



eBraille: a web-based translation program for Japanese text to braille

Sugano, Aki ; Ohta, Mika ; Oda, Tsuyoshi ; Miura, Kenji ; Goto, Shuji ;
Matsuura, Masako ; Maeda, Eiichi ; Ohshima, Toshiko ; Matsumoto, Yuji ...

(Citation)

Internet Research, 20(5):582-592

(Issue Date)

2010

(Resource Type)

journal article

(Version)

Accepted Manuscript

(URL)

<https://hdl.handle.net/20.500.14094/90001466>



eBraille: A web-based translation program for Japanese text to braille

ABSTRACT

Purpose

We developed a program, which we named eBraille, to translate Japanese text into braille and thereby generate braille documents easily. We provided public access to this program to anyone via the Internet.

Design/methodology/approach

eBraille is a CGI program that is accessible via a Web browser. The core of our program is a braille translating engine called the Kobe University Intelligent Braille Engine for ChaSen (KUIC). It is based on Japanese braille transcription rules (2001) developed by the Japanese Braille Committee. To evaluate the translation accuracy of eBraille, we utilized a corpus that we created from ordinary text and braille newspaper articles.

Findings

eBraille translation accuracy is equivalent to or better than that of other stand-alone braille translation programs. This result suggests that our program achieved the goal of being applicable for practical use. In addition, we utilize our program to make Kobe University Hospital brochures in braille for outpatients and inpatients. The brochures are available in the hospital and are favorably accepted by the blind and the visually impaired. This result suggests that our translation program can facilitate accessibility to information for patients.

Originality/value

Our braille translation program is based on a client-server system and is architecture-independent. Moreover, it is a free system for creating braille text files for anyone who has access to a Web browser.

Keywords: accessibility, normalization, assistive technology, braille, visually impaired people, CGI program

1 Introduction

Braille is a reading and writing system for the blind and the visually impaired that uses six tactile raised dots. It was invented by a blind Frenchman, Louis Braille, in 1825 (Jiménez *et al.*, 2009). Worldwide acceptance of the braille system was achieved in 1878 at the Universal Congress for the Amelioration of the Blind and Deaf-mutes held in Paris (Jiménez *et al.*, 2009). In Japan, Kuraji Ishikawa, a teacher at a school for the blind and speech-impaired in Tokyo, adapted the braille alphabet to Japanese, and in 1890 the school adopted his system, which became the basis of the current Japanese braille system (Japanese Braille Committee, 2001). Japanese braille differs from English braille: the Japanese braille alphabet corresponds to Japanese *Kana* characters (phonograms), which are created from *Kanji* (ideograms imported from China about the 5th century), indicators for the alphabet to differentiate it from *Kana*, signs and characters used in Japanese, and so forth. In addition, Japanese texts are usually written by using a combination of *Kanji* and *Kana* characters, but Japanese braille is written mainly by using phonograms, that is, *Kana* characters expressing the pronunciation of *Kanji*, with certain indicators and a unique system called *Wakachigaki* (Japanese Braille Committee, 2001). The indicators are used to designate and initiate character systems other than *Kana*, such as numerals or alphabets, or to express voiced sounds, long vowels, or special sounds in the Japanese language. *Wakachigaki* is a method of writing Japanese with spaces between words or phrases to make braille texts easier to understand; it is similar to spacing between words in English and European languages, which ordinary Japanese texts do not use (Japanese Braille Committee, 2001; Kadota, 1997; Unger, 1984). Such rules for Japanese braille are standardized by the Japanese Braille Committee and are called Japanese braille transcription rules (Japanese Braille Committee, 2001). However, these complex rules are hard to learn. Also, in recent years, the shortage of braille transcribers has become a critical problem in the United States (Corn and Wall, 2002; Emerson *et al.*, 2006) and Japan.

We therefore developed a translating program for Japanese text to braille, named eBraille, so that

even people with no knowledge of the braille writing system can easily create braille documents.

