Evaluation of Reading Support Tools by Reading Comprehension Tests and Reading Speed Tests

Katsunori KOTANI[1], Takehiko YOSHIMI[2]
Takeshi KUTSUMI[3], Ichiko SATA[4], Hitoshi ISAHARA[5]
[1]*National Institute of Information and Communications Technology, 3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0289, Japan*
kat@khn.nict.go.jp
[2]*Ryukoku University, 1-5 Yokotani, Setaoe-cho, Otsu-shi, Shiga, 520-2925, Japan*
yoshimi@rins.ryukoku.ac.jp
[3]*SHARP Corporation, 492 Minosho-cho, Yamatokoriyama-shi, Nara, 639-1186, Japan*
kutsumi.takeshi@sharp.co.jp
[4] *SHARP Corporation, 492 Minosho-cho, Yamatokoriyama-shi, Nara, 639-1186, Japan*
sata@isl.sharp.co.jp
[5] *National Institute of Information and Communications Technology, 3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0289, Japan*
isahara@nict.go.jp

Abstract

This paper introduces our reading process monitoring systems and also presents the experimental results that show the adequacy of our reading data. Our system divides a text into reading areas and records reading time for each area. We conducted two experiments using this tool to verify the adequacy of our reading process data. In the first experiment, we examined whether the reading process can distinguish easy text reading and difficult text reading, and confirmed the adequacy of our reading process data. In the second experiment, we tried to evaluate efficiency of reading support tools such as (i) chunker, (ii) glosser, and (iii) machine translation system, assuming that efficiency of these systems co-relates with text readability. The experimental results show that only the machine translation system effectively supports reading.

Keywords: *reading process, reading support tools, evaluation of machine translation*

1 Introduction

This paper describes our two types of reading process monitoring tools. These tools share the common properties to display a text on a screen and to record reading time for each reading area. Both tools divide a text into reading areas, and one of the differences can be seen in the size of area. While one tool presents a fixed size area, the other can change the area size, depending on a research question. A reader reads a text by accessing each reading area, and the tool records (i) reading position and (ii) reading time for each position.

Our reading process monitoring tool enables us to collect reading process data easily, compared with some other reading process monitoring systems such as eye movement tracking systems or ERP/fMRI, which record brain activities. Although these systems themselves are expensive and require highly controlled experimental circumstances, our tool requires only a PC system.[1] In addition, there is no need to wear some special equipment in using our tool, which means we can observe normal reading behavior.[2]

Since the measuring range is relatively large, our tool is useful for observing text reading processes. By contrast, an eye-movement tracking system takes a small range as a scope, and hence it is effective in watching character/word recognition. The same is true with some brain activity recording systems. These minute transition data are too complex to observe text recognition. Thus, we analyze text-level reading processes on the basis of our reading process data.

In the following section, we will introduce our reading process monitoring tools. In the third section, we will present two experiments in which we measured reading process in order to examine text readability. If our reading data correctly show the readability, the adequacy of the data should be justified. It is confirmed from the first experiment, because the reading data can show different readability between easy text and difficult text. In addition, we examine whether our reading data can evaluate the efficiency of reading support tools such as machine translation systems. The result suggests that our reading process data can be applied into the evaluation of these support tools. In the final section, we summarize this paper and point out the remaining problems.

