

# 人文科学とコンピュータの10年

山田 奨治†

†国際日本文化研究センター

人文科学とコンピュータ研究会 (CH) は、発足以来まる 10 年が経過した。この報告では、これまでの CH の活動を振り返る意味で、過去 10 年間に出版された CH 研究報告論文のアブストラクトにあらわれた名詞の分析等を通して、研究動向の俯瞰と今後の展望をおこなう。

## Ten Years of the SIG Computers and the Humanities

Shoji YAMADA†

†International Research Center for Japanese Studies

Ten years has passed since the establishment of the special interest group (SIG) on computers and the humanities (CH). In this article, we look back on the activities of CH. Through the analysis of nouns appearing in the SIG notes of CH published in the past ten years, we outline the research trend and its future direction.

### 1 はじめに

人文科学とコンピュータ研究会 (CH) は、情報処理学会の研究会のひとつとして 1989 年に発足し、現在でまる 10 年が経過した。CH は人文科学とコンピュータという、一見なんの関係もなさそうな境界領域をあつかう点で、情報処理学会のなかでは当初は異色の存在であった。歴史を振り返ると、現・国立民族学博物館副館長の杉田繁治・初代主査 (主査就任期間: 89 ~ 92 年度) を中心として、日本アイ・ビー・エム株式会社東京基礎研究所の洪政国氏 (初代幹事) らの献身的な努力によって研究会設置が認められた。以後、主査は及川昭文氏 (93 ~ 94 年度)、八村広三郎氏 (95 ~ 96 年度)、筆

者 (97 ~ 98 年度) と受け継がれ、99 年度からは大阪市立大学の柴山守氏が第 5 代目の主査をつとめる。

この間、「人文科学分野へのコンピュータ応用をめざした、①ハード・ソフトの開発・事例、②研究手法 (処理技術) の開発・事例、③学際的研究や理論的研究」を一貫した研究テーマとして活動してきた。98 年度末までに開催した研究会は 41 回、発表された論文数は 322 件をかぞえる。ここ数年の研究会登録者数は、ほぼ 300 名である。研究会への参加者数は平均 30 名ほどで、登録者のほぼ 1 割が研究会にも出席している計算になる。

95 年度からは及川昭文氏を領域代表者として、文部省科学研究費補助金特定領域 (当

初は重点領域) 研究「人文科学とコンピュータ」が認められた。この特定領域研究は、CH 研究会の活動のなかから領域申請の機運が生まれ、及川氏の指導のもと、4年間のプロジェクトが推進された。99年3月をもって特定領域研究は予定の研究期間を完了し、CH 分野としてはつぎのステップを模索する時期に入っているといえる。

この報告では、これまでのCHの活動を振り返る意味で、過去10年間に出版されたCH研究報告論文のアブストラクトにあらわれた名詞の分析等を通して、研究動向の俯瞰と今後の展望をおこなう。

## 2 アブストラクト出現名詞の分析

CHにおける研究の流れをとらえるために、過去10年間の全研究報告論文アブストラクトに出現する名詞を取り出して分析をおこなった。分析は、

1. アブストラクトの全文入力
2. 形態素解析し名詞を抽出
3. 分析対象とする名詞の決定
4. 2年度ごとの集計・主成分分析

という手順で実施した。形態素解析による名詞の抽出は、一太郎8の使用単語一覧機能を使用した。抽出された名詞の出現数の多いものの中から、「研究」「システム」「データ」など研究報告に頻出することが自明で、年度ごとの出現数に大きな変化のない名詞を除き、分野の研究動向をあらわしていると思われる名詞を主観的に選択した。分析対象とした名詞は50音順に、「漢字」「学習」「画像」「検索」「コンピュータ(計算機を含む)」「情報」「テキスト」「データベース」「日本語」「博物館」「メディア」「文字」の12名詞である。これらの中で「コンピュー

