Bill Gordon

November 1, 2000
Jim Breen’s Japanese Page (http://www.csse.monash.edu.au/~jwb/japanese.html) offers an abundance of resources related to Japanese language and culture. The web site’s main features include an online Japanese-English dictionary server, a frontend interface to the dictionary to translate Japanese text in web pages, an FTP archive of software and files related to primarily the Japanese language, an extensive listing of links to other web sites, and a gallery of *ukiyo-e* prints (Japanese multicolored wood-block prints).

This essay considers the effectiveness of Jim Breen’s Japanese Page and his electronic dictionary by evaluating them against standard criteria for web resources and by performing a comparison with other print and electronic resources. The first section compares the features of his electronic dictionary and translation software with print dictionaries, portable electronic dictionaries, and other web dictionaries and translation software. The second part of this essay evaluates his web site by using standard criteria such as authority, scope, organization, content, and reception by target audience. The final section discusses some unique aspects of using the web for the types of resources available on Jim Breen’s Japanese Page.

1. **Comparison With Other Resources**

Japanese-English dictionaries in print have certain useful features that have not yet been reproduced on the web. In addition to print dictionaries, since the early 1990s electronic dictionaries have gained popularity very rapidly because of their speed, portability, and unique features. Besides Breen’s dictionary and web page translation aid, other companies and individuals offer online Japanese-English dictionaries and web page translation. The first two parts of this essay section analyze the strengths and weaknesses of print and portable dictionaries in comparison to Breen’s web-based dictionary. The third part evaluates Breen’s dictionary and translation aid in relation to other web-based dictionaries and translation software.

**Print Dictionaries**

Publishers have produced many types of Japanese-English print dictionaries, with each one having a specific purpose and its own strengths and weaknesses. However, the major classifications are regular dictionaries, kanji dictionaries, and specialized dictionaries. For each of type of dictionary, one or two examples will be compared to the features in Breen’s online dictionary.

**Regular Dictionaries**

Regular dictionaries have two parts, one with Japanese words and English definitions (Japanese-English) and one with English words and Japanese equivalents (English-Japanese). Although some print dictionaries, such as the Random House Japanese-English English-Japanese Dictionary (Nakao 1995), provide only the equivalent word or expression in the other language, the primary strength of a high-quality print dictionary is providing several example sentences and phrases to show the usage of the word. This is especially important for Japanese words, since many words are used with specific particles. For both languages, a
one-to-one correspondence of words does not exist, so the lack of usage examples can sometimes lead a person to an incorrect conclusion regarding the use of a word.

Breen’s online dictionary provides only word and expression equivalents in the other language, so knowing the word does not mean a person will recognize its appropriate usage. However, many examples of usage in a print dictionary result in much bulk and weight. For example, Kodansha’s popular English-Japanese dictionary (1979) and Japanese-English dictionary (1976) take up about 1,550 and 1,250 pages, respectively. Breen’s online dictionary has about 60,000 entries, whereas the Kodansha Japanese-English dictionary has a similar number (about 50,000 entries), but Kodansha’s dictionary also provides about 60,000 example sentences and phrases. Both of these dictionaries have more entries than the typical commercial dictionary aimed at the beginning or intermediate learner of the Japanese language. For example, the Japanese-English half of the Random House dictionary has about 25,000 entries, but they take up less than 300 pages.

**Kanji Dictionaries**

As a supplement to regular dictionaries, kanji dictionaries in print have various uses such as providing individual kanji character pronunciations and meanings and giving a list of words (with their meanings) for words beginning with a specific kanji. The print dictionaries provide methods to locate a kanji such as by the total number of strokes, the number of strokes in the radical, or the pronunciation. One of the most authoritative kanji dictionaries for English-speaking users (Nelson 1974) has over 1,500 pages, about 5,000 characters, and almost 70,000 words and compound words. In comparison, Breen’s online kanji dictionary has 6,355 characters, and his online dictionary has about 60,000 regular words (not including names and technical terms).