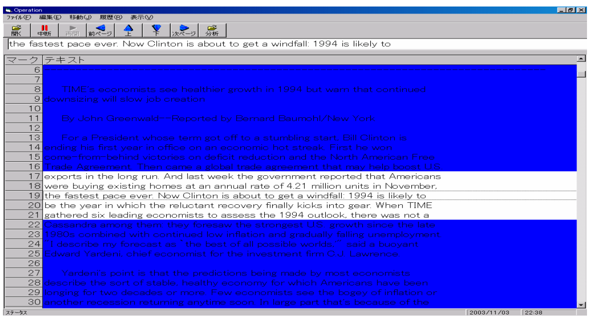
2 Reading Process Monitoring Tool

2.1 Linguistic Effect on Reading Patterns

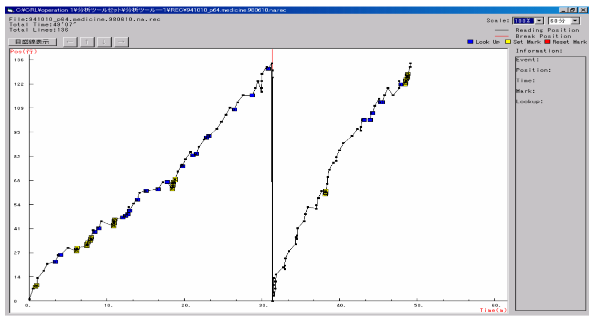
It is well-known that reading shows several patterns such as multiple reading, increase/decrease of reading speed (Bazerman (1988), Dillon (1994), Horney & Anderson-Inman (1994)). We suppose that these patterns occur partially because of linguistic properties. For instance, a reader may read the same part again and again if he/she finds difficult words/phrases to understand the meaning. In order to depict these reading patterns, we developed a reading process monitoring tool.

As shown in the following screenshot (i), a text is divided into two parts. The white background part is a reading area, and the colored background is a non-reading area. The reading area is arbitrarily determined as five lines in a text. In the experiment it is allowed to read the white part, but it is prohibited to read the colored part. The non-reading part remains visible so as to help a reader find where to read back. Thus, the reader can freely read forward or backward by moving the reading part, while the tool records the reading time and position.

The tool not only records the reading data, but also displays a graph of full reading process as shown in the screenshot (ii). The vertical axis shows the reading time, and the horizontal axis represents the reading position. The graph shows the whole reading process for a text. That is, it shows that the reader starts from the opening line of a text, and re-reads the text after reaching the ending line. Thus, this reading has been completed by reading twice. According to the graph below, it takes the less time to finish the second reading.



Screenshot (i) for Display Screen



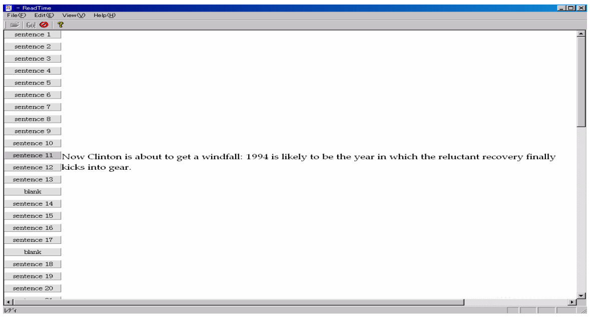
Screenshot (ii) for Reading Process Graph

By analyzing these reading process data, we find the following linguistic effects on reading patterns. First, a reader tends to start backward reading when a reader finds (i) a paragraph break, (ii) summary expression, e.g., *in sum*, and (iii) discourse signal words, e.g., *The plan…* (see Kotani et al. (2003a)). Secondly, the reading speed may increase if a reader sees discourse markers such as *But* (see Kotani et al. (2002); Kotani et al. (2003b)). Finally, the speed may decrease if a reader finds some key words/phrases in the reading area.

2.2 Improvement of the Tool

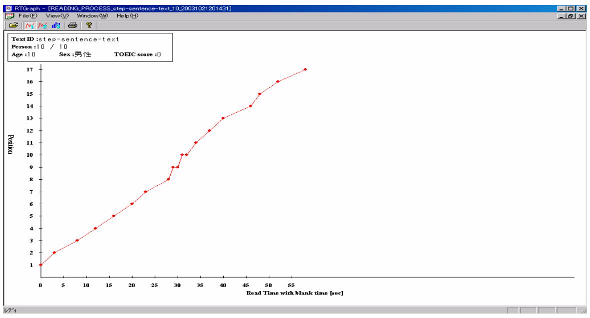
The tool that we have introduced in the previous subsection, however, encounters several problems for analyses in detail, because the reading area is too large to identify particular linguistic effects. In addition, we cannot exclude the possibility that the reading time data may include the time taken when a reader reads the forbidden colored area. We solved these problems by making the non-reading area completely invisible. Furthermore, we changed the fixed reading area into an adjustable reading area so that we can freely determine the size. For example, we can set a single sentence or a paragraph as a reading area, depending on the research question.

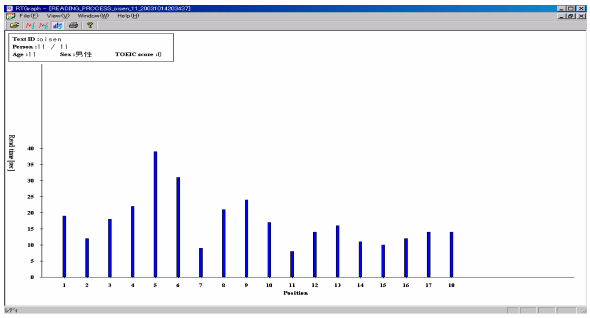
There are reading area icons in the left part of the screen, as shown in the following screenshot (iii). When a cursor is moved onto one of the icons, the reading area appears, and as the cursor moves away, the reading area disappears.



