

日本人の自然観

日本語・英語対照「源氏物語」のテキスト・データベースを使う試み

長瀬真理

東京女子大学情報処理センター

2年がかりのプロジェクトである日本語版と英語版の対照「源氏物語」のテキスト・データベースが完成し、1990年4月より海外ではオックスフォード大学電算サービスが、日本国内では東京大学大型計算機センターがサービスを開始した。

本稿では入力を始めとする、具体的な方法論や制作過程を説明するとともに、テキスト・データベース作成のかかえる様々な問題点の検討と将来について展望する。同時に実際の研究例として、出来上がったデータベースを使って日本人の自然観について若干の考察を試みる。

JAPANESE ATTITUDE TO NATURE

An Experiment to use the Machine-readable Texts of
the Japanese and English Versions of
"The Tale of Genji"

Mari NAGASE

Centre for Information Science of the Tokyo Women's Christian University

Zempukuji 2-6-1, Suginami-ku, Tokyo 167, Japan

The two year project to develop machine-readable texts of the "The Tale of Genji" written by MURASAKI SHIKIBU and its English version translated by E.G.Seidensticker has been recently completed. In this paper, the architecture of constructing machine-readable texts is described and the "Japanese Attitude to Nature" is investigated by using the texts.

1. はじめに

電算機の文字処理能力の高度化と容量の増加と共に、人文系の学問研究においてもコンピュータが利用されるようになった。とりわけ研究の基礎となるテキストや文献・資料の全文を入力するテキスト・データベースの開発が活発になっている。既に欧米では、積極的に大量の文献を機会可読な形にしている。完成したテキスト・データベースは、国際的なネットワークを利用してアクセスできるものから、磁気テープやフロッピー・ディスクなど様々な形で提供されており、自国の研究者のみならず広く国外の研究者にも安価にサービスされている。そのほか商業ベースのテキスト・データベースの販売も始まり、シェイクスピア全集のフロッピー版も登場した。

わが国でも海外で作成されたテキスト・データベースを利用する研究者が増えている。又、国産のテキスト・データベースも作られつつある。しかしその数は非常に少なく、近年輸入超過の非難も聞かれるようになった。

このような状況下、筆者は専門はギリシャ哲学であるが、早くからプラトンの文体研究にコンピュータを利用していったことから、日本語版「源氏物語」（小学館刊行：阿部秋生、秋山虔、今井源衛編）と英語版（タトル社刊行：E. G. サイデンステッカー訳” Tale of Genji”）の対照テキスト・データベースを作成した。1991年4月より国内の研究者には東京大学大型電算機センターがサービスを開始した。しかし残念ながら日本には、国際的なサービスを行なう機関が無い場合、海外についてはオックスフォード大学電算サービス（OUCS：Oxford University Computing Service）に供託サービスを依頼した。OUCSは既に長年に亘って英国内ばかりでなく世界中の研究者にデータベースをサービスしている。

特にOUCS内部にはテキスト・データベースを専門に扱うものとしてOTA(Oxford Text Archive)があり、既に25カ国920冊のテキスト・データベースを世界各国の研究者に提供して

いる。しかしこれまでは日本語のテキストは供託されておらず「源氏物語」が最初のものとなった。

いづれにせよ世界中の研究者が「源氏物語」のテキスト・データベースの恩恵に預かることが出来るようになる訳で、遅滞きながらこの分野での国際的な学术交流へ多少とも寄与出来るようになった。

本稿では、このパイロット・プロジェクトの解説をすると共に、実際にデータベースを使って”日本人の自然観”について考察を行う。

2. 入力形式

英文、和文共に検索システムとしてOUCS(Oxford University Computing Service)が開発した文章解析ソフトであるMicro-OCFの使用を予定しているため、COCOA形式を採用した。この形式は、OCFが最も容易に取扱うテキスト・データ形式で、アトラス研究所時代から使われている。ちなみにCOCOAの名前は word Count Concordance generation on Atlas の大文字部分から採られている。又入力には光学読み取り装置を利用した。

