

# 日本語教育用デジタルコンテンツの開発を終えて

坂谷内 勝 (国立教育政策研究所)

saka@nier.go.jp

## 【要約】

我々は、平成3年に日本語教育支援システム(通称 CASTEL/J:Computer Assisted System for TEaching & Learning / Japanese、以下 CASTEL/J と称する。)研究会を発足し、様々な日本語教育用コンテンツを開発し、提供してきた。これらのコンテンツを活用した様々な研究を紹介し、CASTEL/J 研究会の現状と課題について報告する。

## 1. はじめに

これまで、国立教育政策研究所の研究者が中心となって、日本語教育・日本語研究で利用可能な多種多様なデジタルコンテンツを開発してきた(坂谷内 1998)。

開発したコンテンツは、基本辞書(漢字、筆順、単語、用例、和英、学術)、本(新書等)、台本(寅さんシリーズ全作品)、マルチリンガルデータ(スペイン語、ポルトガル語、ドイツ語、イタリア語)、マルチメディアデータ(音声データ 9,536 件/画像データ 6,829 件)、日本の小中学校用の教材イラストデータ約 1 万件である。辞書の単語等の元データは、『基礎日本語学習辞典(英語版)』(国際交流基金日本語国際センター, 凡人社, 1986)と、『講談社パックス和英辞典』(辞典局, 1991)を使用した。

専門用語は、25 冊の『文部省学術用語集』(大日本図書)から合計 130,296 語を収録した。テキストの総容量は約 70 メガバイトなので、文字数は 3,500 万文字、文の数は 70 万(1 文を 50 文字とする)に相当する。

画像データは、「みんなの教材サイト」(国際交流基金日本語国際センター)からインターネットで提供している(小松・坂谷内 2007)(図 2)。マルチリンガル辞書データ(約 7,000 語の見出し)もインターネットで公開している(坂谷内 2002)。しかし、本、台本等のテキストは、著者からインターネット公開の利用許諾を得ていないため、CD-ROM で、特定非

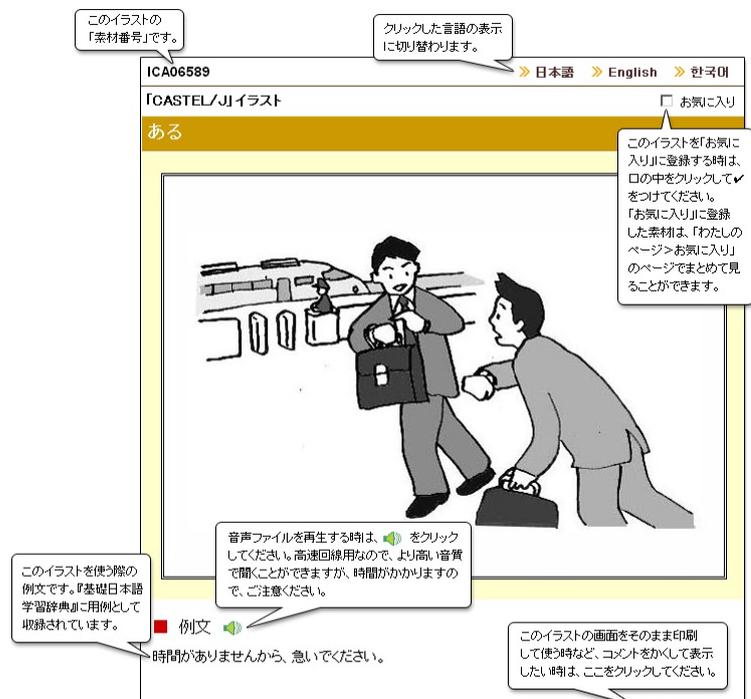


図 1 イラストの例(みんなの教材サイト)

営利活動法人「言語資源協会(GSK)」から日本語教育関係者に限定し、提供している。

また、CASTEL/J 研究会が中心となって、「日本語教育とコンピュータ」(CASTEL/J) 国際会議を開催し、この分野に関心の高い研究者を集って、情報交換を行ってきた。

