

# CCML: Exchanging Marked-up Documents in a Networked Writing Classroom

Hiroaki OGATA, Yoneo YANO and Riko WAKITA\*

*Faculty of Engineering, Tokushima University, Japan*

*\* Faculty of Education, Fukui University, Japan*

**Abstract:** This paper describes an online markup-based composition learning environment system called **CoCoA** (Communicative Collection Assisting System). It allows students and teachers to exchange marked-up documents via Internet, and its environment is very similar to a real world one in which people use pen and paper. In order to record and exchange corrected compositions with marks and comments, this paper proposes **CCML** (Communicative Correction Mark-up Language), which is based on SGML (Standard Generalized Mark-up Language). CCML facilitates teachers to analyze and reuse the marked-up documents for the instruction.

## 1. Introduction

Recently, teacher-centered instructional approaches in traditional writing classrooms are replaced with more active and learner-centered learning approaches with collaborative writing tools (Bonk et al, 1995). These tools can (1) change the way students and teachers interact; (2) enhance collaborative learning opportunities; (3) facilitate class discussion; and (4) move writing from solitary to more active and social learning. Writing compositions includes various sub-processes such as planning, transcribing, and revising, which do not need to occur in any fixed order (Scardamalia & Berierter 1986). In particular, the review process assisted with computer-based writing tools, has recently received much interest (see as examples (Kehagia & Cox 1997; Coniam 1997)).

Many researchers developed online markup systems employing some markup models. However, it is very difficult to analyze and reuse the marked documents which are collected through the writing classroom because the documents do not have a common structure. Therefore, it is necessary to define the generalized format for encoding and exchanging the marked-up documents in order that online markup systems are used easily and widely.

This paper proposes **CCML** (Communicative Correction Mark-up Language) based on SGML (Standard Generalized Mark-up Language) (Herwijnen 1990) which is one of the Internet's key technologies. This structured document allows teachers and students to analyze and reuse it. This paper also describes **CoCoA** (Communicative Collection Assisting System) (Yano et al 1997) which is an application of CCML. This system offers support to learners and teachers to exchange marked-up documents via Internet, and the environment is very similar to a real one in which people use pen and paper. CoCoA allows not only teachers to correct the compositions sent from students by e-mail, but also allows students to see where and why the teacher had corrected them.

We have been investigating technological support for Japanese language learning among

overseas students. For example, CAI systems called Kanji Laboratory (Hayashi & Yano 1994), JUGAME (Yano et al. 1993), GRACILE (Ayala & Yano 1995) and JULLIET (Ochi et al. 1996) were developed to support Japanese language learning. However, an on-line mark-up supporting system for Japanese language learning has not yet been proposed. Usually, in a Japanese writing classroom, teachers have to individually review learners' documents using pen and paper (Sato et al. 1994). It takes a lot of time for teachers to do this. Therefore, we have implemented CoCoA for writing Japanese composition.

This paper is organized as followings. Section 2 summarizes online markup models. In section 3, we describe the experimental results of correcting Japanese compositions written by learners using pen and paper. We propose CCML based on the experimentation in section 4. Section 5 illustrates the design and the implementation of CoCoA and the error analysis of the learner's documents. We describe the experimental evaluation of CCML and CoCoA in section 6. Finally, the concluding remarks are given in section 7.

## 2. Online markup models

There are some editing systems that support teachers to review and correct the students' drafts with online mark-up. Farkas & Poltrock (1995) classified the mark-up models as followings:

- (1) *Silent editing model*: This is the simplest model and it requires no special techniques. However, it is very difficult for the author to check the editor's work. This model is destructive because the editor cannot readily recover the original words once he/she has changed it.
- (2) *Comment model*: This model employs pop-up notes, temporary footnotes, hidden text, and special symbols placed within the text. This model can work for special groups and ad-hoc situations. A system called XyWrite (Kincade & Oppenheim 1994) was proposed with this model.
- (3) *Edit trace model*: In this model, the editor works in the manner of an author, deleting, adding, and moving text as usual. The computer can compare the editor's new version with the original text, and allows the author to view the draft that contains the editor's changes. This model is apt to encourage heavier editing and less regard for the author's original text. Microsoft Word accepts this model.
- (4) *Traditional mark-up model*: This adapts the traditional paper mark-up model to the computer screen. The symbols are both familiar and intuitive for editors and authors; for example, deletion, insertion, and move. For instance, Red Pencil allows the editor to apply a complete set of traditional editing symbols directly to a document. The editor uses "digital ink" to mark a traditional editing symbol along with the words. Moreover, MATE (Hardock et al. 1993) allows the editors to use both digital ink and voice command toward pen and voice computing. In this model, authors and editors can interpret the editor's markings much more readily than in the edit trace model.

