

電子辞書の多漢字コンテンツ作成について

野村英登 (財団法人交流協会)

電子出版でもっとも成功しているのが電子辞書であると言われているが、IC 電子辞書、CD-ROM、オンラインなどのどの媒体でも、漢和辞典などの中国古典の研究では必須のコンテンツが必ずしも十分でない。他方、XML を利用した電子出版のフォーマットが世に出はじめ、汎用性のあるデータ作成が可能な状況が実現しつつあるので、専門性の高い多漢字電子辞書コンテンツ作成を研究者が関わって行う可能性が見えてきたように思われる。

Multi-Kanji Characters of Digital Dictionary

NOMURA Hideto (Interchange Association)

It is widely accepted that the digital dictionary is most succeed case of digital publishing, but some contents like Kanji dictionaries are not enough to study Chinese canon. Another, new digital dictionary formats based XML are spread and it will be realized to create Multi-Kanji Characters dictionary which has technical knowledge.

1 はじめに

電子出版では、もっとも成功しているのが電子辞書だとされる。辞書の基本的性格が本来読むものでなく使うものであるため、高度な検索が可能となる電子化は辞書の性格によく合っているため、当然の結果と言えるだろう。電子辞書の形態には、専用の IC 電子辞書やソフトウェアとしてパソコンにインストールして使うもの、オンラインでデータベースのように利用したり、ウィキペディアのようにインタラクティブに変更が加えられていく辞書など様々だが、本報告では、特に携帯できる漢和辞典の電子化の現状について報告し、発展の可能性を検討したい。

2 IC 電子辞書の場合

IC 電子辞書の発展は近年目覚ましく、紙の辞書が完全収録されるようになっただけでな

く、中国語・フランス語・ドイツ語などの多言語化も進み、一般の学習者から専門家・研究者まで幅広く利用されるようになり、量・質ともに向上してきている ([4]、第5章)。

2.1 字数限定版

さて、小学館の『中日辞典』を搭載した電子辞書の登場により、近現代の中国の文章を読む上で電子辞書の有用性はずいぶん高くなったが、従来、漢籍を対象とする古典研究者にとっては、電子辞書はあまり実用的ではなかった。というのも文献を読むための最初の一步となる漢和辞典が、元の紙の辞書から収録されている親字が削られ、JIS第一、二水準の範囲内に限定されてしまっていたので紙の辞書の代替として利用できなかったのである。例えば、漢和辞典として標準的に搭載されている学研の『漢字源』の場合、紙の辞書で親字が 10,028 字あったのが、電子版では 6,400 字弱に削られていた。熟語もそれに準じて収録数が減ること

になる。

2.2 フルコンテンツ版

ただこうした状況は最近になってようやく改善されつつある。現在、カシオ¹、キヤノン²、シャープ³、セイコー⁴、ソニー⁵の各社から販売されている電子辞書のうち、漢和辞典が搭載されている製品については、セイコーの大修館の『新漢語林』を搭載したSR-E10000 以外は、すべて学研の『漢字源』を搭載している。これらの電子辞書のうち、キヤノンのwordtank V80、G55、G50、V30 が搭載する『改訂新版漢字源』電子辞典版については、書籍版の親字 12,600 字に学研監修の 512 字を加え、合計で JIS 第三、四水準をカバーする 13,112 字を収録している。またセイコーの『新漢語林』モデルは元の紙の辞書と同等の 14,313 字、熟語約 50,000 語を収録し、現在流通している電子辞書でもっとも漢字を収録した辞書となっている。

3 CD-ROM 版漢和辞典

3.1 EPWING フォーマット

パソコンでローカル利用できる漢和辞典については、元データが EPWING のような標準的なフォーマットで作成されていれば、ザウルスや Palm などの PDA で専用のビューワを使うことにより持ち運ぶことができる。ただし、EPWING フォーマットの電子辞書には、岩波の『仏教辞典』⁶のような専門性の高いコンテ

