

日本語教育のための電子化辞書構築

吉岡 亮衛, 小松 幸廣, 坂谷内 勝

国立教育研究所

及川 昭文

総合研究大学院大学

CASTEL/J(Computer Assisted System for TEaching & Learning / Japanese: 日本語教育支援システム) は, 教師と学習者の両面から日本語教育を支援するために開発中のシステムである。本開発研究は, 1987年度文部省科研費による開発に端を発し, 各種辞書データベースおよびテキストデータベースの開発と拡充を行ってきた。また, ソフトウェアに関しては, プロトタイプシステム, OS/2版, マッキントッシュ版と開発を進めてきた。今回, 本システムをCD-ROM化して一般の利用に供する準備が整ったので, 開発の経緯を報告するとともに, データベースの概要について紹介する。

To Build a Electric Dictionary for Teaching Japanese

Ryoei YOSHIOKA Yukihiro KOMATSU Masaru SAKAYAUCHI

National Institute for Educational Research

Akifumi OIKAWA

The Graduate University for Advanced Studies

CASTEL/J(Computer Assisted System for TEaching & Learning / Japanese) is a newly developed system, which supports both teachers of Japanese and learners. This project has been continued since 1987 by Grant-in-Aid for Scientific Research of the Ministry of Education, Science and Culture. Until today, several databases of dictionaries and text have been developed. And the software has been revised two times. In this paper, the details of project and the outline of dictionary databases are reported.

1. はじめに

1. 1 CASTEL/Jの開発の歴史

日本語教育をめぐる状況および社会的環境は、近年急速に変化している。文化庁の調べによれば、1976年（昭和51年）に10,889人であった国内の日本語学習者は、1986年（昭和61年）には35,767人、1991年（平成3年）には62,895人となっている。また、国際交流基金の調査では、海外の日本語教育機関で学ぶ日本語学習者数は、1990年（平成2年）には約98万人となっており、この数は、10年前の約2.5倍となっている。

このような急激な日本語学習者数の増加に対応して、1985年度には、はじめて国立大学に日本語教員養成課程が設置された（筑波大学日本語・日本文化学類、東京外国語大学日本語学科）。また、日本語教育の水準の向上をめざして1987年度（昭和62年度）から日本語教育能力検定試験が実施されている。

本研究の発端となっている研究は、1987年度より3ヵ年の計画で始めた文部省科学研究費補助金（以下科研費と呼ぶ）試験研究「パソコンによる

外国人のための日本語教育支援システムの開発」

（代表者 宮本繁雄：国立教育研究所）である。

そこでは、当時の日本語教育のニーズに応えるべく、日本語教育教師不足、日本語教育教材の不足を改善することをめざし日本語教育支援システムの開発研究を行った。具体的には、各種の日本語教材データベースの構築と、データベース管理および利用プログラムの開発が中心であった。

先の研究が科研費の研究課題として採択された1987年以降に、日本語教育に関連する研究は急増する傾向が見られた。これは、日本語教育についての社会的な関心の高まりと日本語教員養成機関における研究者の充実によると考えられる。その結果、1993年度（平成5年度）には、科研費の分科細目のひとつとして「日本語教育」が登場するまでになっている。

その後、再度科研費試験研究を1991年度（平成3年度）より3ヵ年計画で行った。具体的な研究内容は、先の試験研究（1987～1990）で開発したデータベースの量的拡大と質的拡充、および、プロトタイプシステムのソフトウェアを改良し実用

