New Japanese-English Character Dictionary.

A Semantic Approach to *Kanji* Lexicography

Jack Halpern

1. The Study of *Kanji*

The Japanese script consists of two phonetic syllabaries, called *hiragana* (eg か /ka/) and *katakana* (eg カ /ka/), and thousands of Chinese characters, called *kanji* (eg 本 *hon*). Chinese characters have three basic properties: form, sound, and meaning. Many characters are of complex shape, some having more than twenty or even thirty strokes. Each character may be pronounced according to its Chinese derived *on* reading, or to one of several native Japanese *kun* readings, and each reading may be associated with one or several (sometimes highly) polysemous words and/or morphemes. Moreover, because of the presence of numerous homophones, the Japanese script is highly complex and requires considerable effort to learn.

Japanese has been the subject of various linguistic studies, but little attention has been given to the systematic analysis of its writing system. Kanji are combined with each other to generate countless compound words from a basic stock of a few thousand units. They function as a network of interrelated parts, not as a set of disconnected symbols. Though this is vaguely recognized by educators, it has been largely disregarded in the development of teaching programmes and the compilation of character dictionaries. The traditional approach places emphasis on the rote memorization of characters and compounds, while the meanings and functions of individual characters are often ignored.

Although the number of Japanese language learners has more than quadrupled over the past decade, a character dictionary for non-Japanese users analysing the semantics of kanji has never been compiled. The few existing works are either out-of-date, contain many inaccuracies, or fail to show the meanings and functions of characters on the morphemic level. The lack of effective tools to overcome the difficulties posed by the Japanese script has been one of the principal reasons for the relatively small number of foreigners to have truly mastered the language.

2. New Concept Dictionary

2.1. Dictionary Aims

The *New Japanese-English Character Dictionary* (Kenkyusha, Tokyo 1990) is a completely new work that provides an in-depth understanding of the meanings and functions of all high-frequency characters in contemporary Japanese. It presents 4,421 entry characters (including 834 cross-reference entries) and approximately 60,000 senses for some 42,200 lexical items, which covers about 99% of the characters used in newspapers. Modern linguistic theory and computational lexicography were integrated to produce a work that meets the needs of a broad range of users, including learners, teachers, linguists and developers of teaching materials and courseware. This dictionary aims to serve both as an efficient tool for looking up

characters and as a practical learning aid that provides insight into how kanji function as a system.

To achieve these aims, the dictionary departs from tradition in three important ways: (1) clear and complete character meanings are arranged in a manner that shows their interrelatedness, (2) a new indexing system is used that enables the user to locate entries with great speed and facility, and (3) precise distinctions are given between closely related characters (synonyms and homophones).

Figure 1. Part of entry for 破

2.2. Character Meanings

The most significant contribution of this dictionary is the in-depth treatment of meaning. The principal features that contribute to clarity of meaning are:

1. The **core meaning**, a concise keyword representing a character's central concept, links the principal senses into one conceptual unit. For example, BREAK, the core meaning of 破 (Figure 1), shows how such seemingly unrelated senses as ❶❻ «defeat» (BREAK out) and ❷ «defeat» (BREAK the enemy) are actually variants of the same concept. The core meaning, a unique feature of this dictionary, serves as an effective learning aid in five ways: (1) it concisely conveys the character's fundamental meaning, (2) it serves as a central pivot linking the single senses into an integrated conceptual unit, (3) it provides an instant grasp of the meaning and func-

tion of the character as a morpheme, (4) it is easy to memorize, and (5) it enables the user to grasp the differences and similarities between kanji synonyms.

2. The dictionary departs from traditional character dictionary lexicography by ordering meanings *psychologistically*, rather than historically. A keyword serves as the basis of organization, and the senses are grouped in clusters in a manner that shows their interrelatedness by allowing them to be conceived as a logically-structured unit. As can be seen from Figure 1, the distinct senses of 破 are presented in a manner that clearly shows their differences and similarities.

3. The **character meanings** consist of precise translational equivalents, accompanied by various labels and glosses, that show how each character functions as a free or bound morpheme. The senses are presented in a manner that enables the learner to understand each character's meanings and functions in the generation of compound words. A system of sense division numbers shows the hierarchical and logical relations between main senses and subsenses.