ンツもあるものの、漢和辞典については学研の『漢字源』⁷があるのみで、しかも収録文字数は JIS 第二水準までと少ない。EPWING の文字コードが JIS X 0208 のようなので、外字処理の煩雑さを嫌ってのことだと思うが、漢和辞典としての魅力は少ない。

3.2 独自ソフト方式

一方、『CD-ROM 版字通』⁸は、付属の Windows 用のビューワでしか利用できないものの、親字約 9,500 字、熟語約 220,000 語を収録し、熟語数では国内最大の辞書である。JIS 第一、二水準を超える外字部分は TTF フォントの切り替えで対応しているため、コピー & ペーストでの利用に向いていない。

また『e康熙字典 日本語版』⁹は収録親字数 4,7000 字と最多を誇るが、本文は原典の画像データで表示されるため¹⁰、資料的価値は高いが辞書ツールとしてはほとんど使えないといっている。

漢和辞典でなく、漢-漢辞典になるが、香港で発売された『漢語大詞典 (光盤繁體單機 2.0 版)』¹¹は、親字 18,013 字、熟語 336,385 語を収録し、規模の上ではもっとも大きい。ただし、日本語の Windows では動作に難があり工夫が必要となる¹¹。

以上の三つの辞典については、コンテンツの内容は豊富だが、Windows に利用環境が限定されており、携帯も困難である。

¹ <http://www.casio.co.jp/exword/> 2005 年 7 月 1 日閲覧 (以下同じ)

² <http://www.canon-sales.co.jp/wordtank/>

³ <http://www.sharp.co.jp/e-dic/>

⁴ <http://www.sii.co.jp/cp/>

⁵ <http://www.sony.jp/products/Consumer/DD/>

⁶ <http://www.iwanami.co.jp/BOOKS/13/0/1301490.html>

⁷ http://www.epwing.or.jp/lineup/kokugo/k_04.html

⁸ <http://www.heibonsha.co.jp/jitsuu/>

⁹ http://www.sanseido-publ.co.jp/publ/ep/cd_other_kouki.html

¹⁰ http://www.unihan.com.cn/Jp/JPcpjs_Kx.htm

¹¹ <http://butterflylost.net/index.cgi/HanYuDaCidian>

4 オンライン漢和辞典

4.1 データベース

インターネットを自由に使える環境にあれば、漢籍の読解に際しては、定評ある漢籍の全文データベース、例えば台湾の中央研究院が公開するもの¹²などを利用して、その場その場で用例検索をする場合が多い。しかしことあるごとにデータベースの海に潜るのは煩瑣だし、検索結果が多くありすぎて必要な用例を捜しあてるのに苦労したり、逆に用例が今読んでいる文献のその箇所しかないという場合もままあり、辞書に頼りつつ一字一字検討していく作業は欠かせない。このためオンライン辞書の必要性が減じることはないだろう。

4.2 オンライン辞典

オンラインの漢和辞典については、有償コンテンツとしては『漢字源』¹³と『字通』¹⁴がある。前者は一般的なJIS第一、二水準までの限定版、後者はCD-ROM版と同等の内容だが、外字は画像で置き換えている。

無償のオンライン漢和辞典としては、慶応大学の千田大介氏による著作権の切れた角川の『字源』を電子化したKO字源¹⁵が仮公開されている。また同氏は、漢-漢辞典になるが、オンライン『學生字典』プロジェクト¹⁶も主宰している。文字コードがutf-8なのでほとんどテキストベースで利用できる。

漢-漢辞典としてまた無償のオンライン辞書では、漢和ではなく漢漢になるが、台湾教育部（日本でいう文部科学省）が提供する『國語辭

典』¹⁷は、親字 13,575 字、熟語 160,000 語を収録し、現代から古典まで幅広くカバーし使い勝手がよい。ただし文字コードがbig5なので中国語対応の環境でないと利用できない。

