An Intelligent Partner for Organizing a Paragraph

Hidenobu KUNICHKA\textsuperscript{a}, Chiaki MIYAZAKI\textsuperscript{b}, Yuto MATSUOKI\textsuperscript{b}
Robert CHARTRAND\textsuperscript{c} and Akira TAKEUCHI\textsuperscript{b}
\textsuperscript{a} Department of Creative Informatics, Kyushu Institute of Technology, JAPAN
\textsuperscript{b} Department of Artificial Intelligence, Kyushu Institute of Technology, JAPAN
\textsuperscript{c} Institute of Foreign Language Education, Kurume University, JAPAN

Abstract. This paper presents a support system for organizing an English paragraph for users who have insufficient knowledge of English paragraph writing. The system infers the user’s intention to compose a paragraph, organizes and proposes some outlines of a paragraph, and reports missing information which is necessary for completion of organizing the paragraphs. We have investigated the usefulness of the system by trying it out on subjects who have little knowledge of paragraph writing. Although several points need to be improved, the system was found on trial to be useful for such users.

Keywords. Paragraph Writing, Intelligent CALL Systems, English Learning

Introduction

Writing good English which is persuasive and easy to understand is one of the most difficult tasks for people who use/study English as a foreign language. For writing good English, they should be able to both write grammatically correct sentences, and construct understandable logical structures. In the case of the Japanese language, there is a basic unit of sentences, which is quite different from an English paragraph. Therefore, it may be hard for Japanese language users to apply their native writing skills to English paragraphs. In this paper, we focus on helping the user to construct logical structures, especially for paragraph organization, which is the basic unit of composition [1].

So far, many intelligent computer-assisted language learning (ICALL) systems have been developed. The majority of ICALL systems have aimed at improving a learner’s reading ability, or writing ability focused on grammaticalness [4]. However, there are few systems for supporting construction of logical structures. Criterion [2, 3] and Writer’s Companion [9] give users templates which correspond to typical structures of English paragraphs/essays. BEAR [6] aims to support writing abstracts for literature, and also gives templates. Narita and his colleagues have produced templates by investigating papers printed in well-known technical journals [6]. iWeaver [8] is not limited for learning use, and provides users with a comprehensive environment for gathering what to write and for writing sentences. Target users of these systems are basically people who have sufficient knowledge of English. Thus, these systems just provide environments or indirectly guide users. Moreover, these systems may not
provide enough support for users who have insufficient knowledge of structures of English paragraphs.

The target of our research is to provide support in paragraph organization for users who have insufficient knowledge of English paragraph writing. We think that it is useful for such users to have active and direct support, such as in providing the necessary knowledge or assisting in paragraph organization. This paper presents a paragraph writing support system as “an intelligent partner for organizing a paragraph”. The system is a tool and does not limit the topic of a paragraph. The system has knowledge of paragraph development, which is extracted from the findings in the field of language education. The system infers the user’s intention to compose a paragraph, organizes and proposes some outlines of a paragraph, and reports the missing information which is necessary to complete organizing the paragraphs.

In the following sections, we first describe processes of organizing a paragraph and a support method. Next, we introduce the knowledge of English paragraph development. After that, the implementation of our system and its evaluation are presented. Our conclusions are then given.

1. Organizing a Paragraph and a Support Method

The process of writing roughly consists of four steps: pre-writing, drafting, reviewing and rewriting. Organizing a paragraph is one of the sub-steps of pre-writing. In order to complete paragraph organization, users need to achieve four tasks. First, users decide a topic for a paragraph, gather sufficient ideas related to the topic, and refine them by adding, deleting or classifying their ideas. Second, the users choose suitable ideas which they want to use. Third, the users organize the selected ideas, and compose the outline of a paragraph. Finally, the users carefully reflect on if they should add any other information to each paragraph or if they should, alternatively, delete any ideas.

The difficulty of organizing a paragraph comes from following two kinds of various possibilities to achieve these four tasks. One is selecting ideas. Users need to select an appropriate main idea and supporting ideas from a wide range of ideas. The ideas are fact, opinions, and so on, while the roles of ideas which influence such selections vary depending on viewpoints corresponding to the purposes of the paragraphs. The other is composing the outline of a paragraph. The structure of a paragraph is not unique, and it depends on the purpose of the paragraph. Moreover, a user may stubbornly adhere to a specific paragraph organization, and not be able to think flexibly enough. It is difficult for users who are unfamiliar with paragraph writing to organize a desirable paragraph by considering a number of possibilities without any support.