2 Research Methodology

2.1 Algorithm for eBraille application and architecture

We developed eBraille ver. 1.5 as a Web-based program accessible from anywhere on the Internet (Figure 1). The core of the eBraille is a braille translation engine named KUIC (Kobe University Intelligent eBraille Engine for ChaSen). KUIC generates *Kana* sentences with spaces and converts them into braille. The latest official Japanese braille transcription rules (Japanese Braille Committee, 2001), were implemented via C or Perl languages in KUIC. To construct the eBraille server, we used Apache ver. 2.2.3 for the HTTP daemon and incorporated the Japanese morphological analyzer ChaSen ver. 2.3.3 (Matsumoto *et al.*, 1999) into our program. We installed all those applications on the Sun Microsystems Enterprise Server 3000 (six 336 MHz UltraSPARC II processors, Solaris 10, 6.1 GB RAM, 518 GB HDD).

The translation process of eBraille utilizes mainly ChaSen and KUIC, which undergo morphological analysis; transcription of input sentences in *Kanji* and *Kana* into spaced *Kana* sentences (phonograms), including correction of pronunciations and placement of indicators; and then conversion of those *Kana* sentences into braille (Figure 2). From the morphological analysis to the generation of the *Kana* sentences, transcription details occur in the following order: (1) replacement of characters for several kinds of Japanese long vowel sounds, particles that are pronounced as other written *Kana* characters, and so on, in accordance with braille transcription rules; (2) insertion of spaces in *Kana* sentences (*Wakachigaki*) with reference to the results of morphological analysis, such as Japanese part of speech; (3) correction of certain pronunciations for some words including numerals; and (4) placement of indicators and conversion of character codes. During conversion of *Kana* into braille, (5) options for braille output are prepared, whether

displaying GIF images for braille or converting characters into ASCII codes to generate a braille editor format file called BASE format. To produce braille documents, GIF images are available for tactile graphics embossers and braille styluses and slates, and BASE file is for use with braille printers.

Take in Figure 1

Take in Figure 2

2.2 Implementation of the user interface

We designed a fully automatic mode and a manual mode for eBraille translation via the Web. The fully automatic mode allows users to upload an input text file in *Kanji* and *Kana* characters and download a translated braille text file in BASE format. In the manual mode, the user interface helps users with the following two stages (Figure 2):

(1) The resulted *Kana* (phonograms) sentences are displayed on the screen for users to manually correct the erroneous translation of *Kanji* and/or fix the text segmentation (*Wakachigaki*).

(2) After users transfer the corrected *Kana* sentences to the eBraille server, both GIF images of braille and corresponding *Kana* characters are displayed. In addition, users can download braille sentences in BASE format.

For the second stage, we prepared two options for orientation of GIF images in the form of a pull-down menu: normal (non-mirror) images and mirror images. Non-mirror images are braille images in the normal orientation for reading; mirror images are braille images displayed in the reverse direction, which is convenient for writing with a braille stylus and slate as manual production of braille documents.

2.3 Evaluation of the eBraille program

We first made a corpus for evaluation of braille translation accuracy from 5,191 sentences by randomly selecting articles from a raw corpus of newspaper (“Tenji Mainichi,” Mainichi Newspaper, articles from April to July, 2002). The corpus consisted of pairs of ordinary text written in *Kanji* and *Kana* characters and braille text for each of 233 articles. The former was an input file for the translation program, and the latter was for comparison with the results of the translation. We also made a corpus using the brochures of our university hospital for evaluation.

Second, we defined braille translation accuracy (BTA) as:

$$\text{BTA} = F_{\beta} \times T$$

where F_{β} is the accuracy for text segmentation (*Wakachigaki*) and T is the percent correct translation of *Kanji*, correct indicators, and so on. F_{β} is the F -measure: the harmonic mean of precision (P) and recall (R). P is the percentage of the number of correct *Wakachigaki* units divided by the total number of *Wakachigaki* units in the program output. R is the percentage of the number of correct *Wakachigaki* units divided by the total number of *Wakachigaki* units in the braille text of our corpus. F_{β} is thus calculated by using the following equation, which is derived from the formula of van Rijsbergen (van Rijsbergen, 1979):

$$F_{\beta} = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}$$

$$\beta = 1$$

The calculation of T is as follows:

$$T = \frac{k}{K} \quad (k \leq K)$$

where k is the number of *Wakachigaki* units that are given correct pronunciation by the program and K is the total number of *Wakachigaki* units of the corpus.

$$T = \frac{K - (k - K)}{K} \quad (k > K)$$

Here, we defined the percentage of correct translation by considering the deviation from the correct number of units.