COCOA形式：

COCOA形式では、処理対象となる本文と別に、表題、ページ、版、章、著者、登場人物などの情報をあらかじめ参照部として宣言しておく。そうすることにより、版、ページ、登場人物によって検索したり、並べ変えることが可能になる。又、インデックスや索引作りにも利用できる。

表1は入力されたテキストの章の冒頭の一部で、これを参考に日本語及び英語の参照部の定義、実際の宣言を解説する。

2.1 日本語

日本語の原本は縦書きで、上段に語彙の註、中段に本文、下段は現代語訳と三段に分かれている。

又、所々に挿絵が挿入されている。

テキスト・データベースでは本文のみが横書きで入力され、註、現代語訳、挿絵は除かれた。又本文中に現れる註番号や振り仮名も取り除かれた。なお行の切れ目は原本と同じである。

参照部として今回採用された情報は、作者名(W)、表題(T)、巻数(V)、章(C)、英語の章(E)、小見出し(S)、小見出しに対応する英文ページ(F)、日本文ページ(P)、登場人物(A)である。これらの情報はカテゴリと呼ばれ、それぞれの後ろの()内に示されたようなアルファベットの一字変数として鍵括弧 (<>) でくくって宣言される。

本の場合と違って英文との対応を示す幾つかの情報が付加されている。先ず章に関しては、日本語も英語も各々54章に分かれているため相互に対応する。しかし日本語の章見出しには数字が無い。その為英文から日本文への参照を容易にするため、あえて英文の章番号を挿入した。更に細かいレベルでの対応については、英文には無いが、日本文では章が又いくつかに分けられ、番号順に小見出しが付けられている。この点に着目し、日本文の小見出しに対応する部分についてのみ英文のページが挿入されている。その為小見出しの冒頭の部分は相互に対応させることが出来るようになっていく。

更に本では、総てではないが会話部分について、話者の情報が挿入されている。それを生かすために、登場人物として話者マークのカテゴリ<A>を導入した。その結果、特定の話者の会話部分のみを処理対象に指定することが可能になっている。

原本で挿入のような取扱になっている小見出しの数字に続くタイトル状の解説は { } でくくり、コメント(注釈)とした。又英文の各章のタイトルも注釈扱いとした。OCPは注釈を本文とみなさないが、こうしておけば解析時に処理対象としての宣言することも、あるいは逆に削除することも容易である。

2.2 英語

英文の場合も、入力されたのは本文のみで、脚注や各ページの先頭中央のタイトル及び挿絵は省かれた。

参照部は、一見して明らかな通り日本語より短くなっている。両者の大きな違いは、日本語の解説時にも触れられたが、小見出しが無いことにある。もう一つ、英文では会話部分はダブルコーテーション(””) で囲まれているが、一体誰が話しているか、と云った情報が原本にはない。その為話者マークのカテゴリもない。

日本語と英語の対照として日本語の巻の情報を入れたのは、日本語は全部で6巻に分かれており、各巻1からページ付けがおこなわれていることによる。他方英文は2巻に分かれているが両者には通しページが付いている。

なお既に解説したように、日本文の小見出しの冒頭部のページが参照部として入力されている。番号順にはなっているが、小見出しに続く節が長い場合には、後半部の対応に大きなずれが生じている。

3. 研究例

ここでは実際にコンピュータを使用した研究例を紹介する。せっかくデータベースができて、日本では「一体どのような研究に使えるのか?」、と質問を受けることが多い。そのため門外漢ではあるが、筆者の長年の研究テーマであるギリシャの自然哲学とからめてデータベースを利用してみた。本稿では、無謀にも”日本人の自然観”という専門の源氏研究者が余り取り挙げないような大きなタイトルをつけた。一般に今日われわれが使う”自然”という言葉は明治以降の翻訳語であり、かならずしもギリシャ語のフュシスやラテン語のナトゥラとは合致せず、むしろ天、土、山、川といった万有こそが日本の自然に対応すると言われている。その意味で自然の問題は内容のみならず言語のレベルでも複雑な様相を持つ。探求の困難なことは、これまでまとまった研究がなされていないことから容易に推察できる。しかし自然