Screenshot (iii)

The improved tool can also represent a reading process in a graph style. In addition, the tool can show cumulative reading time for each reading area.





Screenshot (iv) for Reading Process Graph Screenshot (v) for Reading Time Graph

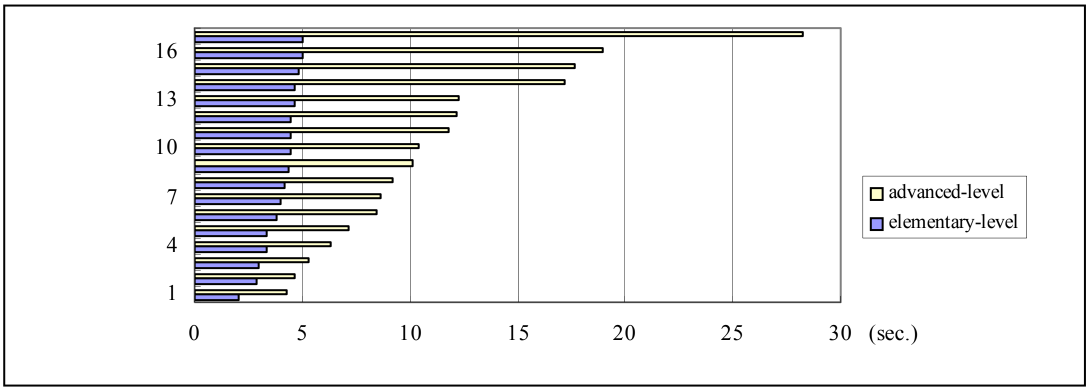
3 Experiments on Readability

3.1 Readability: Easy Text vs. Difficult Text

This sub-section presents our experiment on our improved reading process monitoring tool. Suppose that an elementary-level TOEIC text has higher readability than an advanced-TOEIC text. If our reading process monitoring tool provides the adequate reading data, we can distinguish the readability of these texts by checking the reading data.

17 college students take part in this experiment, and read English sentences using our tool. All the participants' native language is Japanese, and their English ability varies greatly from the elementary level to the advanced level. 15 sentences are assigned to them out of each textbook. Thus, the reading data consist of 17 participants' reading of 30 sentences. On the basis of the reading data, we examine whether the data reflect the text readability between the easy text and the difficult text.

The reading data show that the average reading time for easy text is shorter than the one for difficult text. It takes only 4 seconds to read a single sentence of the easy texts, while it takes more than 11 seconds for the difficult text sentences. Moreover, we find that the reading time of difficult texts is longer than that of easy texts through all the participants as shown in the following Graph (i). The vertical axis shows the participants, and the horizontal axis represents the reading time. Thus, we suppose that the reading time data correctly reflects the text readability.



Graph (i)

3.2 Readability: Evaluation of Reading Support Tools

Given the adequacy of our reading data, we suppose that we can use the data as evaluation criteria for the efficiency of reading support tools such as machine translation systems (henceforth, MT systems). Some studies adopt usability of MT systems as evaluation criteria. Fuji et al. (2002) utilizes TOEIC reading comprehension test score. They compared two MT systems with respect to the percentage of correctly answered questions. In addition to comprehension tests, we also adopt the reading speed data as another criterion. Furthermore, contrary to Fuji et al. (2002), we will compare readability of different support systems: (i) a chunker, (ii) a glosser, and (iii) an MT system.

In this experiment, we prepared TOEIC English texts supported by these systems. As control texts, we further used English texts and human translated Japanese texts. In sum, we prepared the following five types of texts: (i) English-only texts, (ii) English texts with chunking information, (iii) English texts with glossers, (iv) English texts with machine translated sentences, and (v) Japanese-only texts. 25 college students whose reading ability was intermediate participated in this experiment. Two test sets were assigned to each participant. As a single set contains 5 different type texts, each participant was to read 10 TOEIC reading texts and answer the questions.