We evaluated the braille translation accuracy with our corpus as follows: first, eBraille translated ordinary text in the corpus and output braille text in *Kana* characters (phonograms). Second, we compared the output from eBraille with the corresponding braille text in the corpus for every *Wakachigaki* unit. Finally, we counted the number of correct *Wakachigaki* units and pronunciations for *Kanji*, indicators, and so on to calculate the translation accuracy. Our evaluation assistant program efficiently performed those procedures. In addition, we analyzed the translation accuracy of other braille translation programs for comparison: Système de transcription automatique en braille (Zushi, 2002), Oten-chan (ver. 4.10) (Katsunuma, 2003), ibukiTenC (ver. 0.65) (<http://www.ikd.info.gifu-u.ac.jp/ibukiTenC/>), and IBUKI-TEN (ver. 0.56) (Hyodo *et al.*, 2001).

Finally, by utilizing our eBraille program, we made braille brochures for outpatients and inpatients at our university hospital. These brochures were placed in the general reception area, the patient information center, and the reception area of the Division of Ophthalmology for outpatients and inpatients. These brochures were also used for evaluation.

3 Results

3.1 Practicality of the eBraille program

We checked the operation of the CGI program and braille images via the Web browser and examined the availability of the resultant braille file in BASE format, which is a standard for braille editors. The CGI program successfully performed its operation, and braille images were displayed on all Web browsers that we checked. We confirmed that the resultant BASE format

data could be used by most Japanese braille editors for Windows: T-editor (<http://www6.ocn.ne.jp/~t-editor/>), WINB (<http://homepage2.nifty.com/winb/index.html>), IBUKI-TEN (Hyodo *et al.*, 2001), Braille Editing System 4 (Technotools Co. Ltd.), EXTRA (Extra, Ltd.), and Braille Star (New Braille System Co., Ltd.). In addition, we have received no requests for changes to the eBraille interface for the 1.5 years since eBraille has been available via the Internet. We therefore conclude that eBraille provides a sufficient degree of practicality.

3.2 Comparison of eBraille and other braille translation programs

Comparison of braille translation accuracy

With our newspaper corpus (233 articles, 5,191 sentences), the translation accuracy of eBraille measured 91.76 BTA (Table I). In addition, analysis of the distribution of translation accuracy showed that more than 85 percent of the 233 articles had scores higher than 85 BTA. We found that articles with scores less than 85 BTA were about medicine with more technical terms or were interviews that reflected dialects, which may be the reason for the lower accuracy.

Table I shows the comparison of eBraille translation accuracy with translation accuracy of other braille translation software programs. The translation accuracy was calculated in two ways: total accuracy for all 233 articles as one text file, and means for each article \pm standard deviation (SD). The comparison showed that eBraille had a significantly higher accuracy than the other braille translation programs. The Système de transcription automatique en braille (Zushi, 2002), which was developed with reference to our program, had the lowest accuracy: 56.20 BTA. These results thus suggest that eBraille is a practical program with sufficient accuracy.

Comparison of the dictionaries

To investigate differences in translation accuracy, we examined the number of words in the dictionaries of the programs that we used for evaluation. However, we found no correlation

between translation accuracy and dictionary size (Figures 3 and 4). We then analyzed the words and phrases in the dictionaries. The analysis showed that the dictionaries of Oten-chan (Katsunuma, 2003), ibukiTenC and IBUKI-TEN included information of *Wakachigaki*. In Oten-chan dictionary (Katsunuma, 2003), the part of speech tagset is unique to this program, which is specialized for braille translation and does not follow the grammatical systems standardized in Linguistics field. A part of speech tagset denotes a system of part of speech tag classification which is used in natural language processing. The *Kana* translations of compound nouns or proper nouns in *Kanji* such as personal or geographical names are segmented in *Wakachigaki* units in the dictionary of this program. Moreover, numerical values are given to each word in the dictionary. The numerical values denote the degrees of priority to be selected as the most probable *Kana* translation of *Kanji* out of possible choices and to determine the most probable *Wakachigaki* boundary. The higher the values are, the higher the priority. The multiple choices of *Kana* translation are based on the fact that some *Kanji* has multiple *Kana* translations.

