How Sensitive Are Our Eyes to Text Difficulty? Application of Schema Fixation Curves to Japanese Text

Ken HORII*, Kentaro KOTANI*, Yutaka KITAMURA**, Gery d'YDEWALLE***

(Received September 12, 2005)
(Accepted January 30, 2006)

Abstract

This paper discusses the applicability of Schema Fixation Curves to the detection of changes in the behavior of eye movements in accordance with the readability of text. If the eyes are to respond to the degree of difficulty of the given task, we may say that the eyes are an output device of our cognitive activities. Our previous research led us to the notation of Schema Fixation and Schema Fixation Curves, a technique with which graphically analyze the cognitive load the subjects bear when they read texts. The results of our experiments based on this technique show that the eye movement records are a good clue to the detection of text difficulty or readability of texts.

Conventionally, computer-calculated readability indices have been used to predict text readability, but the precision of the prediction may not necessarily be so high. This is because most of these indices use syntactic elements of text such as average sentence length and word length. Difficulty of texts arises from a variety of factors, such as the reader's background knowledge of the passage, the range of vocabulary used in the text, syntactic and semantic ambiguities, etc. In this experiment, we used the Japanese language in order to focus on syntactic effect on readability. Japanese allows much freer syntactic structure than present-day English. For example, the natural, normal, and unstressed word order of English (from amongst the six logical possibilities, SVO, SOV, VSO, VOS, OSV, OVS) is SVO while various combinations are both possible and natural in Japanese. We changed the syntactic order of words in sentences and presented them to the subjects in order to examine the recorded eye movements, and found that different orders produced different levels of readability.

1. Introduction

Readability of texts written in English has long been studied[1] and many formulas have been devised to quantify the difficulty of texts. Today, with the advent of microcomputers, there are many software applications to compute many of these readability formulas[2)~10)]. In Kitamura et al[11)], we introduced Schema Fixation Curves, a methodology for global analysis of eye movements in reading text displayed on the computer monitor screen, to evaluate computer-generated readability indices. Our previous experiment results showed that computer-predicted readability scores for English texts were generally identical with the analysis results obtained by using the schema fixation curve technique provided the

* Department of Systems Management Engineering, Kansai University
** Institute of Foreign Language Education and Research Kansai University
*** Laboratory of Experimental Psychology, Katholieke Universiteit Leuven, Belgium

subjects' first language was English, and suggested that these readability indices can be a rough measure of the difficulty of English texts. At the same time, we learned that eye movements are a good clue to probing the properties of cognitive activities that take place in the human brain.

It may safely be said that readability of English text depends on a lot of interrelated factors. For instance, stylistic difficulty of texts, the reader's readiness for or interest in a particular text etc. Stylistic difficulty may arise from both syntactic and semantic reasons. It is universally agreed that context has much to do with the interpretation of a sentence. As for the hypothesis that contextual information can influence the decisions of the syntactic analysis stage, opposing theories have been proposed, i.e. the "garden path" theory[12), 13)] and the "Incremental Interactive" theory[14)]. A great variety of sentences have been artificially composed as task sentences, and yet it is extremely difficult to avoid the interference of semantic elements on syntactic elements, and vice versa, in the composition of these sentences, because the English language is heavily reliant on word order to convey meaning. English is rather rigid in terms of word order, although, as David Crystal[15)] shows quoting the speech used by the Jedi Master Yoda in the film, 'Return of the Jedi', there is a variation of the SVC word sequence.

1) Sick I've become.

2) Strong with the Force you are.

3) Your father he is.

4) When nine hundred years you reach, look as good you will not.

But Crystal points out that the impact of this strange speech style is ascribed to the rarity of the word sequence.

On the other hand, Japanese is a highly flexible language in terms of word order, because it incorporates the declension of nouns to indicate case attributes of nouns. It is, therefore, possible to change structural elements of a sentence, such as the subject and the object of the sentence, employing exactly the same words without changing the meaning of the sentence. In other words, by using the Japanese language, it becomes possible to measure the effects of syntactic influence upon comprehension of text eliminating the effects of semantic elements. This is a great advantage in designing experiments to detect any change in the amount of cognitive load imposed on the subject in reading.

