Spaces Enhance Word Segmentation and Comprehension in Tacit Reading

Mayumi Toshima[1], Tetsuo Ishikawa[2,3,4], and Ken Mogi[4]

[1] Department of Informatics,
Graduate University for Advanced Studies
2-1-2 Hitotusbashi, Chiyoda-ku, Tokyo, 101-8430, Japan
mamitako@nii.ac.jp
[2] Department of Computational Intelligence and Systems Science,
Tokyo Institute of Technology
G3-46, 4259 Nagatsuta-cho, Midori-ku, Yokohama, 226-8502, Japan
i@brn.dis.titech.ac.jp
[3] Centre for Advanced Research on Logic and Sensibility, Keio University
2nd Floor Mita Toho Bldg. 3-1-7 Mita, Minato-ku, Tokyo, 108-0073, Japan
[4] Sony Computer Science Laboratories, Inc.
Takanawa Muse Bldg. 3-14-13 Shinagawa-ku, Tokyo, 141-0022, Japan
kenmogi@csl.sony.co.jp

Abstract. There are a lot of types of written language systems all over the world. The Japanese writing system uses ideographic and phonetic symbols without spaces between words, where the ability of finding the boundary between words is expected to have a positive correlation with the ability of general language comprehension. Is such a correlation to be found in languages (e.g., English) where words are conventionally separated by spaces between them? Here we report a study on how spaces put on the boundary of two adjacent words enhance tacit reading texts. We conducted a "slash task" where our subjects whose first language was Japanese were instructed to place a slash between adjacent words of sentences written continuously without spaces between the words in typical, short and easy sentences. Our data illuminates the role of "active segmentation" in language comprehension.

Keywords: Language, word, segmentation, learning, tacit reading.

1 Introduction

Languages are important elements in human cognition, and play an essential role in the communication among humans. In the process of cognitive development of one's mother tongue, a progress in the ability of the first language is accompanied by an increasing difficulty in acquiring other languages. Such a cognitive phenomenon occurs because of the characteristics of our brains [1][2][3]. The nature of the first language acquired results in significant differences in the construction of efficient neural network systems, which in turn affect second language acquisition [4][5][6].

Reading words consist of perceiving language symbols visually and assigning them appropriate meanings. Some of the difficulties encountered in second language acquisition are due to the differences in the system of writing between the languages. Among the various languages in the world, there are two distinctive systems of writing, i.e., ideographs and phonograms. In both of these writing systems, sentences are constructed of strings of words. In Japanese writing, the segmentation between the words is given implicitly, where the letters are written continuously without placing spaces between the words; both of the Chinese letters [7] which contain ideograms and phonograms and the Thai letters [8] which contain phonograms do not need spaces as well. In English and other European languages, spaces are placed between the words. It is interesting to compare the cognitive process of language segmentation between these two types of writing systems, i.e., one with spaces and one without spaces between words. Previous studies indicate that unspaced English texts influenced the speed of oral reading [8] and eye movement behaviors [9].

In the domain of speech perception, continuous auditory sequences contain no "spaces" (i.e., pauses). However it is known that (i) 8-months old infants can recognize the word boundaries [10], (ii) ERP N400 is correlated with word onset in continuous auditory stream [11] and (iii) the recursive neural network can find English word boundaries in unspaced texts [12].

Here we report a study on effects of spaces enhancing segmentation in tacit reading. In order to conduct the experiments, we invented "slash tasks"; when one finds word boundary in unspaced text, he or she is required to put a slash down on the boundary. To measure one's reading competency, several types of tests have been conducted. Most of them (e.g., [8]) require reading strings of letters aloud. Therefore, those tests depend not only on reading competence but also on speaking ability. Our slash tasks, on the other hand, only require the subjects to mark slashes between words while reading tacitly. Subjects were university students whose first language was Japanese. The subjects conducted a "slash task" where they were instructed to place a slash between adjacent words of typical English sentences written continuously without spaces between the words.

2   Experiment

Materials and Methods

32 subjects (12 females, mean age 19.6±1.7) whose native language was Japanese participated. All were undergraduate students, right-handed by self report, with normal or corrected to normal vision.