There are many systems that employ traditional mark-up which allow multiple users to mark-up an electronic document as if they were marking up a printed copy of the document. However, such systems do not globally come into practical and wide use in composition writing classes because of their special format. Moreover, it is very difficult to analyze and reuse the marked documents because the marked documents are unstructured. Therefore, the system should provide a generalized and structured format for encoding and interchanging marked-up documents via the Internet.

### 3. Experimental Correction with Pen and Paper

We investigated what kind of marks editors used to revise overseas students' compositions with pen and paper, and what kind of features they needed to do it.

#### 3.1. Users and Tasks

In the experiment, we selected 18 compositions written by 18 students respectively. They were beginners in Japanese language and they came to Japan from various countries; e.g., China, Malaysia, Guinea, and so forth. Then, 13 master course students of Japanese language and one teacher of Japanese education reviewed the compositions of the students. The teacher had four years of teaching experience, including one year of computer programming experience. They reviewed three compositions that have 713 words. The reviewers found and corrected 53 errors in the documents during 90 minutes.

#### 3.2. Experimental results

The experimental results are summarized in the following:

- (1) The Japanese compositions written by overseas students contained some compound errors. Hence, it is very difficult for the reviewers to understand what they meant and how to correct the compositions. Editors cannot expect the support of computerised text analysis programs (Brock 1995) to correct the documents automatically.
  - (2) The respective reviewers used various kinds of marks for the proofreading.
  - (3) The importance of marks was varied in the document.
  - (4) Annotations of the reviewer are very important to give effective instruction to the student.
- Based on these results, we analyzed the marks, the importance of corrections, and the annotations.

JIS (Japanese Industrial Standard) defines 20 marks and 18 sub-marks as the traditional marks for proofreading (Shimano 1986). A mark shows the spot of correction, e.g., text insertion. All the marks were not required in our situation because the documents were generated with a computer. Based on this standardization, six marks and annotations were used in our experiment. Figure 1 shows the frequency of marks in this experiment. The mark for replacing words was used more than any other mark.

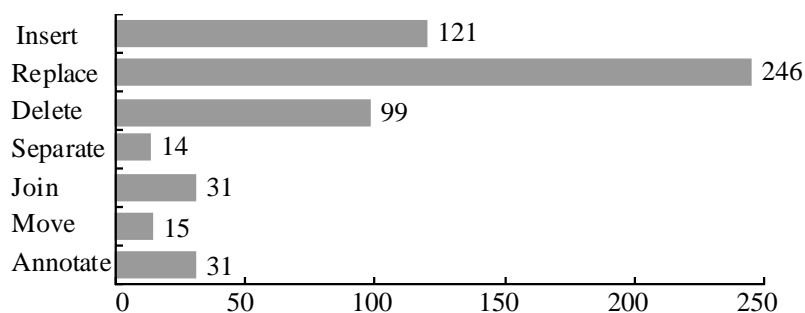


Figure 1: The frequency of marks used by the editors.

### 4. CCML

Based on the experimental results, we propose CCML for exchanging marked-up documents. CCML is an application of SGML which is an ISO standard (ISO 8879:1986), and it supplies a formal notation for the definition of generalized mark-up languages. SGML is a device- and system-independent method of representing texts in electronic form. That is to say, SGML is a set of mark-up conventions used together for encoding texts. A mark-up language must specify

what mark-up is allowed, what mark-up is required, how mark-up is to be distinguished from text and what the mark-up means.