4. Each sense is illustrated by numerous high-frequency **compounds and examples** that provide maximally useful examples of each sense. Unlike conventional character dictionaries, the compounds appear under the senses which they illustrate, enabling the user to know how they are formed from their constituents (Figure 1).

5. If the semantic relation between compound components is not obvious from the way in which the compounds are grouped by meaning, a **compound formation article** describes how each component contributes to the meaning.

2.3. New Indexing System

The lack of an efficient kanji indexing system has long been a source of frustration to character dictionary users. Traditional systems have not achieved the speed and simplicity required to meet learner needs. This dictionary introduces a new scheme, the **System of Kanji Indexing by Patterns (SKIP)**, that enables the user to locate entries as quickly and accurately as in alphabetical dictionaries. The system is based on four easy-to-identify geometrical patterns as shown below:

No.	Pattern	Examples			
1	LEFT-RIGHT	相 4-5	八 1 1	代 2 3	
2	UP-DOWN	高 2 8	声 3 4	赤 3-4	
3	ENCLOSURE	2-sides 迫 3-5	3-sides 同 2 4	4-sides 国 3 5	
4	SOLID	1 top line 下 3-1	2 bottom line 上 3-2	3 through line 中 4 3	4 others 九 2-4

Figure 2. System of Kanji Indexing by Patterns

Characters belonging to the first three categories are referred to as *divisible characters;* characters that cannot be divided by SKIP rules, called the *indivisible characters*, are classified under pattern ■ 4, (solid). The caracters are arranged in ascending order of hyphenated numerals called the *SKIP number*, as illustrated below:

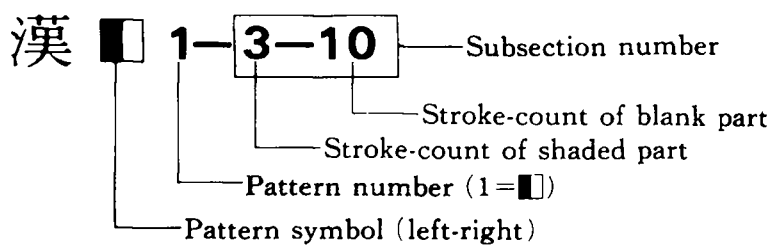

Figure 3. SKIP Number of Divisible Characters

To locate a divisible character, the user first identifies the pattern to which it belongs to determine the first part of the SKIP number, then divides it and counts the strokes of each part to determine the second and third parts. For example,漢 can be divided into left and right parts and is thus classified under pattern ◧ 1. Since it contains three strokes in the shaded part (氵) and ten strokes in the blank part (莫), its SKIP number is ◧ 1-3-10 (Figure 3).

To locate an indivisible character, the user classifies it under pattern ■ 4 (solid), determines the total stroke-count of the character to find the second part of the SKIP number, then identifies the solid subpattern to find the third part. The indivisible character 卜, a three-stroke character containing a top line, is thus classified under pattern ■ 4, subsection 3-1, as shown below:

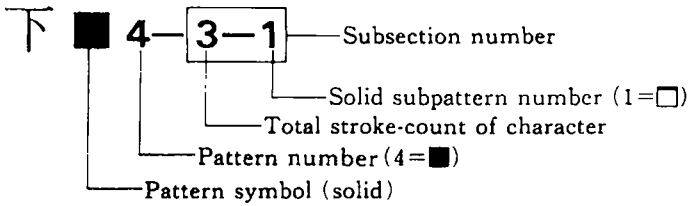

Figure 4. SKIP Number of Indivisible Characters

The character entries in the main part of the dictionary are ordered according to their SKIP numbers. Divisible characters sharing the same SKIP number are divided into subgroups containing a shared element (such as 扌 and 木) for maximum lookup speed. The characters within each subgroup are further subdivided into progressively smaller groups until each character is assigned its own position. The entry characters are listed in the same order in the **Pattern Index** (Figure 5), so that the user can either (1) quickly locate entries in the index or (2) locate the entry directly without using the index.