The dictionaries of ibukiTenC and IBUKI-TEN (Hyodo *et al.*, 2001) included words or segments of multiple morphemes in addition to the words in the part of speech unit. These words or segments are given braille translation notations such as where to separate within the phrases to make *Wakachigaki* units, the priority level of *Kana* translation called “cost” in Hyodo *et al.* (2001), and the parameters for occurrence of sequential voicing in Japanese. The dictionaries of these braille translation programs thus included the words and phrases that are specialized for braille translation although they do not necessarily follow the regular part of speech in Linguistics. In contrast, we use a dictionary solely with the words in a standard part of speech tagset. Thus, the structures of the dictionaries are widely different and it suggests the different translation algorithms. Indeed, the braille translation method between ibukiTenC or IBUKI-TEN and eBraille are different. In ibukiTenC or IBUKI-TEN, the incorporated segment analyzer called ibuki (Hyodo *et al.*, 2001) analyzes *Kanji* and *Kana* texts and outputs segments from the texts. The programs

further divide the segments into *Wakachigaki* units and translate them into braille by referring to the translation rules in the dictionaries or in the programs. In eBraille, ChaSen (Matsumoto *et al*, 1999) outputs morphemes from the texts and the braille translation engine KUIC connects some morphemes to make *Wakachigaki* units and translate them into braille according to the implemented translation rules.

Next we expected that the addition of medical words to the dictionary would provide higher accuracy. We then added 4,330 medical words to the dictionary of our program and we named this revised program eBraille for medicine (eBraille-M). Analysis of eBraille-M translation accuracy showed higher accuracy (Table I). This result thus suggested that accuracy could be improved by enlarging the dictionary for our program.

Take in Figure 3

Take in Figure 4

Table I. Comparison of accuracy of braille translation programs and the distribution of translation accuracy for 233 articles (5,191 sentences)

BTA score	Number of articles for each program with the BTA scores					
	eBraille	eBraille-M (4,330 medical words added)	Système de transcription automatique en braille	Oten-chan	ibukiTenC	IBUKI-TEN
95-100	59	67	0	25	2	18
90-95	114	112	0	121	25	99
85-90	31	32	0	55	77	78
80-85	23	18	1	20	70	29
75-80	5	3	4	8	40	6
<75	1	1	228	4	19	3
Mean BTA ± SD for each article	91.78 ± 4.96	92.18 ± 4.63	56.39 ± 10.97	90.26 ± 4.90	83.49 ± 6.31	89.22 ± 4.72
Total BTA (one file of 233 articles)	91.76	92.28	56.20	90.48	83.93	89.05
Total BTA for hospital brochures	89.01	91.14	61.31	88.58	79.83	83.98

3.3 Use of eBraille for creating braille brochures and evaluation of the brochures

Translating the brochures of our university hospital into braille would result in a several-fold increase in the amount of text, similar to what would happen with text after replacement of *Kanji* with *Kana* characters. Therefore, we modified the brochures by removing repetitive content and simplifying the sentences and reduced the text size. The translated, printed, and bound brochures were placed at the general information counter, general reception window, and reception window in the Division of Ophthalmology in the hospital to respond to requests from patients. They were also placed in the information center for patients and on the bulletin board on the aisle for patients to touch and read freely.

With the brochures of our university hospital, we compared total BTA scores of our programs and others. As a result, eBraille-M showed the highest translation accuracy (Table I). We then investigated the translation errors of each program. We found that eBraille-M showed the smallest number of translation errors for *Kanji* from *Kana* and symbols. This result suggested the reason why our program produced the highest BTA score. We found errors of *Wakachigaki* with compound nouns in eBraille-M outputs.

4 Related Work

Braille translation programs can be classified into proprietary programs and Web-based programs. Examples of the former are Oten-chan, ibukiTenC and IBUKI-TEN that we used for braille translation comparison. These work only on the Windows OS. In contrast, web-based programs including our program are independent of system requirements or OS. Other Web-based programs are Système de transcription automatique en braille (Zushi 2002), which was developed based on our program, and the program by Ono et al (2000). Ono et al's program

(<http://mediaeng.ibe.kagoshima-u.ac.jp/tenji/>) is rule-based and uses case-based reasoning. The case-base is made up of user corrections for braille translation errors. In addition, it is an interactive system that enables editing and provides a list of all clauses having the same ambiguity in the text for translation. Thus, their program implies that users are familiar with braille translation rules, while our program is targeted at users without knowledge of braille translation.