表1 研究の経過

1987.04	科研費試験研究No.62890010（代表 宮本）（～90.03） 「パソコンによる外国人のための日本語教育支援システムの開発」研究開始
1987.04	科研費海外学術No.62041145（代表 及川）（単年度） 「海外における外国人のための日本語教育の実態に関する調査研究」研究開始
1988.04	科研費海外学術No.63041156（代表 及川）（単年度） 「海外における外国人のための日本語教育の実態に関する調査研究」研究開始
1988.10	第1回研究発表会開催（於：教育会館）
1990.03	第2回研究発表会開催（於：教育会館）
1990.04	科研費国際学術研究No.02044153（代表 瀧澤）（～93.03） 「日本語教育・学習支援システムの開発」研究開始
1991.04	●日本語教育支援システム研究会発足
1991.10	科研費試験研究No.03559010（代表 浅木森）（～94.03） 「マルチメディアを利用した日本語教育支援システムの開発」研究開始
1992.08	第3回研究発表会開催（於：金沢工業大学）
1994.02	第4回研究発表会開催（於：国立教育研究所）
1994.04	科研費国際学術研究No.06044221（代表 菱村）（～97.03） 「日本語教育用音声・画像データベースの開発研究」研究開始 科研費総合研究A No.06301098（代表 及川）（～97.03） 「日本語教育支援データベース流通促進のための総合的研究」研究開始

的なシステムを開発すること、さらに音声と画像のデータベースを取り扱うマルチメディアに対応した試験的システムを開発することであった。

その結果、辞書データベースおよびテキストデータベースに関しては当初の予定通りの拡充を行った。また、画像に関してはイラスト画のデータベース、音声に関しては単語の発音のデータベースを構築した。ソフトウェアに関しては、アップル社のマッキントッシュ上で動作するシステムとして、データベース管理および検索が行えるようになった。

これまで二度にわたり試験研究を行ってきたが、その間に1987年度（昭和62年度）、1988年度（昭和63年度）に科研費海外学術研究（学術調査）の、1990年度（平成2年度）から3年間に渡っては国際学術研究（共同研究）の助成を受けて、日本国内ばかりでなく海外における日本語教育機関を実際に訪問し、現地の日本語教育の現状について訪問調査を行った。また、海外の日本語教育機関に在籍する研究者と共同して、現場で必要とされる日本語教育支援システムのデータベースおよびソフトウェアの機能について検討し、プロトタイプシステムの評価を受けた。表1は本研究の経過である。

現在も2つの科研費を得て、研究は継続している。それぞれの研究目的は、総合研究Aは、CASTEL/Jを実際の教育現場あるいは研究に利用可能なように、データベースをCD-ROM化し、流通を図るべく環境および条件の整備を行うことである。また、今後研究を進めていく予定になっているマルチメディア情報に関連して、著作権問題をクリアした日本語教育教材を確保するために、科研費国際学術研究により自前の映像情報、音声情報の収集を行っていく計画である。

1.2 CASTEL/Jの概要

CASTEL/Jを開発するに当たって最も留意した点は、「コンピュータのハードウェア、ソフトウェアのレベルアップにも柔軟に対応できる」、「利用者に対応した利用が可能である」ということで

ある。この結果、CASTEL/Jには他の類似のシステムにないいくつかの特色が見られる。これらのうち主なものを紹介すると以下ようになる。

(1) データベース・オリエンテッドなシステム

この種のシステムの多くは、ソフトウェア・オリエンテッドなものとなっているが、CASTEL/Jは、データベース・オリエンテッドなシステムとなっている。従来はソフトウェアの性能、機能等の仕様を決定し、それに合わせてデータやデータベースの構造等を決めるというソフトウェアを中心としたシステムであった。したがって、データベースはソフトウェアに依存する形でつくられることになり、同じデータベースを他のシステムで利用するという事は多くの場合困難である。これに対してCASTEL/Jはまず日本語教育および学習に必要なデータベースはどのようなもので、コンピュータ上で活用されるためにはどのような構造を持つべきかという観点から、データベースの仕様を決定し、次にそれらを利用するためのソフトウェアの仕様を検討するという開発手順をとった。これにより、データやデータベースのソフトウェアからの独立性が保証され、他のシステムへの移植等も容易に行えることになる。

