

Computer Supported Proofreading Exercise in a Networked Writing Classroom

Hiroaki Ogata Cheng Feng Yoshiaki Hada Yoneo Yano
Department of Information Science and Intelligent Systems, Tokushima University
2-1, Minami-josanjima, Tokushima-shi, 770-8506, Japan
E-mail: ogata@is.tokushima-u.ac.jp
URL: <http://www-yano.is.tokushima-u.ac.jp/ogata/>

Abstract: CoCoA (Commutative Collection Assistant System) developed for supporting students and teachers to exchange online marked-up documents via email. Its environment is very similar to a real one in which people use paper and pen. In order to effectively reuse marked-up digital documents created with CoCoA, this paper proposes a computer supported proofreading exercise environment called CoCoA2. CoCoA2 can reduce the overload of the teacher, and it can make composition learning more effective. CoCoA2 has been developed and used it experimentally.

Keywords: distance education, computer assisted language learning, composition studies, SGML, online markup.

1. Introduction

Recently, researchers in educational systems attempt to provide technological support for cooperative and collaborative learning advocated by educational theories (O'Malley, 1994). The explosive growth of networking, in particular, raises the possibility of widespread collaborative and open-ended learning activities (Harasim, et al., 1995). We have investigated on technological support for open-ended and collaborative learning activities (Ogata, et al., 1996; 1997). In composition studies, coupled with the shift in writing pedagogy toward an open-ended and learner-centered approach, the theoretical view of learning has created a surge of interest in collaborative learning in such way as to prize conversation and interaction among students and among students and teachers.

In computer-based classrooms, writing teachers and students use various software tools to substantiate theories of writing. Writing compositions includes various sub-processes such as planning, transcribing, and revising, which do not need to occur in any fixed order (Scardamalia & Bereiter, 1986). Hence, it is very important for students to receive the instruction through the review. There are some asynchronous editing systems that support teachers to review and correct the students' drafts with online mark-up. These systems adapt the traditional mark-up model to the computer screen. The symbols are both familiar and intuitive for editors and authors; for example, deletion, insertion, and move (Farkas & Poltrock, 1995). For instance, MATE (Hardock, Kurtebach, & Buxton, 1993) allows the editors to use both online markup and comments. In this model, authors and editors can interpret the editor's markings much readily. Although there are many systems that employ markup that allows multiple users to correct an electronic document as if they were marking up a printed copy of the document, the systems do not provide a generalized format for exchanging the documents using Internet. Therefore, such systems do not globally come into practical and wide use in composition writing classes.

CoCoA (Communicative Correction Assistant system) has been developed for supporting foreigners and teachers to exchange marked-up documents by e-mail (Ogata et al., 1997). Its environment is very similar to a real one in which people use paper and pen. CoCoA allows teachers not only to correct the compositions sent from foreigners by E-mail, but also foreigners to see where and why the teacher had corrected them. CoCoA improves the opportunities that foreigners have for writing Japanese compositions and for receiving instructions from teachers. **CCML** (Communicative Correction Mark-up Language) (Ogata et al., 1998) has also been proposed for the representation of marked-up documents, which is based on **SGML** (Standard Generalized Mark-up Language) (Herwijnen, 1990). With CCML, teachers and students can exchange marked-up documents via e-mail (Ogata et al., 1999). In the experimental use of CoCoA, most of users commented that CoCoA was easy for them to understand the mistakes in documents because of the use of marks, and that the optional view of the original, marked or revised text was very useful.

This paper proposes an online markup based proofreading exercise environment called CoCoA2, for Japanese language composition studies. In order to effectively re-use marked-up digital documents created with CoCoA,

CoCoA2 recycles the documents in correcting exercise for other learners. In CoCoA, a learner sends a teacher an original text. The teacher then corrects the text with CoCoA, and returns the learner the marked-up text. The marked-up texts of all the learners are stored into a repository of the teacher. From the students in a networked writing classroom, CoCoA2 chooses an appropriate learner for proofreading exercise by analyzing writing errors in the CCML formatted documents repository. CoCoA2 presents the selected learner with an original text in which there are a number of errors which the learner has to find and correct. This exercise elevates error-correcting skills and self-monitoring skills in writing a composition (Coniam, 1997). After the learner finishes proofreading, CoCoA2 evaluates the result through comparing the correction of the learner with that of the teacher. CoCoA2 creates the opportunity of collaborative learning by connecting the proofreading learner and the learner who wrote the original text. The document can be improved through the collaboration. Because of the recycle of online marked-up digital documents, CoCoA2 can reduce the overload of the teacher, and it can make composition learning more effective.