Another Web-based program is available for music score translation called BrailleMUSE (Braille Music Support Environment) (http://gotoh-lab.jks.ynu.ac.jp/braille_music_score_english/index.html). This program converts music symbols such as music notes or accords on staff notation into braille in BASE format in accordance with different transcription rules (The Ministry of Education, Science and Culture, 1984) from that for literary braille.

5 Discussion

We developed eBraille and made it public via the Internet to help sighted people create braille text to communicate with blind and visually impaired people. We confirmed that eBraille worked with all Web browsers. Moreover, we provided a path to traditional braille editors via BASE format file download as a program redundancy. Thus, the number of times the program was accessed and the number of page views using eBraille have increased during the year since the major eBraille version was made available to the public. We continue to work on minor updates and have provided open access to eBraille-M and to eBraille for Traditional Medicine (eBraille-TM), with an additional 17,290 words related to Japanese traditional medicine.

The dictionaries of other programs included information of *Wakachigaki* segmentation have helped it to achieve higher translation accuracy. In contrast, eBraille had higher translation accuracy with a standard dictionary by using rule-based translation by the braille translation engine

KUIC. Our result suggested that adding words to the dictionary of eBraille would likely increase its translation accuracy. In addition to eBraille-M for hospital use, we developed eBraille-TM α version for blind acupuncturists and massage therapists. Both programs are also accessible via the Internet. eBraille-TM was developed because of the need for translation of the specialized literature for professionals in Japanese traditional medicine into braille. Thus, on the basis of its expandability (related to the algorithm), its redundancy, and its higher translation accuracy, eBraille can be described as a superior braille translation program.

We developed braille brochures for outpatients and inpatients at our university hospital and received favorable comments from visually impaired patients. Blind and visually impaired patients seemed to recognize the significance of the hospital's providing braille documents. Also, because eBraille received mass media recognition seven times during 1 year, we note that development of the program has involved social responsibility.

To evaluate the availability and the significance of our program and the translated braille brochures, we sent out a questionnaire and the brochures to students and instructors in National Kobe Rehabilitation Center for the Visually Disabled. In the 30 respondents, the period of braille learning or usage is: less than one year, 16.7%; over 5 years and less than 10 years, 50%; more than 10 years, 33.3%. Their braille reading proficiency is: primary level of numerals only, 30%; intermediate level of numerals and basic 50-character *Kana* syllabary, 46.7%; advanced level of numerals and all language sounds including special sounds, 23.3%. The summary of the questionnaire with 5-point scale showed that the necessity of a Web-based braille translation program was 4.3 points average and the helpfulness of the braille translation program for users without knowledge of braille translation was 4.4 points average. The helpfulness of the translated brochures for outpatients, inpatients were 4.8 points, 4.7 points on average, respectively. These results indicate that our program is significant and our brochures are helpful for the visually impaired. The questionnaire also suggested that these braille brochures were welcomed and

evaluated favorably as part of the hospital's efforts to improve accessibility. In addition, eBraille received publicity: during 1 year, newspaper articles mentioned eBraille three times, regional or nationwide news broadcasts referred to eBraille three times, and a radio program also mentioned eBraille, which suggests that the social institutions paid attention to our program.

6 Conclusions and Future Research

As a result of our study, eBraille translation accuracy became equivalent to or better than that of other stand-alone braille translation programs. This study suggests that expansion of the dictionary size of eBraille, with additional words from certain fields, will lead to greater translation accuracy. eBraille-M, with 4,330 additional medical words, indeed showed a higher total BTA score and much higher translation accuracy for our hospital brochures (Table I). To improve the accuracy for *Wakachigaki* segmentation, we developed a statistical learning model based on Support Vector Machine (SVM) and analyzed its effectiveness. The result showed that eBraille-M had a higher accuracy than that of our model (data not shown), which means the combination of the existing natural language processing techniques and the resources of the translation rules with the enlarged dictionary still had an advantage in F-measure for *Wakachigaki*. Our statistical learning model, however, showed its availability for partial use for the segmentation of compound nouns for which the current eBraille-M does not rather perform well because some of such errors are improved. Thus, detailed analysis for our model is warranted and left for our next research.

We hope that our program would be coordinate with the medical information system in our university hospital to realize that the braille translation of medical treatment plan or informed consent is handed to visually impaired patients as needed basis. In addition, we are considering developing English braille translation program from English texts with some modification of our program.

