

# The Text Viewer System with XML to Assist Reading Comprehension of Foreign Pupils

Keiko Nakao<sup>†</sup>, Jun-ya Morishita<sup>††</sup>

Many foreign pupils need assistance in reading a Japanese schoolbook. We propose a system for assisting a reading comprehension of foreign pupils in Japan. The schoolbook is composed in blocks. Through finding out the causal relationship between the functional blocks, we realize the visual appeal with “question and answer” blocks in the context of schoolbook. We use XML data format to describe a structures of the schoolbook and we use Xpath to retrieve both pattern matching for sentences and node corresponding to question and answer blocks in semantic structure. We are confident our system must be helpful to assist a reading ability by means of reducing the information amount to foreign pupils.

## 1. Introduction

For these 10 years, many foreign children came to Japan with their parents who are Japanese descent. There are much more than 20,000 pupils now in school in Japan. Since the pupils don't know Japanese language, they have a lot of problems in school life in Japan. The academic Japanese competence is a biggest problem in school life to learn school subject for foreign pupils. Although it becomes easier to communicate in Japanese for them soon, it is still difficult to understand the concept in schoolbook. And even if in spoken language in the class activities, the teachers talk in written academic Japanese of schoolbook to lead pupil's thought as Ishii's study says[1]. So as to understand the new concept of matters through the academic activities in both cases of written and spoken Japanese, they must prepare to know the words to express the abstract concept for grasping the outline of the key concept of the schoolbook before the class activities. However, foreign pupils do not have the ability of reading comprehension enough either in own language or in Japanese, since many pupils came to Japan before

developing their reading comprehension in their mother tongue. So as to understand the abstract concept, their Japanese language teacher

<sup>†</sup>Faculty of Socio-Cultural Studies, Otemae University

<sup>††</sup>Graduate School of Cultural Studies and Human Science, Kobe University should develop their reading skill in supplemental lesson of Japanese or their mother tongue.

But there is hardly helpful material to assist their preparation and developing reading skill. It is very important and so expected to assist their reading skill by pointing the key suggestion and learning point in schoolbook.

In this paper, we propose a text viewer system to emphasize the stream and the point of semantic context in schoolbook, for assisting foreign pupil's reading skill to skim.

We use XML to describe the structure of schoolbook[2]. We defined a minimum unit of text as “block”. It is a node of the structure of schoolbook[3]. We found the function of some expressions that forms a learning process as a semantic structure[4]. We make use of this block

and the functional expressions to emphasize the learning point in the semantic learning process. It is identified the learning process with Xpath, marking these expressions and the relation between “Question and Answer” blocks.

We discuss about searching on the semantic structure that is realized with XML and Xpath[5] [6] [7]. We discuss it in order of (2)Data structure, (3)Measures of visual appeal, (4)Searching method, and (5)Concluding remarks.

## 2. Data structure

For visual appeal of learning point in schoolbook, we use these 2 kinds of structures in a schoolbook. One is a physical structure as a book, and the other is a semantic structure according to the text context. We adopt the following 2 steps: At first we find out the important part in text semantic context, and next using the result that is found in text context beforehand we emphasize the position on the physical structure as book.

For discussing the system, we explain about the structure of schoolbook first.

Here in this section we discuss in this

order below. (1) A schoolbook in Japan, (2)Physical structure description in XML, (3)Semantic structure description in XML, (4)Relation between the physical blocks and semantic patterns of words, (5)Acquisition of the abstract concept in “Answer” block.

### 2.1. A schoolbook in Japan

The fig.1 is an example of schoolbook which is for 11~12 years old in Japan[7]. The recent schoolbook has rich and complicated structure with many photographs, pictures and small units of text stream in any kind of subject. The schoolbook is emphasized with a lot of color and marks putting on the physical structure as section, subsection and text unit. And it is often put one point advice around the text unit to lead into the important concept with some mascot characters.

But these effectiveness with color, mark and mascot character are vague hints that shows the existence of important concept. This effectiveness is not related with learning point buried in text stream directly. The concept of the words and sentences make a form of a



Fig.1 an example of schoolbook in Japan

story as the semantic context.

