既存の漢字索引の効率性の分析およびコード化にもとづいた索引の開発

ヴォロビヨワ・ガリーナ
キルギス民族大学コンピューター技術・インターネット学部　上級日本語講師
国立国語研究所　招聘研究員

キーワード　漢字索引，効率指数，選択係数，漢字のコード化，漢字コードのデータベース

１．漢字索引の種類

漢字辞典の使い方が非漢字系の日本語学習者にとって複雑であることは周知のとおりである。漢字辞典を引く際には一般に活用されている部首索引，総画索引と音訓索引が使用されている。しかし，伝統的な部首による検索法は，部首の抽出が分かりにくいこともあり，総画索引を使用する場合は画数を数える際に間違える危険性がある。また，同画数の漢字がたくさんあり，複雑である。そのため，漢字辞典を引く際には，一般に活用される上記の索引以外にも漢字圏でも,非漢字圏でも多様なタイプの索引が構築され使用されている。例えば,

・「ロシアのグラフィックシステム」にもとづく「漢字の五段排列」（ロゼンベルグ 1916),
・四角号碼（諸橋 1984),
・カタカナ字形分類索引（加納 1998),
・書き出しパターン索引（加納 1998),
・筆順索引（若尾&服部 1989),
・字形索引（坂野・池田・品川・田嶋・渡嘉敷 2009),
・System of Kanji Indexing by Patterns “SKIP”（Halpern 1988),
・Key Words and Primitive Meanings Index (Heisig 2001),
・Kanji Fast Finder (Matthews 2004),
・Index by Radicals (Hadamitzky&Spahn 1981)などである。

２．漢字索引の効率指数

これらの多様な既存の漢字索引については，各々の効率の評価と比較分析が必要であると考えた。漢字索引の効率の比較にあたって，本研究ではコンピュータデータにおける処理の効率を表す「選択性」(selectivity)という概念を用いることにした（ヴォロビヨワ 2009)。(http://www.akadia.com/services/ora_index_selectivity.html)
そのため漢字索引の効率指数「選択係数」(CS, Coefficient of Selectivity)という概念を以下のように定義した。CS=V/N×100%。
ここでＮは索引に入っている漢字の数で，Ｖは索引の中で漢字が所属するグループの数である。グループは同じ部首に所属する漢字群，画数が同じ漢字群などである。例えば総画索引の場合，Ｖは索引に含まれる漢字の最大書記素数，部首索引の場合，Ｖは部首の種類の数である。これまで開発された索引と比較するため，ここでは常用漢字のみを扱うことにする。1945字の常用漢字群に含まれる漢字の画数によるグループは 23 であり，使用される部首の種類の数は 201 であると明らかになった。

例　総画索引　V=23，N=1945，CS=23/1945×100%=1.2%
　　部首索引　V=201，N=1945，CS＝201/1945×100%=10.3%

選択係数の計算をもとに既存のタイプの漢字索引の効率を比較評価した。その結果，漢字の構造にもとづく索引の選択係数は 1.2～15.4%と低いことが明らかになった。その理由は一般の漢字索引は主に一つのみの漢字の要素か性質にもとづいていることである。例えば，部首索引は部首だけ，総画索引は画数だけにもとづいている。カタカナ字形分類索引と書き出しパターン索引は最初の書記素の形だけにもとづいている。その結果，個々の部首や総画数，書き出しの書記素にはたくさんの漢字が属していることになる。

読みと意味に基づいた索引の選択性は高くても，効果的に調べるためには予めたくさんの漢字の読み方か意味を覚える必要がある。

3. 漢字のコードにもとづく新しいタイプの索引の開発

筆者はその評価と比較分析を検討し，選択性が高い漢字索引の開発の案を出した（ヴォロビヨワ 2009，2011)。それは字体を適切に表す漢字のコード化にもとづく選択性が高い索引である。新常用漢字の 3 種類の漢字コードのデータベースを構築し，3 種類の文字・数字のコードデータを辞書編集上の順番で並べ替え，漢字のアルファベット・コード索引とシンボル・コード索引とセマンティック・コード索引を開発した。この新しいタイプの索引を使用すれば，調べるための労力は表音文字の辞書と同等のものとなり，漢字辞典の調べ方をより効果的にできる。また，その場合は，漢字の構造に対する学習者の理解が深くなり，機械的な覚え方から解放されることも期待できよう。しかし，このコードのシステムを習得するためには努力が必要である。新しいタイプの索引を使用するには特に大事なことは漢字の書記素と筆順及び構成要素の知識である。

参考文献

ヴォロビヨワ・ガリーナ（2009)「選択性が高い漢字索引の開発」『日本語教育方法研究会誌』Vol. 16 No 1, 72-73

ヴォロビヨワ・ガリーナ（2011)「構造分析とコード化に基づく漢字字体情報処理システムの開発」『日本語教育』149 号 207-214

加納喜光（1998A)『常用漢字ミラクルマスター辞典』小学館

坂野永理，池田庸子，品川恭子，田嶋香織，渡嘉敷恭子（2009)『イメージで覚える「げんき」な漢字 512』The Japan Times

諸橋轍次(1984)『大漢和辞典』大修館書店

ロゼンベルグ・オ.(1916)『五段排列漢字典』(O.Rosenberg. *Arrangement of the Chinese Characters according to an Alphabetical System being the Introduction to a Japanese Dictionary of 8000 Characters and List of 22000 Characters*) 東京 興文社

若尾俊平，服部大超（1989)『くずし解読字典』栢書房

Hadamitzky, W.＆Spahn, M. (1981) *Kanji &Kana Revised Edition A Handbook of the Japanese Writing System* Tuttle Language Library

Halpern, J. (1988/1990) *New Japanese-English Character Dictionary*. Tokyo: Kenkyusha.

Heisig, J. (1977/2001) *Remembering the Kanji*. *Vol. 1*. Tokyo: Japan Publications Trading Co. Ltd.

Matthews, L. (2004) *Kanji Fast Finder* 漢字早引き辞典 Tuttle Publishing

http://www.akadia.com/services/ora_index_selectivity.html（Selectivity）