References

- Corn, A. L. and Wall, R. S. (2002), "Training and availability of braille transcribers in the United States", *Journal of Visual Impairment and Blindness*, Vol. 96 No. 4, pp. 223-32.
- Emerson, R. W., Corn, A. and Siller, M. A. (2006), "Trends in Braille and large-print production in the United States: 2000-2004", *Journal of Visual Impairment and Blindness*, Vol. 100 No. 3, pp. 137-51.
- Extra, Ltd. "EXTRA", available at: <http://www.extra.co.jp/extra/index.html> (accessed 18 September 2009).
- Hyodo, Y., Yokohira, K., Hayakawa, S., Murakami, Y. and Ikeda, T. (2001), "IBUKI-TEN: Japanese braille translation system with error detection function" (in Japanese), *Transactions of the Institute of Electronics, Information and Communication Engineers*, Vol. J84-D-1 No.7, pp. 1102-11.
- Japanese Braille Committee (2001), *Japanese braille transcription rules 2001* (in Japanese), Daikatsuji Co. Ltd., Tokyo, Japan.
- Jiménez, J., Olea, J., Torres, J., Alonso, I., Harder, D. and Fischer, K. (2009), "Biography of Louis Braille and invention of the braille alphabet", *Survey of Ophthalmology*, Vol. 54 No. 1, pp. 142-9.
- Kadota, M. (1997), "Japanese braille tutorial", available at: <http://www.hi.sfc.keio.ac.jp/access/arc/NetBraille/etc/brttrl.html> (accessed 18 September 2009).
- Katsunuma, S. (2003), "Oten-chan", available at: <http://www17.plala.or.jp/otengan/> (accessed 18 September 2009).
- Matsumoto, Y., Kitauchi, A., Yamashita, T. and Hirano, Y. (1999), "Japanese morphological analysis system ChaSen version 2.0 manual", *Information Science Technical Report TR99009*,

Nara Institute of Science and Technology, available at: <http://hdl.handle.net/10061/3138>

(accessed 19 September 2009).

New Braille System Co., Ltd. “Braille Star”, available at: <http://www.nbs.co.jp/bsw.htm> (accessed 18 September 2009).

Ono, S., Hamada, Y., Takagi, Y., Nishihara, S. and Mizuno K. (2000), “Interactive Japanese-to-Braille translation using case-based knowledge on the web”, in Mizoguchi, R. and Slaney, J. (Eds.), *PRICAI 2000 Topics in Artificial Intelligence*, Springer-Verlag, Berlin, pp. 634-646.

Technotools Co., Ltd. “Braille editing system 4”, available at:

<http://www.ttools.co.jp/product/eyes/BES4.html> (accessed 18 September 2009).

The Ministry of Education, Science and Culture (1984), *The guide for braille music score* (in Japanese), Nippon Lighthouse, Osaka, Japan.

Unger, J. M. (1984), “Japanese braille”, *Visible Language*, Vol. 18 No. 3, pp. 254-66.

van Rijsbergen, C. J. (1979), *Information Retrieval (2nd ed.)*, Butterworths, London; available at:

<http://www.dcs.gla.ac.uk/Keith/Preface.html> (accessed 25 September 2009)

Zushi, N. (2002) “Système de transcription automatique en braille”, available at:

<http://www.muzik.gr.jp/tenji/default.asp> (accessed 18 September 2009).

Figure Titles and Notes

Figure 1. Characteristics of the eBraille system

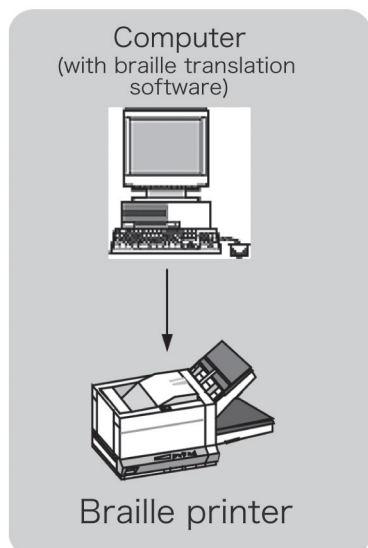
Figure 2. Algorithm for the eBraille program

Figure 3. Comparison of dictionary sizes of braille translation programs evaluated