		瀬	808
		瀧	809
3-17	扌	壌	810
	女	嬢	811
3-18	扌	擶	812
		攝	813
3-20		黴	814
3-21		囑	815
3-22		灣	816
4-1	扌	札	817
	礻	礼	818
4-2	扌	朴	819
		机	820
		朽	821
		权	822

Figure 5. Pattern Index

SKIP is a product of seven years of computer-assisted research and experimentation on how kanji elements are intuitively perceived. Since it can be learned in a very short time and does not require prior knowledge of kanji elements, many users and educators consider it an important breakthrough in Chinese character lexicography.

2.4. Precise Distinctions

2.4.1. Kanji Synonyms

This dictionary presents a detailed analysis of **kanji synonyms** for the first time, which is especially useful to the advanced user and linguist. Kanji synonyms along with their English keywords and cross-references are given for the principal senses of main entry character (Figure 6). The headword at the beginning of each group expresses the semantic relationship between the group members, and thus shows how they resemble each other, while the keywords show how the members differ from each other. The main sense relationships are class-inclusion (superordinate followed by hyponyms), synonymy, part-whole, and complementarity.

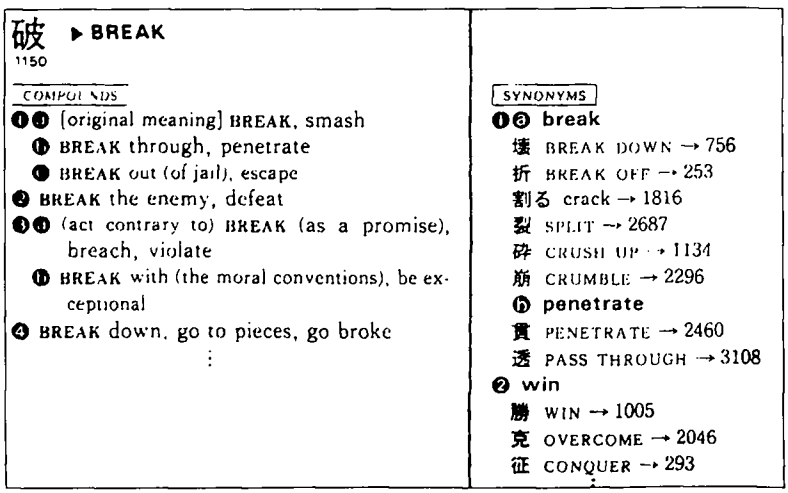

Figure 6. Kanji Synonyms for 破

In the above table, the core meaning and English equivalents show that the central concept represented by 破 is BREAK. Referring to sense ❶❸ of the SYNONYMS section, which corresponds to sense ❶❸ of the COMPOUNDS section, we find the headword **break**, which indicates the meaning *shared* by the group members. By comparing the individual keywords, such as 壊 BREAK DOWN, 折 BREAK OFF, and 裂 SPLIT, we can see how they differ from each other. Studying the core meaning, English equivalents, headwords, and keywords together provides a good understanding of the distinctive features of each group member. The user can gain a deeper understanding of each character by turning to the entry numbers indicated by the cross-references, where detailed meanings and examples for each group member appear.

2.4.2. Usage Notes

Japanese has a large number of *kun* homophones, which are native Japanese words pronounced alike but written differently and usually differing in meaning. Since many homophones are often very close or even identical in meaning, they are a source of confusion to Japanese and non-Japanese alike.

Easily Distinguished		Easily Confused	
hashi		*noboru*	
橋	bridge	上る	go up (steps, a hill)
端	end, edge	登る	climb, scale
箸	chopsticks	昇る	ascend, rise (up to the sky)

Figure 7. *Kun* Homophones

Homophones in Japanese present the following difficulties:

1. There are many homophones in Japanese.
2. The differences between homophones are often subtle and confusing.
3. There are numerous orthographic variants.
4. Usage is sometimes contrary to expectations.
5. Writing in Japanese often involves uncertainty over orthography.
6. The treatment of homophones in existing works is inadequate.

To help the user overcome these difficulties, this dictionary presents, for the first time in English, hundreds of **usage articles** that provide complete guidance on the discrimination of all one-character *kun* homophones in current use. By comparing the English equivalents, the user can accurately grasp the differences and similiarities in shades of meaning and in usage between easily confused homophones.