We have been investigating about technological support for Japanese language learning among overseas students. For example, CAI systems called Kanji Laboratory (Hayashi & Yano, 1994), JUGAME (Yano, et al., 1993), and JULLIET (Ochi, et al; 1996) were developed to assist Kanji learning. Moreover, intelligent agents based on a CSCL environment called GRACILE (Ayala & Yano, 1995) which supports collaborative writing focusing on Japanese composition conjugation. However, an online mark-up based collaborative learning system for Japanese language learning has not been developed. Therefore, we have designed and implemented CoCoA2 for Japanese writing class.

This paper is organized as the following. First, section 2 presents overview of CoCoA and CCML. A proofreading exercise environment with CoCoA2 is proposed in section 3. Section 4 illustrates the implementation of CoCoA2 and section 5 describes the experimental use of CoCoA2. Finally, the concluding remarks are given in section 6.

2.Overview of CoCoA and CCML

2.1.Learning processes using CoCoA

CoCoA consists of CoCoA-Editor and CoCoA-Viewer. A teacher corrects the document of foreigners 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. While CoCoA-Viewer corresponds to a WWW browser like Netscape Navigator, CoCoA-Editor has the same features of an HTML editor. By using CoCoA, a learner receives the instruction about Japanese composition from a teacher with the following processes:

- (1) A learner writes an original text with his/her familiar editor.
- (2) The learner sends the document to his/her teacher with his/her email tool.
- (3) Using CoCoA-Editor, the teacher corrects the document with marks and annotations. The system makes the documents double-spaced and it 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 with his/her email tool.
- (5) CoCoA-Viewer provides the learner the marked text after interpreting its CCML document. Then, the system allows the learner to select the important level to see the important part of the marked-up 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.

2.2.Interface of CoCoA

Figure 1 shows the interface of CoCoA. By selecting a mark from the mark palette shown in the window (A), a teacher can correct learners' documents. Moreover, the teacher can annotate in the document using the annotation palette, and he/she classifies the marks using a window which looks like (B). The "undo" button erases corrections. In figure 1, the teacher changes the words which means "attend", and determines the important level of the correction is three.

A learner obtains a marked text in a window (C). In this case, CoCoA-Viewer provides only the marks over the level two of importance. "?" denotes a question and "*" shows a explanation from the teacher. The system displays the contents of the annotation after clicking the mark. As shown in the message window, the

learner can reply to the message and send it after revising the document.

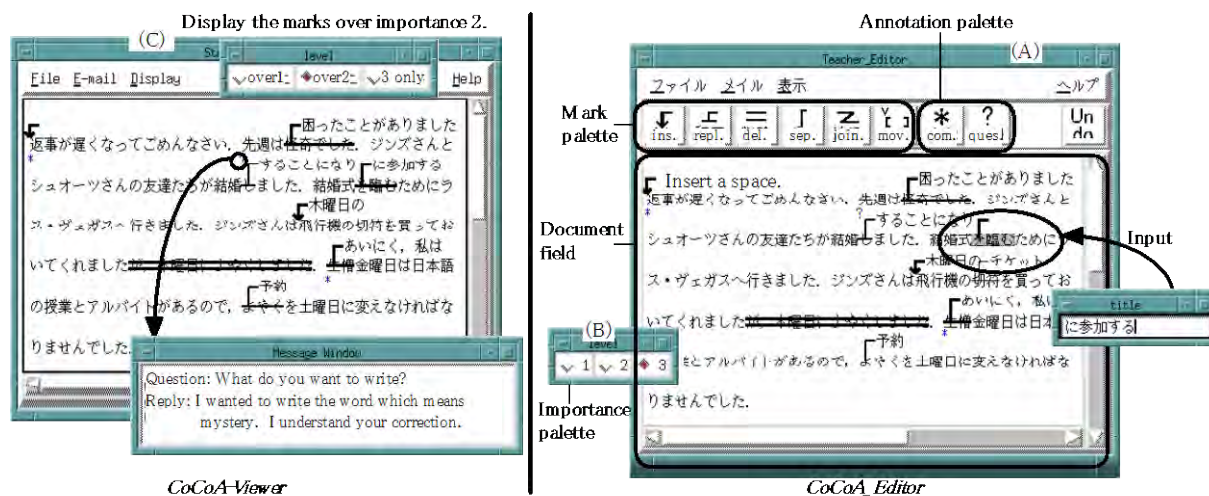


Figure 1: Screen snapshots of CoCoA-Viewer and CoCoA-Editor.