Figure 4. Braille translation accuracy scores of translation programs evaluated

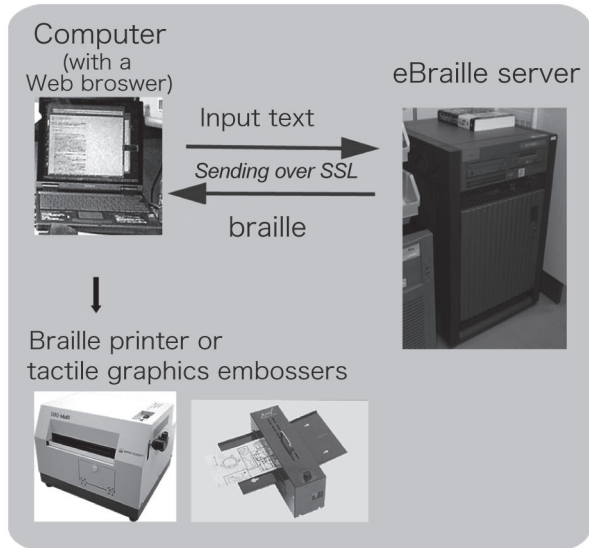
Notes: $*p < 0.05$; $***p < 0.001$; NS, not significant

Other systems



Need to install a software dependent on computer architecture and OS

Our system on the Internet



Need only a Web browser, architecture-independent

Figure 1

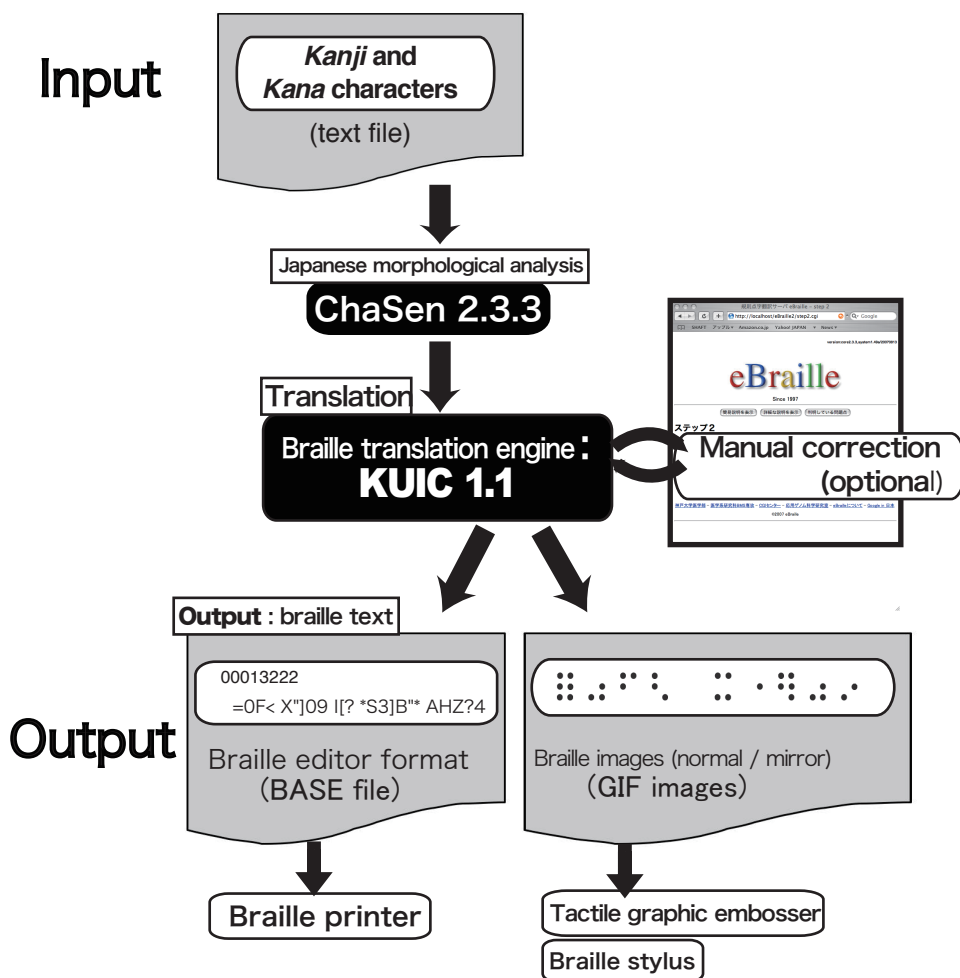


Figure 2

