

# A New E-learning Paradigm through Annotating Operations

Hiroaki Saito, Kyoko Ohara, Kengo Sato, Kazunari Ito, Shinsuke Hizuka  
Masaya Soga, Tomoya Nishino, Yuji Nomura, Hideaki Shirakawa, Hiroyuki Okamoto

Dept. of Computer Science, Keio University  
3-14-1, Hiyoshi, Kohoku-ku, Yokohama, 223-8522, Japan  
hxs@ics.keio.ac.jp

## Abstract

This paper proposes a new e-learning paradigm which enables the user to type in arbitrary sentences. The current NLP technologies, however, are not matured enough to perform full-automatic semantic or discourse analysis. Thus, we take a different approach; an instructor corrects the contents and its correction is embedded into the contents in an XML format called KeML. The key/mouse operations of the user are also embedded as annotations. Thus, the contents can be incomplete at the initial stage and become solid gradually as being utilized. We have implemented an e-learning system of group discussion for a foreign language class, which is demonstrated at the workshop.

## 1 Introduction

Many old e-learning systems asked the user to click the button or to type a unique answer. The domains of the e-learning systems were limited by their approaches and their contents were fixed. This paper introduces a new approach to expand the domain. Namely, the user himself annotates the contents in addition to the instructor.

Annotated contents can be used for further learning such as example-based learning and group learning. The burden of building e-learning contents are so heavy that this “annotated contents become other contents” scheme is important for practical applications. Annotations are attached in an XML format.

This project can be considered as another application of XML technologies, like MATE [1] [2] or Anvil [3] to name a few. The principal difference is that some annotations are implicitly attached and used for NLP.

## 2 System Overview

Here we consider the debate discussion for a foreign language class as an example. This course was originally taught in a regular classroom and through an electronic chatting board supervised by a human instructor. One student posts his/her thought in English and others express positive or negative responses to it. Since this is a foreign language apprehension class, the instructor corrects the students’ English if necessary. Since students express various opinions, we cannot prepare for them in advance.

The instructor was occupied with correcting syntactic errors, therefore, the instructor could not thoroughly pay attention to the flow of the debate or to whether students had appropriately expressed their opinions.

In Figure 1 example debate submissions are shown on the topic “English should be taught at an elementary school.” #*n* indicates the submitted order and P1, P2, ... stand for the identifier of the debaters. (We will further explain Figure 1 later.)

Our system is designed for multi-user discussion, not for self-learning. Thus, we divide the system into the server and the client machines as shown in Figure 2 (only one client machine is drawn in the figure). The server machine manages the contents and handles computationally heavy NLP, while each client machine is responsible for user interface.

We have developed an e-learning system which offers the process above. Here we describe five important modules:

- Sentence analysis module: In this module the input sentences are parsed and tagged syntactically and semantically in the GDA (Global Document Annotation) [4] format. We have adopted the Charniak parser [5], which is customized so that the head word is identified because the GDA tagging requires the attachment direction between phrases. The GDA tagger consults the WordNet [6] to find the root form of a word. GDA tags can be utilized for such further NLP as high quality summarization or information retrieval [7].
- KWIC module: Novice English learners often make such mistakes in collocations and in phrasal verbs. These word-usage mistakes can be effectively resolved by looking at example sentences rather than by consulting a regular dictionary. This module presents the corpus sentences in the KWIC (KeyWord in Context) format in which the specified words are included. Although any corpus will do for KWIC, we have chosen the EDR English corpus [8] which consists of 125,820 tagged sentences. Because the root form of each word is described as a tag, conjugated forms are also searched.
- Annotation module: The instructor corrects the wrong usages of the students’ English. This op-