Two independent variables were designed in the experiments. (1) Spacing: spaced or unspaced texts, all written in English; (2) Coherence: coherent or incoherent texts (Fig. 1). The order of tasks was counterbalanced with coherence.

The subjects were asked to put slashes between words in a coherent/incoherent text with/without spaces within four minutes. Reading time was recorded by self-report. After that, all the subjects took a word recognition test, making an

Coherent and Spaced (COH-SPA) text

When I first met her she was on the beach/
She was looking toward the sea and she was cute/
Her mother and father went somewhere at that time then/
Before they left the island they said to the boy/
It was just before the last long big war ended/
Another day passed and still another day passed so quickly/

Coherent and Unspaced (COH-UNS) text

Iamteachingalanguageinaverysmalltown/
Youcannotimaginethelifeinaruralcountry/
Itisdifferentfromthelifeinamoderncity/
Peoplearoundheredonotworrysomuchabouttime/
Ifabusislatetheywillwaitforone/
Itisimportantforpeopletolearnanofficiallanguage/

Incoherent and Spaced (INC-SPA) text

grew day visit was in that would and but started/
it daughters happened had it said surely the it it/
mountain its mountain on A the The have lonely ages/
And very every Nothing said spring that bird There Then/
place daughters to bird been would a for mountain many/
that small one mountain speaking it once flew a lonely/

Incoherent and Unspaced (INC-UNS) text

tocouldinamsendHethisandheAnd/
himfatherMytotraveltraveldiedbearfather'svery/
lateddywantagemyButatIworld/
ofnowplaceinstudentforty-fivelastwantedaroundwas/
thenjuniorhewasamuchtocancerInot/
busywantheabroadthehighdoctorschoolwinterJapan/

Fig. 1. Stimuli (4 types of texts)

"old" or "new" judgment. Random sampled words were presented in this test and the subjects were required to judge whether the words had been shown in the previous slash task text ("old") or not ("new"). Four sessions were conducted without intervals. The same subjects were required to take another type of test in twenty minutes; this contained thirty questions taken from a TOEIC problem book. TOEIC (Test of English for International Communication) is a widely used test of English to measure learners' ability to use English as a language system (grammar, vocabulary and usages.).

The data were examined using ANOVAs with Space and Coherence as factors. Then, Games-Howell post-hot multiple comparisons were conducted. The significance level was set at 0.05.

3   Results

We found that the main effects of Coherence ($p = 0.031$) and Space ($p < 0.001$) show significances on the reading time, but their interaction indicates no significance ($p = 0.087$). The mean reading time of COH-SPA (the Coherent-Spaced condition) and INC-SPA (the Incoherent-Spaced condition) were not significantly different ($p = 0.73$). Those two conditions required shorter mean time than COH-UNS (the Coherent-Unspaced condition) ($p < 0.001$) and INC-UNS (the Incoherent-Unspaced condition) ($p < 0.001$). The mean reading time of COH-UNS and INC-UNS were rather comparable ($p = 0.28$) (Fig. 2).
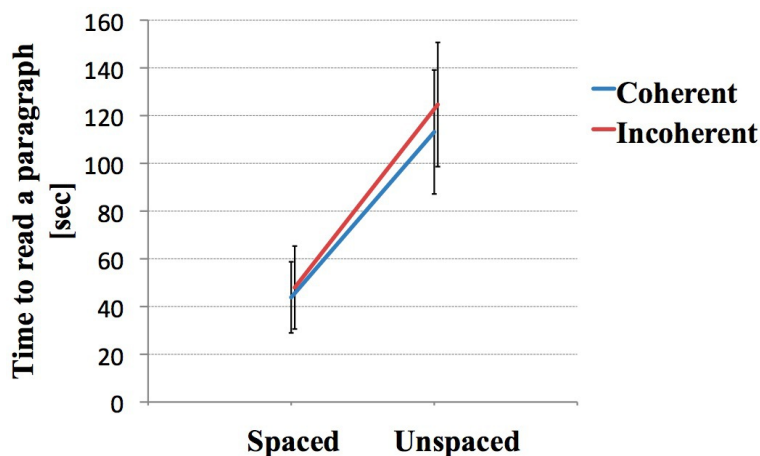