#### **4.1. Features of CCML**

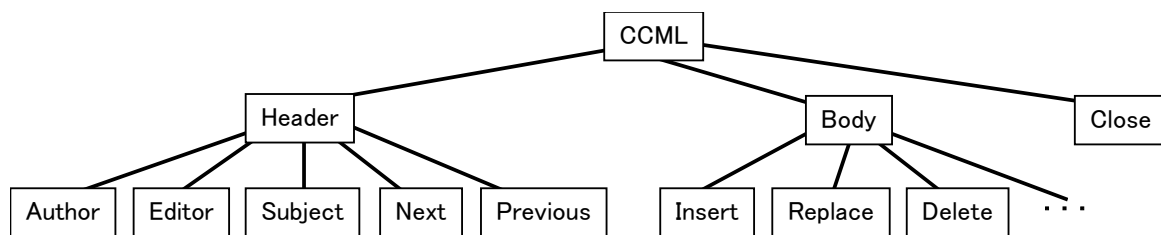
The main characteristics of CCML are:

- (1) Based on the experiment, CCML presents six marks and annotation CCML tags.
- (2) The marks have three degrees of importance levels against respective corrections.
- (3) The original text is generated by removing all the CCML tags.
- (4) The revised text is derived from the CCML document.
- (5) Because CCML documents are text-formated, it is easy to send them by e-mail.
- (6) CCML documents easily make up a full-text databases.

Needless to say, CCML inherits its features from SGML.

#### **4.2. CCML structure**

As shown in figure 2, CCML documents consist of three parts: header, body and close. “Header” represents additional information about the document. For instance, “next” tag denotes the next version of the document. The marks for review are shown in the “Body” as CCML tags. “Close” shows the editor’s comments.



**Figure 2:** The tree structure of CCML documents.

Table 1 shows the six marks and one annotation. In one sentence, “insert,” “replace” and “delete” marks were used, while “join,” “separate” and “move” marks were used over two sentences. The part between the start tag and the end tag denotes the learner’s mistakes. The “string” attribute represents the revised part of the document.

**Table 1: Marks and CCML tags.**

	Correction	Mark	Tag
Single sentence	1. Insert		<Insert string="text">
	2. Replace		<Replace string="text"> text </Replace>
	3. Delete		<Delete> text </Delete>
Multiple sentences	4. Separate paragraph		<Separate>
	5. Join paragraph		<Join>
	6. Move		<MoveFrom refid="ID">
			<MoveTo id="ID"> text </MoveTo>
7. Annotate		<Annotate type="question"   "explanation" string="text">	

#### 4.3. Level of marks

We found that the marks do not have the same level of importance. We identify corrections on the following levels:

- (1) Weak correction: The learner does not need to revise the document.
- (2) Normal correction: The learner should correct the document.
- (3) Strong correction: The learner must correct the document.

The strong corrections denote the important part of marks to be revised in the document. Using the importance level that the teacher had entered, the system provides the learner with the marks he/she wants to see. Therefore, the learner can avoid information overload from the reviewed documents. Every tag in table 1 has an attribute "level" that a teacher gives a number from one to three. Its default is two as normal correction.

#### 4.4. Level of annotations

It is very important for a teacher to annotate the marked text for instruction in composition. For example, PREP Editor (Neuwirth et al, 1996) is a word processor that allows writers and reviewers to create electronic margins, or columns, in which they can write and communicate through their annotations. We identify the following different kinds of annotations:

- (1) Explanation: This is used for explaining the reason of a correction.
- (2) Question: This is used for asking the learner a question; e.g., what do you want to write?
- (3) Comment: This shows the educational view of the teacher with respect to the document.

### 5. CoCoA

CoCoA consists of CoCoA\_Editor and CoCoA\_Viewer. A teacher corrects the document of overseas students with CoCoA\_Editor and the system saves the marked document in CCML format. CoCoA\_Viewer interprets the CCML document, and shows the learner the marked document.