丘
3495

USAGE

oka

丘

[sometimes also 岡] *hill, hillock, mound*

岡

① [usu. 丘] hill, hillock, mound—used chiefly in proper names

② [sometimes also 傍-] [in compounds] outsider, third party, bystander

陸

land, shore

傍-

[usu. 岡] outsider, third party, bystander

★丘 and 岡 have the same meaning. The former is used in both common nouns and place names, while the latter is used chiefly in the writing of proper names.

HOMOPHONES

oka ⇨ 岡 2997 陸 543 傍 147

Figure 8. Usage Note

3. Other Features

Various other features help the user acquire an in-depth understanding of kanji:

1. Six retrieval methods make the dictionary a highly efficient lookup tool.
2. A wide range of character forms, styles and readings make the dictionary a convenient tool for reading Japanese.
3. Such features as ample data on orthography and synonym/homophone discrimination make the work an effective manual for writing Japanese.
4. The entry-head lists useful reference data, such as stroke order diagrams, a full range of forms and readings including Chinese, radical data, etc.
5. A system of labels indicate style, function, level of formality, etc., and distinguish frequent meanings from rare/archaic ones, free forms from bound forms, etc.
6. Typographical devices and labels indicate the degree of importance of each sense for four levels of study. Frequency statistics enable the beginner and advanced student to use the dictionary with equal facility.
7. A network of cross-references directs the user to a wealth of useful information.
8. Eleven appendixes provide quick access to ample reference data, such as frequency lists, historical tables and orthographic rules.
9. A visually attractive layout makes the dictionary highly user-friendly.
10. A semantically classified list of kanji synonyms makes it possible to

use the dictionary as a simple thesaurus and to locate characters from their core meanings or keywords.

4. Compilation Methods

4.1. Editorial Policy

This dictionary is committed to the descriptive approach and modern linguistics. Although statistical data and hundreds of reference sources were consulted, character meanings were extracted from actual occurrences. Unlike other character dictionaries, usage is recorded as it occurs in the living language, not on the authority of other works. This led to some unexpected results. We have found, for example, that the heavy reliance of existing works on classical Chinese resulted in their inclusion of many archaic meanings, while newer meanings are often missing. An important contribution of this dictionary is the presentation of accurate and up-to-date meanings, especially many senses that have never appeared in previous works.

4.2. Semantic Analysis

Character meanings were determined by an exhaustive semantic analysis that underwent revision by experts in Japanese, English, and Chinese. The meanings were analysed by such techniques as componential analysis and the study of near-synonyms, with the aim of achieving a structured presentation in which the character meaning is divided by numbered divisions and subdivisions to give the various senses a clear and cogent order.

The analysis of character meanings was inseparably linked to the analysis of kanji synonyms. As Ladislav Zgusta points out (*Manual of Lexicography*, 1971, p. 98), the study of near-synonyms and the analysis of their differences is one of the most outstanding duties of the lexicographer, because, he asserts, one cannot really know the precise meaning of a word if it is not examined in comparison and in contrast with its near-synonyms. In this dictionary, the analysis of character meanings was firmly based on such a policy. Each meaning was carefully analysed down to its single senses, and the connotation and range of application of each component of the meaning and its semantic relationships to other members of the same synonym group were analysed and compared. The analysis of *kanji* synonyms thus served as a powerful technique for establishing character meanings of high precision and clarity.

4.3. Spinoffs

This is the first kanji dictionary edited and produced entirely by computer. The project has cost more than US $ 1,600,000 and required some 66 man-years for completion over a period of sixteen years. Great pains were taken to ensure accuracy and up-to-dateness: some 700 programs were written specifically for editing and proofreading the data, while Japanese language experts in the U.S. and Japan have reviewed the manuscript and confirmed its accuracy. The dictionary will serve as a standard

source of data for a series of spinoff projects such as learners' dictionaries, reference manuals and courseware. Four projects are now under way: a German edition, a concise edition, a synonyms dictionary and a CALL application. (Readers are invited to submit project proposals.)

The dictionary has attracted worldwide attention among educators for being the first systematic treatment of kanji compiled by computational lexicography on the basis of a firm theoretical foundation. Because of its many distinctive features, it promises to become a standard reference work for the study of Japanese.