破壊など厳しい環境問題に直面する現在、日本人の持つ自然観の源泉を探求することはそれなりに意義のあることであろう。日本ではコンピュータを使った研究はまだ認知されていないので、今後専門のみならず多くの人々にデータベースを利用してもらうことを願って、あえて現在誰にでも関心のあるテーマを取り上げた次第である。実際はまだ研究の途中でもあり問題の所在を展望し、コンピュータ利用の有効性を探るにとどめる。なお参考の為に、コンピュータを利用した他の研究例も紹介する。

3.1 自然を巡る問題

紀元前のギリシャでは、近世以降の主観・客観の区別に基づく、人間に対立する対象としての自然観と異なり、ピタゴラスの自然観に代表される自然との調和や輪廻転生の宗教観が信奉されていた。日本でも同様の考え方は神統の自然信仰や仏教的輪廻等に見いだせる。しかしこの問題はギリシャでは、原理や論理に裏付けられた思想大系として凝縮される為、日本とギリシャの単純な比較は許されない。しかしギリシャにとどまらず、アメリカン・インディアンにも類似した発想を見られ、基本的には人類にとって普遍的な自然観の一つとして人間生活に影響を与えている。しかしこのような素朴な了解を待つまでもなく、身近なところにも様々な自然観の反映を見ることが出来る。顕著なものとしてしばしば庭園が挙げられるが、幾何学的な枠構造のなかにシンメトリカルに作成される西洋の庭園と、庭の垣根を越え、はるか遠くに見える山までを考慮した日本庭園では自ずと自然に対する考え方が異なることが解る。前者では人工の庭園と自然との境が強調され、後者では自然との継続・連続あるいは融合が見られ、そこには人間と自然の対立は感じられない。

このような観点にたって「源氏物語」を読み、サイデステッカーの翻訳等と比較すると色々な問題点を発見する。

例えば英語のIt's raining.は客観的状況しか表

さないが、「源氏物語」では、雨や霧といった自然現象の描写が、登場人物の心理、即ち嘆き、悔い、煩惱を表す。これはデータベースの原本の編者である秋山氏も指摘しておられるが、単に直載的な自然描写のみならず、「ゆかし」、「いたはし」、「いとほし」「はづかし」などといった形容詞や「ここちす」といった動詞にも主観性・客観性の融合がみられるということである。「源氏物語」は「万葉集」の5倍の言語量を持ち、しかも1万2千語という平安朝の物語のなかで最大の語彙量を誇り、日本語・日本文学に与えた影響は計り知れない。これまでも専門の研究者のみならず多くの人々に多様な読みを可能ならしめてきたが、どうやら”自然”というテーマにおいても期待を裏切ることはなく、我々は日本人の自然観の原形質をみることができそうである。

表2はMicro-OCFを使って、本居宣長以来「源氏物語」の基調語として有名な”あはれ”を第4章の「夕顔」から文脈1行と共に出力させた結果である。この章だけでも27箇所あり、しかも用例は実に多種多様で、単に心情をを表現する際に試用されるのみならず、空といった自然現象の形容にも使われている。

”あはれ”の使用は「源氏物語」全体で1024箇所にもものぼるが、空の他に、霧、雨、野火、夜などにも使われる他、森羅万象のすべてに使われており、このように一つの語が多様に使われる例は、他の言語では類をみない。

翻ってコンピュータの利用について言えば、このようにテキストで多用されている語彙の検索に大変重宝することは自明である。更に出力をうまく調整すれば、これまで研究者が悩まされていた大量のノートやカード作りの手間も軽減してくれる。又結果がすぐ出ることから試行錯誤もやりやすい上、研究者は出力結果の吟味にすぐとりかかることが出来る。”自然観”については、まだ結論が出せる段階ではないが、少なくとも語彙研究におけるコンピュータ利用の有効性は示されたのではないかと思う。