2. Short Introduction to the Modern Japanese Language

Nouns and pronouns in the modern Japanese language do not have genders, nor are modified either by definite or definite articles. Nouns do not have singular or plural forms. There is no redundancy in terms of numbers in Japanese, while English is highly redundant in this particular case as follows.

5) There is a forest near the town.

6) There are many forests near the town.

Plurality in 6) is expressed by three elements in the sentence: i) the plural form 'are"; ii) the adjective "many" ; and iii) the plural s-ending. In Japanese, the plurality at issue is marked only by employing a word that expresses the idea "many" (in Japanese the word is used adverbially not as an adjective as in English), and no other elements of the sentence are affected.

As mentioned earlier, cases are signified by the declension of nouns. To be more precise, an auxiliary particle is added to the end of a noun to decline. Although Japanese basically takes the (S)OV word order compared to the English SVO word order, all the other combinations are possible and natural. For instance, in English, your love for the person you are talking to can be confessed by a sentence whose word order cannot be changed:

7) I love you.

In Japanese, the following word orders are possible to confess your love almost to the same effect.

8) (I) love you.

9) (I) you love.

10) You (I) love.

11) You love (I),

12) Love (I) you.

13) Love you (I).

Of course there are preferred and usual word orders among these, but the choice of any particular word order is a matter of individual taste. As the English nominative and accusative forms of the second person singular (and also plural) are identical in their forms and pronunciations, the combinations 9) through 13) create great confusion or simply fail to convey what is meant by the speaker or writer. Among many other features of the Japanese language, it is worthy of note that a statement can be converted into a question by adding an auxiliary particle at the end of the statement, and that negation occurs also at the end of a sentence, which gives the speaker or writer a great freedom to change his or her initial intent of utterance in the midst of discourse so that unnecessary misunderstanding or unfavorable effect of utterance can be avoided. This flexibility in syntax is a most outstanding feature of this language, and it is this property that we will make use of in designing the experiment at issue.

3. Experiment and Method

3.1 Apparatus

Two different eye-movement recorders, NAC's Eye Movement Recorder Model V (hereafter abbreviated as EMR V) and NAC's Eye Movement Recorder Model VI (hereafter abbreviated EMR VI) were employed to record eye movements while subjects were reading Japanese texts displayed on the CRT, which was placed at a distance of 30 cm from the viewing position. The reading materials were displayed on the CRT with a resolution of 640 pixels x 400 pixels. At that distance, one Japanese character regardless of the types of characters whether *kana* syllables or *kanji* ideograms subtended one degree of visual angle. The eye-movement data were recorded every 33 ms by EMR V, whose programs for data collection and analysis were developed at our laboratory.

3.2 Materials

Three pairs of short Japanese passages (Texts A-1, A-2; Texts Bl-1, Bl-2; Texts C-1, C-2), consisting of four lines when displayed on the CRT, were prepared for the experiment. Each pair comprised two versions of a passage that were exactly the same, both in content and in the number of characters and differing only in word order, thus completely eliminating semantic interference. The texts, labeled Text *-1, have a rather entangled word order, while those labeled Text *-2 are supposed to be normal in their word order. The passages are shown in Appendix A.

3.3 Subjects

Ten Kansai University students with normal vision were paid to participate in the experiment, in which EMR V was used. Authors had initially recruited more than 30 students from whom they selected five subjects for our experiment using EMR VI, but we had to give up on most of the applicants simply because EMR VI was extremely selective of subjects. This was because EMR VI adopted the reflected-light method for amplifying the eye movements, which not only resulted in imposing a finely-tuned calibration on the experimenters' side, but also in requiring a certain smooth curvature of the eyeball on the subjects' side. An unexpectedly large proportion of the subjects were inappropriate for EMR VI. Therefore, we mainly used the former model EMR V, which allows easier calibration, so that we could gather eye-movement data extensively from more subjects.