One significant advantage of Breen’s web interface to the kanji dictionary is the ability to find words containing a specific kanji either at the beginning of a word or any place within the word, whereas a print dictionary generally only gives words beginning with a particular kanji character. As a second advantage over a print kanji dictionary, a separate interface program allows a user to search the kanji dictionary by clicking on multiple radicals contained within the kanji character. Third, the entries in the online dictionary have about 20 codes that indicate various information about the kanji character (e.g., index number for various classification schemes, frequency-of-use ranking). This advantage however applies to only a few specialists, since most users will never use any of these codes. Breen explains in the dictionary documentation that the file “has information in it which is not much use for people who are not studying and researching Japanese orthography” (Breen, KANJIDIC, 2000).

**Specialized Dictionaries**

Specialized dictionaries play an important role in communicating between English and Japanese. People who use Japanese often need a dictionary to determine how to read the kanji for a person’s family or given name and for names of places such as cities.
Breen’s online dictionary excels in names by giving users over 160,000 entries, whereas one of the best Japanese-name print dictionaries for English users (O’Neill 1989) has only about 36,000 entries. Breen’s dictionary provides great value in determining the pronunciation of names with rare or non-standard readings.

In addition to a dictionary with names, many people require dictionaries for words used in specific fields of work or study since regular dictionaries often do not include many of these technical terms. Breen has tried to add various specialized dictionary files to allow a user to search all of the files at once to try to find technical terms. For example, there are special files for words used in life sciences, aviation, finance, pulp & paper industry, computing and telecommunications, life sciences, business & marketing, law, concrete, and geology (Breen, WWWJDIC, 2000). As a test of the completeness of the business words in Breen’s dictionary files, several words were selected from a print dictionary with about 3,000 business terms (Akiyama and Akiyama 1988), but Breen’s online dictionary did not contain many of the terms. Based on this limited test, users may want to retain their specialized dictionaries in print, but Breen’s effort to gather together specialized terms from various fields should benefit many users.

**Electronic Dictionaries**

Since the early 1990s, electronic dictionaries have provided almost all of the functionality of regular Japanese-English dictionaries and kanji dictionaries. For example, the popular Canon Wordtank (Model IDX-9500) provides an Japanese-English dictionary, English-Japanese dictionary, and a kanji dictionary all in one machine. The instruction manual (Canon 1993, 64) states the three dictionaries contain a total of about 650,000 entries. For example, the Japanese-English dictionary contains about 90,000 word entries and over 50,000 usage entries. The kanji dictionary contains 6,353 kanji characters, about the same number as Breen’s online kanji dictionary. The Canon Wordtank allows the user to jump from one dictionary to another. Other valuable features include a wild card function for searching, the capacity to store up to 500 words for later review, and the capability to insert additional IC cards with specialized dictionaries for business terms and current terms. When traveling away from home or office, most English speakers who speak and read Japanese at an intermediate or advanced level prefer an electronic dictionary to print dictionaries. Many people also prefer an electronic dictionary even when print dictionaries are readily available at home or office.

The speed, portability, and capabilities of electronic dictionaries can lead to the infrequent use of an online dictionary such as Breen’s. An electronic dictionary turns on almost instantaneously, whereas it takes time to boot up a PC. Moreover, the limited bandwidth of Internet connections described by Mitchell (1995, 17, 121) five years ago continues to be a problem today for the majority of users, so the slow download times of the Internet can not yet match the almost instantaneous access to information in electronic dictionaries. The lack of portability of most Internet access devices also can discourage
users of Breen’s online dictionary. However, within the last couple of years, online access through portable devices has grown rapidly. Although few in the US access the Internet through portable devices, more than a third of the Japanese web users access the Internet using wireless technology (primarily with by Internet-enabled cellular phones produced primarily by NTT DoCoMo) (Stevenson 2000). Breen has already developed a cellular phone interface for his dictionary.