### ***5.1.Learning processes using CoCoA***

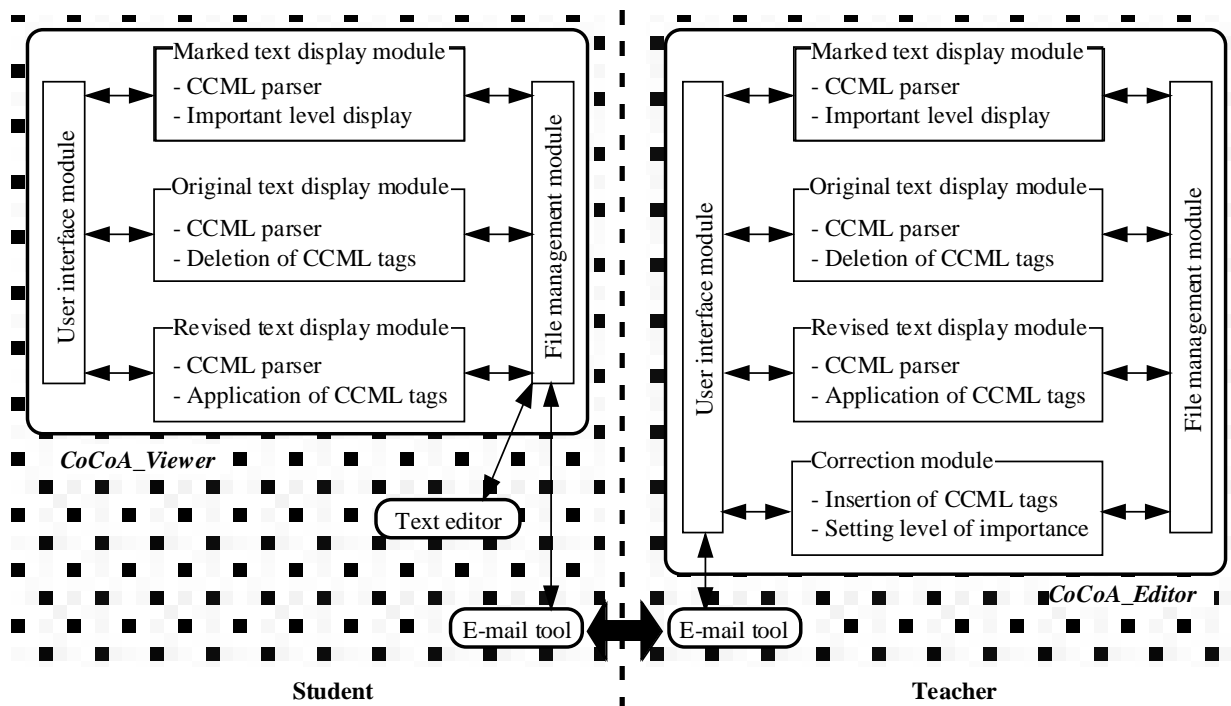
By using CoCoA, a learner receives instruction about a Japanese composition from a teacher with the following processes:

- (1) The learner writes an original text with his/her familiar editor.
- (2) The learner sends the document to his/her teacher with his/her own e-mail tool.
- (3) CoCoA\_Editor makes the document double-spaced. The teacher corrects the document with online marks and annotations. Then, the system allows the teacher to set the importance level to the marks in the document.
- (4) After CoCoA\_Editor saves the marked text as a CCML document, the teacher sends it to the learner by e-mail.
- (5) CoCoA\_Viewer provides the learner with the marked text after interpreting the CCML document. Then, the system allows the learner to select the importance level to see the important part of the marked text.
- (6) CoCoA\_Viewer automatically generates both the original text and the revised one from the CCML document. After editing the revised text, the learner can send it again to the teacher and continue refining the text.
- (7) CoCoA maintains the version of the document, if the learner wants to revise the same document.

### ***5.2.System configuration***

Figure 3 depicts the learning environment of CoCoA.

- (1) Marked text display module: This module analyzes CCML documents using the CCML parser after reading them through the file management module. Then, it provides the results of correction according to the level of importance of marks.
- (2) Correction module: This module inserts CCML tags into the learner's document, according to the revision of the teacher. After saving the marked text, the teacher sends it by e-mail to the learner.
- (3) Original text display module: This module generates the original text from the CCML document by removing all the CCML tags.
- (4) Revised text display module: This module generates the revised text by applying CCML tags.
- (5) File management module: This module manages the versions of the documents. When the learner sends the teacher the revised document, the system creates a new CCML document, inserts the "next" tag into the old CCML document, and also enters the "previous" tag into the new CCML document.



**Figure 3:** System configuration of CoCoA.

### 5.3. User interface

Figure 4 shows the screen snapshot of CoCoA. First, the learner writes a Japanese composition with a word processor. After that, the learner sends the document to the teacher by e-mail. By selecting a mark from the mark palette shown in window (A), the teacher can revise the document. Moreover, the teacher can annotate the document using the annotation palette, and he/she can classify the marks using a window which looks like (B). The “undo” button erases corrections. In figure 5, after the teacher substitutes “他国” with “留学先” which means “foreign countries”, the user enters three (strong correction) into the important level of the correction. After saving the marked document as a CCML text (see appendix A), the teacher can send it to the student by e-mail.

