博士（文　学）　高　田　智　和

学位論文題名

漢字情報処理のための基礎的研究

学位論文内容の要旨

第１章で研究の目的と方法が述べられる。明治以降、『和玉篇』や『康熙字典』の流れをくむ辞書が数多く編纂され、その中に上田万年・岡田正之・飯島忠夫・栄田猛猪・飯田伝一編『大字典』がある。和訓を多く採録し、親字に通し番号を付すなど新たな試みを遂げているが、その評価や検証は未だ多くなされていない。国語学の分野では、調査に携帯され、索引作成や漢字字体処理に用いられる等の経緯があり、その利用価値は今なお失われるものではない。

本論は、『大字典』の掲出字（18,000 字）を取り上げ、漢字集合としての規模と質の検証を試みる。方法として、『大字典』の全掲出字をデータベース化する。これによって、『大字典』を検証し、文献の電子化における問題点について考察している。『大字典』諸版も対象とし、『大字典』改版史を記述する。近代漢和辞典を考察する上での『大字典』の研究であると同時に、将来に向けての電子化テキストの利用を視野に入れつつ、漢字情報処理のための基礎研究をなすことを目的としている。

第２章で『大字典』改版史が記述される。

『大字典』の版種は 13 種を数え、改版によって文字の追加や熟語の増補を行なっている。大規模に掲出字を増補しているのは、縮刷版（大正 9[1920]）と復興版（大正 13[1924]）である。大正期における『大字典』の改版は、文字の収集と増補が第一目的であり、これが改訂の基本方針である。全体の頁数を 2,602 頁に保つなど、出版上の制約のもとで掲出字の増補が行なわれていることからも、文字増補に重点を置いていたことが裏づけられる。

縮刷版の改訂は、漢籍を読むための漢和辞典を目指した文字増補がなされ、復興版の改訂では、当代の中国で使用している文字を含めて、「漢字」で書かれた文献を読むための漢和辞典を作り上げようとしている。中国と日本で共通する漢籍の文字用法と、日本でのみ行なわれてきた文字用法に加えて、当代中国での文字用法までをも一冊の辞書で記述しようとする試みである。漢字文化圏の古典から当代までを包含しうる辞書を実現しようとしたのである。

文字の増補は番号を付与しない掲出字の記載形式をとり、番号つき掲出字には手を加えないのが原則である。掲出字に対する文字番号は検索の効率化を図って付与されたものであるが、一々の文字を他と区別する文字番号の意義を『大字典』編者はおおむね理解していたと思われる。華語増補版（昭和 15[1940]）の附録「華語時文辭典」には、本編掲出字の文字番号を付し、検索機能に加えて参照機能も見出している。文字番号をコードとして利用する、すなわち文字番号コードブックの萌芽であるとしている。

第３章で『大字典』データベースをつくっている。

『大字典』掲出字の符号化は、主としてJIS X 0208:1997（JIS漢字）で行い、JIS包摂規準等の適用による符号化文字集合の最大限の運用を試み、『大字典』を電子化テキストに再構築した。

データベース作成によって把握しえた『大字典』の特徴は、『康熙字典』との相違、活字文化の反映、「國字」「異体字」に対する門戸の拡大の3点である。『康熙字典』との相違は、『大字典』が漢籍を読むための漢和辞典以外の側面を重視していたことの証である。文字の採取先として、国書を漢籍同様に扱ったことから生じた差異である。辞書の収録字数に比して、「國字」が多数あることもこれに起因する。活字の字体や「俗用」の「異体字」の採録は、規範よりも記述重視という編纂理念からくるものと考えられる。この点で、『大字典』は記述的性格の強い辞書であるといえる。「俗用」の文字、あるいは日本的な用法・用字まで記述しようとしたため、『大字典』は書名に「漢和」を名乗らなかったのであろうか。漢和辞典の草創期において、国書や「國字」、「俗用」の「異体字」へと広く門戸を広げたことが、漢和辞典史上の『大字典』の意義である。

第4章で『大字典』データベースをつかう実際の処理が記述されている。

コンピュータによる漢字処理では、「異体字」の扱いが普遍的な命題である。本論は、「異体字」処理に対して、漢字シソーラスおよび字種と字体の二重コードの運用を提唱し、『大字典』コードブックによる漢字処理の実践を行っている。現在は字体による整理が主流であるが、「異体字」関係にある複数字体をひとつにまとめた字種の概念を導入することによって、漢字を字種と字体の面から立体的にとらえると共に、資料間での照合を効率的に行なうことができるようになる。

『大字典』の掲出字に、字体コードにあたる文字番号のほかに、複数字体をひとつにまとめた字種コードを与える。『大字典』18,000字を字種を基準に再編成すると、12,500字種の漢字集合となる。これを基礎集合として、「図書寮本類聚名義抄掲出字索引」「石塚漢字字体資料」「郵便報知新聞社説コーパス」を処理すると、ほぼすべての用例を扱うことができる。漢字処理における、字種と字体の二重管理および『大字典』12,500字種の有効性を実証し得たものと考えられる。

学位論文審査の要旨

主　査　教　授　石　塚　晴　通
副　査　助教授　池　田　証　壽
副　査　助教授　弥　　和　順

学位論文題名

漢字情報処理のための基礎的研究

本論文は、国語学者の経験を生かして編纂された辞書である上田万年・岡田正之・飯島忠夫・栄田猛猪・飯田伝一編『大字典』について、改版史を検証して、その全掲出字を取り上げ、漢字集合としての規模と質の検証を行ったものである。全掲出字をデータベース化し、『大字典』の検証を行い、文献の電子化における問題点を考察し、漢字情報処理のための基礎的研究を行っている。

本論文で丹念に検証された『大字典』の改版史は、長年に亘り広く流布した辞典であるが故に、量的多大さと散逸の容易さとが伴う困難を排して、手際よく記述された労作である。

作成された『大字典』データベースは、掲出字の符号化を主としてJIS X 0208:1997（JIS漢字）で行い、JIS包摂規準の適用による符号化文字集合の最大限の運用を試み、『大字典』を電子化テキストに再構築したもので、極めて有用なものである。

このデータベースの活用により、漢字の基礎集合として、種々の歴史的文献の漢字処理が可能となり、今後漢字字体論・漢字集合論・辞書論等の分野で、有効に活用されて行くものと判断される。

以上より、当委員会は、全員一致して、本論文を博士（文学）の学位授与に相応しいものと認定する。