- ・第1回：CASTEL/J '95、1995年、イタリア・パヴィア、発表件数15件
- ・第2回：CASTEL/J '99、1999年、カナダ・トロント、発表件数44件
- ・第3回：CASTEL/J 2002、2002年、米国・サンディエゴ、発表件数71件
- ・第4回：CASTEL/J 2007、2007年、米国・ハワイ、発表件数80件
- ・第5回：CASTEL/J 2012、2012年、日本・名古屋、発表件数85件

## 2. コンテンツの活用事例

我々が開発した日本語教育用コンテンツは、様々な研究に活用された。まず、我々が開発したシステムは、映画を見ながら日本語を学習するシステムである(図2参照)。光ディスクから流れる映画の台詞をパソコンに表示するシステムである。

台詞の中に出てくる単語がわからないときは、その単語をマウスでクリックすると、その意味が出てくる。漢字がわからないときは、漢字をクリックすると漢字の読み方や筆順等が出てくる。映画のシーンは、自由に選べるので、例えば「病院」を指定すると、病院のシーンを見ることができる。

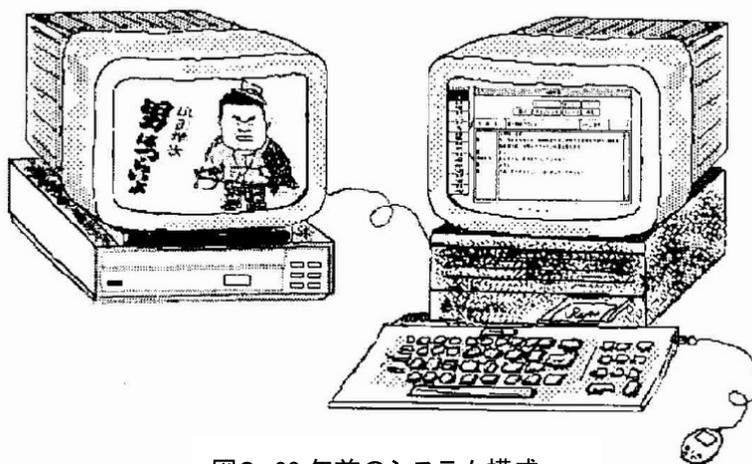


図2 20年前のシステム構成

以下、様々な研究成果が報告されているので、その一部を紹介する。

### ①「CASTEL/Jの教材データベースに出現する漢字(羽田野1994)

CASTEL/Jの教材データベース(リーディング用テキスト)を分析し、テキスト(本)によって、出現する漢字に特有の傾向があることを明らかにした。また、テキストに出現する漢字の70%以上は初級漢字であることがわかった。

### ②CASTEL/Jの用例辞書の増補について(藤田1994)

CASTEL/Jの単語辞書、和英辞書、用例辞書を分析し、単語辞書、和英辞書に用例のない単語を調べ、これらの単語に用例を増補する試みを行った。

### ③教材開発支援システムの開発(山元・他1996)

UNIXの通信ツールやシェルツールとCASTEL/Jのデータベース群を使い、日本語教材開発を進めるためのシステムを開発した。

### ④読書支援システム「新書ライブラリー」の開発(鈴木1998)

CASTEL/J CD-ROMに収録されている新書の文章を素材として、上級の日本語学習者のための読解教材「読書支援システム『新書ライブラリー』」をコンピュータ上に開発した。このシステムは、わからない単語をクリックすれば、単語の意味の説明、例文、関連語(反対語、類義語、派生語)を表示する

ことができる。

#### ⑤機能語の出現頻度調査（鈴木 2001）

CASTEL/J が所収する 8 種類の文章 85 万字を対象に、機能語 384 項目の用例を検索し、そのデータをもとに文章中の出現頻度が高い項目、低い項目を明らかにした。中級前半に教えると良い機能語として、出現頻度が高い項目である「～とともに」、「～によって」等が該当する。