**Other Web Dictionaries and Translation Software**

Several online dictionaries exist on the web, but almost all use Breen’s dictionary files (e.g., FreeDict; Friedl 2000; Graf 1999; Heartful Dictionary; Pacific Software Publishing 2000; Palevich 1996; Peterson; Rudick 1997). Chan (2000) states that “all Web-accessible Japanese dictionaries . . . that I’ve come across seem to use Jim Breen’s Edict, Kanjidict, or XJDIC [dictionary files] as the base.” Many of these other web sites have developed software to extend the uses and functionality of the dictionary files (e.g., one site uses active HTML to display a translation of Japanese words when the mouse pointer is placed on them) or to provide their own web interface.

Eijiro on the Web (2000) is a dictionary that does not use Breen’s dictionary files. However, this dictionary is directed more to native Japanese speakers since furigana (used to determine pronunciation) is not provided for the kanji characters. Based on limited testing, this dictionary appears quite large, with numerous business-related terms not found in Breen’s dictionary. A few other web-based Japanese-English dictionaries are directed at native Japanese speakers, but none of those reviewed had the comprehensiveness of Breen’s dictionary or Eijiro on the Web.

Besides the ability to look up individual words in the dictionary file, Breen also provides software with a web interface to translate words in Japanese text on web pages or text input by the user. He makes clear in the documentation that the system does not try to translate the entire Japanese text into English. Instead, “it simply attempts to identify the words in the text and to display the translations of those words. The user is expected to know enough Japanese grammar to make sense of the results” (Breen, WWWJDIC, 2000). Breen provides links to two online text translation systems, elingo and Alis Technologies, but he does not “think much of the quality of the results.” A search of the web did not identify any other online Japanese-English text translation systems, but several companies and individuals do sell translation software (e.g., Internet Language Company sells bidirectional Japanese translation bundle for $1,395). These commercial translation systems have not been examined in this paper.

In order to test the features of Breen’s translation aid and the other two online text translation systems, a typical sentence found in the business section of a large Japanese daily newspaper (*Asahi Shimbun* October 28, 2000) was selected and then translated into English:
On the US stock market on the 27th, the Dow Jones Industrial Average, which shows the price movements in large company stocks, rose sharply because economic indicators showing a business slowdown were announced, and a sense of relief spread among investors who felt that monetary tightening will end.

Next, the three translation systems were used for this sample Japanese sentence. Breen’s translation aid seemed very easy to use, with the web page shown in the left-hand frame and the Japanese word, pronunciation (hiragana), and definition shown in the right-hand frame. However, the translation aid’s weaknesses became readily apparent with this limited test. Breen’s system did not translate several of the Japanese compound words (e.g., “stock market,” “Dow Jones Industrial Average,” “economic indicators”) in the sentence, because his dictionary contained only the individual words making up the compound word. However, the dictionary did provide the meanings of some compound words (“monetary tightening,” “large company”). The dictionary probably also has difficulty handling one-kanji prefixes and suffixes. For example, the kanji prefix for “US” (bei) is the same as the kanji for “uncooked rice” (kome), but the dictionary only showed the meaning of “uncooked rice.” Finally, Breen’s translation aid produced a couple of incorrect readings of the text. “Dow” was translated as “down.” “Spread” (hirogatta) was translated as the name “Hiroshi,” since the software program apparently did not recognize the verb ending (~gatta) of the verb for “spread” even though the Japanese verb hirogaru is in the dictionary.

Both online text translation systems, elingo and Alis Technologies, gave unintelligible results for the sample sentence. The Alis Technologies software provided the following translation:

The dhow industrial stock average by which the fluctuation in prices of the brand of the big enterprise is shown from the extension of a safe feeling that economic indicators which show the business deceleration are announced, and tighter credit ends rises greatly in the US stock market on the 27th.

The elingo system gave the following translation:

As for the American stock market on the 27th, the economic index which shows business deceleration announces When it is done, the tight money ends, the relief impression had to spread Empty and large The Dow Jones industrial stocks average which shows the value movement of brand of enterprise substantially rise.