これらオンラインの辞典については、通信環境の良否が足枷となる他にも、文字コードの問題をクリアするのに手間がかかるし、携帯性はずいぶん劣ることになる。ただ『字通』については、文字コードがEUC-JPと日本語環境に限定されていることが幸いして、利用可能な端末の選択肢が広い。Javascript対応のフルブラウザがあれば携帯電話からでも利用することもできるだろう。

5 IC 電子辞書で多漢字を

5.1 BBeB で電子辞書

電子辞書を販売している各社のうち、ソニーの取り組みがもっとも興味深い。ソニーが現在販売している電子辞書は、同社が提唱する電子ブックの標準規格BBeB規格¹⁸のうち、BBeB Dictionaryフォーマットに準拠した電子書籍端末という位置付けとなっている。したがって他の電子書籍端末、LIBRIe¹⁹でも電子辞書コンテンツが利用可能であるし、またBBeB Dictionary専用ビューワーソフトとコンテンツを同梱したCLIEも発売されている²⁰。

BBeBのフォーマットでは、XMLを中間フォーマットとしてコンテンツの主要部分をテキストで記述し、配布にあたって著作権保護機能をほどこしたバイナリ形式に変換している（[3] 図1参照）。BBeBの著作権保護機能はOpenMGを利用した強力なもので、まったく

¹² <http://www.sinica.edu.tw/~tdbproj/handy1/>

¹³ http://books.bitway.ne.jp/online/mt-ao_detail/trid-brws/ccid-07/cont_id-00020001.html

¹⁴ <http://moji.japanknowledge.com/>

¹⁵ <http://wagang.econ.hc.keio.ac.jp/zigen/>

¹⁶ <http://wagang.econ.hc.keio.ac.jp/xszd/wiki.cgi>

¹⁷ <http://140.111.34.46/dict/?open>

¹⁸ <http://ps.canon-sol.jp/bc/bc-kikaku.html>

¹⁹ <http://www.sony.jp/products/Consumer/LIBRIE/>

²⁰ <http://www.sony.jp/products/Consumer/PEG/PEG-TH55/feat5.html>

何の制限もかけない状態から利用期間 / 範囲、印刷範囲 / 回数などを限定するところまで幅広く設定できるようになっている ([3])。

BBeBのコンテンツを作成するためには専用のオーサリングツールが必要となる。LIBRIeとWindows用PCビューワで読むための、BBeB Bookフォーマット、つまり電子書籍についてはツールが市販されており²¹、多漢字コンテンツ制作の実現性がある ([5])。他方、BBeB Dictionaryフォーマットについては、ソニーとライセンス契約を結んでツールの提供を受ける必要があり、学術利用が可能かどうかはソニーと交渉する必要がある。

しかしながら、簡単なコンテンツであれば、最近市販された製品に同梱されているソフトウェア『辞書工房』²²を利用することで、自作の電子辞書コンテンツを作成できる。このソフトで作成した辞書コンテンツは、ソフトが同梱されている機種以外の他のBBeB Dictionaryフォーマット対応端末、もちろんLIBRIeでも利用できる。

ただし、『辞書工房』では Shift-JIS 以外の文字は「=」に置き換えてしまうので、多漢字コンテンツの作成には向いていない。BBeB 規格の標準文字セットは 14375 文字あるし [3]、専用のオーサリングツールの方では何らかのかたちで外字処理ができるようになっているはずである。

5.2 XMDF で電子辞書

辞書コンテンツの作成については、シャープが推進している電子書籍の形式、XMDF (モバイル・ドキュメント・フォーマット)²³でも

可能である。BBeBより先行して展開されてきたXMDFは、マルチプラットフォーム路線を推進していて、XMDF形式の電子ブックの閲覧は、WindowsPC、PDA、携帯電話と様々なハードで可能であり、携帯できる電子ブックフォーマットとしては現在もっとも普及しているといつてよいだろう。

XMDF も、中間フォーマットを XML とすることで、データの汎用性を高め、バイナリ形式で配布する際に著作権保護機能を追加している。著作権保護は暗号化とフットプリントによる発信元の特定に止まっている。