タ」「情報」は情報処理の研究報告に自明な名詞であるが、出現の少ない年度があるため、あえて分析対象に加えることにした。

12名詞の2年度ごとの出現数集計をもとに主成分分析をおこなって、主成分負荷量を2次元でプロットしたものが図1である。第2主成分までの累積寄与率は0.78である。

これらの名詞を図1の象限ごとにグルーピングすると、表1のようになる。

表 1: 名詞グルーピング

グループ	名詞
1	検索 データベース 情報
2	博物館 コンピュータ
3	漢字 学習 テキスト
4	画像 メディア 文字 日本語

各名詞グループごとに出現数の推移をみたものが、図2～5である。グループ1は初期に多く、93-94年ころ一時減少したがふたたび多くなってきている名詞である。グループ2は、初期に多かったがその後減少した名詞である。グループ3は、93-94年に急激に増加し、その後減少している名詞である。グループ4は、95-96年ころに増加したが、97-98年にはふたたび減少している名詞である。これらの推移をみれば、研究テーマの盛衰をおおよそみてとることができる。

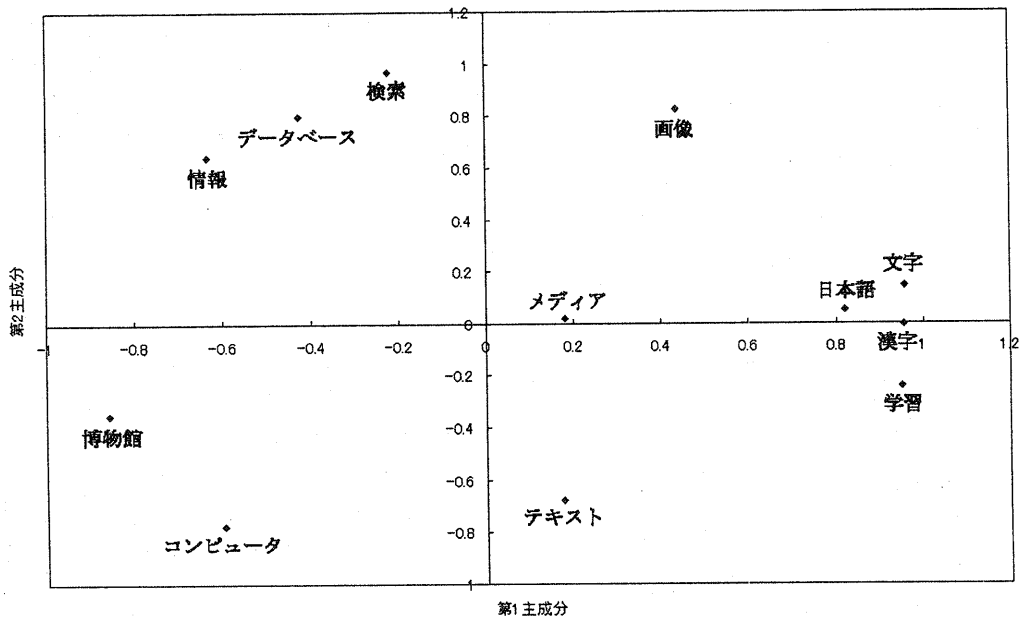


図 1: 主成分負荷量

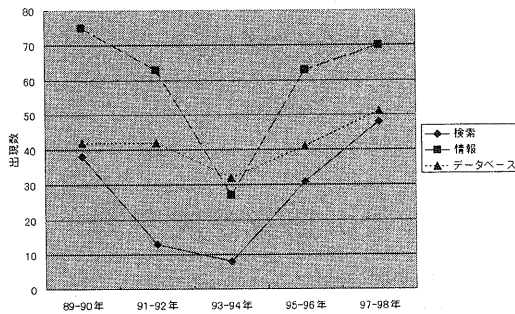


図 2: 名詞グループ1の推移

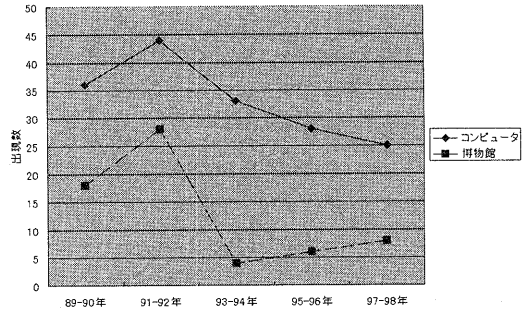


図 3: 名詞グループ2の推移

つづいて主成分得点をプロットしたものが図6である。各象限を代表する名詞を、図6中に並記した。これからいえるのは、最初の2年間の研究内容は「検索」「情報」「データベース」といった言葉で代表される分野からはじまって、つぎの2年間では「博物館」「コンピュータ」へシフトし、「漢字」「学習」「テキスト」をへて「画像」「メディア」「文字」「日本語」と変化し、ここ2年間は

ふたたび「検索」「情報」「データベース」に回帰するという変遷をたどっていることである。