The results from these two online translation systems suggest that it may be more useful to have only individual words translated, like in Breen’s translation aid, rather than getting back a confusing and misleading translation that may be impossible or very difficult to untangle its basic meaning.
2. Site Evaluation Using Standard Criteria

As the Internet grew rapidly in the 1990s, certain standard criteria have been developed to evaluate the quality of Internet resources. Many of these criteria mirror the ones used to evaluate the quality of print resources. In deciding on the criteria to use in evaluating Breen’s web site, the following four lists of evaluation criteria received special attention since they seem to apply directly to a site such as Breen’s, which has a more academic flavor: Librarians’ Index to the Internet (2000), Social Science Information Gateway (2000), Alastair Smith (1997), and Wesleyan University Web Literacy Class (2000). Although differences exist between various lists of criteria for evaluating web sites, most organizations and individuals agree that the criteria should include authority of the author, scope, organization/design, content, and reception by target audience. These five key criteria will be used to evaluate Jim Breen’s Home Page.

Authority

Jim Breen provides on his home page a description of his background and qualifications, which shows he has the credentials necessary to write authoritatively regarding electronic dictionaries and related language software. He serves as an associate professor in the School of Computer Science and Software Engineering at Monash University in Australia. His research and classes are in the field of digital and data communications, and he works on electronic dictionaries, especially those related to the Japanese language. He worked in the field of programming and data communications from 1968 to 1985, and he has been a university professor since 1985. He served as a member of the Australian Internet Working Group, which established the first Internet backbone in Australia.

Breen’s web site contains a disclaimer that states the home page is a personal page published by the author and not sponsored by Monash University. Despite this site being a personal home page, the site contains many resources of value to other academics and to persons (from beginners to experts) interested in the Japanese language and culture. Others recognize the author as an expert in his field, and many have praised the quality of the resources at his site (see further discussion in section on “Reception by Target Audience” below).

Scope

The top page of Jim Breen’s Japanese Page lacks a clear description of the site’s purpose and scope, but the table of contents gives the user some idea of the focus of the site. Also, the meta information in the HTML for the page has a description of “information about the EDICT/KANJIDIC/JMdict [names of dictionary files] projects, the WWWJDIC dictionary server [Japanese-English translation aid], and many links dealing with Japanese WWW resources: software, education, dictionaries, culture, literature, etc.” The site’s contents reflect this description, but the site also has an extensive archive of software and files related to Japanese dictionaries, Japanese language software, and various aspects of Japanese culture. The archive includes software for many different types of operating
systems (e.g., Windows, Unix/Linux, Macintosh, MS-DOS, and even palmtops) and includes word processing, dictionaries, educational software, and other utility programs.

This web site has multiple targeted audiences. First, the site contains much information for other professionals working on dictionary files, translation software, and Japanese-English translation. This accounts for the technical language used in much of the dictionary documentation. Second, the site assists English-speaking people who study or use Japanese from the beginning to the advanced level. Finally, Breen provides an abundance of information and resources for a more general audience who have some interest or have a specific question on Japan. The numerous links to Japan information and the Ukiyo-e Gallery target this third audience.

**Organization and Design**

Jim Breen’s Japanese Page has the appearance of a web site put together by a professor, rather than a professional organization. The pages contain almost no color, images (other than the Ukiyo-e Gallery), or side margins, and many of the pages are extremely long. For example, the front page (i.e., Jim Breen’s Japanese Page) provides over ten pages (when printed) of links to web sites relating to Japan and Japanese language information. The front page has a table of contents for the site, but most of the ten main categories have no description, and some categories have acronyms that a new user will not know. For example, one of the ten main categories is “THE JMDICT (NEW EDICT) PROJECT,” with no further description. The site lacks a map, so the user can sometimes get lost.

Despite the organizational weaknesses of the table of contents, the front page’s links, which make up about 90 percent of the front page’s content, have been logically organized into easy-to-understand categories. Breen provides clear, concise annotations for the links and marks with a “New!” image any link added within the prior month or so.

