# Interactive Paraphrasing Based on Linguistic Annotation

**Ryuichiro Higashinaka**
**Keio Research Institute at SFC**
**5322 Endo, Fujisawa-shi, Kanagawa 252-8520, Japan**
rh@sfc.keio.ac.jp

**Katashi Nagao**
**Dept. of Information Engineering, Nagoya University**
**Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan**
nagao@nuie.nagoya-u.ac.jp

## 1. Introduction

**Background :** Large number of documents of great diversity on the Web is making some of the documents difficult to understand for some users due to lack of background knowledge.

**Conventional Method :** Automatic consultation of online dictionaries with a popup window to show word definitions, usually triggered by mouse movements.

**Problems :**
1. **Difficulty of word sense disambiguation**
   In case of polysemic words, all possible word sense candidates are usually displayed, forcing the viewer to choose the correct meaning.
2. **Dictionary lookup results are independent definition statements**
   The user has to reread the original document, taking the word definition into account.
3. **NLP is not directly applicable to dictionary lookup results**
   Natural language processing techniques such as summarization, translation, and voice synthesis cannot be easily applied to the results.

> The ``Interactive Paraphrasing'', making use of linguistic annotation associated with the document and online dictionaries, paraphrases words on user demand solving above three problems.

## 2. Example



**Original Document**

After paraphrasing ``agents'' and ``deployed''

**The conventional method : popup window, showing the word definition**

After paraphrasing ``authority'' in the already paraphrased area

## 3. Linguistic Annotation

Linguistic annotation consists of word sense annotation and syntactic annotation.

### Word Sense Annotation

We annotate polysemic words in the document with word sense, so that their word definitions can be retrieved uniquely from online dictionaries. We use WordNet for word sense notation.

```
<su id="id1-0" ptb="S">
 <ajp ptb="DT">the</ajp>
 <n bf="power" ptb="NN"
 sense="power-noun-1" >power</n>
 <adp>
  <ad ptb="TO">to</ad>
  <vp>
   <v bf="make" ptb="VB">make</v>
   <np ptb="NPL NNS">decisions</np>
  </vp>
 </adp>
.</su>
```

**Example of an annotated text**

### Syntactic Annotation

Syntactic annotation gives a syntactic structure to the document on the basis of a new tag set proposed by the GDA (Global Document Annotation) project, enabling the intelligent processing of documents. We also use PennTreeBank tag set as a syntactic feature.



Using our *Annotation Editor*, users can interactively annotate documents with linguistic structure (syntactic and semantic structure) and word senses. The editor is capable of fundamental natural language processing and interactive disambiguation.

# Interactive Paraphrasing Based on Linguistic Annotation

Ryuichiro Higashinaka
Keio Research Institute at SFC
5322 Endo, Fujisawa-shi, Kanagawa 252-8520, Japan
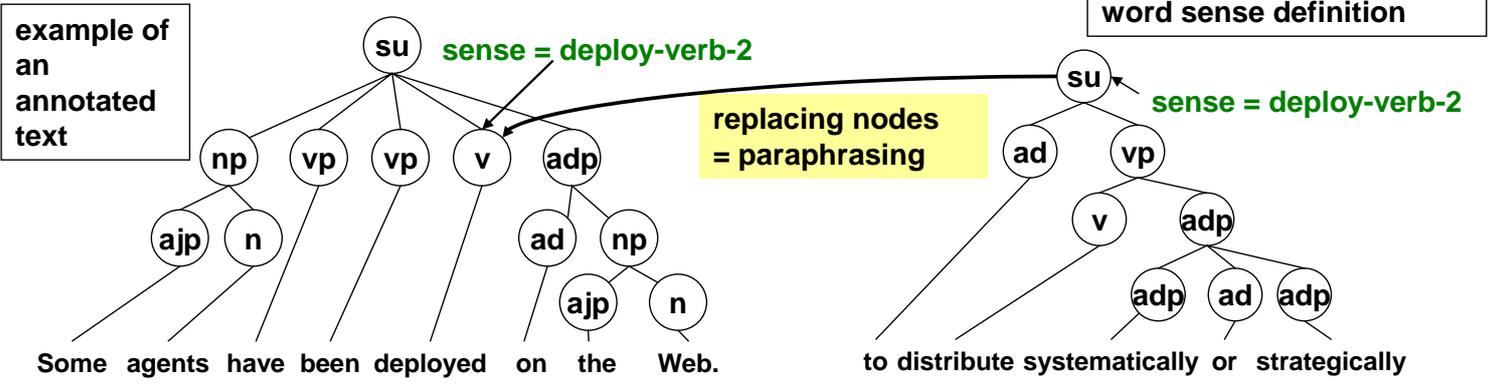rh@sfc.keio.ac.jp