(2) ユーザのニーズに応じたデータベース構築

教師が必要あるいは作成しようとする教材は、その教師自身の教授法や採用している教科書、カリキュラム等に基づいたもので、100人の教師がいれば、100の異なる教材が存在することになる。また、学習者にとっても、学習レベルや学習環境によって、必要とする教材は異なってくる。したがって、すべての教師や学習者が満足するような教材データベースを作成することは、ほとんど不可能である。CASTEL/Jは、教師や学習者が必要とする教材を、それぞれの要求に基づいて作るのに十分な教材源（辞書、用例文、試験問題、新書・台本等のテキスト等）をデータベースとして構築し、それらを利用するための道具（ソフトウェア）を合わせて提供しようとするものである。こ

れによって利用者は、自分のニーズに合った教材を作成することができるようになる。

等の印刷物
→ビデオや語学テープ、CDやLD等の磁気媒体

(3)データベースの構築と利用の分離

本システムで取り扱う教材データベースは、多種多様かつ大量のデータである。そのため、データベースの構築には大型汎用コンピュータの環境が不可欠である。一方、教材データベースの利用（利用者の日常的な利用）は、ワークステーション（パソコンを含む）で行う形態をとった。それゆえ、本システムは、教材データベースも含めて可搬性の高いものとなる。

(2)素材からの分類

- ・紙に印刷されたもの
- ・ビデオテープ
- ・語学テープ
- ・CD、LD
- ・郷土玩具など「物」そのもの

(3)情報の形態からの分類

- ・文字や数字等 → テキスト情報
- ・絵、写真やスライド等 → イメージ情報（静止画）
- ・ビデオ、映画等 → イメージ情報（動画）+音声情報
- ・語学テープ等 → 音声情報
- ・CD、LD等 → マルチメディア情報

2. 辞書データベース

2.1 全体的な説明

現在、日本語教育で実際に利用されている教材にはもともと教材として作成されたものだけでなく、小説、新聞記事、レストランのメニュー、新聞・雑誌等の広告等、実に多種多様である。これらの教材をいくつかの観点から分類すると以下のようになる。

このようにいくつかの観点からの分類が可能であり、データベースを構築する場合には、それぞれの特性を十分に検討しておく必要がある。特に、作られたデータベースがどのように利用されるのか、あるいはどのように利用したいのかということについては、時間をかけて検討しておかなければならない。

(1)利用形態からの分類

- ・教師自身が独自の教材を作成するために利用するもの
 - 辞書・事典類、用例集、試験問題集
- ・学習のために利用するもの
 - 教科書、辞書・事典類、小説、新聞記事

CASTEL/Jでは、データベースデータは、タグ番号と項目値を1レコードとする形で構造を統一し

表2 辞書データベース

電子化辞書名	備考（データ量および材料）
漢字辞書	約7千字（JIS第一・二水準の漢字）
漢字筆順フォント辞書	約4千字（JIS第一水準漢字）
単語辞書	●基礎日本語学習辞典（約3千語） ●PAX和英辞典（約3万語） ●学術用語（約10万語）
用例辞書	約6千用例（基礎日本語学習辞典中の用例）
音声辞書	約3千語（基礎日本語学習辞典の見出し語）
画像辞書	約3千語（基礎日本語学習辞典の見出し語）

ている。本システムで構築したデータベースは、大きくわけて辞書データベースとテキストデータベースの2つになる。辞書データベースには、表2のようなものがある。本章では、構成される漢字、単語、用例の各データベースについて概説し、マルチメディアに対応した音声・画像辞書については、次章で述べる。

2. 2 漢字データベース

CASTEL/Jの漢字の漢字データベースには、JIS第一・第二水準のすべての漢字6,349字が収められている。それぞれの漢字には、次のような属性項目が付いている。

漢字ID番号, シフトJISコード, JIS漢字コード, JEF漢字コード, 画数, 部首番号, 漢字レベル1, 漢字レベル2, 漢字レベル3, 音読み登録数, 訓読み登録数, 人名読み登録数, 音読み, 訓読み, 人名読み, 熟語, コンポーネント(漢字), コンポーネント(コード), 属性, 異体字, 筆順フォント情報