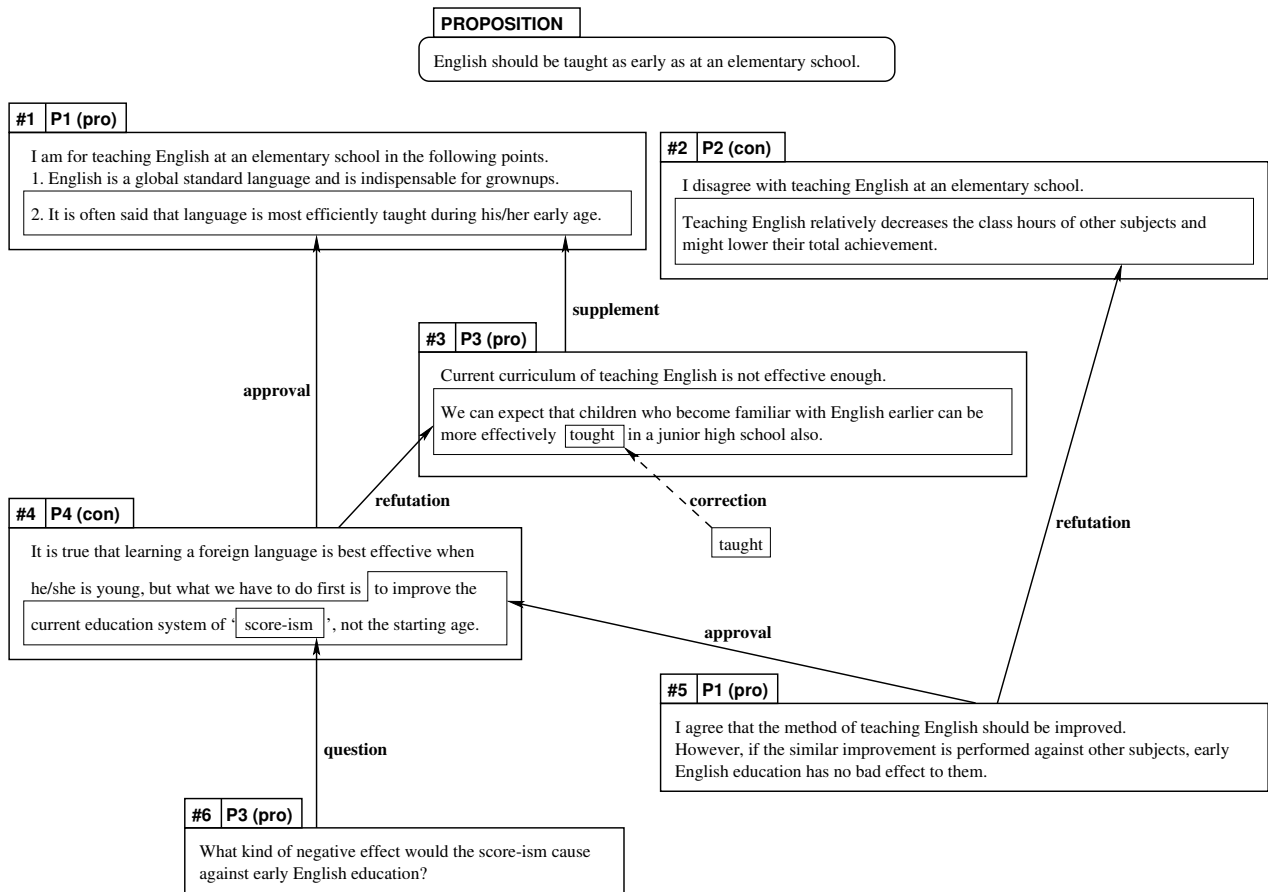


Figure 1: Submission Statements and their Relations

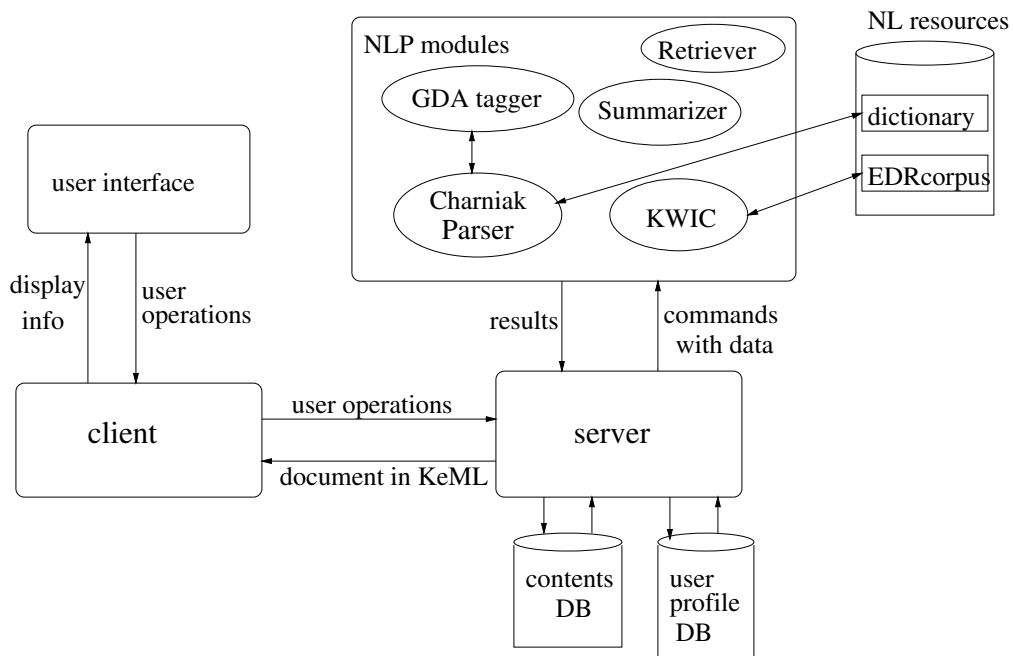


Figure 2: System Architecture

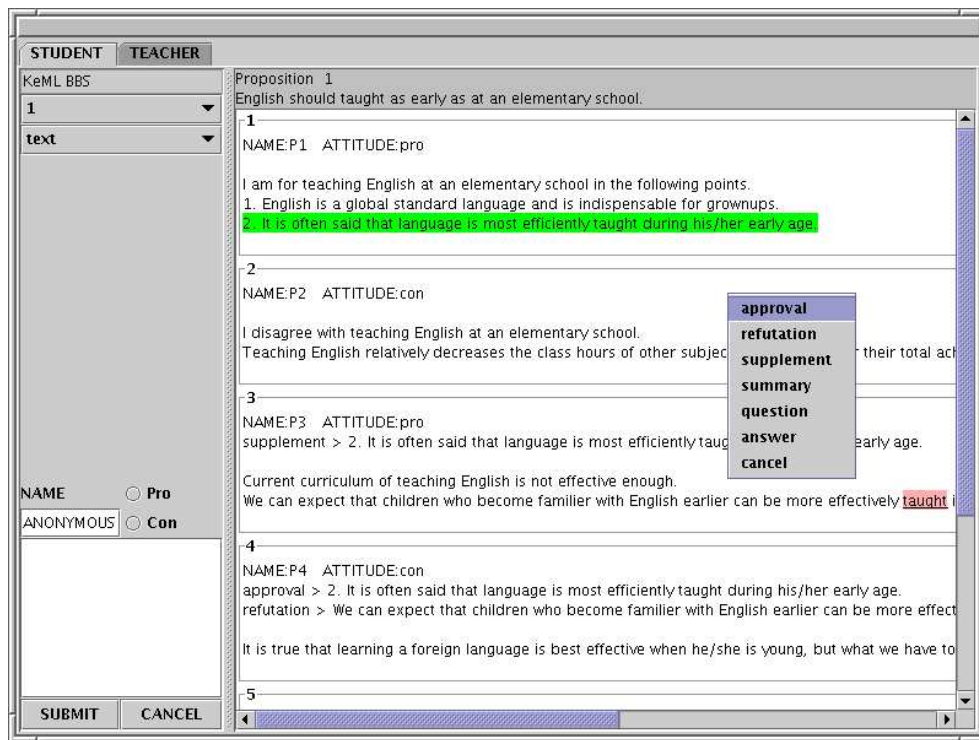


Figure 3: A Snapshot of Interface Window

eration is recorded as annotation, not overwriting the originals. Preserving the originals is effective for education; it can prevent other students from making the similar mistakes. When the debater expresses his opinion against/for someone else's, that operation is also observed and attached to the contents, which will be exploited by NLP.

- Interface module: This module enables the user to type in sentences, specify what part he is arguing about, express his/her attitude, etc, efficiently. This module displays the contents effectively according to the needs of the user with the help of annotations. Our current interface snapshot is shown in Figure 3.
- Debate Flow module  
It is important to know the debate flow when one expresses his/her opinion. Since the relations among statements are annotated, precise analysis of the debate flow is possible.

In the following sections, the annotation module is explained deeply.