The information required for input into the web interface for Breen’s dictionary is easily understood, and the software has the capability to handle romaji (Latin letter) variations in the spelling of Japanese words. The results are displayed with a maximum of ten entries per page, and the “continue” button must be pressed to see additional entries. The display of the entries is organized well on the web page, but codes are displayed for each word without any code explanations on the display page. For example, each name is assigned at least one code, such as s (surname), p (place name), f (female given name), m (male given name), and g (given name, as yet not classified by sex) (Breen, ENAMDICT, 2000). Other words may have any one of over 40 codes, such as abbr (abbreviation) or oK (word containing outdated kanji). The user must go to the file documentation to obtain the meanings of the codes, but it is difficult to find quickly the meanings of the codes within the documentation. The addition of the codes’ meanings to the display page (especially for names) or a link directly to where the codes’ meanings could be easily found would improve the understandability of the dictionary entries. In addition to the codes described previously, each word has a two-letter code at the end to indicate the dictionary file that it originates from (e.g., LS = Life Science
dictionary file). The site’s usability could be improved by the elimination of these two-letter codes from the display page. Many other sites that use Breen’s dictionary files do not include these two-letter codes on their display pages.

**Content**

This section considers the accuracy, currency, and writing quality of Jim Breen’s Japanese Page. The entire site seems to contain highly accurate information. In the dictionary’s documentation, Breen acknowledges the corrections and proofreading performed by others to the dictionary entries over a period of several years. He encourages feedback, “I will be delighted if people send me corrections, suggestions, and ESPECIALLY additions” (Breen, EDICT Documentation).

Breen constantly works on keeping the contents of his site current and on enhancing the functionality of his dictionary files. For example, when reviewing the site, no dead links were encountered. He provides a detailed history of changes to software, files, and even the external links on the top page. A scan of these histories reveals how frequently improvements and changes have been made to the site. A couple of the many significant planned improvements include provision for inclusion of examples of word usage and provision for cross-references to related entries.

Breen writes with a clear, direct, and concise style. This makes the link annotations on his top page easy to understand. However, the site contains much technical information, especially in the documentation for the various dictionary files, so a user without a good background in programming, computers, or Japanese will find some of the explanations extremely difficult to follow. Even those people with a good background in these fields will need some time to familiarize themselves with some of the acronyms and abbreviations.

**Reception by Target Audience**

Web users frequently go to Jim Breen’s Japanese Page because of the comprehensiveness and high quality of its resources. As evidence of this, the Google search engine, which uses a sophisticated system to rank a web page based on an analysis of the quality of sites that link to the page (Google 2000), consistently ranks the pages of Jim Breen and other pages that use his dictionary files near the top of its searches. For example, a search for “Japanese” and “dictionary” shows the top ten listings of 173,000 to be based on Breen’s dictionary files. Google’s Web Directory’s “Language and Linguistics > Natural Languages > Japanese” category ranks Jim Breen’s Japanese Page as the number one resource (Google Web Directory 2000). Even Jim Breen’s Ukiyo-e Gallery outranks 8,600 other entries for a Google search on the keyword “ukiyo-e.” His dictionary files and other software have five fully-operational mirror sites around the world, which provides additional confirmation of the popularity of his resources.

Praise abounds for Jim Breen’s Japanese Page. Adjectives used to describe Jim Breen and his site include “Our hero!”, “amazing,” “outstanding,” “interesting and useful”
(University of Washington 1999; Eicher 1999; Masden; Lunde 1999). An Asahi Evening News article (Horvat), which reports on the dearth of high-quality web sites on the Japanese language, calls Jim Breen’s site “perhaps the most often praised Japan-related site.” One Japanese language organization describes his Japanese Page as “quite possibly the definitive source for Japanese language reference on the net. Jim Breen has assembled an incredible collection of resources for anyone who wishes to use Japanese on the net” (nihongo.org).