For assisting in gathering ideas, cluster diagrams which consist of ideas and links are introduced in some learning materials [e.g., 7, 10]. In the diagrams, although users can manually link one idea to another to express relations between ideas, these kinds of relations are implicit. When users select ideas from gathered ideas or organize a paragraph by referring to such a cluster diagram, the users will become confused with the implicit relationships.

In order to solve this problem, we first propose a labeled cluster diagram, in which the relations between each set of two ideas are explicitly expressed as labels attached to the both ideas connected by a link. Moreover, the importance of the idea, the writing order and an optional free memo of each idea can be expressed as attributes of the idea.
The importance is expressed as a numeric number and used for indicating how much
the user wants to use the idea. The writing order indicates the partial-order among the
ideas, which have the same parent.

We also propose a support system for organizing a paragraph. By referring to a
user’s labeled cluster diagram, the system organizes and offers several outlines of a
paragraph from various viewpoints with missing information to complete each outline.
We think that this capability brings the following effects:

− Offering proposals stimulates users to carefully consider several possibilities on
  selecting ideas and composing outlines.
− Users increase their knowledge of English paragraph structures by seeing
different types of paragraph structures.
− Users complete their own outline to their satisfaction.

2. Paragraph Development Schemata

In order to realize the support system for organizing a paragraph as a partner, the
system needs to have knowledge of structures of paragraphs; that is, various kinds of
typical structures, components of each kind of structure, and restrictions of/among
components. We have, therefore, defined the knowledge of paragraph structures which
can be understood by computers. There are different types of paragraphs depending on
the purposes people want to write in the paragraph. We have examined fifteen text
books for paragraph writing [e.g., 1, 7, 10], and obtained ten types of paragraphs:
Listing, Example, Comparison / Contrast, Objective Analysis, Cause and Effect,
Definition, Classification, Opinion and Reason, Process and Direction, and Personal
Description. Although a paragraph basically has three parts: an introduction part, a
supporting part and a concluding part, the details vary according to the type. Each type
of paragraph has a typical structure. Also, there are words and phrases frequently used
in a particular type of paragraph. In order to allow computers to treat them, we have
formally defined “paragraph development schemata (in short, PDS)” [5] for the ten
types of paragraphs. Moreover, PDSs not only have information for computers, but also
for humans to compose paragraphs.

A PDS consists of the following five kinds of information.

− **Schema**: The structure of the paragraph. Each schema in PDS consists of what to
  write and the order.
− **Explanation**: What should be described in each component of the schema.
− **Tip**: Important points for writing a good paragraph and/or matters that require
  attention.
− **Words and phrases**: Information on words and phrases frequently used in a type of
  paragraph. The information consists of words, the part of speech, the component
  where the words are used, and restrictions in use.
− **Dependence**: Dependence among the components in a paragraph.

As an example, we present the *schema* and *dependence* for Comparison / Contrast
paragraph which can be used for describing the similarities between topic A and topic
B, the differences between topic A and topic B, or the advantages and disadvantages of
a topic. Figure 1 shows the *schema* and *dependence*. In the figure, 0, 1, * and + express
the number of repetition. * and + mean “zero or more” and “one or more”, respectively.
There are two types of *schema*: one is Block organization and the other is Point-by-
Point organization. Selecting one of the schemata depends on the number of items or the aim of a paragraph. For example, when many items are described, Point-by-Point organization is suitable. Dependence expresses such relationships.

3. Paragraph Organization Support System

This section describes how to realize the following four support functions of our system: helping to draw labeled cluster diagrams, selecting the appropriate schema and ideas, organizing and proposing the outlines of a paragraph, and reporting missing ideas to complete the outline of the paragraph. In order to encourage users to consider various possibilities for organizing a paragraph, the system gives them the three top-ranked proposals.

(1) Helping to Draw Labeled Cluster Diagrams

As described before, the system provides users a tool to draw labeled cluster diagrams¹.

¹ Users can freely draw diagrams, but only tree diagrams can be accepted for other support functions at present.
Figure 2 shows an example of a diagram. A labeled cluster diagram consists of ideas, links, labels, importance, order to write and memo.

A user chooses a label for an idea by selecting from a list of labels. Labels correspond to the names of components of schemata. Table 1 shows all the labels. Some labels have an “is-a” relationship; e.g. “Item (reason)” is a label as subclass of “Item”. Users are expected to choose more specific labels if possible because such labels may concretely express what they intend.