### 3 Annotation by the Instructor and Students

When a student expresses his/her opinion in response to someone else's, he can specify and denote what part he is arguing about. This linkage is annotated by the user and recorded in the contents. The corrections/comments by the instructor are also

stored in the learning contents as annotations. Arrows in Figure 1 show the relation of statements, where a dotted line expresses the linkage denoted by the instructor, and solid lines mean that the debater specified those relations.

### 4 The Tag Set for Debate

We have defined a tag set for annotating debates in an XML format called KeML (Keio e-learning Markup Language). Here we describe our tag set along with how each tag is attached through operations by the instructor or students.

<debate> encloses the whole debate and is attached when a new debate starts. No attribute is allowed. Possible child-nodes are one <proposition> and zero or more <statement>s.

<proposition> is attached when a new proposition is submitted. Its mandatory attribute is 'id' and whose value is always '0'. Its child-node is <su> of GDA. The instructor or students should remark the proposition as pros or cons.

<statement> This tag is attached when a statement to a proposition or other statements is submitted by the instructor or students. Its mandatory attributes are 'attitude' whose value would be pro or con, 'person' whose value indicates who submitted that statement, 'time' which indicates when that statement was given, and 'id' number (an integer). The values of the first two attributes are given by the user explic-

itly, while those of the last two are filled by the system. Its optional attributes are ‘approval’, ‘refutation’, ‘supplement’, ‘summary’, ‘question’, and ‘answer’ (some of those attributes appear in Figure 1). They are expressed as “approval=target\_id” for example. Its child-node is <su> of GDA.

Such tags below the <su> level as <np> or <v> are attached by the parser according to the GDA specifications. Every tag must have ‘id’ attribute and its value is filled automatically by the server.

Appendix shows the annotated contents of the debate example in Figure 1.

## 5 Preserving Corrected Contents

In order that a novice student could observe mistakes by other students, our system preserves the original contents and shows them effectively when needed. While some mistakes are obvious, others are not. Only the instructor can correct or comment those errors and KeML offers two levels of correction preservation. Obvious mistakes are stored as the value of ‘original’ attribute; ‘<np original=”tought” ..>taught</np>’ for instance. Unobvious mistakes are commented in the value of ‘comment’ attribute; ‘<su comment=”This is a comment for this sentence.” ....</su>’ for example. When further correction is made against already corrected contents, only the very first version is preserved. Our current implementation allows the correction/comments under <su> nodes.

## 6 Conclusions

We have implemented an e-learning system which facilitates group-discussion for second language learning. Plain texts become solid as being used because of the embedded explicit and implicit annotations by the instructor and students. Accumulated contents will be a good resource for statistical analysis and example-based learning.

## References

- [1] MATE Workbench Homepage:  
<http://www.cogsci.ed.ac.uk/~dmck/MateCode/>
- [2] MATE Homepage:  
<http://mate.nis.sdu.dk>
- [3] Anvil Homepage:  
<http://www.dfki.de/~kipp/anvil/>
- [4] The GDA Tag Set Homepage:  
<http://www.i-content.org/gda/>
- [5] Charniak, E. “A Maximum-Entropy-Inspired Parser”, NAACL 2000. (For software, see <http://www.cs.brown.edu/people/ec/>)
- [6] WordNet  
<http://www.cogsci.princeton.edu/~wn/>
- [7] Miyata, T. and Hasida, K. “Information Retrieval Based on Linguistic Structure” in Proceedings of the Japanese-German Workshop on Natural Language Processing, July 2003.

- [8] EDR Electric Dictionary, EDR English Corpus  
<http://www2.crl.go.jp/kk/e416/EDR/>