CASTEL/Jでは、内部的に漢字ID番号を決定して、各漢字に付与している。ID番号は、ユニークな番号となっており、常用漢字表の漢字に掲載順に番号を振ったあと、それ以外のJIS第一水準漢字のJISコード順に付け、さらにJIS第二水準漢字のコード順につけた。この番号は、CASTEL/Jで決定したもので、外との対応関係は上述の関係以上のものはない。

漢字レベルについては、わが国での学習の順序、初級500漢字、分野別使用頻度の3種類がある。

漢字の読みについては、角川の「漢和中辞典」を参照して、データのチェックを行っており、人名読みも、同辞典の人名要覧にもとづいて読みを付与している。また、各読みには、ローマ字読みを付けた。

CASTEL/Jでは、ローマ字は発音を表すものと考え、特に長母音の表記が通例とは異なっている。これは外国人学習者に配慮したものであり、

CASTEL/Jのデータベースのひとつの特徴である。

熟語は、単語データベース中に収録された単語の中で漢字を2字以上含むものについて、その含まれる漢字に対して単語IDを付与した。

コンポーネントは、部首以外の漢字の部分に着目して検索することを可能にするためのものである。ある漢字の一部が漢字として存在するものと、漢字として存在しない部分で特徴的な形(コード)のものがある。

属性(造語成分, 否定語)は、それぞれ特別な働きをする漢字に対する情報である。

2. 3 単語データベース

単語データベースのデータベース項目には、次の項目がある。

単語ID番号, 漢字見出し, 仮名見出し, ローマ字見出し, 音訓許容外, 使用漢字レベル, 語源, 品詞, 自立語品詞区分ID, 意味分類番号, 語義, 語義英語, 単語レベル, 専門分野, 合成語, 用例, 用法説明, 用法説明英語, 参照語, 反対語

データベース化の作業は、最初に「基礎日本語学習辞典」を基に行った。重要とされる項目をすべて満たしている見出し語は、「基礎日本語学習辞典」の中に2,905語あった。それらの見出し語に対する全項目の内容の見直しを行い、データベース化を行った。

単語ID番号は、CASTEL/J独自のID番号であり、仮名見出しのあいいうえお順に付与した。この番号はCASTEL/Jのシステム内でユニークである。見出し語として、漢字だけではなく、平仮名見出し、ローマ字見出しを付けることとし、漢字見出しには普通の表記のとおり片仮名語は片仮名を表記した。一方、平仮名見出しではすべて平仮名で表記した。これは、平仮名から学習を始める学習者に配慮するためである。前節で述べたように、ローマ字はあくまで読みを表すものとして考えているので、ローマ字見出しはCASTEL/Jのローマ字表

記にしたがって表記している。

日和(ひより)、大和(やまと)など、常用漢字表の音訓以外の慣用的な読みは、音訓許容外としてマークを付ける。ただし、先の2,905語には、これに該当するものはなかった。使用漢字レベルは、漢字見出しの1字ずつについて、対応する漢字のレベルを示すものである。語源は、特に外来語などで、由来する国名が分かる場合に付与することとした。

ひとつの品詞に2つ以上の字義が区別されている場合には、いくつかの項目は繰り返しとなる。意味分類番号は繰り返しのカウンターの役割を果たす項目である。

CASTEL/Jでは品詞は、「基礎日本語学習辞典」に習って、名詞、代名詞、I型動詞、II型動詞、III型動詞、形容詞、形容動詞、副詞、連体詞、接続詞、感動詞、助動詞、助詞、接頭語、接尾語、連語の16種類に分類されている。自立語品詞区分IDは、活用語の活用語尾の変化のパターンを分類し、自動的に活用語尾を発生させることを意図したものである。

語義には、単語の意味が入り、語義英語には語義の英訳語が入る。基本的にCASTEL/Jでは、英語を対照語としてデータベースを構築しており、特に単語データベースには、英訳語の未整備な語は登録されていない。