In order to get hold the buried learning concept in semantic context from the text stream, it is necessary to skim the key sentence patterns and the words that lead into the important learning point. The key pattern means the suggestive sentence and the answer of that suggestion. These make a stream of learning process in the text stream. The suggestive sentence can be clear with the ending form of it. It is treated in 2.3.

Foreign pupils have less ability of Japanese language especially for reading. It is difficult for them to grasp the buried stream in the text part under the many emphasizing effectiveness in a mess. Our system supports them to lead into the semantic context marking the suggestive sentence and it's counter part, corresponding to an Answer of assignment.

## 2.2 Physical structure description in XML

The fig.2 is a model of one page of a natural science schoolbook. "Block" on the picture means a minimum text stream. We defined that it is an elementary unit of the text part in schoolbook. These blocks

make one class on physical structure of schoolbook and it can be under the class of section, subsection and itself, too. As the fig.2 shows, the text part as "block" is in places on each class of the physical structure. The block forms a part of physical structure of schoolbook. And it is a different unit from the semantic unit that consists a stream of semantic context. We realize this physical structure of schoolbook, putting the tags for each classes. The tags are "textbook", "section", "subsection", "subsubsection", "block", "paragraph", "sentence" and "word". These tags show the class. In other words these classes indicate a constituents where the elements of schoolbook place like photos, sentences, words and etc.

## 2.3 Semantic structure description in XML

Any kind of text has a story. And a story has a consequent stream. Even if the explanation of the fact, it is logically correct. The schoolbook has also a logically story for leading pupils into learning concept. The schoolbook indicates a learning process under it's text stream. It can be called an intended

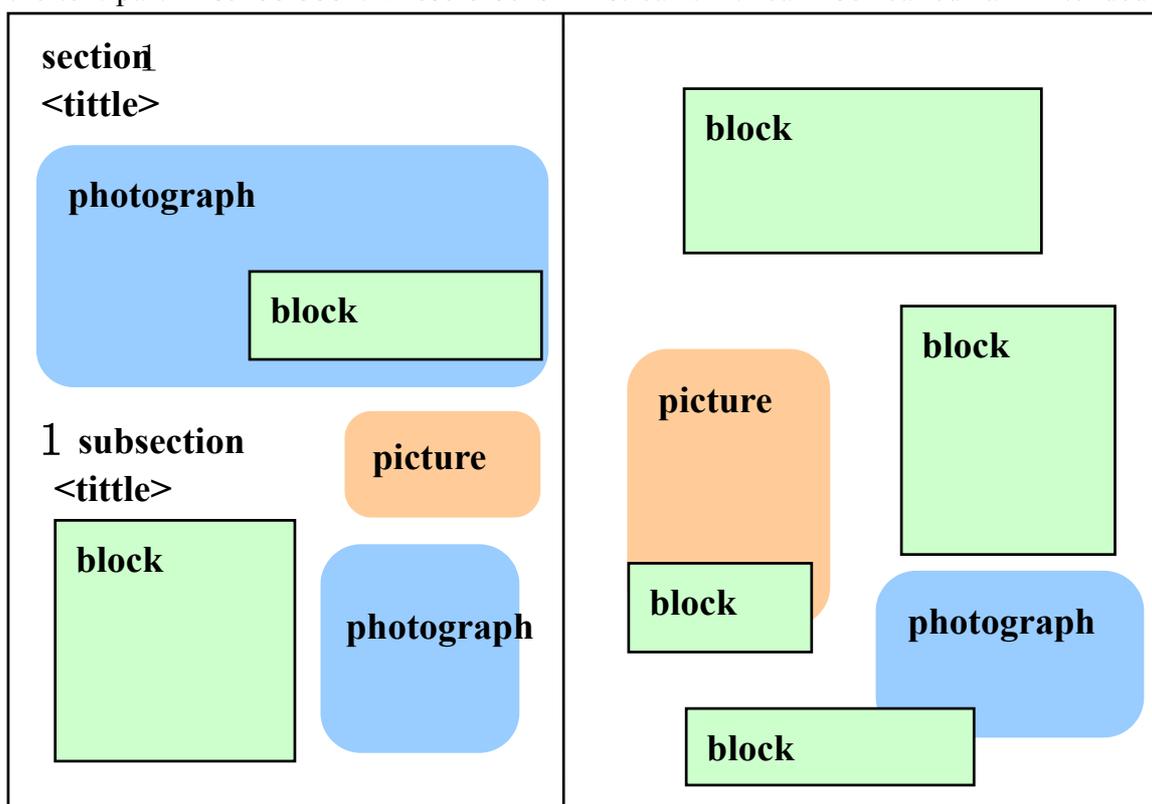


Fig.2: a model of schoolbook

learning process on schoolbook. It is necessary to grasp the key to lead into the learning concept, for understanding the context of text stream which indicates the learning process behind.