2.3. CCML

Japanese Industrial Standard (JIS) defines 20 marks and 18 sub-marks as the marks for proofreading (Shimano, 1986). A mark shows the spot of correction, e.g., text insertion. All the marks are not required in our situation because documents are inputted with a computer. Based on this standardization, seven marks were selected through our experiments. Table 1 shows the seven marks; insert, replace, delete, separate, join, move and annotation. “join,” “separate,” and “move” marks are used over two sentences, and the others are used in one sentence.

Table 1: Marks for correction and CCML tags.

Correction	Mark	Example		
		Marked sentence	CCML sentence	Meaning in English
1. Insert		英語の勉強して ^を います。	英語の勉強<Insert level="3" string="を">しています。	Learning English.
2. Replace		僕のクラスメイト 僕のクラスメー ド	僕の<Replace level="3" string="クラスメイト"> クラスメー</Replace>	My classmate.
3. Delete		写真 に ありがとう	写真<Delete level="2">は</Delete>ありがとう	Thanks for photograph.
4. Separate		寒かった。今は	寒かった。<Separate>今は	It's cold now.
5. Join		をしています。私の仕事は、たのしい	をしています。<Join> 私の仕事は、たのしい	It's joyful in doing my work.
6. Move		① 運賃が人数で計算します。 ② [子供は大人の半額です。]	<Movefrom refid="1">運賃が人数で計算します。 <Moveto id="1">子供は大人の半額です。 </Moveto>	Child's ticket is in half, how much will pay is depend on the number of people.
7. Annotate		前半 [?] 試験	<Annotate type="question" string="何の試験ですか">	What is previous examination.

Based on SGML, CCML for exchanging marked-up documents has been proposed. SGML is an ISO standard (ISO 8879:1986) which supplies a formal notation for the definition of generalized mark-up languages. SGML is device-independent methods of representing texts in electronic form. We mean 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.

CCML document consists of three parts: header, body and close. A header part has an editor name, an author name, the revision time. A close part shows the editor's comments. A body part consists of the six marks and the annotations. Every tag has an attribute “level” that a teacher gives a number from one to three; 1:weak correction, 2:normal correction, and 3:strong correction. 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. In CCML, annotations are classified into three categories: explanation, question, and comment.

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.
- Needless to say, CCML inherits its features from SGML.

3.CoCoA2: On-line Markup Based Proofreading Exercise

3.1.Learning process

Correcting exercise can help learners understand reasons for writing errors. Exercise materials can also transform the learning experience and enhance learner motivation. In making full use of every corrected document and writing errors data processed, in CoCoA, the correcting exercise method is suggested to assist learners for enhancing their writing ability. Figure 2 shows the method of providing learner with some similar documents for learner's correcting exercises.

- (1) Learner A writes an original text and sends it to a teacher by e-mail.
- (2) The teacher corrects the text with CoCoA-Editor.
- (3) The teacher sends learner B the marked-up text by e-mail.
- (4) The teacher stores the marked-up text of learner B into the CCML document database.
- (5) CoCoA2 selects a suitable learner for correcting exercises using the marked-up text of learner A. Analyzing the document database, the system selects a candidate student who wrote texts containing similar errors to learner A. In figure 2, the system recommends learner B for the exercise.
- (6) Teacher sends learner B the CCML document of learner A, who can then modify those compositions while referring to teacher's correction before.
- (7) CoCoA generated the original text from the CCML text, and provides learner B the text that includes similar writing errors of learner A. Learner B corrects the text with CoCoA.
- (8) After learner B finished correction, CoCoA2 evaluates it through comparing the correction of the learner with that of the teacher. Moreover, learner B can be helped to re-correct the text with some advice of learner A who received the corrected text by the teacher.

In CoCoA2, it is very important how the system can support process 5 and 8. These designs are described in the following section.