3.4 Procedure

The subjects were first asked to read several sample texts displayed on the CRT on different occasions so that they could get accustomed to the equipment and the experiment procedure. This phase of training was also necessary for the operators of the eye movement recorders, for they had to spend some considerable time before they could learn the conditions of curvature of individual subjects' eyes to ensure better calibration. The subjects were asked to visit our lab some days later so that any eye strain from the training phase of the experiment was completely healed. For the stability of the head coordinance, a chin and forehead rest was adopted. The onset of the display of the assignment text was controlled by the subject, who presses a button placed on the desk right in front when he/she is ready to start the display.

First, one of the text versions *-1 and then the corresponding *-2 version were given. The subjects were instructed to read the texts for meaning, and were asked not to employ any special speed reading techniques, such as reading a text from beginning to end by only skimming the *kanji* characters.

3.5 Japanese Readability Index

Reading Grade Value (hereafter abbreviated as RGV), a formula to predict the readability of Japanese texts was developed by Y. Asano and K. Ogawa[16]. The formula was given as follows:

$$RGV = -0.17ph - 0.28pk - 3.49pc + 27.62$$

where:

RGV = assigned grade level

ph = percentage of *hiragana* characters

pk = percentage of *katakana* characters

pc = percentage of end punctuation marks

It should be noted that the three RGV variables are all typographical features of Japanese text. It is obvious that RGV is not in the least sensitive to any change in readability or difficulty due to the variations of the word order of the assignment sentences in Japanese, for the relative frequencies of the three variables remain exactly the same and unaltered. We used this index as a rough measure of the stylistic difficulty of texts we used in this experiment.

3.6 Schema Fixation Curves and Reading Schematic Load

Kitamura et al.[11] introduced the Schema Fixation Curve, a technique for global analysis of eye-movement record, and showed that the stylistic difficulty of English texts affected the eye movements while reading. This global analysis technique treats the boundary-time condition for the determination of fixation as a variable rather than a constant. Instead of choosing a specific duration of time as the boundary-time, which is conventionally determined by measuring the average duration of fixations for all possible durations, i.e. n times the minimum resolution of the sampling time of the eye-movement recorder-cluster occurrences are counted, which satisfy the conditions whereby a fixation should form a cluster of recorded eye positions within one degree of arc deriving from the fixation immediately preceding, and that a series of such positions should last beyond the boundary-time. By plotting boundary-times with an increment of the minimum resolution of the eye-movement recorder on the x-axis and the corresponding occurrences of fixations on the y-axis, we obtained a schema fixation curve.

The technique proved to be highly sensitive to the change in the amount of cognitive load when the subject is reading a passage displayed on the CRT. Since the schema fixation curves employed graphical representations to show the analysis result, quantitative comparisons were not possible. Kitamura and Horii[17] quantified the schema fixation curves to obtain an index called Reading Schematic Load (abbreviated hereafter as RSL), so that one may directly compare the difficulty of texts in terms of the cognitive load that affected the eye movements while the subject was reading text. The index value increases as the readability of text decreases.

The following is the formula for RSL:

$$RSL = \sum_{i=1}^{n} \frac{y(s_i) + y(s_{i+1})}{2} * (s_{i+1} - s_i)$$

where $y(s_i)$ is the frequency of occurrences of fixations determined at a boundary time of s_i ms. The index value increases with decrease in the readability of the text.

4. Results

We first processed the eye-movement data obtained by EMR-V in the conventional way in order to determine the fixations and saccadic distances. A fixation was determined as a cluster of recorded eye positions within one degree of arc deriving from the fixation immediately preceding and taking place beyond the boundary minimum time of 99ms. The center of mathematical gravity in such a cluster was treated as a fixation point. Tables 1-1 and 1-2 show the results of the analyses for the data obtained from 10 subjects for two versions of three different texts.

Table 1-1. Average span of progressive saccades (unit: degrees). Ten subjects read three different texts (A, B, C), each with two versions (*- 1, *- 2), whose respective Reading Grade Values are listed in the column 'rgv'.