外字については、XMDF の場合、仕様上は文字コードとして utf-8 を採用しているものの、幅広い端末での利用を考えてオーサリングツールの方で Shift-JIS 以外の文字を使用できないようにしている。必要な文字はビットマップフォントを作成し、外字タグとして埋め込む必要がある ([1]、図 2 参照)。

オーサリングツールとしては、シャープが開発した GUI により XMDF 作成を行える『XMDFビルダ』があり、デジブックジャパンがライセンス販売している²⁴。こちらも出版社等を前提としたライセンスなので、学術利用を考える際には交渉が必要となるだろう。

XMDFでは、ニンテンドーDSの『DS楽引辞典』²⁵にも採用されているように、電子辞書も作成できる仕様になっている。IC電子辞書では上位機種のシャープのPW-C8000²⁶でしか利用できないが、携帯電話で利用できるコンテンツを作成すれば携帯性は非常に高くなるだろう。

²¹ <http://ps.canon-sol.jp/bc/>

²² http://www.sony.jp/products/Consumer/DD/KENSAKU/kensaku_0.html#koubou

²³ <http://www.spacetown.ne.jp/menu21/books/what/xmdf.html>

²⁴ <http://www.dbook.co.jp/>

²⁵ <http://www.nintendo.co.jp/ds/adjj/>

²⁶ <http://www.sharp.co.jp/e-dic/lineup/product/pw-c8000/index.html>

6 . まとめ

電子辞書において元々の紙の辞書で定評のあるものが優位に立っていることから明らかだが([4]、第5章) 基礎的なツールである辞書については、常に批判の目を向けずに済むようすでに評価を経たものである方がより望ましい。ウィキペディアなどの試みは既存の辞書を補完するものとはなっても、本質的に辞書の中心とはなりづらい。したがって、古典研究のような専門性の高い辞書が長年に渡っていくつも出版されてきた分野においては、電子化にあたって新たに最初から辞書を作成していくことも今後の研究の発展を考えると確かに重要だが、著作権保護を考慮されたフォーマットを利用して、既成の辞書を電子化していくことができれば、研究者にとってのメリットはその方が大きいと思われる。

参考文献

- [1]北村義弘・岩崎圭介・田中秀明「電子出版とXMDF技術」、『シャープ技報』第84号、2002年12月
- [2]シャープ株式会社「XMDF電子書籍フォーマット」
- [3]ソニー株式会社 e-book ビジネス推進室「「BBEB規格」概要説明 Ver.1.00」
- [4]高木利弘『電子書籍ビジネス調査報告書2004』、インプレス、2004年
- [5]野村英登「リブリエできるかな 多漢字電子ブックの試み」、『東洋学へのコンピュータ利用第16回研究セミナー』、2005年3月