<table>
<thead>
<tr>
<th>Part</th>
<th>Labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction part</td>
<td>Introductory sentence, Topic sentence, Topic sentence (definition / cause / opinion / on topic A and topic B)</td>
</tr>
<tr>
<td>Supporting part</td>
<td>Item, Item (example / on topic A / on topic B / viewpoint / support component / effect / category / reason / process / information gotten by five senses), Explanation of the item, Explanation of the item (example / on topic A / on topic B / viewpoint / support component / effect / category / reason / process / information gotten by five senses), Transition, Word for enumeration</td>
</tr>
<tr>
<td>Concluding part</td>
<td>Concluding sentence</td>
</tr>
</tbody>
</table>

Table 1. Labels for ideas

More concretely, the system searches the three top-ranked schemata which match a labeled cluster diagram and the system searches for ideas which fit each schema as follows. (a) The system adds the weight and importance of each label expressed in the user’s labeled diagram to the score of each schema in which the label is used. The weight of each label has been decided depending on the frequency of the label in all the schemata through an experiment. Table 2 shows the weight for each label. (b) The system selects the schema which received the highest score as the best match; the selected schema is used by another function “organizing and proposing the outline of a paragraph”. (c) The system picks up ideas whose labels appear in the selected schema as “available ideas”. Here, we call available ideas connected by links, “group”. (d) The system finds the biggest “group”, and identifies the “group” as the ideas to be proposed. (e) The system repeats from (b) to (d) twice more.

Figure 3 shows an example of selecting the first place of schema and ideas from a labeled cluster diagram. The system has calculated the score of each schema, and here we assume Schema 2 received the highest score. In this case, the available ideas for

<table>
<thead>
<tr>
<th>Weight</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Introductory sentence, Concluding sentence</td>
</tr>
<tr>
<td>2</td>
<td>Topic sentence</td>
</tr>
<tr>
<td>1</td>
<td>Transition, Word for enumeration</td>
</tr>
<tr>
<td>0</td>
<td>Topic sentence (on topic A and topic B), Item, Explanation of the item (on topic A, on topic B)</td>
</tr>
<tr>
<td>6</td>
<td>the others</td>
</tr>
</tbody>
</table>

Table 2. Weights of labels
Schema 2 are Ideas 3, 4, 5, 6, 8 and 9. According to the connection between the ideas, there are three groups in the available ideas, that is, [Ideas 3, 4, 5, 6], [Idea 8] and [Idea 9], and the biggest group is the first one. The group [Ideas 3, 4, 5, 6] is finally selected.

(3) Organizing and Proposing the Outlines of a Paragraph
In order to stimulate users to seriously consider proposed outlines, the structure of each outline should be based on the knowledge of paragraphs. Offering such outlines to users who have insufficient knowledge will also encourage them to acquire new knowledge, or to be aware of incorrect or missing knowledge that users have. The system, therefore, organizes outlines of a paragraph by referring to schemata, which are extracted from the findings in the field of language education, and proposes the outlines as follows. The system first puts each idea of the selected group onto the place of the appropriate component of a selected schema by considering the order to write each idea if it is set. That is to say, when there are ideas attached to the same label and the order of each idea is set, the system put them in the set order. Following that, the system shows the schema including the ideas to a user as a proposal. In this way, the system gives users the following proposals: a schema which corresponds to a point of view, ideas which should be described in a paragraph, and the outline of the paragraph.

(4) Reporting Missing Ideas
A user may miss necessary ideas to complete an outline or attach incorrect labels to ideas in the user’s labeled cluster diagram. In order to solve the problems, the system reports missing components of a schema. As mentioned before, the system shows a selected schema in which all the selected ideas are put in suitable components of the schema. Components in which no ideas are put are missing components. Though optional components may also be reported, users can recognize them by marks indicating the number of repetitions of components mentioned in the second section. As a result, after the user considers if the missing components in the schema are really necessary, the user will add new ideas or attach another label to ideas to complete the user’s outline.

4. The Usefulness of our system
Giving a user outlines of paragraphs, which reflect the user’s intention, aims not only to (1) stimulate the user to consider the outlines carefully, but also to (2) teach some knowledge of paragraph structures. We, moreover, expect that the user will (3) finally complete his/her own outline with satisfaction. We have investigated these three points.
through trying out the system on subjects who have little knowledge of paragraph writing.