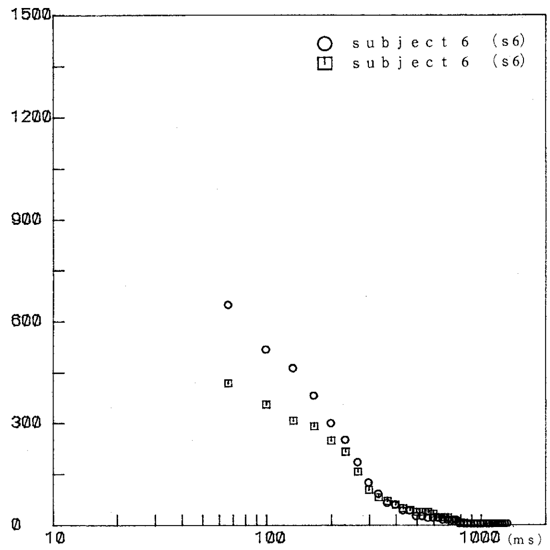
	rgv	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	avr
A-1	13	2.9	3.7	3.8	2.6	3.5	2.7	3.6	3.4	3.5	3.5	3.3
A-2	13	2.5	3.6	2.6	1.9	4.4	2.6	4.7	2.7	5.3	3.1	3.2
B-1	10	3	4.1	3.2	2.6	4.7	2.5	5.1	2.8	3.3	3	3.3
B-2	10	2.8	5	2.4	2	3	2.4	6.4	2.8	4	3.2	3.3
C-1	9	2.2	3.7	4.9	2.7	3.3	3.9	4.4	3.2	3.3	3.8	3.4
C-2	9	2.5	5.4	4.1	2.2	4.5	3.5	5.2	3.1	4.7	3.8	3.5

Table 1-2. Average span of regressive saccades (unit: degrees).

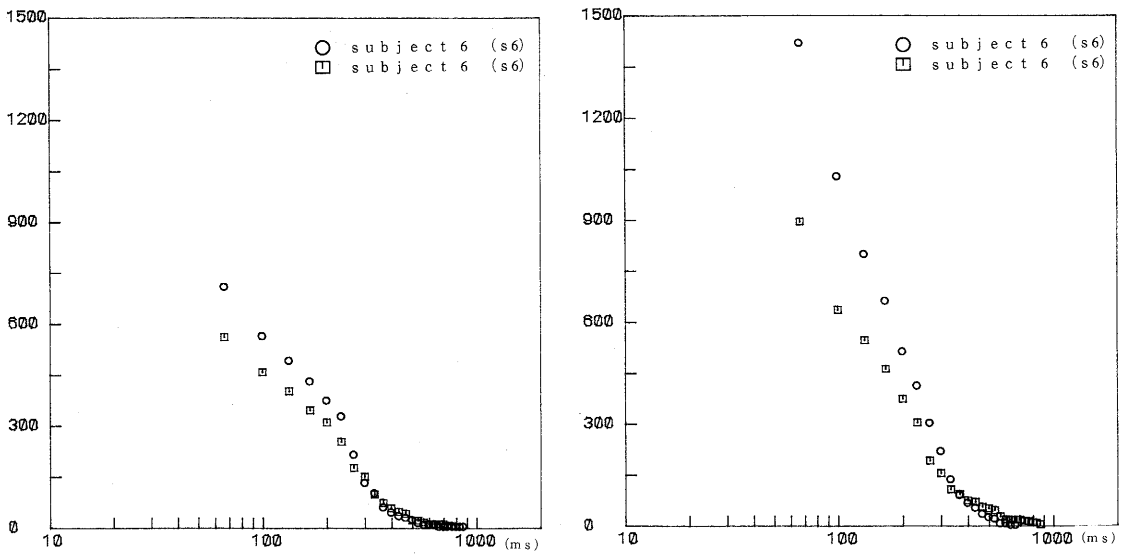
	rgv	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	avr
A-1	13	1.8	3	2.3	2.4	1.8	1.7	2	3.2	2.9	2	2.4
A-2	13	1.7	2.7	2.3	1.8	2.7	2.1	2.9	2	2.2	1.6	2.3
B-1	10	1.5	2.3	2.4	1.9	1.7	1.5	2.4	2.3	2.5	2.6	2.3
B-2	10	1.2	2.1	1.9	2.2	3	1.5	2.2	1.9	1	2.5	2.2
C-1	9	1.8	2.9	3.3	3.3	1.3	2.3	3.6	3.2	3.1	3.5	2.5
C-2	9	1.4	3.7	1.9	2.4	2.4	2.2	3.5	2.7	2.4	1.7	2.5