Fig. 2. Average reading time of 4 types of the texts (Mean±SD)

The number of errors in reading was the number of failures to find word boundaries. The main effects of Coherence ($p = 0.007$) and Space ($p < 0.001$) on the number of errors and also their interaction ($p < 0.001$) were found. According to the multiple comparisons, the amounts of the errors indicated significant
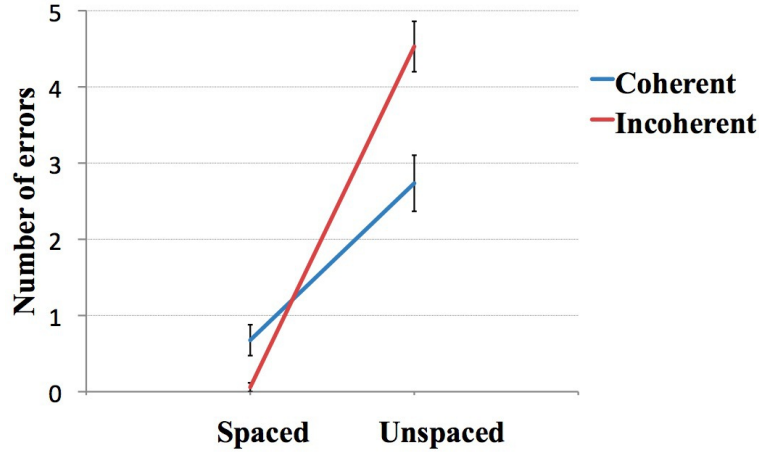
Fig. 3. Errors in reading of 4 types of the texts (Mean±SEM)

differences among all of the combinations of each condition: INC-SPA < COH-SPA < COH-UNS < INC-UNS (INC-SPA vs. COH-SPA: p = 0.03, COH-SPA vs. COH-UNS: p < 0.001, COH-UNS vs. INC-UNS: p < 0.001) (Fig. 3).
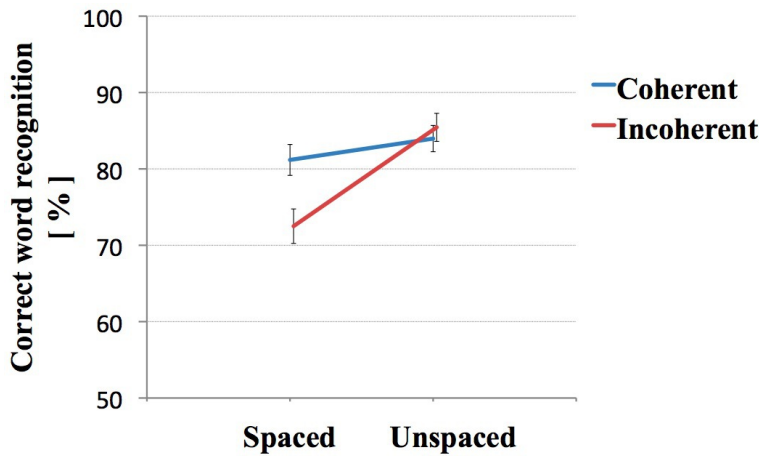

Fig. 4. "Old" or "new" judgment performances (Mean±SEM)

The correct rate of word recognition was defined as the rate of correct-old ("hit") and correct-new ("correct rejection") responses. Note that the chance level of word recognition accuracy was 50 percent because of the two-alternative ("old" or "new") forced choice. The main effect of Space on the correct rate of word recognition was significant (p < 0.001), but that of Coherence was not significant (p = 0.078). The interaction between them was significant (p = 0.003).

The multiple comparisons showed that the INC-SPA correct rate of word recognition, the old and new judgment performance was lower than other conditions (INC-SPA vs. COH-SPA: p = 0.03, INC-SPA vs. COH-UNS: p = 0.001, INC-SPA vs. INC-UNS: p < 0.001) (Fig. 4).

4   Discussion

Spaced texts and coherent texts were read faster than unspaced texts and incoherent texts, respectively. It is suggested that spaces work efficiently as a sort of punctuation systems. Without spaces, the subjects had to find the blocks of letters with meanings; it was necessary for the subjects to segment the words attentively and intentionally.

