

第17回 公開シンポジウム

人文科学とデータベース

「データ」を読む・観る・解く

2012年1月7日

主催：第17回公開シンポジウム実行委員会
後援：人文系データベース協議会

目 次

特別講演

- ・聴覚障がい者の受療時コミュニケーション支援用手話アニメーションの開発…………… 1
森本一成（京都工芸繊維大学）

一般講演

- ・文末表現を考慮した文章の特徴量を用いた質問回答文の因子得点の推定…………… 9
横山友也・宝珍輝尚・野宮浩揮（京都工芸繊維大学）
- ・感性のあいまい性を考慮した評価に向けて…………… 21
赤井俊介・宝珍輝尚・野宮浩揮（京都工芸繊維大学）
- ・個人の好みに関する感性のモデル化に関する研究…………… 29
荻野晃大（京都産業大学）
- ・介護場面の日本語学習に向けた例文データベースの構築支援技術…………… 35
竹内和広（大阪電気通信大学）
- ・文化遺産のバーチャルアーカイブのデジタルコンテンツ化に関する研究…………… 43
渡邊俊祐（同志社大学）
- ・奈良文化財研究所におけるデータベース…………… 51
森本 晋（奈良文化財研究所）
- ・西日本縄文・弥生時代集落のGISデータベース化と時空間動態評価…………… 55
山口雄治（同志社大学）
- ・箸墓古墳実測図に関する一考察…………… 63
小沢一雅（大阪電通大学）

- 公開シンポジウム「人文科学とデータベース」プログラム…………… 71

聴覚障がい者の受療時コミュニケーション支援用 手話アニメーションの開発

Development of Sign Animation for Hearing-Impaired People to Support Communication in a Hospital

森本 一成

Kazunari Morimoto

京都工芸繊維大学 大学院工芸科学研究科, 京都市左京区松ヶ崎

Graduate School of Science and Technology, Kyoto Institute of Technology
Matsugasaki, Sakyo-ku, Kyoto

あらまし:聴覚障がい者は病院に手話通訳者がいないと受療を拒む傾向のあることが知られている。これを解消するための一手段として、手話通訳者に代わる手話アニメーションシステムの開発を行っている。手話アニメーションの生成と評価について説明する。特に、病院窓口での会話支援システムや胃部レントゲン検査用手話アニメーションシステムを例に開発の問題点を検討する。

Summary:It is well known that hearing-impaired people tend to refuse to go to hospital, because sign language interpreters for support interaction among doctors, technical staffs, nurses and them are very few in a hospital. We are developing a sign animation system to grow up smooth communication in a hospital. The feature of the sign animation for reception service and X-ray examination is introduced.

キーワード:手話, 手話アニメーション, レントゲン検査

Keywords:sign language, sign animation, X-ray examination

1. まえがき

聴覚障がい者と健聴者とのコミュニケーションには、筆談を用いるか、手話通訳者を介している場合が多い。筆談の場合、情報の伝達速度は遅くその量も限られるため、感情の伝達が難しいなどの問題がある。したがって、講演、講義、施設での受付などでは、手話通訳者を介する場合が多い。

医療現場でのコミュニケーションでは、受療行為の抑制や健聴者の医師や看護師に正確に症状を伝えられないなどといった大きな問題が指摘されている。また、手話通訳者を介す場合は、プライバシーの問題に関わってくる問題もある。

また、聴覚障がい者の数は手話通訳者の数に比べて圧倒的に不足しているという問題もある。専属の手話通訳者をおいている病院は非常に少ない。その結果、聴覚障がい者は手話通訳者のいる病院へ集中し、手話通訳者に

精神的・肉体的に多大な負担がかかり顎肩腕障害などの健康障害が見られたという報告もある[埜田ら, 1996]。このように聴覚障がい者にとって医師や看護師との意思疎通が困難な状況は様々な問題を起こしている[北原ら, 2001]。

こうした問題点を解決するための一手法として、手話・日本語間相互翻訳の機械化がある。これが実現すれば、聴覚障がい者と健聴者の両者が手話通訳者を介さずに、円滑なコミュニケーションを行うことが期待出来るようになる。筆者らは日本語を中間型手話に翻訳し、それを人物モデルで表現する手話アニメーションを使った手話・日本語間相互翻訳の開発・研究に取り組んでいる[森本ら, 2005]。