The authors used to adhere to the hypothesis that the easier the text is, the greater will become the span of progressive saccades. Table 1-1, however, does not indicate any such remarkable tendency, nor can we derive any trend from the average regressive saccadic distance of each text, as shown in Table 1-2. Subjective impressions of the subjects were unanimous that the two versions of the three texts were definitely different in terms of readability. The results of t-tests show that there was no significant difference at p=0.05, or even at p=0.1, in the average of the progressive saccade spans between versions A-1 and A-2; B-1 and B-2; and C-1 and C-2.

But when we compare the schema fixation curves, it can be easily seen that there is a difference between the two versions of each Japanese passage. See Fig. 1 for an example. It represents a typical example of global analysis of eye-movement records using a Log-normal Schema Fixation Curve. In the graphs A, B, and C, two curves are plotted, one with small circles and the other with small boxes with a dot inside. The curves represented by small circles are the Schema Fixation Curves for Texts *-1 and those represented by small squares with a dot inside are for Texts *-2. Our data from past experiments convinced us that the higher the position of a Schema Fixation Curve, the more difficult the text is to comprehend. In graphs A, B, and C in Fig. 1, the schema fixation curves for Texts *-1, that is, those texts with higher syntactic difficulty, come above the schema fixation curves for Texts *-2, which have normal syntactic structure. The graphs in Fig. 1 are the analysis results of the eye-movement records of subject No. 6.



(C)



(A) (B)

Fig.1 Schema fixation curve of Subject 6 reading the assignment texts.

The RSL values for the assignment texts of this subject are listed in Table 2, together with those of the other subjects. Since these values are obtained by quantifying the schema fixation curves, the comparison of RSL values of these subjects will of course show the same result.

The t-test results for the RSL values in Table 2 showed that two pairs A-1, A-2 and B-1, B-2 belonged to the same group, but that there was a significant difference between C-1 and C-2 at p=0.05. But at p=0.1, all three pairs (A-1, A-2; B-1, B-2; and C-1, C-2) were shown to belong to the same group. Note that the saccadic distance failed to detect the change of text difficulty even at p=0.1. Here we have an interesting problem: Text C alone is an arithmetic problem while the other two are ordinary prose passages. The simple comparison of RSL averages for C-1 and C-2 convinces us that the two are significantly different. Does the problem-solving activity during reading affect the eye movement? At present, the authors do not have an answer to this question, but the statistical output suggests it may.

Table 2. RSL values of the ten subjects when they read texts A-1, A-2; B-1, B-2; C- 1 and C-2.

	rgv	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	avr
A-1	13	2795	5819	2512	4233	3196	3284	2494	2685	1641	2440	3109.9
A-2	13	2494	1797	2132	3387	1975	2787	2849	2759	1756	851	2278.7
B-1	10	2774	4293	2350	2336	2568	3126	3284	2755	2007	2415	2790.8
B-2	10	2616	1852	2327	3984	1564	2413	2490	1767	1162	2338	2251.3
C-1	9	4258	4203	4929	4279	1877	5120	5830	7480	2734	6613	4732.3
C-2	9	2911	1576	2170	3336	2205	3760	4839	2977	1256	4310	2934

We were equally interested to find that the order of presentation of two different versions of assignment texts has much to do with the cognitive load at the time of reading. When easier versions are shown prior to their corresponding difficult versions, the difficult versions can be read with far less difficulty. Once what is written is understood, the understanding of the text seems to reduce the difficulty of decoding syntactic information in the text. This phenomenon is well-illustrated in Fig. 2.