We have received feedback through questionnaires. The subjects are thirteen graduate and undergraduate students of our institute. All the subjects answered they had little knowledge of paragraph structures. More concretely, one subject knows only the names of some paragraph structures but not more specifically, six subjects know only the existence of paragraph structures, and six subjects do not know anything about this topic. We asked them as follows; to write a labeled cluster diagram using the system, to write an outline without using the system (We call the outline “outline A”), to answer a pre-questionnaire, to get proposals of outlines from the system, to refine outline A if necessary and complete it as the final outline (We call the outline “outline B”), and to answer a post-questionnaire. Table 3 shows investigation points of each questionnaire, the number of questions for each point, and an example of a question. The subjects were asked to answer most of the questions in a five-point Likert scale ranging from 5 for most positive to 1 for most negative. It took about two hours, including instructions on how to use the system.

Table 3. A part of the contents of the questionnaires

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Investigation point</th>
<th>Number of questions</th>
<th>Example of question</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre-questionnaire</td>
<td>(1) Careful consideration to outlines</td>
<td>1</td>
<td>Did you compose and compare more than one outline, and select one of them?</td>
</tr>
<tr>
<td></td>
<td>(2) Existence of knowledge</td>
<td>1</td>
<td>What types of paragraph structures do you know in detail?</td>
</tr>
<tr>
<td>post-questionnaire</td>
<td>(1) Appropriateness to the diagram &amp; Careful consideration to outlines</td>
<td>3</td>
<td>Did the proposed outlines reflect your labeled cluster diagram?</td>
</tr>
<tr>
<td></td>
<td>(2) Knowledge acquisition</td>
<td>3</td>
<td>Were you aware of some components of lacking in your outline?</td>
</tr>
<tr>
<td></td>
<td>(3) Satisfaction at the final outline</td>
<td>1</td>
<td>Have you satisfied the final outline?</td>
</tr>
</tbody>
</table>

As for the result of (1) in the post-questionnaire, the average score of the reflection of each subject’s labeled cluster diagram in proposed outlines was 4.0, and twelve subjects (92%) answered that the first place of the proposed outline is the best. We can say that the system can propose good outlines. We think that one of the main reasons is using related ideas and their importance expressed in a user’s own labeled cluster diagram. The subjects who answered proposed outlines did not reflect their own labeled cluster diagrams had incorrectly attached labels to ideas in the diagrams. We, therefore, need to consider a method to support attaching labels to ideas. Next, although four subjects (31%) composed and compared more than one outline when subjects wrote outline A as the result of the pre-questionnaire, we have found that all the subjects considered and compared outline A and proposed outlines after being offered the outlines. Thus, we can say that offering proposals stimulates users to carefully consider outlines.

As for result of (2), the average score of acquiring knowledge of paragraph structures was approximately 3.5. More specifically, seven subjects have acquired new knowledge, nine were aware of the incorrectness of their own knowledge, and seven were aware of that a part of their knowledge was lacking. We have discovered that almost all subjects have increased their knowledge. We, moreover, need to confirm if they keep acquiring knowledge even in the future. Only one subject did not answer that their knowledge has increased. We expected that subjects can indirectly learn paragraph structures by being offered reasonable outlines, but we also need to equip the system with a persuasive instructional function on paragraph structures.
Lastly, the average score of satisfaction of outline B was approximately 4.2 as the result of (3). We can, therefore, say that almost all subjects composed satisfactory outlines because they were enabled to carefully compare and consider outlines.

5. Conclusions

We have described a method to help in composing an outline of an English paragraph for users who are unfamiliar with paragraph writing. Our system proposes possible outlines of a paragraph with a part of the ideas gathered by users. Though several points need to be improved, the system was found on trial to be useful for users who have little knowledge.

The system proposes outlines of paragraphs; it should be easily extendable for larger and more complex units. For example, it will be usable for composing sections. The organization of a section is very similar to that of a paragraph [7]. First, it is necessary to define the knowledge of section development. After that, allowing users to write main ideas of paragraphs into a cluster diagram, the system will be able to provide support for organizing sections.

As future tasks, we need to consider a method to support selecting appropriate labels for each idea, investigate if users keep acquiring knowledge even after a period of time, and realize a persuasive instructional function on paragraph structures. In addition, we have a plan to extend the system for organizing sections, chapters, and so on. For this task, it is necessary to define schemata for section or chapter development.

References