2. 手話・日本語間相互翻訳システム

2.1 手話アニメーションの規則合成

本研究で使われている手話アニメーションアバターは、TALKFILE と呼ばれる手話単語コードが記述されたファイルを読み込ませることで、ファイルに記述されたコードを規則合成して生成され、画面に表示される。本章では手話を表現する人物モデルの構造、TALKFILE ならびにアニメーション生成のアルゴリズムについて述べる。

人物モデルには骨格構造が組み込まれており（図 1）、この骨格モデル上に対応する体表データを割り当て、ポリゴンで描画している。

図 1 の黒点は人間の関節に相当しており、各部位間の接合点を表している。またこの骨格は階層構造で管理されており、各部位はそれが属する上位の部位の動作に連動して状態を変化させる。これらの接合点の回転角や傾きを変更することで、人物モデルに様々なポーズを取らせることができる。肩や肘の位置は任意に指定できる。

骨格モデルの上に対応する体表データは、複数の描画命令をオブジェクト化し、そのオブジェクトの移動、回転などの操作が容易にかつ高速に行える機能を持つ OpenGL（3次元グラフィックス用のライブラリ）のディスプレイリスト機能を用いて作成されている。顔面の一部は頂点をそれぞれ管理する方法で描画している。これらの頂点を動かすことによって眉の上下、眉間の幅、眉の傾き、目の開きなどの動きを可能とし、これらを組み合わせることで人物モデルの表情の変化が可能となる。

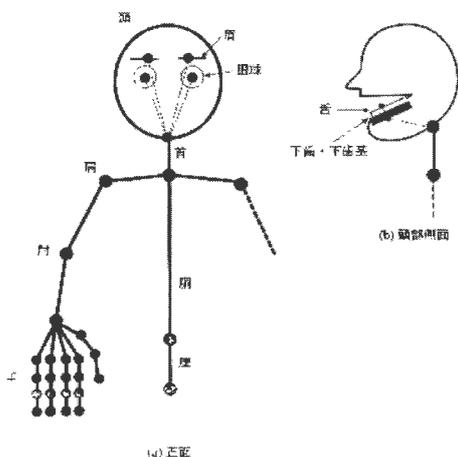


図 1 手話アバターの骨格構造

筆者らが用いている TALKFILE とはアニメーション表示する手話単語ごとに、人物モデルの各部位の動作をコード化したものを記述したファイルであり、キーフレームごとにパラメータのそれぞれの状態を表すコードを記述する仕様となっている。

キーフレームとは動作の基点となるフレームのことで、各部位がいままでと異なる動作をするときに、その次の状態を順次記述したものである。キーフレーム間の時間は自由に設定することができ、また動作速度も自由に変わることができる。TALKFILE で用いるコードには、手指の動きを表す手指コードとして、動作終点での手の位置、方向、形状を制御するコード、動作の始点から終点までの運動軌跡、変位速度、運動時間を制御するコードがある。

運動軌跡には円や波型などの様々な動きがあり、コードによって 3次元空間のあらゆる方向の平面上で人物モデルの手首動作を表現できる。また、手指動作の運動補間にはメリハリのある動作を実現するために、スプライン曲線補間を使用している。

人物モデルの手指以外の動きを表すコード（非手指コード）として、手指動作が行われている間の目、眉、首、腰などの動きも記述可能である。また、聴覚障がい者の協力のもとに作られた基本口形の口形パラメータを使って、口話もコード化され（口話コード）アニメーションで表現できるようにしている。

2.2 アニメーション生成と描画

まず、光源やカメラ位置、人物モデルの初期ポーズなどの設定を読み込まれ、静止状態の人物モデルが表示される。この状態から TALKFILE を読み込ませると、規則合成によるアニメーション生成が行われる。

アニメーション生成の準備として、まずすべてのキーフレーム間を補間し、フレームごとに肘、肩以外の各部位がどのような状態にあるかを算出する。その計算の後、肩位置と肘位置において TALKFILE の指定が無い場合は、フレームごとの胴体の状態から肩位置を、肩と手首の位置から肘位置を算出することで、上腕と前腕の状態がわかる。ここで初めてすべての部位を描画する。