On the other hand, the learner obtains the same marked text in a window (D) which the teacher revised. In this case, CoCoA\_Viewer provides only the marks of strong correction (see window (E)). “?” denotes a question and “\*” shows an explanation from the teacher. The system displays the contents of the annotation as in window (F) after clicking the mark. The message window allows the teacher and the learner to communicate with respect to the document. For example, the messages are the learner’s questions, or the teacher’s diagnosis. The learner can reply to the teacher’s comments and collaboratively write a composition with the teacher.

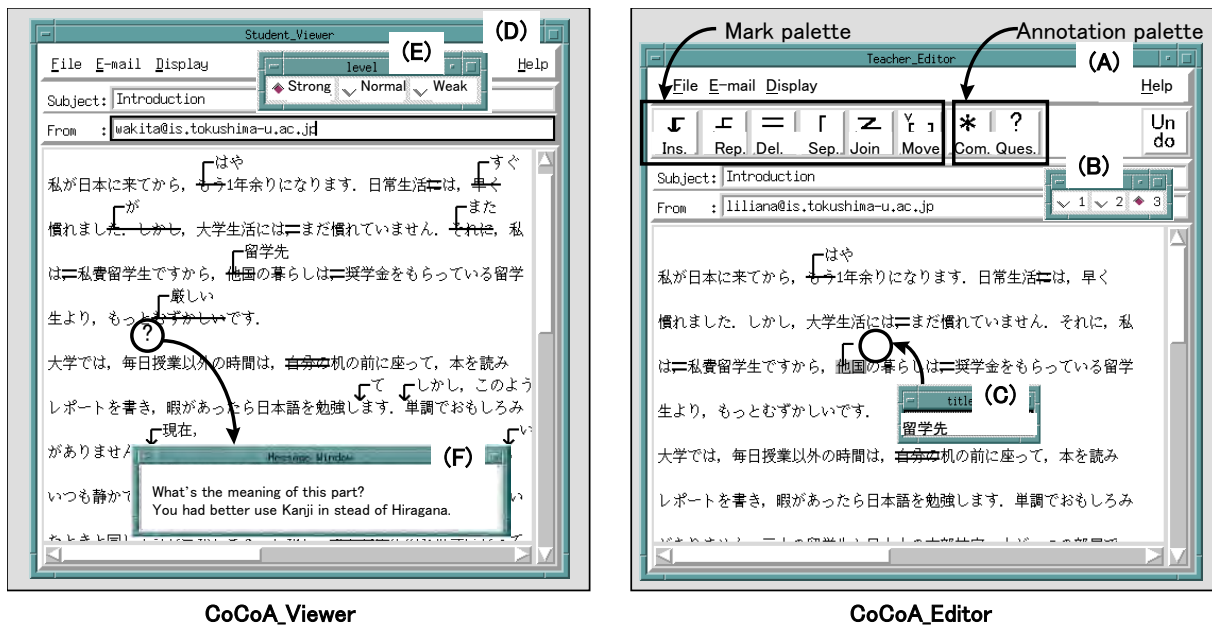


Figure 4: Screen snapshots of CoCoA\_Viewer and CoCoA\_Editor.

#### 5.4. Error analysis of CCML documents

In language learning support systems, it is very important to understand where he went wrong, what grammar rules the student misunderstood, and so on. Error analysis of the student's documents is very difficult if the corpus is ill-formed and include a lot of errors (Schwind 1995). Since the documents of CoCoA users are structured with CCML tags and include both correction and errors, the teacher can easily find out what kind of mistakes the learner made often, what the learner does not understand well in writing composition and so on. CoCoA allows the teacher to make the following queries in each of the learner's documents:

- (1) In what kind of morphemes did the learner make mistakes?
- (2) In what kind of representation did the learner make mistakes; Hiragana, Katakana, or Kanji characters?
- (3) What kind of importance level of marks did the learner's document have?
- (4) What kind of marks did the learner's document include, e.g., insert, replace or delete?

According to the teacher's query, CoCoA analyzes the learner's documents and it represents the result of the analysis graphically.

Figure 5 depicts the result of the CCML document analysis. We analyzed the morphological error of 18 compositions mentioned in section 3 with ChaSen (Matsumoto et al. 1997). ChaSen is a Japanese morphological analysis system and it can classify the errors into 91 categories in detail. In this analysis, CoCoA finds out the corrected parts with the CCML tags, e.g., replace, deletion, insertion tags, and it analyzes the corrected part with ChaSen. This analysis assists teachers to give the learner adequate diagnosis. For example, the teacher can advise the learner to pay attention to the usage of post-positional particles. Moreover, this system can store the pairs of errors and corrections into the database and it allows the teachers and learners to retrieve the errors or corrected sentences from the database.