派生語は、名詞や副詞がサ変動詞となるものには「～する」の形を入れている。合成語は、見出し語とともに合成語となる単語が単語データベースの中に存在した場合に、その単語のIDを入れることとした。

普通、辞書には見出し語の用例がついている。つまり、見出し語に付随する形で用例が扱われているわけである。CASTEL/Jでは、データを効率よく利用するために用例は用例データベースとしてまとめて構築している。そこで単語データベースの用例の項目には、用例データベース中の用例ID番号が入っている。

用法説明は、単語の使用法の説明であり、用法説明英語はその英訳である。これらのデータは、

基にしている辞書に載っている場合にのみそれを取り込んだ。

参照語、および、反対語は、見出し語に対応するそれぞれの単語の単語IDが入っており、それらの語が単語データベースの見出し語に無い場合には付けられることはない。したがって、印刷物の辞書にある「～を見よ」の～が存在しないというケースは皆無となる。

2. 4 用例データベース

用例を単語辞書から分離するメリットは、ひとつひとつの用例が有効利用できるようになることである。つまり、ある用例には複数の単語が使用されており、その用例はひとつの単語の用例にとどまらず、複数の単語のどの用例にもなりうるはずである。したがって、少ない用例数でもたくさんの用例があるように感じられるようにすることは可能である。もちろんたくさんの用例が整備されれば、印刷物の辞書では経済的な面から載せきれない量の用例を引き出すことが可能となる。用例データベースのデータベース項目は以下の項目である。

用例ID番号, 用例文, 英訳, 文属性1, 文属性2

現在、用例データベース中には、「基礎日本語学習辞典」から抽出した6,427用例が蓄積されている。すべての用例文には、平仮名読み下し文を付け、さらにCASTEL/Jのローマ字表記にしたがったローマ字文を付けてある。

収録する用例は、必ず対訳(英訳)がついているものを収録することになっているので、先の実例文はすべて対訳付である。文属性項目は、用例を属性検索できるようにするためのもので、現在、文の完成度と諺について情報を入れている。今後、目的の用例を検索する手掛かりとして役立つものを順次整備拡張していく予定である。

3. 音声・画像データベース

文字データだけでは伝えることが困難な事柄、あるいは音声や画像情報があったほうが理解しやすい事柄は多く存在する。特に日本語の学習者にとっては、多様な情報が提示されたほうが理解しやすいことは言うまでもない。そこでCASTEL/Jでは、マルチメディア化の第一歩として、国際交流基金の「基礎日本語学習辞典」の見出し語（約3,000語）といくつかの用例について、音声辞書と画像辞書をデータベースとして構築した。データベース構築に際しては、CASTEL/Jの流通を考えて、標準的な機器構成で利用できること、データ量はCD-ROMの容量内に納まることの2つの要件を満足させることを重視した。

3. 1 音声・画像データベースの構築要件・方法

(1)音声データベース

日本語教育での利用を考えた場合、音声辞書データは、良い音質で標準的な発音が取り込まれることが重要である。そのため、音声データのデジタル化については、CDと同等の音質（サンプリング速度44.1kHz、量子化幅16bit）で行うこととした。この場合のデータ量は1分で約5MB、ステレオでは倍の10MBとなる。標準的なアナウンサーの会話速度は一語平均1秒弱であり、これをもとに計算すると音声辞書に登録する情報の録音時間は約50分となる。ステレオ録音では、音声辞書だけでCD-ROMが満杯となるため、必要最小限にデータ量を止めるために、特に音像をイメージすることが必要な場合以外はモノラル録音とした。これによって全体の音声データ量は、約300MB程度となる。

音声データの収録には、プロの女性アナウンサーを起用した。また、アクセントの確認には「NHK編日本語発音アクセント辞典」（日本放送出版協会）を用いた。音声は、DAT(Digital Audio Taperecorder)を使ってデジタル・オーディオ・テープに収録し、そのデータをオーディオ・カード(Sound Designer II)を装着したMacintosh