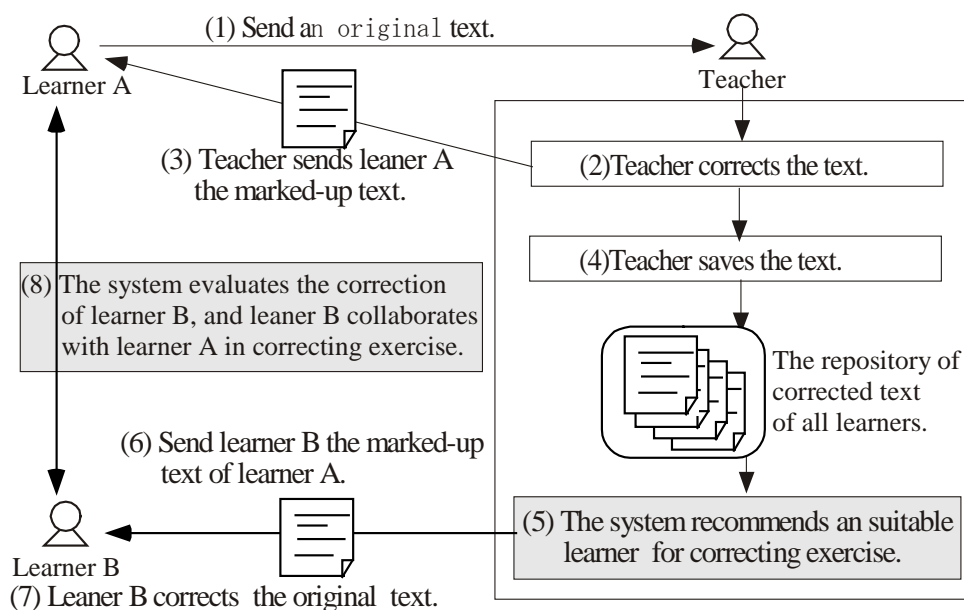


Figure 2: Learning process of proofreading exercise in CoCoA2.

3.2. Selecting a target learner for correcting exercise

In order to select appropriate learners for proofreading exercise, CoCoA2 analyzes the CCML document using natural language processing technologies (Feng, Ogata & Yano, 1998). Error types are identified by a morpheme analyzer called Chasen (Matsumoto, 1997). Then, by comparing original text with corrected texts, morphemes of writing errors can be gotten. Statistical analysis of these errors can help the system to find target learners who made similar types of errors to an original learner who wrote the text. Target learners can also refer to teacher correction. Using a TF-IDF *vector space model* (Salton et al., 1994), writing errors are classified. TF-IDF is one of the most successful and well-tested techniques in Information Retrieval (IR). A document is represented as a vector of weighted terms. The computation of the weights reflects empirical observations regarding text. Terms that appear frequently in one document (TF = term frequency), but rarely on the outside (IDF = inverse document frequency), are more likely to be relevant to the topic of the document. Therefore, the TF-IDF weight of a term in one document is the product of its term-frequency (TF) and the inverse of its document frequency (IDF). In addition, to prevent longer documents from having a better chance of retrieval, the weighted term vectors are normalized to unit length. With this method, CoCoA2 compares the number of each morpheme of the learner's documents with that of other learners' compositions, and CoCoA2 finds similar documents including the same errors as the learner's.

3.3. Evaluation of learner's correction

The system evaluates learner's correction as follows:

1. Standardizing the corrected documents;
2. Comparing correction of the respective document;
3. Scoring a point on the learner's correction.
4. Advising the learner about the wrong correction.

The standardization is needed to make it easy to compare the two corrections, because the learner may correct the text with the different marks from the teacher. The standardization is based on the following two rules:

- (1) Replacement rule: If there is an insert mark after delete mark, or if there is an delete mark after insert mark, then the two makes change one replacement mark.
- (2) Union rule: If the two same marks adjoin, they are united. For instance, the delete mark and the neighboring delete mark are combined into one delete mark.

After the standardization, the system compares the correction of the learner with the model correction of the teacher. Correction marks of two standardized documents are classified into the following states:

1. Right correction: Learner completely corrected the same words/sentences as the teacher.
2. Wrong correction: The correction of learner is different from that of the teacher. For instance, corrected words are different, or learner was not aware of writing errors.

In order to make it easy for learner to understand the evaluation result, the system gives a score on the proofreading exercise. The point of right answer (X) is proposed as follows:

$$X = \frac{\sum_{i=1}^n M_i}{C_i} \times 100(\%)$$

n is the total number of teacher's correction. M_i shows the correspondent number of correction between the learner and the teacher. C_i shows the number of the teacher's correction. The more similar to the teacher the correction of the learner, the higher X .