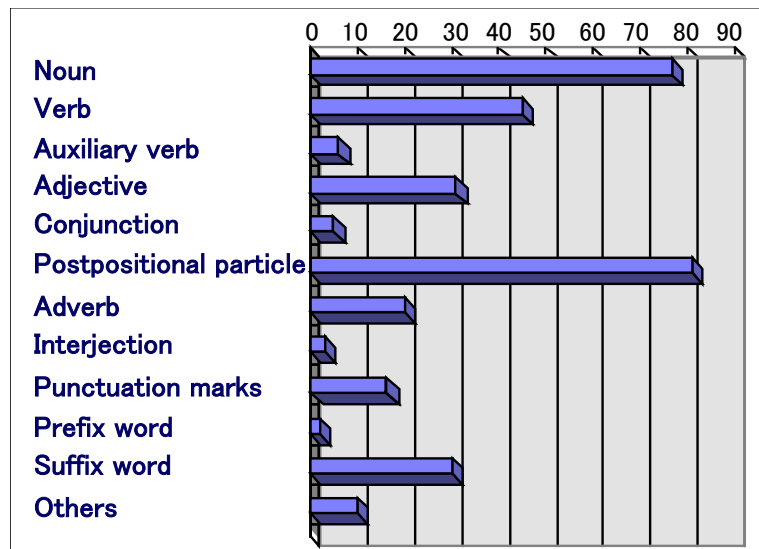


Figure 5: Frequency of errors by each morpheme.

## 6. Experimentation With CoCoA In Japanese Language Learning

### 6.1. Evaluation of CCML and CoCoA

We evaluated CCML and CoCoA by some questionnaires. The users of the system were the same members whom we described in section 3 and we chose the same quantity of documents. While the users corrected the documents with pen and paper in the former experiment, now they did it with computers. Users gave a number between one to five to each one of three questions, the results of which are shown in Table 2. These results prove clearly that the learning environment of CoCoA is enough to correct the documents.

Table 2: The result of the questionnaire.

Question	Ave.
(A) Are the seven marks enough for correcting?	4.9
(B) Does CoCoA facilitate to correct documents?	4.7
(C) Is it easy to understand the marked texts?	4.6

### 6.2. Practical use of CoCoA

CoCoA was used experimentally in a Japanese language learning class at Fukui Prefectural University in Japan for one month. The teacher of the class belonged to Fukui University that was connected to Fukui Prefectural University via the Internet. The number of overseas students was six. For example, the teacher gave the students a written homework about their own hobby. Each learner wrote a composition in Japanese and sent it to the teacher by e-mail. The teacher corrected it with CoCoA\_Editor. The teacher argued that the environment of CoCoA was enough to correct the Japanese compositions of the students. However, it was not easy for the teacher to use the user interface of CoCoA\_Editor. The learners saw the results of the teacher's proofreading through CoCoA\_Viewer. They commented that it was easy for them to understand the mistakes in the document because of the use of marks, and that the optional view of the original, marked or revised text was very useful.

## 7. Conclusions

In this paper, we proposed a computer mediated language learning system called CoCoA and CCML for exchanging electronic marked documents. Although we tried to use CoCoA in Japanese language learning in this paper, CoCoA is independent of the domain of language. If W3C (World Wide Web Consortium) publish XML (Extensible Markup Language) that is the extension of SGML and HTML, it will be easy to adopt CCML into Web browsers. Therefore, CoCoA will be able to be used for learning any language in an open-ended writing classroom. In our future research, we will investigate how to classify students' bugs in their drafts, and how to assist a review process with AI technologies. Moreover, we will attempt that a CCML document can include multimedia data, e.g., movies and voices. We have implemented CoCoA in Tcl/Tk on both Microsoft Windows 95/NT and Sun workstations.

## Acknowledgment

This research was supported in part by the Grant-in-Aid for Scientific Research (B) (2) No.09558017, No.09480036, and No.09230214 from the Ministry of Education, Science, Sports and Culture in Japan.