3.2 語彙研究の応用

3.2.1 統計的手法を使った研究例

戦後まもなく武田宗俊は54帖配列の謎をめぐるの明快な成立論を展開し一躍脚光を浴びた。彼は源氏物語の第一部(源氏を三部構成とし具体的には33巻までの光源氏の将来についての予見が完全に実現するまで)を17帖と16帖の2系列に分け、前者を紫上系、後者を玉葛系と命名した。「紫上系の物語は独立し、完全な統一を持つものとして後者に無関係である。即ち源氏物語第一部から玉葛系16帖を除いても何等欠けることのない物語となっている。玉葛系の物語は紫上の」物語を背景とし、その系の物を取り入れているが、それはただ影を落とすのみで、その物語を玉葛の巻に於いて発展させてこれを紫上系にかえすことはない。玉葛系の物語は松にからみついた藤のように、外見は一体をなしているが、その本を異にし、付加的結合で、有機的な融合とはなっていない。」と主張した。更に紫上系17帖が先ず構想記述され、これに付加して玉葛系が後記挿入されたのだと推定したのである。

この武田の源氏物語成立説は定説としては受け入れられていないが、学界に大きな影響を与え、現在でも賛否両論の活発な論議があり重要な研究領域になっている。

この観点に着目して安本美典氏は紫上系と玉葛系の「比喩多用型か比喩節用型」どうかであるかどうかという観点から因子分析を行い、武田氏の説が妥当なものであるとの結論を出している。

3.3.2 「宇治10帖」の真偽問題について

西欧ではシェイクスピアやプラトンの真偽問題を扱う文献研究にコンピュータが成果を挙げている。「源氏物語」でも同様の問題があり、最後の10帖と他の44帖との関係、いわゆる”宇治10帖問題”については、長年議論がなされている。

上記の安本氏は「歌物語」「作り物語」といっ

た因子や、名詞の頻度、和歌の用いられている度合いなどを調査して、この問題にも挑戦している。そして10帖は他の44帖に比べ「作り物語」的で且つ又「比喩節用型」であり、両者の文体には差があるとの興味深い結論をだしている。

他方コンピュータを使わない研究も勿論盛んで、サイデンスティッカーは、シェイクスピア=ベイコン説にならって、この問題を”Genji-Baconian Theory”あるいは”Genji Baconians”と命名し興味深い議論を展開している。氏は、42帖の「匂宮」の冒頭で源氏が亡くなるところで物語は終り、42とそれに続く43の「紅梅」、44の「竹河」の3帖を変わり目、あるいは経過の部分とし、45以降から新しい物語がはじまると主張する。

この問題はまだまだ論議が続くと思われるが、テキスト・データベースが公開されれば、シェイクスピアの真偽問題ではコンピュータが華々しい成果をあげていることから、この分野の研究はもっと盛んになるとと思われる。

4. コンピュータと文学研究の将来

4.1 パラレル・テキスト・データベースの将来

日本研究が世界各地で盛んになり、もはや日本に行かなければ日本研究は出来ないという時代は終わった。その意味で研究と教育の両方に役立つ、日本語と英語の相互対照が可能で、いわゆるパラレル・テキスト・データベースの需要は多い。我々が様々な外国文学研究で苦労するのと同様、海外の日本研究者や学生は日本語に苦労している。日本語と翻訳の両方の比較はもちろん、語彙の研究にもコンピュータは大いに威力を発揮するであろう。その意味で今後より多くの日本文学や古典及びその翻訳書のテキスト・データベース作成が必要であろう。併せて日本国内にもテキスト・データベースを海外にサービスする機関の設置が急務となる。

4.2 複数のテキストの比較研究の可能性

一つのテキスト・データベースを作成すると必ずその選択に不満を持つ研究者がでてくる。例えば「源氏物語」でも青表紙本や河内本、別本を始め、全ての校訂本や写本までをもいれるべきだ、という意見は当然予想される。しかしこれは今日のコンピュータの能力を考えると不可能なことではない。現に英国では、写本研究や比較文献学、あるいは学生の教育用に、幾つもの版を入力したハイパー・テキストやソフトが開発されており、普通では入手しにくい貴重な文献に誰でもアクセスすることが出来るようになっている。