sentence and the block which contains these functional sentences and give an structure of the learning process with that functional meaning. This structure for context can be called as a semantic

1	instruction-suggestion	だろう。／だろうか。
2	instruction-order	なさい。
3	instruction-question	か。
4	instruction-lead	・う。
5	remark-warn	なければならぬ／てはいけない
6	remark-sumup	動詞＋ます／ました／ません／ませんでした
7	remark-find	動詞＋た／形容詞＋た／名詞＋だっ＋た
8	remark-explain	動詞／名詞＋だ／名詞-形容詞語幹＋だ／形容詞／名詞等＋である
9	remark-advice	よ。／ね。／わ。
10	remark-lead	(と)よい。

Fig.3:subjective expressions in schoolbook

The schoolbook mainly states about the fact in the world or disserts the fact in the world. But there are a lot of subjective expressions that shows the speakers attitude in addition to the tentative base. These subjective expressions are called as modality expressions in the philological study. And these expressions has specific ending form indicating advice, leading, suggestion and warning . We found that some of these subjective expressions work as the instruction and something intention for making a learning process of schoolbook. So, we define the function of those expressions according to it's work in our previous study in 2000, as "instruction-suggestion", "instruction-lead", "instruction-order", "instruction-question", "remark-warn", "remark-sumup", "remark-find", "remark-explain", "remark-lead" and "remark-advice" indicated in the fig.3.

We applied these definitions to the

structure by construct of the physical structure as a book. XSLT searches the specific form of these functional sentences on the semantic structure by the search according to the . And it emphasizes with the color the suggestive sentence and the answer sentence on the context stream through searching .

#### 2.4 Relation between the physical blocks and the semantic patterns of words

The physical structure is emphasized with some effectiveness like coloring on the schoolbook. But the physical structure is unrelated with the learning process in context. However, there is a key on the block of physical structure which corresponds to the semantic process. We make use of these relations between the semantic structure and the physical structure for identifying the position on

them. Marking the function of suggestive sentence and the concrete form of sentences, it is possible to realize the semantic structure on the physical structure visibly.

**2.5 Acquisition of the abstract concept in “Answer” block.**

We characterized a block according to its ending form of sentences. And we look over the appearance pattern of the characterized block. And we made clear the learning process in schoolbook and also we found that there are 2 types of the function of the instructive expressions from the result of looking over the pattern as the fig.4 in our previous study in 2000.

This fig.4 shows the beginning of one

section until the end of it. The blocks are in places here and there in section. But the blocks are divided into 2 types in suggestive instruction and in

supplemental instruction. When we look over the 2 types separately, we found that the suggestive blocks make a process of learning in subsection. These blocks form a stream in sequence as the fig.4 shows. The semantic characters of suggestive block are in order of “story”, “instruction”, “experiment” and “story”, or “summary”. And the block “Story” following with the block “Instruction” shows the learning point of that subsection.