#### ⑥「～テモラウ」の機能についての分析（山本 2002）

「～テモラウ」の機能分析に CASTEL/J CD-ROM に収録されている「寅さん台本」を使用した。

#### ⑦補文標識「の」「こと」の分析（秋本 2006）

CASTEL/J CD-ROM に収録されている 3 冊の書籍データを使用して、補文標識「の」「こと」の分析を行った。

#### ⑧認識動詞の非情主語受身文の分析（志波 2009）

CASTEL/J CD-ROM に収録されている 6 作品から、認識動詞の非情主語受身文の用例を抽出し分析した。

#### ⑨コリゲーションと形態統語情報の調査（千葉 2009a）

類型的言語特徴とコロケーション（特にコリゲーション）の分析に、CASTEL/J CD-ROM に収録されている講談社ブルーボックスなど約 290 万語のデータを使用した。

#### ⑩アノテートされた大規模コーパスを用いた分析ツールの開発（千葉 2009b）

CASTEL/J CD-ROM に収録されている講談社ブルーボックスなど約 290 万語のデータを使用して、形態素境界を XML データとして登録し、検索に利用した。

#### ⑪時間節および時間句「時」「頃」の用法調査（前田 2011）

調査対象は、評論を中心に 49 冊を集約した CASTEL/J CD-ROM を使用した。位相的な変異の少ない、かたい書きことばにおいて、「時」と「頃」がどのように使われているかを調査した。その結果、「時」は出現頻度が多いので初級で取り上げるが、「頃」は中級以降で取り上げと良いことがわかった。

#### ⑫「ている」の論理的な文章中での使われ方の調査（江田 2011）

CASTEL/J が開発した日本語教育用データとデータベースより自然科学，社会科学の新書をそれぞれ 4 冊選び，その中から採用する部分に偏りがないようにして 5 万字ずつとり，各 20 万字のコーパスを作った。

### 3. 現状と課題

過去を振り返ると、コンピュータのハードウェア、コンピュータ言語、アプリケーションソフトが幾度も変化した。今後も、この変化し続ける状況は変わらないと考える。しかし、この変化に関係なく、CASTEL/J のコンテンツが利用できるのは、開発当初から一貫して我々は「データ重視のシステム開発」を行ってきたからであると確信している。

現在なお、我々が開発したコンテンツの利用を要望する日本語教育関係者がいる。また、本研究会が中心となって約 4 年毎に「日本語教育とコンピュータ」国際会議を開催し、世界の研究者が集い、情報交換も行っている。参加者から、今後もこのような会議を開催して欲しいという意見が多く寄せられている。したがって、CASTEL/J のコンテンツを今後とも継続して提供するために、そして、「日本語教育とコンピュータ」国際会議を継続的に開催するために、CASTEL/J 研究会（現会長：當作靖彦・カリフォルニア大学サンディエゴ校教授）を解散することはできない。

このような現状で、CASTEL/J 研究会はいくつかの課題も抱えている。

第1は、デジタルコンテンツ開発に係る経費の課題である。時代の変化に応じて、日本語教育用コンテンツの量的拡大、つまり、より多くの著作物を日本語教育で利用できるようにしたいのであるが、単なる本や台本のデジタル化は研究の対象にならない（研究費を獲得することが困難である）。

第2は、著作権の課題である。著作物を尊重するという意味で、我々は著作者から著作物の利用許諾をもらっている。このとき、インターネット公開については、公開して欲しくないという著作者が多い。また、しかるべき著作権料を支払って欲しいという意見もある。映画台本をテキスト化する際には、映画会社と何回も協議し、限られた範囲での利用許諾と、システム開発に対しての協力を支払った経緯がある。

平成13年1月16日

日本語教育支援システム研究会  
会長 水谷 修 殿

国際交流基金日本語国際センター  
副所長 小松 詩悦

『基礎日本語学習辞典(英語版)』デジタル・コンテンツに関する利用許可について

貴会の前身の研究グループである、研究グループ『マルチリンガル日本語教育支援データベースシステムの開発』(代表者:坂谷内勝・国立教育研究所教育情報研究協力室長)よりの平成12年9月25日付の標記に関する依頼文書に対し、国際交流基金は、当基金が著作権を所有する『基礎日本語学習辞典(英語版)』のデジタル・コンテンツを、貴会製作 CD-ROM に収録する件について、貴会会員のみによる研究用と限定して、これの利用を許可致します。