Compare Figs. 2 and 1. It can be seen that not only the difficulty of Texts *-1 decreased because of the easy versions first read, but the difficulty also dropped to such an extent that the schema fixation curves for the difficult versions came below those of difficult versions in graphs A and B. This implies that second reading is, generally speaking, much easier than first reading even when the syntactic structure of the texts for second reading are changed to increase difficulty.

Here, we cannot deny that there is a chance for the subject to become a little negligent in reading the text which has the same content with a different syntactic structure, and simply run the eye over lines of a displayed passage without fully trying to understand it. Kitamura, Horii and d'Ydewalle[18] developed the Delta-Schema Fixation Curve technique to detect the skating of the eye along the lines of text, as shown in Figure 3. It should also be noted that in graph C of Figure 2, the schema fixation curve for the second reading stands higher than that for the first reading. From this, we can deduce that the reasoning that takes place at the time of reading has an influence on the reader's reading pattern.

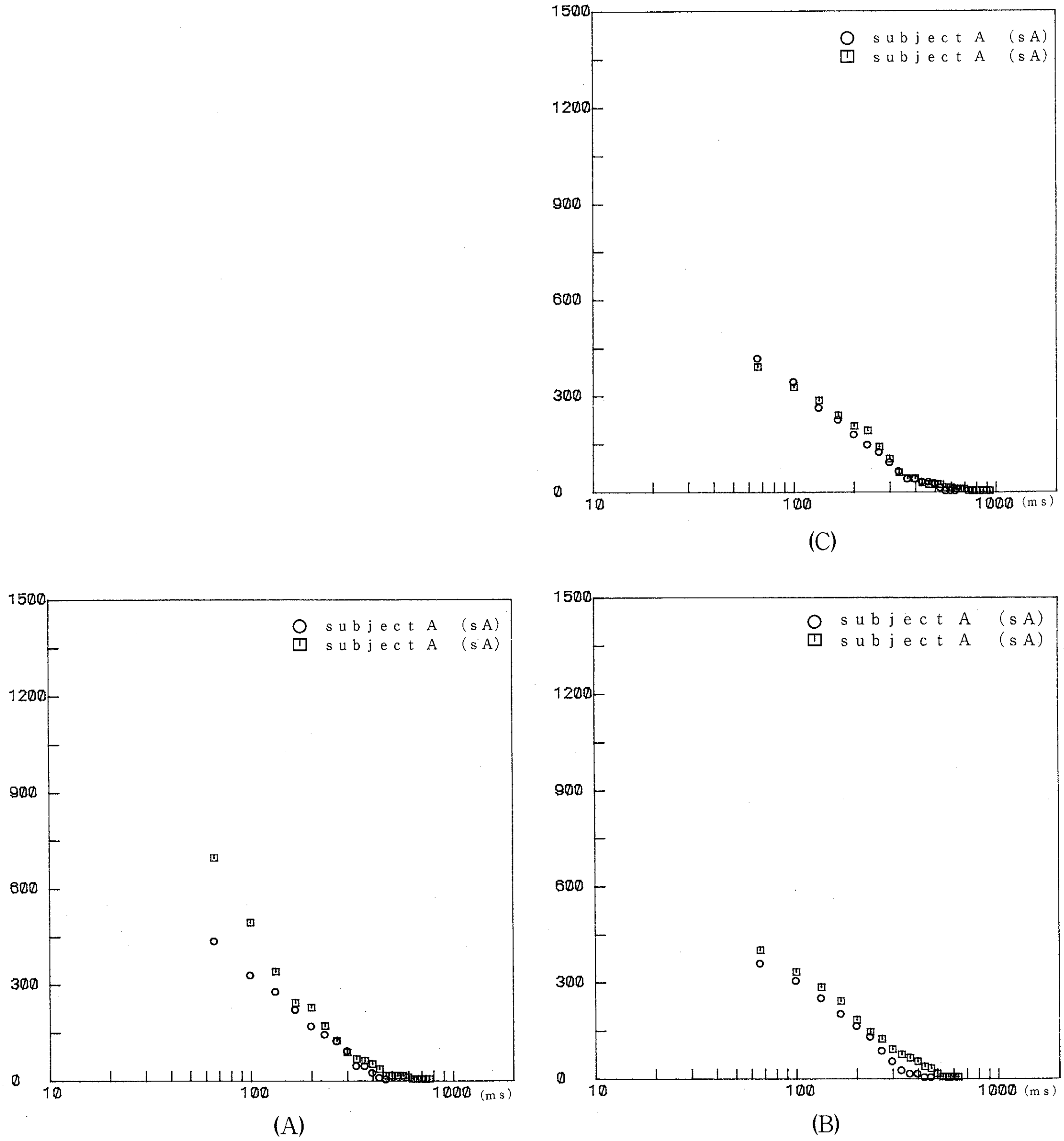