このように上腕と前腕の状態を1フレームずつ算出しながら、生成した画像を1秒間に15回順次描画していく。規則合成の具体的な手順を以下に示す。

1. 手話単語内の手指のコードを全てデコードして、手指動作のキーフレームを規則合成する
2. 手指動作のキーフレームをもとにして、キーフレーム間の運動を補間して求める
3. 時間コード、手の動いた距離をもとにして手話単語全体を再生するのに要するフレーム数を求める
4. 非手指動作コードをデコードし、1でデコードした手のコードと同期させながら、非手指動作のキーフレームを作る
5. 非手指動作のキーフレームをもとにして、キーフレーム間の表情変化を補間して求める
6. 口話コードをデコードして、口話に要する時間を計算する
7. 3と6で求めた手指動作と口話に要する時間を用いて、手の動作と口話の終了が同期するように、手話アニメーションを合成する

このとき、手指動作の速さはコードによって書き込むことが出来るが、口話の速さは一定であるため（ただし1フレームだけ増減可能）、口話の時間より短い時間で手指動作を終えることはできないという制約を持つ。

3. 病院でのコミュニケーション支援

3.1 受付での手話アニメーションを用いたコミュニケーション支援

受付担当者用のディスプレイ1台と、患者用のタッチパネルディスプレイ1台を用いる（図2）。受付担当者用のディスプレイでは手話アニメーションを操作し、患者からの返答を表示できる。患者用のタッチパネルディスプレイには、受付担当者が指定した手話アニメーション

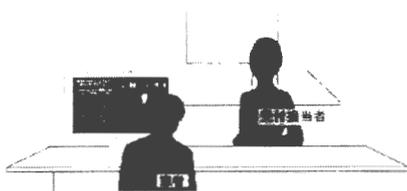


図2 手話アニメーションを利用した受付システム

が流れ、それに対する返答用のボタンが表示される。

受付担当者はあらかじめ調査した受付業務における標準的な手続きの流れにそって、ディスプレイの操作ボタンを使って指示、質問及び説明の手話アニメーションを流す。患者は表示された指示などに対する返答をタッチパネルディスプレイに表示された返答用ボタンから選択する。受付担当者は患者からの返答を確認し、それを受けて次の指示、質問および説明を流す。

こうして作成した病院受付操作画面のユーザビリティ評価（ペーパープロトタイプング法による評価）を行い、得られた結果から病院受付システムの改良を行った後、聴覚障がい者による評価実験を実施した。

実験内容等の詳しい説明は紙幅の関係で省略するが、提案した新しい受付システムを病院に導入した際に期待できる効果は、病院での受付手続きをスムーズに行えるということが第一に挙げられることがわかった。このシステムを用いると、従来から病院の受付手続きで用いられていた筆談の短所である、やり取りに時間がかかる、発言の流れがつかみにくい、崩し文字が読みにくいといった点を解消できると考えられる。また、筆談でのコミュニケーションよりも手話アニメーションでのコミュニケーションにより患者・受付担当者に与える精神的・肉体的な疲労も軽減できると考えられる。

3.2 胃部レントゲン検査時の指示用手話アニメーションシステム

このシステムを開発するために一般的な胃部レントゲン検査の手順を調べた。検査ではまず、受検者が検査室に入り発泡剤（胃を膨らませるための薬）とバリウムを飲む。次に、受検者が胃部レントゲン専用の台に乗った後、台が倒されて仰向けやうつ伏せなどの姿勢や、腰の上げ下げなどの動作の指示がある。その後レントゲン写真を撮るために呼吸を止めるよう指示される。撮影完了後、検査室から出て下剤をもらう。

このように検査の時の動作行動は簡単なように思われるが、胃部レントゲン検査では受検者の姿勢の微妙な調整を技師が行うことがある。聴覚障がい者が受検する場合は姿勢の調整をほとんど指示だけで行う。特に聴覚

障がい者の場合、医者の指示は通訳者を通して聴覚障がい者に伝える方法もあるが検査時間が長くなり効率も落ちる。また、手話通訳者に対する被爆量が高くなることもある。よく使われている方法は技師自身が身振りで示したり、紙に書いたものを提示する方法が取られている。また、検査の概要を示すガイドビデオを受検者に見せて、その内容を覚えてもらうといった方法が取られることもある。このように聴覚障がい者の受検時のコミュニケーション手段は健聴者のそれと比べて情報量が少なく、また、レントゲンを取る時の合図がわからない。レントゲンを開始の合図がわからない。レントゲンの時に適切な指示がない。そして、息を止めるタイミングが分からない、といったことがあるため聴覚障がい者における不安感は大きい。