この変化を当時の状況を勘案しながら解釈してみよう。研究会発足当初の89-90年ころは、人文科学方面へコンピュータをどのように使っていけばよいかの模索の時期であったといえよう。そこでさしあたりの問題として、大量のデータをどのようにデー

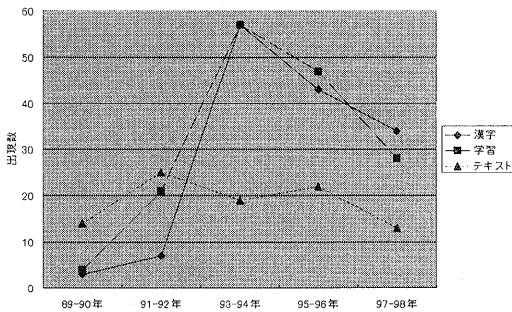


図 4: 名詞グループ 3 の推移

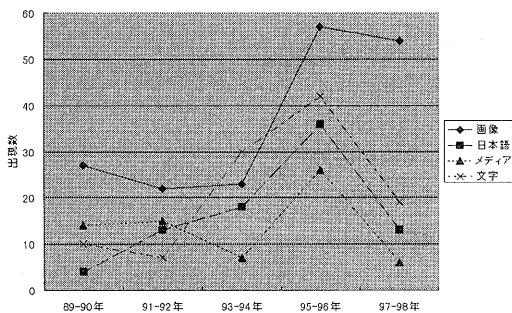


図 5: 名詞グループ 4 の推移

データベース化していけばよいか、検索をどうすればよいかの提案的研究や先駆的なシステム構築事例報告がなされていた。91-92年ころになると、博物館とコンピュータといったことが注目されてきた。大量のデータを所蔵しているのは博物館・美術館・図書館といった機関であるため、その方面への応用が期待されはじめた。「博物館」という名詞が増加したことは、博物館関係の特集が組まれたことが大きいのだが、この時期に特集が組まれたということは、博物館への応用に期待が集まったことをあらわすものである。93-94年ころは、漢字学習にコンピュータを使用する研究と同時に、テキストデータベース作成作業の進行とともに漢字フォントが足りない、メーカーによって微妙に文字が異なるなどの問題が表面化してきた時期であった。95-96年は、パソコン

の急激な価格低下・高性能化という外因によって、それまで個人の計算機環境では不可能であった画像処理・メディア処理が可能になってきたことを受けて、画像・メディアに関する研究が盛んになった。同時にコンピュータで使える日本語・文字の問題が、いっそうの深刻さを増していった。これはすなわち、画像・メディアという新しい研究分野が生まれた一方で、テキストを扱う研究者や業界は、研究に有効なデータベース作りを着々と進行させていたことをうかがわせるものである。そして97-98年になるとふたたび情報・検索・データベースの研究が盛んとなってきた。しかし個々の研究内容をみると、89-90年ころの研究内容とは大きく異なっていることがわかる。すなわち、初期のころはどのようにデータベースを作り検索していけばよいかの研究であったが、近年はすでに完成した大規模データベースをどのように活用していくかの研究にかわってきている。