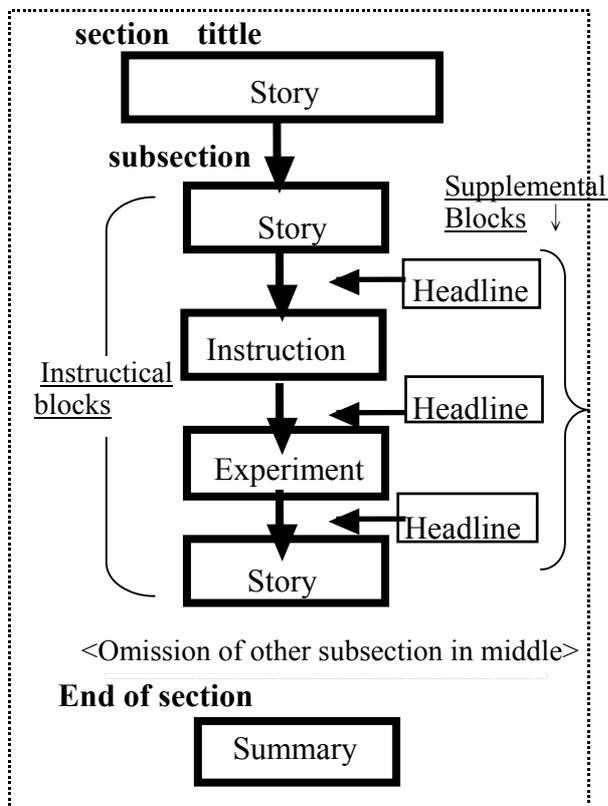


Fig4. :The blocks in section

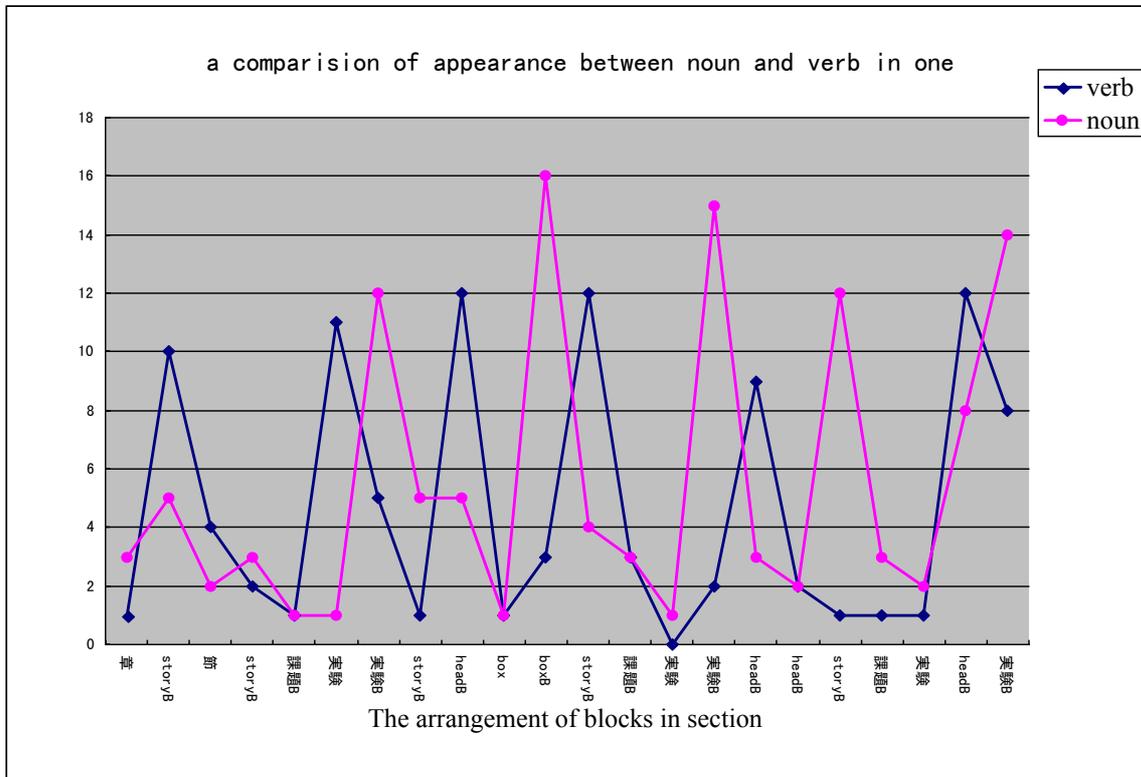


Fig.5:the appearance of the noun and verb in one section

We also reconfirmed the relevancy between the blocks at the word class in our previous study in the year of 2001. We found that there is a pattern of appearance between nouns and verbs based on the semantic structure of the schoolbook as the fig.5. Especially the block “suggestion” and the following “story” correspond to “Question and Answer”. The learning point of subsection bring up with the words in blocks of “Answer” as counter part of

“Question”. And we can say that the words in “Answer” block shows abstract concept of the learning point of that section or subsection as the fig.6.

We make use of these relations in context based on the semantic structure for pointing out the learning point on the physical structure on our system.