さらに、胃部レントゲン検査における身体の位置に関する指示は微妙なものが多、何度も指示が出されることが多い。また、息を止めるタイミングなど健聴者にとっては何でもない指示が聴覚障がい者にとっては非常に困難なものとなっている。聴覚障がい者が受検する場合に至っては、姿勢のほとんどを技師が調整することになる。この場合、技師がレントゲンのカメラをみながら姿勢を直す指示を出すことは不可能になり、技師の負担が増やすばかりか、受検者に不安を与えることになる。

これらの問題を改善するためのシステムを開発している(図3)。このシステムは、操作室に置かれた指示用のインタフェースをレントゲン技師が操作し、聴覚障がい者の受検者にその指示を手話アニメーションにより伝え、受検者はその指示を受けて応答することにより、レントゲン技師と受検者双方のコミュニケーションを図る。

提案システムではレントゲン技師から聴覚障がい者の受検者への指示に手話アニメーションを用いるが、それを受検者の顔に装着させたシースルーの FMD (フェイス・マウント・ディスプレイ) に表示させる。これを本システムの表示デバイスとして選んだ理由は次の通りである。胃部レントゲン検査は他のレントゲン検査に比べ、レントゲン受検者に対して姿勢変化などの指示が多



図3 胃部レントゲン検査指示用手話アニメーションシステム

い。そのため、頭、体の向きが一点に定まらず色々な方向を向く事になる。つまり、液晶モニターなどのある特定の場所に固定された表示デバイスでは、受検者は無理な姿勢でその表示デバイスを見る必要性がでてくる。一方、FMD は受検者がどのような姿勢であっても、受検者の目の前に表示ディスプレイ(手話アニメーション)を確認することができる。また、FMD は小型・軽量で半透明である事から、表示画面の向こう側を確認することができる。そのため、レントゲン技師からの様々な姿勢の変化姿勢の変化指示に安全に対応することができる。

手話画像に用いる胃部レントゲン検査の指示内容の選定には、健聴者が実際に胃部レントゲン検査を受けている場面を記録したビデオテープを資料とした。検査技師が受検者に与えた指示を抜き出し、重複した指示内容や技師が受検者にかけた気遣いの言葉を除くと 55 種類の文で構成されていた。それらの中から胃部レントゲン検査の手順を最初から最後まで表現するのに最低限必要と考えられる 27 種類の指示文を用いることとした(表1)。

手話アニメーションの評価実験方法：

評価実験では制作者の名にちなんで、中田アニメーション、内川アニメーション、川村アニメーション、ならびに手話通訳者が行った手話の4つの胃部レントゲン検査における手話表現による指示文を比較した。

表1 検査技師から受検者への指示文

文番号	指示内容
1	胃を膨らませるために薬を飲んでください
2	ゲップはしないでください
3	薬をすべて飲んでください
4	左右の棒をしっかりと持ってください
5	台を倒します
6	手を下ろしてください
7	力を抜いてください
8	右腰を上げてください
9	ゆっくり腰を戻してください
10	右から腹ばいになってください
11	顔は左を向いてください
12	ゆっくり仰向けになってください
13	少し息を吸ってください
14	お腹を膨らませて、しっかりと息を止めてください
15	案にしてください
16	足を伸ばして、下につけてください
17	台を起こします
18	棒でお腹を触ります
19	右肩を少し前に出してください
20	コップを左手に持ってください
21	少し口に含んでください
22	合図したら飲み込んでください
23	はい、飲み込んでください
24	アゴを少しあげてください
25	少し揺らします
26	右を向いて前のめりになってください
27	全て終わりました

被験者を液晶ディスプレイの前に座らせ、実験方法やレントゲンの受検状況の説明を紙面で行った。被験者にこれから見せる映像が胃部レントゲン検査の指示を手話で表現したものであることを説明し、また、被験者がレントゲン検査の現場を想定しやすいように、実際の検査の状況がどのようなものであるかを、レントゲン検査の現場を撮影した写真によって説明した。

1) 一文毎のわかりやすさの評価

被験者には液晶ディスプレイ上に再生される手話による指示文を読み取らせ、その内容を予め配っておいた回答用紙に記入させた。これを一文毎に正答であるかどうかを判断した。なお、手話文は1文につき2度表示され、1度再生する度に回答用紙に記入させた。