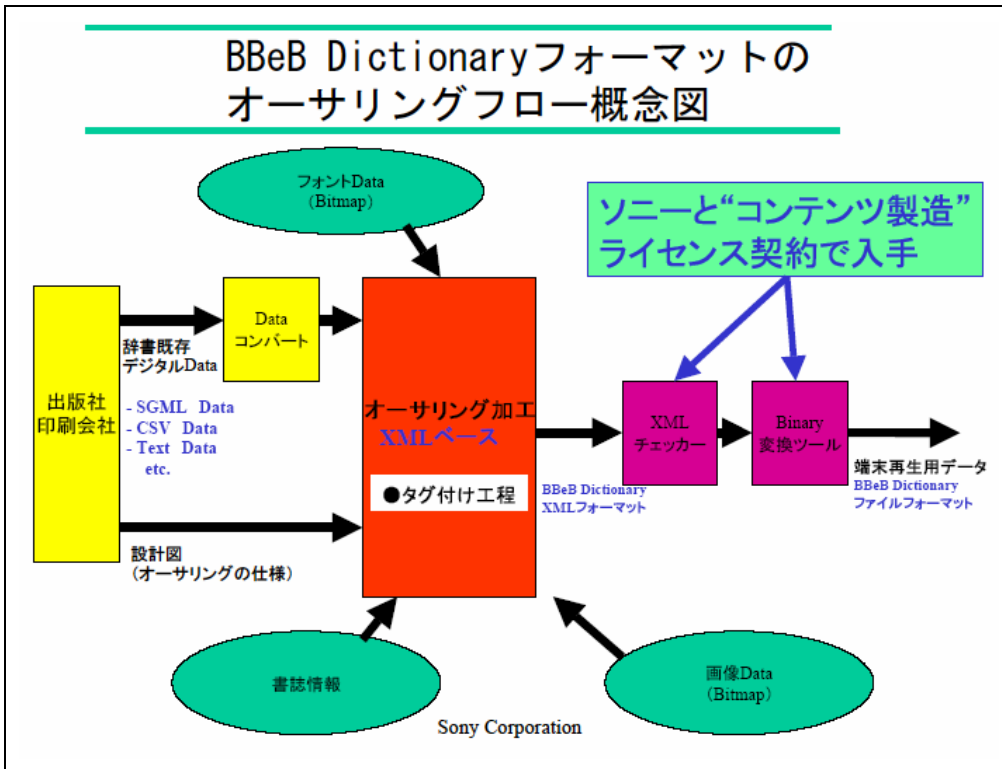


図 1 : BBeB Dictionary フォーマットのオーサリング ([3])

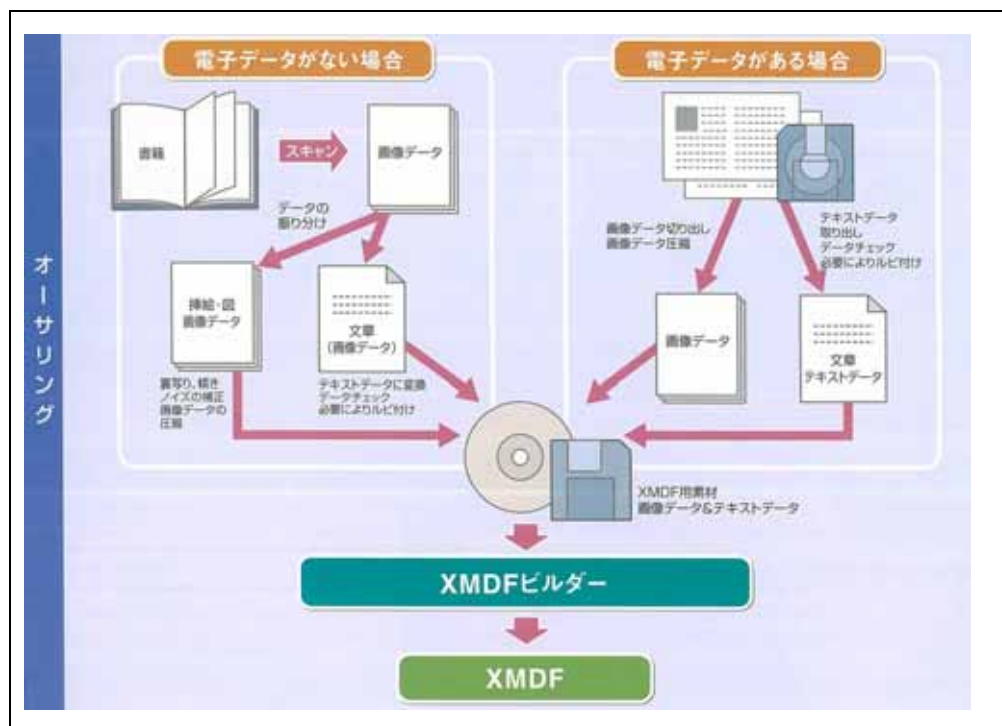


図 2 : XPDF のオーサリング ([2])