Fig.2 Schema fixation curves of Subject 7 when the subject was reading texts with easy versions displayed first and difficult ones later.

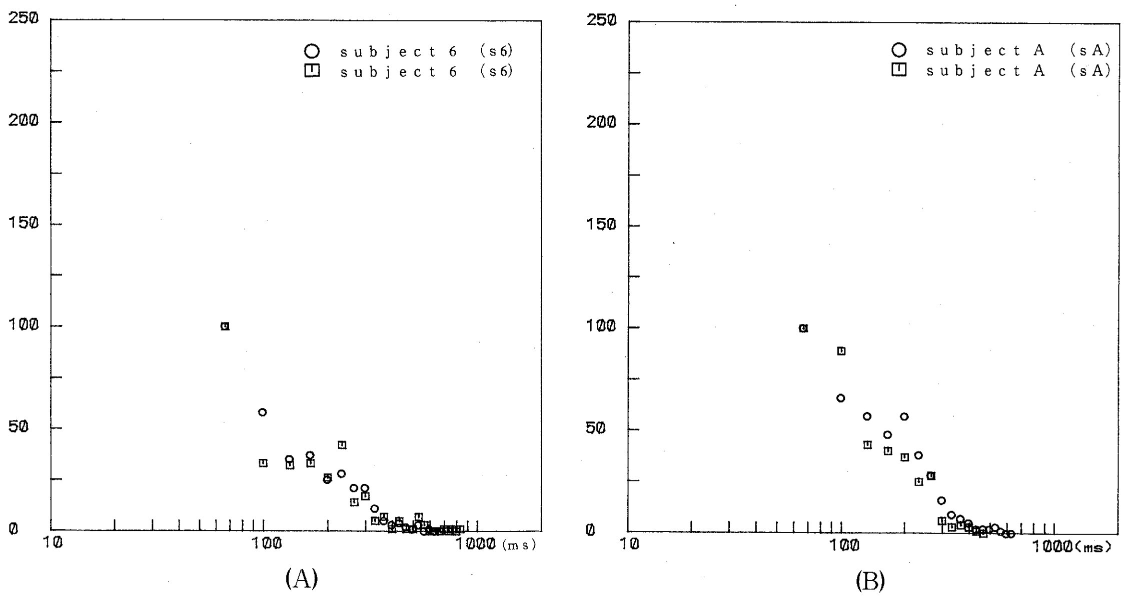


Fig.3 Delta schema fixation curves. Graph A is an example of reading with attention, and Graph B is an example of the eye skating along the text without attention.

5. Conclusion

We would like to point out that reading involves such delicate matters as degree of understanding while reading, which has something to do with the instruction given to the subjects. The subjects, being Japanese, were able to read C-1 and C-2 for meaning with ease because the text was written in Japanese. But when it came to the matters of understanding and reasoning, individual subjects responded to the instruction to read the displayed text for meaning. Some seem to have tried to get to only the surface meaning of the passage, while others tried to solve the arithmetic problem, but only a few could give the correct answer (although they were not asked to do so).

All the more important is the fact that there exist levels of understanding, which we cannot directly measure, and that the level affects reading behavior and eye movements greatly. Unfortunately, we must conclude that studying the main tendencies of fixation durations and saccadic distances are not sufficient to investigate this problem.