2) 一文毎の手話文についての主観評価。

手話のわかりやすさの程度を評価した。評価項目は以下の3つであった。1文毎に5段階で評価をさせた。評価項目と評価尺度値の内容は以下の5項目である。

- ・手話の意味が読み取れたか
 - 1 読み取れなかった ～ 5 読み取れた
- ・単語の切れ目は分りやすかった
 - 1 分りにくかった ～ 5 分りやすかった
- ・手の動きの速さは適切だったか

1 遅かった ～ 3 適切だった ～ 5 速かった

また、手話表現について何か気付いた点、あるいはどこをどの様に改良したら分りやすい手話表現になるかについて任意で記入させた。

被験者は20代から60代の読み書きのできる聴覚障がい者32名。手話暦は約2年以上から50年以上の男性10名、女性22名であった。被験者32名を1グループ2名として16グループに無作為に分け、各被験者は1から27までの指示文を番号順に回答と評価を行った。同じ指示文番号毎に、評価したい3種類の手話アニメーションと手話通訳者が行う手話の計4種類の手話表現による指示が存在する。この4種類の手話表現の中から筆者が適当に選んだものを1グループに評価させた。

実験を始める前に、被験者に実験内容を紙面により説明し、アンケート用紙を配った。次に、ディスプレイに映し出される手話を読み取った内容を記入させるための回答用紙を配り、記入が終わると、その手話内容を記したものと主観評価用紙を配った。

実験結果：

(a) 指示文の正答率

全27文の平均正答率は、手話通訳者による手話表現では約70%、川村アニメーションが約56%、内川アニメーションが約49%、中田アニメーションが約26%であった。

指示の1文毎の平均正答率を図4に示す。川村アニメーションによる手話表現と手話通訳者による手話表現の正答率を比較する。川村アニメーションの方が手話通訳者に比べて6文(1, 9, 12, 17, 18, 25)において正答率が上回っていた。また、正答率が同じであったのが7文(3, 5, 11, 13, 20, 25, 27)であり、川村アニメーションの方が手話通訳者に比べて正答率が下回っていたのは残りの14文であった。次に、川村アニメーションと内川アニメーションの正答率を比較する。正答率は7文において同じであり(2, 13, 14, 15, 22, 25, 27)、川村アニメーションの方が内川アニメーションに比べて正答率が下回っていたのは7文(4, 8, 21, 23, 24, 28)であった。これ

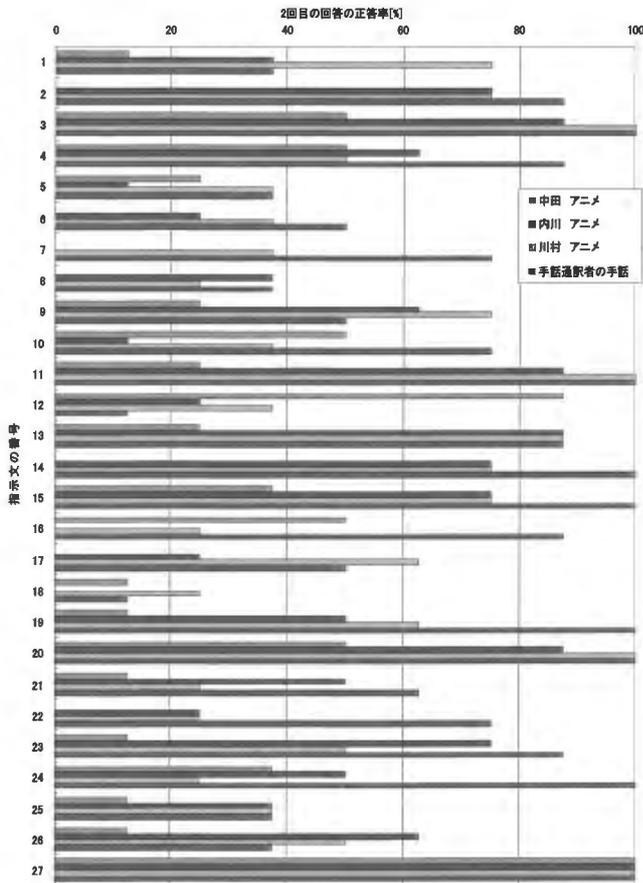


図4 指示の一文毎の平均正答率

ら以外の14文において川村アニメーションの方が内川アニメーションに比べて正答率が上回っていた。

さらに、川村アニメーションと中田アニメーションによる手話表現の正答率を比較する。川村アニメーションの方が中田アニメーションに比べて正答率が下回っていたのは4文であった。22文において川村アニメーションの方が中田アニメーションに比べて正答率が上回っていた。