4. Implementation of CoCoA2

Teacher selects the document for proofreading exercise with. CoCoA2 shows the proofreading learner the uncorrected text that is generated from the CCML document. After finishing correction, CoCoA2 evaluates it and connects the learner and the original writer.

CoCoA2 is extended from CoCoA-Editor. Whenever learner wants to study about compositions, s/he can start a proofreading exercise with CoCoA2 and correct a text with (A). The text is generated from the CCML document that teacher sent to the learner. This system begins to evaluate the results of learner's correction by pushing "evaluate" button. The results are shown in (B). The right answer rate is 53% in this case, comparing the learner's correction with the teacher's correction. "Show answer" button displays the teacher's correction as an answer. "Right answer" button shows only the same corrections of learner as teacher's correction. On the other hand, "wrong answer" button displays the different corrections of learner from

teacher's corrections. "Advise" button shows helpful hints in (C), e.g. indicating the places to correct. "Collaboration" button creates a communication channel between the learner and the original writer of the document. The learner reviews and re-correct in collaboration with the original writer. They can communicate with an online chat tool or email. If the original learner agrees with the learner's correction, the exercise will finish.

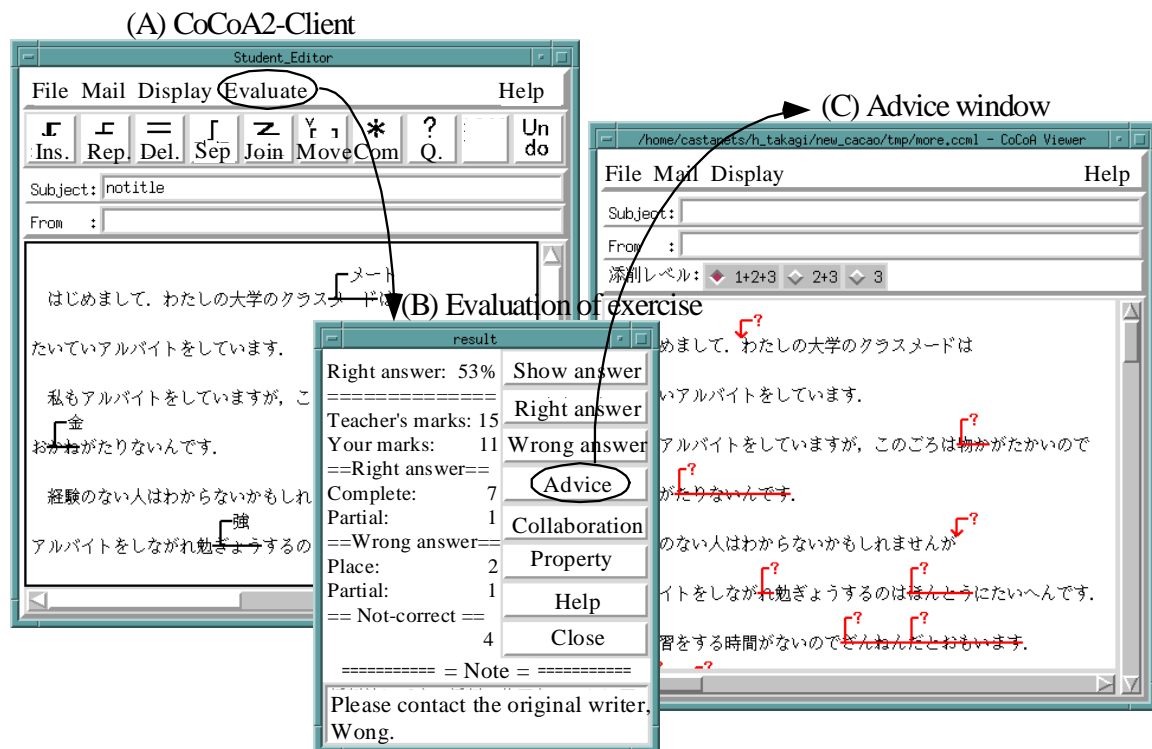


Figure 4: Interface of CoCoA2-Client.