4.3 電子化辞書の必要性

コンピュータによる古典研究の環境を整備するためには、内在資料としてのテキスト・データベースや検索ソフトの開発が必要であるばかりでなく、様々な背景への理解を助ける為の外在的資料として、同時代に書かれた他の作品のテキスト・データベースが必要であると同時に、電子化古語辞典が必要不可欠である。既に海外では様々なテキスト・データベース作成と同時に、辞書作成が盛んに行なわれており、OED(Oxford English Dictionary)のCD-ROM版は大きな成果をあげている。これを使うと、語源がアラビア語あるいはヘブライ語の言語を一八世紀のテキストの中から探す、あるいはディケンズのテキストから探す、といった昔なら一人の研究者が何年もかかるような作業があつという間に出来る。辞書を一番使うのはいうまでもなく研究者である。このような辞書を見ると末端のユーザーである文学研究者の意見がいかに良く反映されているかが解る。欧米では辞書の権威は学者・研究者の使用に足るものかどうかで決まる。質の高い辞書学(lexicography)の長い伝統の蓄積と最先端のコンピュータによるデータベース作成技術の見事な成功例といえるだろう。

日本でも将来この種の電子化辞書は確実に必要

となるであろう。言葉は数字と違って曖昧である。重要な語や語句程、意味も広く、多様であり、沢山の用例研究を必要とする。日本文化の重要なキーワードである、“わび”、“さび”なども、時代の流れ、解釈により意味は多様化している。それら語の用例が日本の古典の中から全て検索することが可能になれば、意味の変遷の研究に大いに役立つであろう。先に挙げた“あはれ”にしても、「源氏物語」だけでなく、「伊勢物語」や「古今集」等を網羅した電子化辞書が出来れば、研究者のみならず多くの人々が恩恵を受けるであろう。一方、消滅してしまって全く使われなくなった語句も多い。変化し崩れつつある日本語を保存するためにも古語辞典の電子化プロジェクトは必要であろう。

以上「源氏物語」のテキスト・データベースの作成を中心に、関連する様々な問題の検討と将来について解説した。出来るだけ多くの人々にテキスト・データベースを利用して頂くことを期待したい。又御批判や助言は大歓迎である。

<参考文献>

- 1) 池田勉：日本文学における文芸意識の構造、国文学試論、1930
- 2) 武田宗俊：源氏物語の最初の形態、文学、1950
- 3) 安本美典：文学作品を因子分析する、計量国語学、No.14、1968
- 4) E.G.Seidensticker:How many People Wrote The Tale of Genji,An investigation to Japan's Leterature, 1974
- 5) 柳父章：翻訳の思想、平凡社、1981
- 6) 長瀬：日本語-英語対照「源氏物語」のテキスト・データベースの作成に関する基礎的研究、情報知識学会誌、1990
- 7) 長瀬：日本語-英語対照「源氏物語」のテキスト・データベースの暫定公開にあたって、東京大学大型計算機センターニュース、Vol.22、1990

表1 第39章「夕霧」冒頭のテキスト・データベース

<p><W Murasaki Shikibu>{Translated by <T The Tale of Genji> <K 4> <C 39>{Evening Mist} <N 1> <P 676> Making full use of his name for pro that he thought the Second Princess the world that he was only being fa many a solemn visit, and came to fe months went by that the situation w mother thought him the kindest of g from the loneliness and monotony of romantic intentions, and it would n must go on being kind, and the time princess would invite overtures. He sion presented itself, of her manne He was still awaiting his chanc clutches of an evil and very stubbo Ono. A saintly priest who had long renown as a healer had gone into se never to return to the city. He wou the mountain, and it was for that r provided the carriage and escort fo too busy with their own affairs to of them, had taken an interest in t which she had greeted evidence that made him feel unwelcome. Yu-giri ha ing his intentions to himself. When <P 677> see to the vestments and offerings was too ill to thank him. The women insisted that, given he would not be pleased with a note must answer. And so she did present good, and the single line of poetry</p>	<p><C 夕ぎり> <E 39>{Evening Mist} <S 1> {夕霧、小野の山荘の落葉の宮を思慕 <F 676> <P 383> まめ人の名をとりてさかしがりたまふ大将 この一条宮の御ありさまをなほあらまほし と心にとどめて、おほかたの人目には昔を 忘れぬ用意に見せつつ、いとねむごろにと ふ。下の心には、かくてはやむまじくなむ まさりたまひける。御息所も、あはれにあ もあるかなと、今はいよいよものさびしき えず訪れたまふに慰めたまふことども多か はじめより懸想びても聞こえたまはざり 懸想ばみなまめかむもまばゆし。ただ深き まつりて、うちとけたまふをりもあらじや さるべき事につけても、宮の御けはひあり <P 384> みづからなど聞こえたまふ ことはさらになし。いかな らむついでに、思ふことを もまほに聞こえ知らせて、 人の御けはひを見む、と思 しわたるに、御息所、物の怪にいたうわづ 野といふわたりに山里持たまへるに渡りた 御祈祷の師にて、物の怪など払ひ棄てける 里に出でじと誓ひたるを、麓近くて、請じ なりけり。御車よりはじめて、御前など、 たまへるを、なかなかまことの昔の近きゆ 事わざしげきおのがじしの世の営みに紛れ 出できこえたまはず。弁の君、はた、思ふ で気色ばみけるに、事の外なる御もてなし ひてえまでとぶらひたまはずなりにたり。 <P 385> この君は、いとかしこう、さりげなくて にためり。修法などせさせたまふと聞きて</p>
--	--