## Appendix. Annotated contents (a head portion)

```
<?xml version="1.0" encoding="UTF-8"
standalone="yes"?> <debate> <proposition
id="0" time="Sun Jun 13 22:08:30 JST
2004"> <su id="0.1"> <segs id="0.1.1"
mph="ptb;NNP;;English;">English</segs>
<v id="0.1.2"> <v id="0.1.2.1"
mph="ptb;MD;;should;">should</v> <vp id="0.1.2.2">
</vp> </v> <v id="0.1.2.2.1" mph="ptb;VBN;;teach;">taught</v>
<adp id="0.1.2.2.2"> <ad id="0.1.2.2.2.1"> <adp
id="0.1.2.2.2.1.1" mph="ptb;RB;;as;">as</adp> <ad
id="0.1.2.2.2.1.2" mph="ptb;RB;;early;">early</ad>
</ad> <adp id="0.1.2.2.2.2"> <ad id="0.1.2.2.2.2.1"
mph="ptb;IN;;as;">as</ad> <adp id="0.1.2.2.2.2.2">
<ad id="0.1.2.2.2.2.2.1" mph="ptb;IN;;at;">at</ad>
<np id="0.1.2.2.2.2.2.2"> <adp id="0.1.2.2.2.2.2.2.1"
mph="ptb;DT;;an;">an</adp> <ajp id="0.1.2.2.2.2.2.2.2"
mph="ptb;JJ;;elementary;">elementary</ajp> <n
id="0.1.2.2.2.2.2.2.3" mph="ptb;NN;;school;">school</n>
</np> </adp> </adp> </adp> </vp> </v>.</su>
</proposition> <statement id="1" attitude="pro"
person="P1" time="Sun Jun 13 22:10:41 JST 2004">
<su id="1.1"> <np id="1.1.1" mph="ptb;PRP;;I;"
eq="p1">I</np> <v id="1.1.2"> <v id="1.1.2.1"
mph="ptb;VBP;;be;">am</v> <adp id="1.1.2.2">
<ad id="1.1.2.2.1" mph="ptb;IN;;for;">for</ad>
<vp id="1.1.2.2.2"> <v id="1.1.2.2.2.1"> <v
id="1.1.2.2.2.1.1" mph="ptb;VBG;;teach;">teaching</v>
<segs id="1.1.2.2.2.1.2"
mph="ptb;NNP;;English;">English</segs> <adp
id="1.1.2.2.2.1.3"> <ad id="1.1.2.2.2.1.3.1"
mph="ptb;IN;;at;">at</ad> <np
id="1.1.2.2.2.1.3.2"> <n id="1.1.2.2.2.1.3.2.1"> <adp
id="1.1.2.2.2.1.3.2.1.1" mph="ptb;DT;;an;">an</adp>
<ajp id="1.1.2.2.2.1.3.2.1.2"
mph="ptb;JJ;;elementary;">elementary</ajp>
<n id="1.1.2.2.2.1.3.2.1.3"
mph="ptb;NN;;school;">school</n>
</np> <adp id="1.1.2.2.2.1.3.2.2"> <ad
id="1.1.2.2.2.1.3.2.2.1" mph="ptb;IN;;in;">in</ad> <np
id="1.1.2.2.2.1.3.2.2.2"> <adp id="1.1.2.2.2.1.3.2.2.2.1"
mph="ptb;DT;;the;">the</adp> <vp
id="1.1.2.2.2.1.3.2.2.2.2" mph="ptb;VBG;;follow;">follo
wing</vp> <n id="1.1.2.2.2.1.3.2.2.2.3"
mph="ptb;NNS;;point;">points</n>
</np> </adp> </np> </adp> </v> </vp> </adp>
</v>.</su> <su id="1.2"> <np id="1.2.1"> <num
id="1.2.1.1" mph="ptb;CD;;1;">1.</num> <segs
id="1.2.1.2" mph="ptb;NNP;;English;">English</segs>
</np> <v id="1.2.2"> <v id="1.2.2.1"> <v
id="1.2.2.1.1" mph="ptb;VBZ;;be;">is</v>
<np id="1.2.2.1.2"> <adp id="1.2.2.1.2.1"
mph="ptb;DT;;a;">a</adp> <ajp id="1.2.2.1.2.2"
mph="ptb;JJ;;global;">global</ajp> <ajp id="1.2.2.1.2.3"
mph="ptb;JJ;;standard;">standard</ajp> <n
id="1.2.2.1.2.4" mph="ptb;NN;;language;">language</n>
</np> </v> <segs id="1.2.2.2"
mph="ptb;CC;;and;">and</segs> <vp id="1.2.2.3">
<v id="1.2.2.3.1" mph="ptb;VBZ;;be;">is</v>
<ajp id="1.2.2.3.2"> <aj id="1.2.2.3.2.1"
mph="ptb;JJ;;indispensable;">indispensable</aj>
<adp id="1.2.2.3.2.2"> <ad id="1.2.2.3.2.2.1"
mph="ptb;IN;;for;">for</ad> <np id="1.2.2.3.2.2.2"
mph="ptb;NNS;;grownup;">grownups</np>
</adp> </vp> </v>.</su>
```