### 3 今後の研究動向予測

情報処理学会の表彰制度のなかに、山下記念研究賞というものがある。これは研究会での優れた発表に対して、研究会からの推薦を受けて学会が表彰するものである。山下記念研究賞受賞発表の流れをみると、その研究分野がその時代に何を志向していたかがよくわかる。表2は、CHから推薦された山下記念研究賞の発表タイトル一覧である。大きな流れとしては、画像関係の研究からテキスト研究へと移行している。また表には記していないが、発表形態を調べると最近の2回は情報科学者と人文科学者の共同発表に対して賞が贈られている。

ふりかえてみると、人文科学方面への

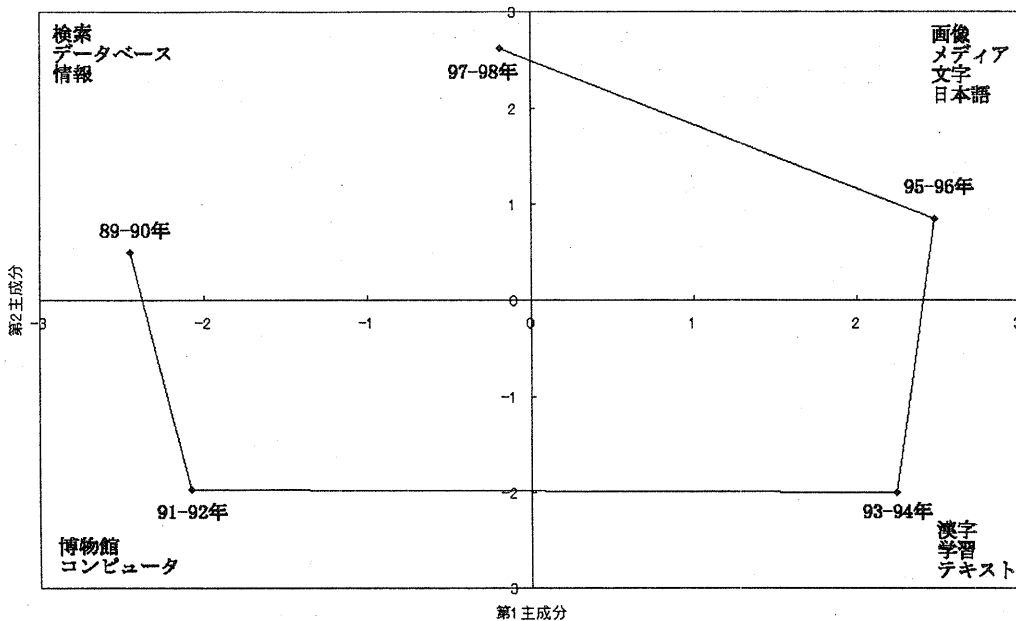


図 6: 主成分得点

コンピュータ応用は、文芸や言語をあつかう研究者が自己のあつかうデータを機械可読化していったことから始まっている。初期はいかにして大量の文字を機械可読化するかが大きな問題であって、検索は単純なもので十分であった時代がづいた。テキスト中心から、しだいに画像や音声・動画などのマルチメディア情報をデータベース化し検索する方法の研究に主流が移ったかのようにみえた時期もあったが、実はその間に機械可読化されたテキスト情報の蓄積が大きく前進していた。実に多種の全文テキストデータが CD-ROM で販売されたりインターネット上で公開されたりしている今日の状況をみれば、テキストの機械可読化という地道な作業の積み重ねがいかに大きかったかがよくわかる。テキストからマルチメディアへと変化した研究の流れは、ふたたびテキストに回帰しているように思える。しかしこれからのテキスト研究におい