この結果から、今回作成した手話アニメーションによる指示は、改良前の手話アニメーション（中田・内川アニメーション）による指示よりわかりやすかったことがわかる。しかし、手話通訳者の手話に比べるとまだまだ多くの指示文において手話アニメーションによる指示が、十分にはわかりやすいものになっていないことが窺える。

川村アニメーションと内川アニメーションの正答率

を比較する。内川アニメーションを改良した川村アニメーションの正答率が上がったのは、1, 3, 5, 6, 7, 9, 10, 11, 12, 16, 17, 18, 19, 20の指示文であり、変化の幅は12.5%~37.5%であった。正答率に変化がなかったのは、2, 13, 14, 15, 22, 25, 27の指示文であった。一方、正答率が下がったのは、4, 8, 21, 23, 24, 26の指示文であり、変化の幅は12.5%~37.5%であった。正答率が低下したこの6つの文は全てが体や腕を使った身振りで表現されていた文であった。

なお、今回の川村アニメーションを生成するうえで大きな改良をしなかった指示文3, 14, 27について述べる。これらの指示文は内川アニメーションの評価実験において、高い正答率（80%以上の正答率）が得られた指示文であるため、川村アニメーションを生成するうえで特に大きな改良を行わなかった。この3つの指示文を無視すると、24文中13文の正答率が上がり、5文の正答率に変化がなく、6文の正答率が下がった。

(b) 主観評価尺度値

胃部レントゲン検査における指示文の手話アニメーションと手話通訳者の手話について、27文の主観評価の平均を取ったものを表2に示す。

主観評価項目「手話の意味が読み取れたか」、「単語の切れ目はわかりやすかったか」、「手の動きの速さは適切だったか」について、1文毎の平均結果を見ると、手話通訳者、川村アニメーション、内川アニメーション、中田アニメーションの順に単語の意味が読み取りやすい、あるいは単語の切れ目が分りやすかったという評価であった。

また、主観評価項目「手の動きの速さは適切だったか」については、手話通訳者、川村アニメーション、中田アニメーション、内川アニメーションの順に手の動きの

表2 主観評価尺度値

	意味が読み取りやすいか	単語の切れ目は分かりやすいか	手の動きの速さは適切か
中田	2.3	2.5	3.3
内川	2.9	3.0	3.3
川村	3.1	3.1	3.2
手話者	3.8	3.9	3.2

速さが適切であった。川村アニメーションの正答率と主観評価の関係について述べる。先述のように、内川アニメーションよりも正答率が低下した指示文は4, 8, 21, 23, 24, 26であった。指示文23以外の指示文4, 8, 21, 24, 26において、内川アニメーションよりも意味が読み取りにくく、さらに単語の切れ目がわかりにくかった。また、指示文26においては川村アニメーションが他のどの3つの手話よりも手話動作が速いという結果であった。

(c) 考察

手話アニメーションの改良により全体的に正答率や主観評価は良くなったが、依然として改良前のアニメーションから正答率が伸び悩む単語がある。具体的には、指示文5「台を倒します」、6「手を下ろしてください」、8「右腰を上げてください」、16「足を延ばして、下に付けてください」、18「棒でお腹を触ります」、25「少し揺らします」である。どの単語も正答率が40%を下回っており、わかりやすいアニメーションというには程遠い。これらの単語は先ほど述べたように、指示内容を身振りにより伝えようとする指示文であり、身振りが行う内容がわからなければ指示文を理解しにくいためであると考えられる。

手話アニメーションの場合に、手話通訳者の行う身振りによる表現内容が伝わりにくい要因として次の事が考えられる。手話アニメーションは手話通訳者が行う身振りのように細かな動きを表現しにくい。さらに、人間である手話通訳者に比べて口頃見慣れない手話アニメーションの表情や表現が不自然で読み取りにくかったものと考えられる。被験者からの意見として、手話アニメーションによる手話動作はぎこちなく不自然であるとの声があった。