Quadra800に入力し編集、加工を行った。

(2)画像データベース

今回の画像データは、静止画像を対象とした。そのため、動作や状態も視覚的に分かりやすく表現されていることが重要となる。そこで対象を的確に表現できるイラストを多用した。また、教師が教材作成時に素材として印刷や加工利用が容易に行えるようモノクロ表現とした。

原画の作成は、イラストレータに依頼した。見出し語のイメージ化には、基礎日本語学習辞典の用例の情景を手掛かりとし、具体的な画面構成はイラストレータに任せた。

画像データの制作は、手がきのあるものをスキャナーを使ってパソコンに取り込む方法と、ペイント系の描画ソフトを使って直接パソコンに入力する方法の二つの方法がある。方法の選択はイラストレータ自身に任せた。

イラストの大きさはB6版(180×128mm)を基本とし、画素は72dpiとした。このため、線画表現のイラスト1枚のデータ量は、平均20KBである。画像データは全体で約50MB程度となった。

3. 2 音声・画像データベースの構造

(1)音声データベースの構造

音声データベースの項目は以下の項目である。

音声ID番号、音声漢字見出し、音声仮名見出し、音声ローマ字見出し、音声英訳、品詞、用例ID番号、音声データ

音声データベースの項目中で、音声ID番号は音声データに付けられたユニークな値である。この値は「基礎日本語学習辞典」の見出し語の順(ローマ字名称のアルファベット順)になっている。また、音声漢字見出し、音声仮名見出し、音声ローマ字見出し、音声英訳、品詞の各項目は、音声データベースを単独で管理することを考慮して、単語辞書の対応項目をコピーしたものである。用例ID番号は、音声データが用例である場合に対応する用例データベースのID番号が付けられる。音

声データ項目は、音声データファイルと連携をとるための項目で、音声データファイル名が入る。

(2) 画像データベースの構造

画像ID番号、画像漢字見出し、画像仮名見出し、画像ローマ字見出し、画像英訳、品詞、用例ID番号、キーワード1、キーワード2、キーワード3、画像データ

画像データベースの項目は、ほぼ音声データベースの項目と一致している。ただし、キーワード1～3項目は、画像データベースに独特の項目であり、画像の内容がどのような状態を表しているかによって、登場するもの、動き、時間についてキーワードを付加している。

3. 3 画像・音声データベースの検索

音声辞書および画像辞書はCASTEL/Jの辞書検索機能のうち、単語および用例を検索することにより表示することができる。

このとき、音声あるいは画像情報として表示されるのは、検索対象となっている単語又は用例に対応した情報である。通常、単語あるいは用例の表示ウインドウに示された機能ボタンを選択することによって出力・表示できる。音声の出力が終了すると他の機能が選択出来るようになる。表示された画像は、ウインドウをクリックすることによって消去され、他の機能の選択が可能となる。