When doing the tasks without spaces, most of the subjects were observed leaned toward the texts and putting the tip of the pen closer to the written letters, as if they were tracing invisible lines drawn under the strings of letters. These behaviors were not observed when they were doing the tasks with spaces. Without spaces, the subjects had to be more careful to find the segments of the letters with meanings; therefore, their postures were different from those of when doing the tasks with spaces.

For the all subjects, unspaced English sentences are rather unfamiliar even if they see Japanese (their first language) sentences which have no spaces between characters and letters; they are sufficiently used to reading Japanese with its ideographic and phonetic symbols. Their scores of reading "Hiragana" and "Katakana" (phonetic symbols) and "Kanji" (ideographic symbols) are not significantly varied. The relationship between the first language and the second language word segmentation competence should be clarified in the further researches.

Errors were observed in all of the conditions. No subjects got perfect scores in all tasks. The least errors were observed when doing "Incoherent and Spaced (INC-SPA)" task, not "Coherent and Spaced (COH-SPA)" one. This may suggest that the subjects should pay more attention when they are given more unfamiliar or unusual texts than familiar and usual texts.

The correlation between "the speed and correctness" and "TOEIC scores" showed no significance in the present study. However, according to preceding researches (data not shown), it is possible to think that the ability of recognizing the blocks of words that we call it "active segmentation" should be correlated with certain cognitive and comprehension system when reading written languages.

The slash task has a possibility of application to language education and clinical assessment of aphasia and dyslexia [13][14] as it is not explicitly influenced by articulating or speaking ability. It also has the following advantages: (i) Only a pen and a bundle of paper are needed, and no complicated and expensive devices such as eye movement trackers are necessary. (ii) It requires at most about 120 seconds. (iii) It is possible to conduct classroom experiments with many participants simultaneously.

References

1. Deacon, T.W.: The symbolic species: the co-evolution of language and the brain. W.W. Norton & Co., New York (1997)
2. Crinion, J., Turner, R., Grogan, A., Hanakawa, T., Noppeney, U., Devlin, J.T., Aso, T., Urayama, S., Fukuyama, H., Stockton, K., Usui, K., Green, D.W., Price, C.J.: Language control in the bilingual brain. Science 312, 1537–1540 (2006)
3. Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen, I., Frenck-Mestre, C., Hirschensohn, J.: Second-language learning and changes in the brain. Journal of Neurolinguistics 21(6), 509–521 (2008)
4. Cummins, J.: The cross-lingual dimensions of language proficiency: Implications for bilingual education and the optimal age issue. TESOL Quarterly 4, 75–87 (1980)
5. Cummins, J.: Bilingualism and minority language children. Ontario Institute for Studies in Education, Toronto (1981)
6. Nakada, T., Fujii, Y., Kwee, I.L.: Brain strategies for reading in the second language are determined by the first language. Neuroscience Research 40(4), 351–358 (2001)
7. Bai, X., Yan, G., Liversedge, S.P., Zang, C., Rayner, K.: Reading spaced and unspaced Chinese text: evidence from eye movements. Journal of Experimental Psychology: Human Perception and Performance 34(5), 1277–1287 (2008)
8. Kohsom, C., Gobet, F.: Adding spaces to Thai and English: Effects on reading. Proceedings of the Cognitive Science Society 19, 388–393 (1997)
9. Rayner, K., Fischer, M.H., Pollatsek, A.: Unspaced text interferes with both word identification and eye movement control. Vision Research 38(8), 1129–1144 (1998)
10. Saffran, J.R., Aslin, R.N., Newport, E.L.: Statistical Learning by 8-Month-Old Infants. Science 274, 1926–1928 (1996)
11. Abla, D., Katahira, K., Okanoya, K.: On-line Assessment of Statistical Learning by Event-related Potentials. Journal of Cognitive Neuroscience 20, 952–964 (2008)
12. Elman, J.L.: Distributed Representations, Simple Recurrent Networks, And Grammatical Structure. Machine Learning 7(2), 195–225 (1991)
13. Eden, G.: Presentation at Fluency Conference. Dyslexia Research Foundation, Crete (2000)
14. Wolf, M.: Proust and the Squid – the story and science of the reading brain. Harper, New York (2007)