表2 第4章「夕顔」出力結果：KWIC

あはれ 27

夕が 1	57	210	3	のはかなき住まひを、あはれに、いづこかさしてと思ほ
夕が 2	59	212	6	ものしたまへば、いとあはれに口惜しうなん。命長くて
夕が 2	59	213	1	目くはず。君はいとあはれと思ほして、「いはけなか
夕が 6	62	219	1	こえ顔なる片つ方人を、あはれと思さぬにしもあらねど、
夕が 6	62	219	13	たけれど、人のためはあはれと思しなさる。むすめ
夕が 6	62	220	8	べきふし加へなどして、あはれと思しぬべき人のけはひな
夕が 8	63	224	12	のたまひて、もしかのあはれに忘れざりし人にや、と思
夕が 9	65	226	8	まどはしつつ、さすがにあはれに、見ではえあるまじく、
夕が 9	65	229	2	たふるに 従ふ心はいとあはれげなる人、と見たまふに、
夕が 9	65	229	7	し移ろふことあらむこそあはれなるべけれ、とさへ思しけ
夕が 10	67	229	13	声々、目覚まして、「あはれ、いと寒しや」、「今年こ
夕が 10	67	229	15	ひかはすも聞こゆ。いとあはれなるおのがじしの営みに、
夕が 10	67	232	8	たへがたげに行ふ。いとあはれに、朝の露にことならぬ世
夕が 10	67	232	11	は思はざりけり」と、あはれがりたまひて、優婆塞が
夕が 11	68	237	12	何心もなきさし向ひをあはれと思すまに、あまり心深
夕が 15	75	252	8	へる尼の住まひ、いとあはれなり。御燈明の影ほのかに
夕が 15	75	253	4	のほどに心を尽くしてあはれに思ほえしを、うち棄てて
夕が 17	78	258	15	したまひけめと、思ふもあはれになむ。またうち返しつら
夕が 17	78	259	2	などさしも心にしみてあはれとおぼえたまひけん。なほ
夕が 17	78	260	8	しあはせて、いよいよあはれまさりぬ。「幼き人まどは
夕が 17	78	261	4	かなるに、空の気色いとあはれに、御前の前裁枯れ枯れに
夕が 17	78	262	8	人の心には従はんなむあはれにて、わが心のままとり
夕が 18	81	263	14	めづらしきに、これもあはれ忘れたまはず、「生ける
夕が 19	82	266	2	まふ。その人となくて、あはれと思ひし人の、はかなきさ
夕が 19	82	266	4	に譲りきこゆるよし、あはれげに書き出でたまへれば、
夕が 19	82	266	13	しやりつつ、念誦をいとあはれにしたまふ。頭中將を見た
夕が 21	83	269	5	ぐれて、空のけしきいとあはれなり。ながめ暮らしたまひ

TOTAL WORDS READ	=	3936
TOTAL WORDS SELECTED	=	3936
TOTAL WORDS PICKED	=	27
TOTAL WORDS SAMPLED	=	27
TOTAL WORDS KEPT	=	27
TOTAL VOCABULARY	=	1