4. 辞書の利用

図1はシステムの利用例である。本システムのデータベース間の連携は図のようになる。

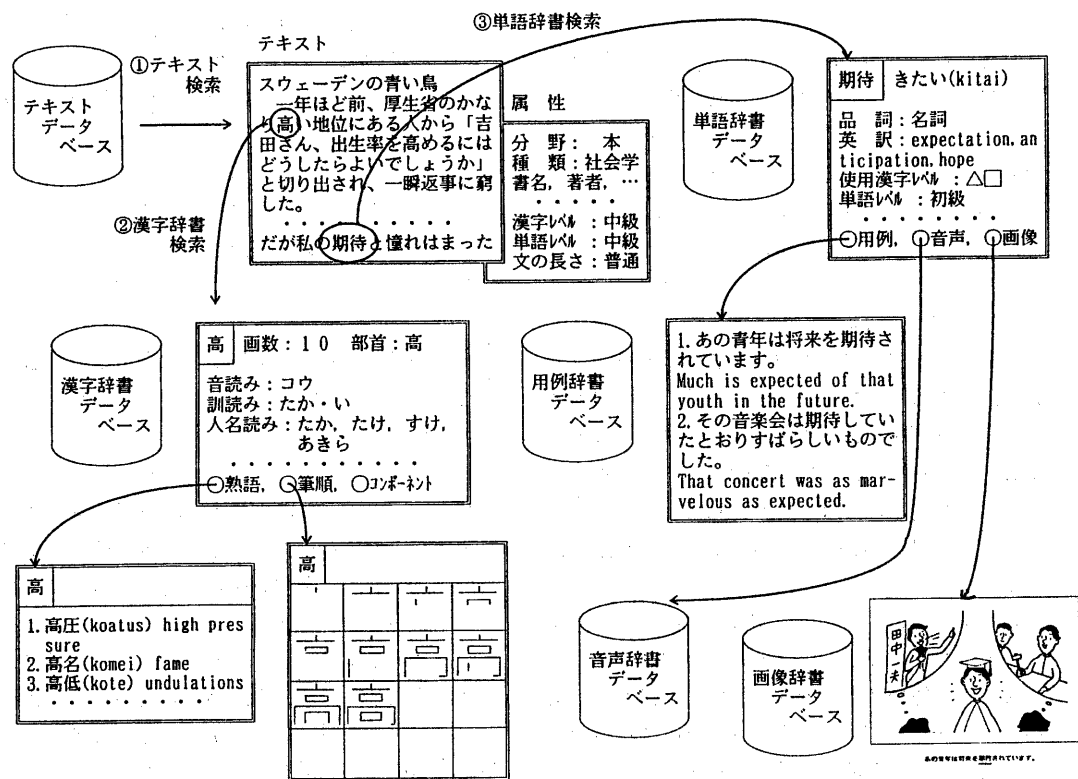


図1 CASTEL/J システム利用例

5. 今後の課題

本プロジェクトは開始以来すでに9年目を迎えている。ようやくデータベースのCD-ROM化までたどり着いたが、本来ならばもっと早く実用化されているはずであった。ここまでの最大の障壁は、データベース構築の際に直面する様々な問題点や疑問点に答えてくれる有益な研究が少ないことであった。たとえば、文法が統一されていない。分かち書きの規則が確定していない。等々、日本語自体が生きていることも原因の一つかと思われるが、多くの基礎的な研究が行われて、その上に立ってデータベースの構築を行うことができれば、もっと短時間にデータベースを完成させることができたと思われる。

CASTEL/Jは実用化をめざして開発を進めてきたが、これには「ソフトウェアの流通」と「データベースの流通」という2つの側面がある。前者については、マニュアルや保守体制の整備など研究者のみでは困難であることから、開発開始時よりメーカなどの民間との共同開発を進めてきており、その流通は基本的には民間側の役割であるという立場をとっている。他方、後者のデータベースについては、ビジネスとしてはペイしない、専門家が主体とならなければデータベースの質の維持は困難である、著作権の許諾を得やすいなどの理由から、研究者側でその流通のための体制を整えるべきであると考えている。

現在、CASTEL/Jのデータベースは、研究者が利用する限りにおいてはほぼ実用化は達成できており、これらのデータベースの普及は日本語教育関係者から強く要望されている。しかし、「流通」および「日本語教育の現場での利用」という観点からみた場合、以下の要件を満足している必要がある。

- ①流通・保守のための体制が整備されていること。
- ②著作権の許諾を得ていること。
- ③データベースの構造が明示的、かつ単純で加工し易いこと、また可能な限りの統一がなされていること。
- ④質・量ともに利用者の要求を満たしていること。

- ⑤基本的な利用ソフトウェアが準備されていること。

これらの他に安価であること、マニュアルが整備されていることなども重要である。

これらの要件に照らした場合、①については「日本語教育支援システム研究会」を組織しており、また、②についてはすべてのデータベースについて著作権の許諾を得たものをデータベース化している。しかし、その他の点については、例えば、