Katashi Nagao
Dept. of Information Engineering, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan
nagao@nuie.nagoya-u.ac.jp

## 4. Interactive Paraphrasing

**Features :**

1. Users can interactively select words to paraphrase by mouse clicks on normal Web browsers.
2. Paraphrasing history is stored for profile-based paraphrasing, which automatically selects words to paraphrase based on user's knowledge. (not yet implemented)
3. Resulting sentence can also be the target for the next paraphrase. By allowing incremental operation, users can interact with the document until there are no paraphrasable words in the document.

**Paraphrasing Mechanism :**

example of an annotated word sense definition

example of an annotated text

sense = deploy-verb-2

replacing nodes = paraphrasing

sense = deploy-verb-2

su
np vp vp v adp
ajp n ad np
ajp n

Some agents have been deployed on the Web.

su
ad vp
v adp
adp ad adp

to distribute systematically or strategically

## 5. Paraphrasing Rules

• Paraphrasing rules are node replacing rules.
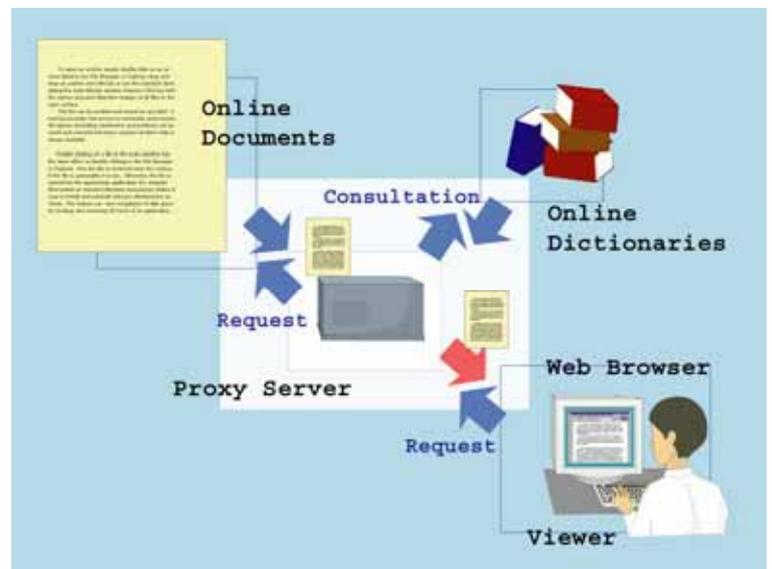• There are two types of rules.

1. **Global rules** are applied to any pair of nodes.

   e.g.,
   • Avoid double negation, which increases the complexity of the sentence.
   • To avoid redundancy, remove the same case-marked structure found in both structures from the definition nodes.

2. **Local rules** are applied depending on the syntactic features of the nodes.

   e.g.,
   • When replacing verb node(A) with verb node(B), apply A's conjugation to B and replace A with B.
   • When replacing noun node(A) with noun node(B), replace A with B agreeing in number.

## 6. Implementation

Online Documents

Consultation

Online Dictionaries

Request

Proxy Server

Web Browser

Request

Viewer

## 7. Summary and Future Plans

We have presented a method, ``Interactive Paraphrasing", which enables users to interactively paraphrase words in a document by their definitions, making use of syntactic annotation and word sense annotation. ``Interactive Paraphrasing" enhances the accessibility and the reuse of documents.

Our future plans include: *reduction of the annotation cost, realization of profile-based paraphrasing using personal paraphrasing history, and retrieval and semantic merging of similar documents using linguistic annotation.*