5. Experimentation

To evaluate the effectiveness of the CoCoA2, an investigation was undertaken in an intermediate level Japanese class, with one Japanese teacher and six overseas students who contributed thirty-six compositions. In each of composition, there are about 400 words, the results of experimentation are:

1. CoCoA2 stored each learner's documents after teacher finished correcting each of documents. The average numbers of writing errors are eighteen in these documents.
2. Writing errors were classified into thirty-three morphemes. For example, learners' writing errors were in verbs (24.1%), nouns (21.5%) and postpositional particles (17.6%). According to these results, teacher could give leanness adequate advice and helpful comments.
3. System provided documents for proofreading exercise for each learner, according to morpheme type.
4. Students found out and corrected more than 85% writing errors. Most of learners commented out that correcting exercises were very interesting and effective for language learning, because of reviewing others' documents.

Under viewing the results of experimentation, CoCoA not only could provide teachers and learners with an environment for teacher correcting learner's composition, but also the CoCoA2 could recycle online marked-up documents by proofreading exercise and support to review and polish documents by collaborative learning.

6. Conclusions

This paper proposed a computer supported proofreading exercise environment called CoCoA2 where a teacher and learners can exchange electronic marked-up documents. Although we tried to use CoCoA2 in Japanese language learning, CoCoA2 is independent of a language domain. Therefore, CoCoA2 will be able to be used for learning any language. In the future work, we will evaluate CoCoA2 in detail during long term, and we will also investigate how to support multi-lingual language learning in a cyber space.

CoCoA2 have been implemented in Tcl/Tk on both Microsoft Windows 95/98/NT and Sun workstations. A Japanese version of this software is distributed to other researchers and developers for demonstration. For further information, our home page (<http://www-yano.is.tokushima-u.ac.jp/research/cocoa/cocoa.html>) is available.

Acknowledgment

This research was supported in part by the Grant from the Telecommunications Advancement Foundation and the Grant for Scientific Research (A) No.11780125, No.09558017, No.09480036 and No.11878032 from the Ministry of Education 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.
- Brock, N. M. (1995). "Computerized text analysis: Roots and research," *Computer Assisted Language Learning*, Vol.8, No.2-3, pp.227-258.
- Coniam, D. (1997). "A Computerized English Language Proofing Close 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 Trans. in Professional Communication*, 38(2), pp.110-117.
- Feng, C., Ogata, H. and Yano, Y. (1998). "Error Analyzer for Supporting Collaborative Japanese Language Learning," *Proc. of International Conference on Computers in Education*, Vol.1, pp.588-594.
- Hardock, G., Kurtebach, G., & Buxton, W. (1993). "A marking based interface for collaborative writing," *Proc. of the ACM Symposium on User Interface Software Technonogy*, pp.259-266.
- Herwijnen, E. (1990). *Practical SGML*, Kluwer Academic Publishers.
- Harasim, L., Hiltz, R. S., Teles, L., & Turoff, M. (1995). *Learning Networks -A field guide to teaching and learning online-*, The MIT Press.
- 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.
- 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). "JULIET: Interactive learning environment for Kanji compounds learning," *Proc. of ED-Media '96*, pp.539-544.
- Ogata, H., Matsuura, K. & Yano, Y. (1996). Knowledge awareness: Bridging between shared knowledge space and collaboration in Sharlok, *Proc. of Educational Telecommunications*, 232-237.
- Ogata, H., & Yano, Y. (1997). "Active knowledge awareness filtering toward efficient collaborative learning," *Proc. of AI-ED 97*, Kobe, Japan. (in press)
- Ogata, H. and Yano, Y. (1997). "CoCoA: Communicative Correction Assisting System for Composition Studies", *Proc. of International Conference on Computers in Education*, pp.461-468.
- Ogata, H. , Yano, Y., and Wakita, R. (1998). "CCML: Exchanging Marked-up Documents in a Networked Writing Classroom," *Computer Assisted Language Learning*, Vol.11, No.2, pp.201-214.
- Ogata, H. , Feng, C., Hada, Y. and Yano, Y. (1999). "Supporting E-mail Based Correction in Computer Assisted Language Learning," *Proc. of Int'l Conf. on Computer Processing of Oriental Languages*, pp.241-244.
- O'Malley, C. (1994). *Computer supported collaborative learning*, *NATO ASI Series, F: Computer & Systems Sciences*, Vol. 128.
- Salton, G., Allan, J. and Buckley, C. (1994). "Automatic structuring and retrieval of large text files." *Commun. ACM*, 17(2), pp.97-108.
- Scardamalia, M., & Bereiter, C. (1986). "Research on written composition," M. Wittrock (Ed.), *Handbook of research on teaching* (third ed.), Macmilian, pp.778-801.
- 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 ICCE '93*, pp.411-413.