Appendix

The following texts are coarse translations of the Japanese texts used in this experiment.

Text A:

Heisuke, an apprentice cook, overheard Gen, who is always causing quarrels by speaking ill of somebody or something as he gossips among other cooks who come to the market to get foodstuffs for meals over the New Year's holiday. He says, "Our master is a good for nothing, for he is always oversleeping."

Text B:

Today, I introduced Hanako Taro, who has known me well since childhood, and had long before asked me to introduce some nice male friend to her at Etoile, a coffee house near the station, where a wide variety of music, ranging from classical to Latin music, can be enjoyed from morning till night.

Text C:

This is a mathematical problem. Jiro, who had brought two satsuma tangerines and an apple for his lunch, gave one of the tangerines to Taro, whose mother had asked him to buy three tangerines for thirty yen each, five apples for fifty yen each, and four pears for about fifty yen each. How much money did Taro save?

References

1) Klare, G. R., '*The measurement of readability*', Iowa State University Press, Ames, IO, 1963.

2) Coke, E. & Rothkopf, E., 'Note on a simple algorithm for a computer produced-reading ease score', *Journal of Applied Psychology, 54*, 209-210, 1970.

3) Coleman, M. & Liau, T., 'A computer readability formula designed for machine scoring', Journal of *Applied Psychology*, 60, 283-284, 1975.

4) Danielson. W. & Bryan, S., 'Computer automation of two readability formulas', *Journalism Quarterly, 40*, 201-206, 1963.

5) Fang, I., 'By computer Flesch's reading ease score and a syllable counter', *Behavioral Science, 13*, 249-251, 1968.

6) Gross, P. & Sadowski, K., 'FOG INDEX: a readability formula for microcomputers, *Journal of Reading, 28*, 614-618, 1985.

7) Keller, P., 'Maryland micro: a prototype readability formula for small computers', *The Reading Teacher, 35*, 778-782, 1982.

8) Klare, G. R., Rowe, P., St. John, M. & Stolurow, L., 'Automation of the Flesch reading easereadability formula with various options', *Reading Research Quarterly, 4*, 550-559, 1969.

9) Kretschmer, J., 'Computerizing and comparing the Rix readability index', *Journal of Reading, 27*, 490-499, 1984.

10) Strong, S., 'An algorithm for generating structural surrogates of English text', *Journal of the American Society for Information Science, 25*, 10-24, 1974.

11) Kitamura, Y., Horii, K., Tomoda, Y. & Akai, Y., 'Evaluation of readability formulae in terms of eye movements: derivation of schema fixation curves', in G. d'Ydewalle & J. Van Rensbergen(eds.), *Perception and cognition: Advances in Eye Movement Research* (pp. 265-274), Elsevier Science Publishers B.V., Amsterdam, 1993.

12) Clifton. C. & Ferreira, F., 'Ambiguity in context', *Language and Cognitive Processes, 4*, 77-104, 1989.

13) Rayner, K., Carlson, M. & Frazier, L., 'The interaction of syntax and semantics during sentence processing', *Journal of Verbal Learning and Verbal Behavior, 22*, 358-374, 1983.

14) Altmann, C. T. M. & Steedman, M. J., 'Interaction with context during human sentence processing', *Cognition, 30*, 191-238, 1988.

15) Crystal, D., '*The Cambridge Encyclopedia of Languag*', Cambridge University Press, Cambridge, 1987.

16) Asano, C., & Ogawa, K., 'Analysis of VDT text reading style', in H. J. Bullinger (ed.), '*Human Aspects in Computing: Design and Use of Interactive Systems and Work with Terminals*', Elsevier Science Publishers B. V., Amsterdam, 1989.

17) Kitamura, Y., & Horii, K., '*Application of schema fixation curves to the prediction of reader's competence; a multi-language stud*', paper presented at the Third International Conference on Visual Search, UK, 1992.

18) Kitamura, Y., Horii, K. & d'Ydewalle, G..., '*Is fixation a point or an area?: Properties of fixation*', paper presented at ECEM7, UK, 1993.