We predicted that the participants would get the best marks in the case of Japanese texts, and the worst marks in the case of English texts. We further supposed that the scores for text with some support would appear in-between them. This prediction is not borne out. Although the Japanese texts get the highest score, i.e., 75.9%, the English texts get the second highest score, i.e., 59.0%, the chunked texts get 45.8%, glossed texts, 39.5%, and the MT texts, 32.7%. On the basis of the comprehension test results, we might conclude that all the reading support systems may decrease the readability of a text because their scores are lower than English text score. However, by closely examining the texts, we find that English texts are relatively shorter than the other texts. In addition, English texts take the form of an advertisement, while the other texts are newspaper/magazine articles. The text type may have affected the test scores. Thus, we contend that the comprehension test results correctly reflect the text readability and that they constitute another piece of evidence for the comprehension test-based evaluation method.

The reading speed tests bring an interesting result that our prediction on readability is borne out here. As the following Table (i) shows, the average reading speed ranking matches with our prediction. That is, the reading speed of English text is the slowest, and that of Japanese text is the fastest. In addition, the speed of the supported texts appears in-between these texts.

TEXT-TYPE	**ENG-TXT**	**CHUNKED-TXT**	**GLOSSED-TXT**	**MT-TXT**	**JPN-TXT**
Speed(WPM)	64.2	64.1	78.9	103.0	124.6

Table (i)

We statistically compare the variation in reading speed among the five text types, and find that English text reading is similar to the reading of chunked texts and that of glossed texts. On the contrary, it is significantly different from Japanese text and MT text reading. By comparing with the Japanese text reading with other texts reading, we also confirm the association between Japanese text and MT text reading, and the disassociation of the Japanese reading from other texts reading. Thus, we conclude that English text reading belongs to the same class with chunked text and glossed text reading, while Japanese text reading constitutes another class with MT text reading. This classification suggests that MT systems are more efficient than the other systems.

4 Summary

In this paper we introduced our reading process monitoring tools, and the experimental results. The results roughly suggest that the reading time/speed data reveal the readability of texts.

As the remaining problem, we have to account for the relation between comprehensive tests and reading speed tests. In our experiment, these tests exhibit different readability. For instance, MT texts exhibit low readability on the basis of the comprehension tests, while the texts show relatively high readability.

Acknowledgements

We would like to express our gratefulness to the Program Committee members of IWLeL 2004. We would also like to thank the participants for their active involvement in the discussions and their comments. All the remaining inadequacies are of course our own.

Notes

1. Yoshimi et al. (in press) describes the system itself in more detail.

2. Although some studies report the problems of reading text from electronic displays, e.g., O'Hara & Sellen (1997), we leave this issue open in this paper.

References

Bazerman, C. 1988. *Shaping Written Knowledge*. The University of Wisconsin Press, Madison, Wisconsin.

Dillon, A. 1994. *Designing Usable Electronic Text*. Taylor & Francis, London.

Fuji, H., N. Hatanaka, E. Ito,, S. Kamei, H. Kumai, T. Sukehiro, T. Yoshimi and H. Isahara. 2002. Kikai honyaku sisutemu yuukoosei-hyooka: Dono yoo na hito-ni-totte MT-wa yaku-ni tatu-no-ka. *Proceedings of the Eighth Annual Meeting of the Association for Natural Language Processing*, 41-44.

Horney, M. and L. Anderson-Inman. 1994. The electro text project: Hypertext reading patterns of middle school students. *Journal of Educational Multimedia and Hypermedia*, 3(1), 71-91.

Kotani, K., T. Yoshimi, T. Kutsumi, I. Sata and H. Isahara. 2002. Danwa-hyosiki: Dokkai-katei-ni-okeru koosatu. *Proceedings of the Pragmatics Society of Japan*, 41-44.

Kotani, K., Yoshimi, T., Kutsumi, K., Sata, I. and H. Isahara. 2003a. Reading processing units: Application of a measurement tool Operation. *Proceedings of the 2nd International Conference on Speech, Writing and Context*, 96-99.

Kotani, K., T. Yoshimi, T. Kutsumi, I. Sata and H. Isahara. 2003b. Research on reading process: Discourse marker effect. *Proceedings of the Conference Pacific Association for Computational Linguistics 2003*, 241-244.

O'Hara, K. and A. Sellen. 1997. A comparison of reading paper and on-line documents. *Proceedings of Conference on Human Factors in Computing Systems-97*, 335-342.

Yoshimi, T., K. Kotani, T. Kutsumi, I. Sata, and H. Isahara. In press. A method of measuring reading time for assessing EFL-learners' reading ability. *Transactions of Japanese Society for Information and Systems in Education*.