今後は、会員への配布枚数、研究利用内容等につき変化のあり次第、随時お知らせ頂ければありがたく存じます。

以上

### 図3 利用許可に関する許諾書

第3は、サーバの維持管理の課題である。サーバの維持管理者がいなくなると、デジタルコンテンツは提供できなくなる。永久にサービスするためには、それなりの組織が必要である。特に、改良・更新作業が伴うデジタルコンテンツは大変である。利用者は只で利用していても、サーバの維持管理は決して只ではない。開発費用、維持管理費用、トラブル対策費用などがかかる。最近、サーバを無償で貸してくれる業者がある。しかし、このような業者を利用すると、広告や宣伝が多かったり、フリーメールが突然使えなくなったり、サーバにアクセスが集中して応答しなくなったりするというトラブルは過去に何度も起きている。

## 4. おわりに

最後に、日本語教育用コンテンツの開発は、科学研究費補助金の助成を得て、多くの研究者の協力を得て行われた。研究プロジェクトの発足は及川昭文氏(当時、国立教育研究所室長)、辞書データの開発は吉岡亮衛氏(国立教育政策研究所総括研究官)、音声・画像データの開発は小松幸廣氏(当時、国立教育政策研究所総括研究官)が中心となった。科研費プロジェクト皆様に感謝の意を表する。

[発足当時のメンバー(敬称略)]

大曾美恵子、加納千恵子、木村捨雄、小山揚子、鈴木庸子、中野洋、羽田野洋子、藤田正春、山元啓史、吉岡亮衛、(財)国際文化フォーラム(牛島通彦、原島陽子、地挽里麻、中野佳代子)

[国際会議の実行委員長(敬称略)]

アルド・トリーニ、中島和子、當作靖彦、品川覚、野山広

## 参考文献

- 秋本瞳 (2006) 「補文標識『の』『こと』の共時的観点からみた使い分けについて」、麗澤大学言語研究センター ワークショップ。
- 江田すみれ (2011) 「『ている』の論理的な文章中での使われ方—『効力持続』『長期的な動作継続』を重点にして—」、『国立国語研究所論集 2』, 19-47.
- 小松幸廣・坂谷内勝 (2007) 「イラスト教材データベースの構築」、『日本語教育連絡会議報告発表論文集 Vol.19』, 43-48.
- 坂谷内勝 (1998) 「日本語教育とコンピュータ—CASTEL/Jの開発—」、『日本語教育連絡会議第10回総合報告書』, 52-55.
- 坂谷内勝 (2002) 「日本語教育用マルチリンガル電子辞書の開発」、『第14回日本語教育連絡会議報告発表論文集』, 64-67.
- 志波彩子 (2009) 「認識動詞の非情主語受身文」、東京外国語大学、『日本研究教育年報13』, 1-24.
- 鈴木庸子 (1998) 「上級日本語読解教材『親書ライブラリー』の表現調査」、『情報処理学会人文科学とコンピュータ研究会報告 98(51)』, 87-94.
- 鈴木庸子 (2001) 「CASTEL/Jを利用した機能語の出現頻度調査」、『ICU日本語教育研究センター紀要10』, 13-27.
- 千葉庄寿 (2009a) 「コリゲーションと形態統語情報—類型論的観点から—」、『電子化された言語資料と個別言語研究』, 37-54.
- 千葉庄寿 (2009b) 「アノテートされた大規模コーパスを用いた分析ツールの現状と今後の方向性」、『電子化された言語資料と個別言語研究』, 55-70.
- 羽田野洋子 (1994) 「CASTEL/Jの教材データベースのリーディング用テキストに出現する漢字」、『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』, 115-120.
- 藤田正春 (1994) 「CASTEL/Jの用例辞書の増補について」、『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』, 107-114.
- 前田直子 (2011) 「時間節および時間句『時』『頃』の用法」、『学習院大学文学部研究年報 第58輯』.
- 山元啓史・他 (1996) 「日本語教育支援データベース CASTEL/Jを利用した教材開発支援システム」、『日本科学教育学会年会論文集 20』, 105-106.
- 山本裕子 (2002) 「『～テモラウ』の機能について—『～テクレル』と対比して—」、『名古屋女子大学紀要 48 (人・社)』, 263-276.