3. Unique Aspects of Using Web

The resources of Breen’s web site illustrate some of the unique aspects of using the Internet, such as the collaboration of experts from around the world and the strong interconnections between Breen’s resources and other material on the web.

Collaboration

Although Breen has assembled a remarkable collection of resource material, his site reflects in numerous ways the intensely collaborative nature of developing an electronic dictionary and of assembling a huge number of resources on Japanese-related software. O’Donnell (1998, 63-64) describes a world connected by the Internet where we can bring “together all the world's available talent to solve a given problem” and where “it is not quite clear who is the author of a collective, cumulative, and collaborative work of scholarship.” These statements accurately describe the collaborative character of Breen’s web site.

Numerous people provided materials or other support to develop the Japanese-English dictionary and other resources at Breen’s web site. Breen personally acknowledges the contributions of over 150 people who played a part in the development of the dictionary, and he describes the contributions made by several of these individuals. For example, a Japanese professor at an English university proofread the entire dictionary file. A person at Sony, who had put together a large online dictionary, contributed many entries for the dictionary. He describes the support of several other individuals who informed him of errors, provided entries, identified obscure kanji compounds, or keyed in entries. Without the Internet, this type of collaboration between experts around the world would have been impossible. Breen acknowledges this, “Virtually the entire compilation process has been carried out using electronic mail and file transfers, and indeed the project would never have occurred without the services provided by the Internet” (EDICT Information, 2000).

But Breen’s site is not the only place where collaboration occurs in Japanese-English dictionaries. In May 2000, volunteers formed a group to create a free online Japanese-English dictionary (jeKai 2000). The new dictionary tries to address some of the limitations of print dictionaries and Breen’s online dictionary files by incorporating the following features:

- definitions that explain the meaning of a words as completely as possible
- as many examples as possible of each word in real contexts
- photographs and other illustrations, especially for entries about uniquely Japanese things
• no restrictions on the type or range of vocabulary
• no restrictions on the length of entries

This dictionary currently has only about 200 entries, which limits its usefulness. However, Breen has already provided links if one of his dictionary entries has a corresponding jeKai entry.

**Links to Other Resources**

Landow (1998, 105) explains another form of collaboration in which a work on the web always exists in relation to other works in a way that print material can not do. “First, any document placed on any networked system that supports electronically linked materials potentially exists in collaboration with any and all other documents on that system; second, any document electronically linked to any other document collaborates with it.” The links from numerous other sites to Jim Breen’s Japanese Page demonstrate their dependence and interaction with the many resources found at his site. Breen’s extensive links to other resources on the same subject matter quickly shows a user that his site does not stand as an independent work. For example, as described in the previous section, Breen has links from words in his dictionary to corresponding words in the jeKai dictionary. As another example, his information page on individual kanji has links to four other databases on the web.

**Conclusion**

Jim Breen’s Japanese Page represents an extremely valuable resource for its multiple audiences: translation and software specialists, Japanese language users from beginners to experts, and persons with a general interest in Japan. Although some shortfalls can be identified in the site’s organization and design, the comprehensiveness and overall high quality of the resources and links make this site invaluable to any English speaker with an interest in the language and culture of Japan. When comparing Breen’s online resources to print books and to electronic dictionaries, both advantages and disadvantages can be identified. However, Breen’s online dictionary has no peer on the web, and his translation aid based on the dictionary files provides better results than free translation software available online.

The Internet enables authors to realize certain unique benefits that they could not accomplish in print. Many experts from around the world can collaborate on a project that could never be achieved by one person alone. Also, an author can create works independently that are interconnected through the web with other works, and this results in a collaboration between many authors that can provide users with quick access to more in-depth and more comprehensive material on a specific subject. Jim Breen’s Japanese Page exemplifies these unique benefits of web-based material.
Works Cited


We have been thinking a lot about how we connect with our clients, what content we create and curate, and how we share it. It seemed such an easy task at first - we haven’t considered we would need to create a special strategy for this sort of things.