ては、単なるデータベース化や検索ではなく、すでに大量に蓄積されたデータのなかからいかにして今までにない知見を見いだすことができるか、その過程において情報科学的方法をいかに適用していけばよいか、これからの重要な研究テーマになってこよう。

いまひとつ、興味深い現象に着目しておく。図3にみられるように、アブストラクトから「コンピュータ」という名詞がしだいに減少しているのである。一方で「情報」という名詞は図2のように93-94年に特異的に減少したが、それをのぞけば高い頻度で使用されつづけている。これはコンピュータはあたりまえの道具として位置づけられるようになり、あえてコンピュータという語を使うことが少なくなっていることを示唆するものである。「人文科学とコンピュータ」の名称も、そろそろ再検討をする時期がきているのかもしれない。

表 2: 山下記念研究賞一覧 (CH 発表分)

発表年度	受賞発表タイトル
90	形状情報を用いた画像の類似検索システムの試作
92	画像処理による節用集 (日用百科書) の使用実態の分析
93	ビデオによる古文書画像の効率的入力法と接続処理
94	古代景観モデラー (ASM)
95	電子文書化された三浦梅園の主著『玄語』
98	和歌データベースからの類似歌の自動抽出

## 4 おわりに

最後に、CH 研究会の課題ということで私見を 2 点あげておく。第 1 は研究会登録者数の減少傾向である。登録者数は 96 年度の 320 名をピークに、97 年度：304 名、98 年度：278 名と減少してきている (図 7)。要因としては、学会員数そのものが減少していることと、96 年度はその前年にスタートした文部省科学研究費特定領域研究プロジェクトのおかげで一時的に登録者数が増加したことが考えられる。しかし特定領域研究によって増加した登録者数を維持できずにきたことは、大きな反省点である。第 2 は研究会取り扱いテーマの見直しである。CH 発足当初からつづく取り扱いテーマは「人文科学分野へのコンピュータ応用をめざした、①ハード・ソフトの開発・事例、②研究手法 (処理技術) の開発・事例、③学際的研究や理論的研究」であるが、今日的な目でみるとコンテンツ研究が抜け落ちている。研究会があつかうべきテーマは、時代の要請にしたがって柔軟に変更していくことが必要ではないだろうか。

CH では昨年度、八村広三郎氏を編集委員長として情報処理学会論文誌のゲストエディタ制による特集号「人文科学とコンピュータ」を企画し、99 年 3 月号として発行され

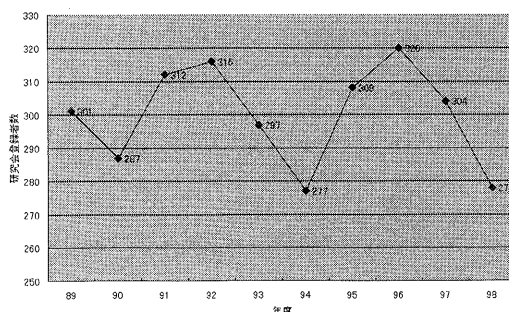


図 7: CH 登録者数の推移

た。特集号の掲載論文には、情報処理学会誌としては異色のものが多いように思える。今後、それに対してどのような反響がよせられるかを注目したい。また今年度からは、研究会主催シンポジウムである「人文科学とコンピュータシンポジウム (じんもんこん:-)1999)」を定期開催する。第 1 回を本年 9 月に国立民族学博物館において開催するのだが、このシンポジウムでは「ショートペーパー」にくわえて「査読付きフルペーパー」というカテゴリーをもうけて、複数査読者による査読をへた一定の質を保った論文を採録していくことを狙っている。こういった場を多いに活用して、研究者間のネットワークづくりに役立ててもらいたいと、企画者のひとりとして切に希望する。