### 3. Measures of visual appeal

Making use of the characters of schoolbook that is cleared by previous

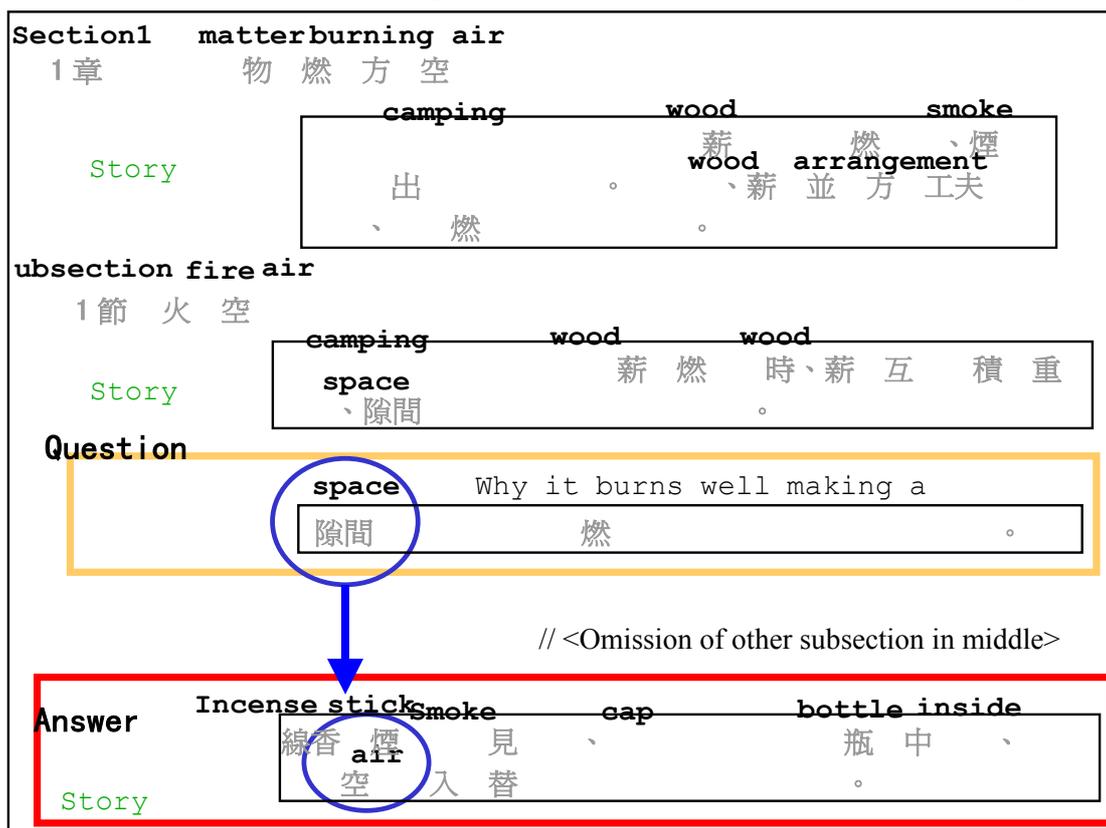


Fig.6: Relation between the blocks as "Question and Answer"

our study, we made a system to assist the reading comprehension for foreign pupils. Here we explain the construction of our system.

### 3.1. Process of getting the semantic structure

The searching target is the learning point in context. It is the indicator of semantic structure in the management of system that the searching target. And according to the result of search on the semantic structure, it should be reflected to the physical structure for pointing the part of Question and Answer of the learning process. In this system we realize the pattern matching, node searching and reflecting the result of both searching with XML to the physical structure in order of below.

1. XSLT(Xpath) searches for the indicator of the semantic structure in the text data.
2. XSLT annotates the results of searching to the text data.
3. XSLT gives the emphasized view to the text data

### 3.2. Process of searching and reflecting

The text data is classified in “section”, “subsection” and “block”. A book structure is composed of these classes for placing the contents in it as the fig.7 shows. Under the class of “sentence” in fig.7, we give a semantic structure through the morpheme analysis system Chasen 2.2.8[9]. So, it is separated to words under the class of “sentence”. And each word has attribute of semantic function as the fig.7 shows through the morphological analysis. The searching process is in this turn below.

