus (paraphrase corpora)
- Paraphrasing of light-verb constructions (LVC)
  - 多くの人々が彼の演説に感動を受けた。（A lot of people were impressed by his talk.）
  - 彼はハンマーで壁を壊した。（He broke the wall with a hammer.）
- Transitivity alternation (Trans.alt)
  - 多くの人々が彼の演説に感動を受けた。（A lot of people received a good impression by his talk.）
  - 彼はハンマーで壁を壊した。（He broke the wall with a hammer.）

4.3 correct examples / 1 hour
2.8 correct examples / 1 hour