- ・データの種類が多く、またそれぞれの開発工程が異なるため構造的に不統一であったり矛盾する部分がある。
 - ・研究初期に開発したものについてはバージョンアップの必要性がある。
 - ・校正や編集を含め項目の追加など、内容的に手を加える必要のあるデータベースがある。
 - ・経費、時間などの面から作成にとりかかれなかった「文法辞書」などのような、利用者からの要望の高いデータベースがある。
- などの問題が残されている。

これらの問題を含め、流通のための要件を満たすためには、日本語教育の研究者だけでなく、情報科学やデータベースの専門家が共同して対応していかなければ、その解決はおぼつかない。CASTEL/Jで構築しているような日本語教育支援のためのデータベースの流通促進が、日本語教育に大いに貢献するであろうことは多くの人の認めるところであり、今後は各分野の研究者が協力して前述のような諸問題を解決し、データベースの流通促進のための諸条件を整えること重要である。

今後のプロジェクトの進展に関して、当面の課題は次のようなものである。

データベース構造の標準化

すべてのデータベースを可能な限り統一し、かつ単純な構造になるように、各データベースの加工・編集を行う。特にテキストデータベースについては、日本語教育だけでなく、テキスト研究などでの利用についても考慮した構造を検討する。

新規データベースの作成

日本語教師から要望が高い「文法辞書」および「概念辞書(仮称)」を新規に作成する。この概念辞書は、関連することばが階層的に整理されて構成されるものである。

テキストデータベースの拡充

初級、中級、上級といった学習のレベル、および文学、歴史、経済、物理などといった専門分野について、それぞれもれのないように拡充を図る。

学術論文は本研究組織の分担者の著作のみが入力されており、かなり偏りがあるため、広く他の分野の研究者に呼びかけることを計画している。また、新書類についてはさらに著作権の許諾を得た段階で随時追加していく予定である。

テキストデータベースとの連携

テキスト中の用語で、特に活用のある語に対して、辞書機能を発揮させるための工夫が必要である。本機能を辞書にもたせるのか、あるいは検索ソフトにもたせるのか専門的な立場からの議論が必要である。

マルチメディア対応

これまでのシステムはモノクロ静止画データを対象にしていたけれども、今後のマルチメディア対応を考えた場合にカラー画像さらに動画の取り込みが要求されるようになると考えられる。辞書の見出し語に限っても動画あるいはカラー画像が決定的に有利な面があり、この面に対する対応を考慮する必要がある。

音声に関しては、テキストの音声出力が最終的な目標となるが、当面はハードウェアによる対応の可能性を追求することになる。

【参考文献】

- 1) 小松幸廣, 「マルチメディア・データベース・システム」, 日本語教育支援システムの開発, 文部省科学研究費補助金国際学術研究(課題番号02044153)研究成果報告書, pp62-67, 1993
- 2) 坂谷内勝, 「CASTEL/Jのシステムとテキストデータベースについて」, 日本語教育支援システムの開発, 文部省科学研究費補助金国際学術研究(課題番号02044153)研究成果報告書,

pp25-42, 1993

- 3) 吉岡亮衛, 「CASTEL/Jの辞書データベースについて, 日本語教育支援システムの開発」, 文部省科学研究費補助金国際学術研究(課題番号02044153)研究成果報告書, pp43-53, 1993
- 4) 坂谷内勝, 吉岡亮衛, 小松幸廣, 「CASTEL/Jの開発(1)-開発の経緯とシステム概要」, 教育学関連学協会連合第4回全国大会講演論文集(第一分冊), pp. 327-328, 1994.10
- 5) 吉岡亮衛, 小松幸廣, 坂谷内勝, 「CASTEL/Jの開発(2)-辞書・テキストデータベース」, 教育学関連学協会連合第4回全国大会講演論文集(第一分冊), pp. 329-330, 1994.10
- 6) 小松幸廣, 坂谷内勝, 吉岡亮衛, 「CASTEL/Jの開発(3)-マルチメディアデータベース」, 教育学関連学協会連合第4回全国大会講演論文集(第一分冊), pp. 331-334, 1994.10