## References

- Ayala, G. & Yano, Y. (1995). "GRACILE: A framework for collaborative intelligent learning environments", *Journal of the Japanese Society of Artificial Intelligence*, vol.10. no. 6. pp.156-170.
- Bonk, J. C. & King, S. K. (1995). "Computer conferencing and collaborative writing tools: Starting a dialogue about student dialogue," 1995 Proc. of Computer Supported Collaborative Learning.
- Brock, N. M. (1995). "Computerised text analysis: Roots and research," *Computer Assisted Language Learning*, vol.8, no.2-3, pp.227-258.
- Coniam, D. (1997) "A computerised English language proofing cloze program," *Computer Assisted Language Learning*, vol. 10, no. 1, pp.83-97.
- Farkas K. D., & Poltrock, E. S. (1995). "Online editing mark-up models, and the workspace lives of editors and writers," *IEEE Transactions on Professional Communication*, 38(2), pp.110-117.
- Hardock, G., Kurtecbach, G., & Buxton, W. (1993). "A marking based interface for collaborative writing," *Proc. of the ACM Symposium on User Interface Software Technology*, pp.259-266.
- Hayashi, T., & Yano, Y. (1994). "Kanji Laboratory: An environmental ICAI system for Kanji learning," *IEICE Trans. on Information and Systems*, E77-D(1), pp.80-88.
- Herwijnen, E. (1990). *Practical SGML*, Kluwer Academic Publishers.
- Kehagia, O. & Cox, M. (1997). "Revision changes when using word processors in an English as foreign languages context," *Computer Assisted Language Learning*, vol. 10, no. 3, pp.239-253.
- Kincade, D. & Oppenheim, L (1994) "Marking it up as we go along: Into editorial production's electronic future," *Journal of Scholarly Publishing*, Vol. 25, No. 3, pp.233-242.
- Matsumoto, Y, Kitauchi, A., Yamashita, T., Hirano, Y., Imaichi, O. & Imamura, T. (1997) "Japanese Morphological Analysis System ChaSen Manual," Nara Institute of Science and Technology Technical Report NAIST-IS-TR 97007. (in Japanese)
- Neuwirth, M. C., & Wojahn, G. P. (1996). "Learning to write: Computer support for a cooperative process," *CSSL: Theory and practice of an emerging paradigm*, Koshmann, T. (Ed.), Lawrence Erlbaum Associates, Publishers, pp.147-170.
- Ochi, Y., Yano, Y., & Hayashi, T. (1996). "JULLIET: Interactive learning environment for Kanji compounds learning," *Proc. of Educational Multimedia and Hypermedia 1996*,

pp.539-544.

- Sato, M., Kano, C., Tanabe, K., & Nishimura, Y. (1994). *Practical Japanese Compositions*, Bonjin-sya. (in Japanese)
- Scardamalia, M., & Bereiter, C. (1986). "Research on written composition," M. Wittrock (Ed.), *Handbook of research on teaching* (third ed.), Macmillan, pp.778-801.
- Schwind, B. C. (1995) "Error analysis and explanation in knowledge based language tutoring," *Computer Assisted Language Learning*, vol.8, no.4, pp.295-324.
- Shimano, H. (1986). *Handbook of proofreading*, Miki Publishers, Japan, in Japanese.
- Yano, Y., Miyoshi, K., & Hayashi, T. (1993). "Development of a game style environmental ICAI system for Kanji idiom learning," *Proc. of International Conference on Computers in Education '93*, pp.411-413.
- Yano, Y., Ogata, H., Sakakibara, R., & Wakita R. (1997). "CoCoA: Communicative correction assisting system for learning Japanese compositions", *Transactions of Japanese Society for Information and Systems in Education*, vol. 14, no.3, pp.21-28. (in Japanese)

#### Appendix A: CCML document in figure 4.

```
<CCML>
<Header>
<Author> liliana@is.tokushima-u.ac.jp
<Editor> wakita@is.tokushima-u.ac.jp
<Subject> Introduction
...
</Header>
<Body>
私が日本に来てから、<Replace string="はや">もう </Replace> 1年になります。日常生活 <Delete
level=1> に</Delete> は、<Replace string="すぐ" level=1> 早く </Replace> 慣れました <Replace
string="が" level=1> . しかし、<Replace> 大学生活には <Delete level=1>, </Delete> まだ慣れていませ
ん。 <Replace string="また" level=1> それに</Replace>, 私は<Delete level=1>, </Delete> 留学生です
から、<Replace string="留学先">他国</Replace>の暮らしは、奨学金をもらっている留学生より、もっと
<Replace string="厳しい" level=3>むずかしい</Replace>です。
...
</Body>
</CCML>
```