1. When user selects the condition, XSLT searches the expressions as indicator on the word class. After finding it, XSLT goes to defined the parent block which the expression belongs to.
2. When XSLT recognize the parent block, continuously it goes to find out the “Answer” block which follows after the starting block.
3. And XSLT gives emphasizing tags to the words in these blocks.
4. In case of that the user need supplemental suggestion, XSLT

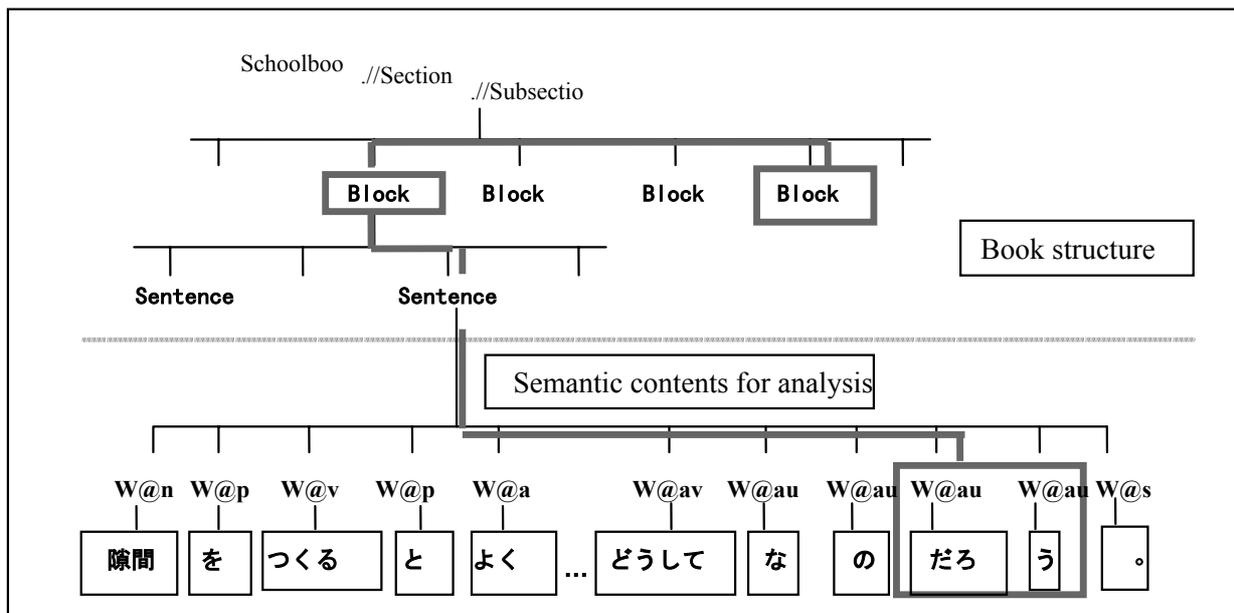


Fig.7:Classifier of text structure

gives emphasizing tag only to the words on the supplemental sentence. Because supplemental suggestion block is unrelated with another block.

### 3.3. Text data on the browser

The user looks the text data with visual appeal through the browser as the fig.8 shows. The fig.8 is an example of searching result of the viewer.

The grade of schoolbook, the text data and the search condition are chose from pop-up list beside. The structure is realized with nested quadrilateral.

According to the user's want, the color of blocks changes into yellow or red. These colors show the relation between the blocks and also important part in the subsection

semantically. When the condition is suggestive expression, the color of blocks "Question and Answer" change into yellow and red respectively. And also, the size and the color of nouns and verbs change into blue and green respectively in these blocks. If the condition is supplemental expression, only the nouns and verbs on the sentence of that expression is emphasized with the bigger size and the color of blue and green.

### 4. Effectiveness of our applying to schoolbook

We constructed the text viewer system using the text data which is structured with XML and is incorporated the result

of morphological analysis as the attribute. This system realizes the view appeal to text data with the searching the indicator of the learning process and the transformation of tags on text data. The result of searching by Xpath formula from the text data is transformed to another formula by XSLT for view appeal. Making use of Xpath, it becomes possible to search the pattern matching according

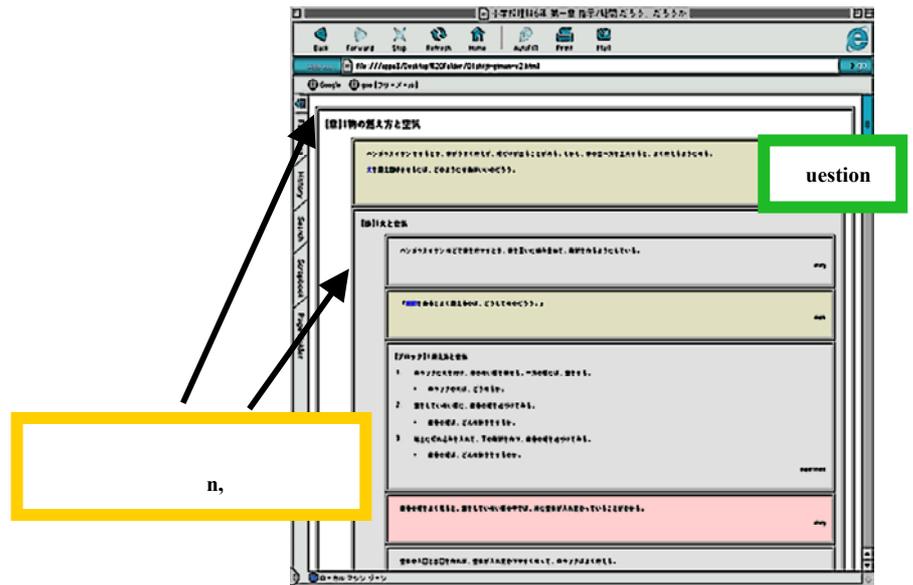


FIG. 8. EXAMPLE OF SCHOOLBOOK ON BROWSER

to the semantic function of the contents and to refer the physical structure of schoolbook at same time.

We made use of an aspect of annotating function of XML to manage our data. The annotative description corresponds to mark-up depends on the contents like Hyper Text. This description is like an "<A >" tag, a "<B>" tag for decorating, an "<I>" tag or an "<E>" tag of HTML. Giving the annotation tag according to contents is giving a particular "view" to data. We made use of conventional Well-formed XML in order to our study for giving a particular view depends on the special needs of foreign pupils.

The mark-up language like XML

possess these function as the description of structure and annotation. There are a lot of proposition of usage and improvement to the description in XML.

But there is few proposition of the annotative description. However, the usage of putting a structure to another structured data is high productivity. This aspect of XML is able to make up the problem of tag collision. We try to examine the usage of annotative description more after this for data holding jointly.

## 5. Conclusion remarks

We proposed a view system for assisting reading comprehension by searching the important blocks of context based on the semantic structure. We selected XML with Xpath to realize both the pattern matching for sentence and the node searching corresponding to "Question and Answer". It should verify using actually, but we are confident our system must be helpful to improve reading ability of foreign pupils. And we would like to make a profound study for apply annotating with XML.

## References

- [1] Ishii,E., *Kyoshitsu Danwa no Fukusu no Bunmyaku*, "Nihongogaku" , Meiji-shoin,1997
- [2] Nakao,K.,Morishita,J., *A Method for Structuring Context with XML and Morphological Analyzer*,IPSJ Symposium Series Vol.2000,No.17,pp,177 ▪ 184, 2000.12.16
- [3] Nakao,K.,Morishita,J., Text database for textbooks of primary school to probe the sequential patterns of the instructical expressions and it's analysis with XML,IPSJ Symposium Series Vol.2001,No.18,pp,173 ▪ 180, 2001.12.15
- [4] Nakao,K.,Morishita,J., An analysis of the instructical expressions in the

textbooks of primary school probing the text construction for the acquisition of reading skills, Cross-Cultural Study 4,pp,133 ▪ 151, Kobe Univ.Cross-Cultural Society, 2001.5

[5] <http://www.w3c.org/XML/>

[6] <http://www.w3.org/TR/xslt.html>

[7]

<http://www.alphaworks.ibm.com/tech/LotusXSL>

[8] Ooki.Michinori, 33  
others.,*ShinteiRika6*,vol1,pp.20-21,S  
hinko Shuppansha-Keirinkan

[9]

<http://chasen.aist-nara.ac.jp/index.html>