従来の手話アニメーションに改良点を施す事により、以前よりもわかりやすいアニメーションになる事が、今回の実験で確かめることができた。しかし、改良したアニメーションは手話通訳者が行う手話と比べるとまだまだわかりにくい。

さらにわかりやすいアニメーションを生成するため

の課題としては、指示文の内容を身振りにより伝えようとするアニメーションの表現方法を変更するか、あるいは手話表現を加えることや、細かな表情や表現が可能となるようなアニメーションを作成するために、手話アニメーションモデルの表現力を上げることがある。そのため、手話の時空間構造に関する研究もなされている[川村ら, 2005]。また、アニメーションシステム自体の問題として、口話の際の口の開きが小さい、あるいは頬を膨らませることが難しいということがある。たとえば、指示文 21「口に含む」などの手話単語が伝わりにくいので、頬を膨らませるといった手話アニメーションモデル自体の表現力を上げる必要がある。また、当然ながら自然で分かりやすい表現を手話アニメーションに導入する必要がある[morimoto et al., 2006]。

(3) マンモグラフィ検査時の指示用アニメーションの開発

マンモグラフィ検査時の検査技師と聴覚障がい者のコミュニケーションを調査し、マンモグラフィ検査用の指示文を求めた。手話アニメーションの作成にはそのマンモグラフィ検査用指示文を手話者に表現してもらい、それをビデオで撮影した。手話を表現したのは普段から手話をコミュニケーション手段として用いている女性の聾者である。この時、表現してもらった手話は指示文の文章そのままでは無く、指示文の意味に一番近くなるような表現をお願いした。よって、単に手話単語だけではなく、ジェスチャー表現も多用されている。

この撮影した映像を基に手話者が行った手話や顔の表情に近づくように手話辞書などを参考にし、手話アニメーションを生成した。その手話アニメーションを聴覚障がい者にとって分かりやすいものにするために聾者と手話通訳士に手話アニメーションを見せて、改良のための意見をもらった。その際に得られた問題点を考慮し、各手話アニメーションの修正を行った。さらに、各指示文の手話アニメーションの読み取り実験を行い、正答率等を求めた[Takahashi et al., 2009]。

被験者 24 人の平均正答率は 40.5%であった。用いた 39 文の中で最も正答率が高かったものは指示文「大丈夫

ですか」であり、正答率は100%であった。逆に、最も正答率の低かったものは指示文「線のところに立ってください」で、正答率は0%であった。被験者ごとの正答率では、最も正答率が高かった被験者は正答率69.2%であり、最も正答率の低かった被験者は正答率15.4%であった。

正答率や主観評価の結果を総合すると、手話アニメーションの正答率を更に向上させるには以下の改良点が指摘できた。まず、細かな表情や表現が可能な手話アニメーションを作成するために、手話アニメーションアバターの表現力を上げる必要がある。現状のシステムでは指先の細かい動きなどが表現できないとか、口話の時間より短い時間で手指動作を終えることができないという問題がある。

次に、手話表現を自然で分かりやすいものにするには、実験結果の意見を参考にして聴覚障がい者が普段使っている手話表現を用いて、手話表現をより自然なものにしていく必要がある。

また、正答率の極端に低い指示文に関しては、手話表現の変更や指示文自体の変更も考慮しなければならないと考える。正答率が0%であった指示文「線のところに立ってください」はジェスチャーで表現したが、手話単語でも表現することは可能である。そちらの表現へ変更することも考慮する必要があると考えられる。また、被験者の意見にもあったように、なるべく単純で分かりやすい意味の指示文を作成し、よりの確なコミュニケーションが取れる手話アニメーションにする必要がある。

今後はこれらの点に対しても修正を加え、実際にマンモグラフィ検査に使用できるように、確実に指示が伝わるように改良しなければならない。改良方法の一つとしては、受検者が検査の状況や指示を理解しやすくするために、手話アニメーションの指示の中に検査機械など実際の医療現場にあるものを画面に映すことで、ジェスチャー表現や検査状況などを理解するための補助をすることができると考える。また、ほとんどの被験者から字幕が欲しいとの意見があった。これも手話や口話を読み取れなかった場合の補助として重要だと考えられるの

で、実際の現場での利用に際しては文字情報も提示するのがよいと考える。

4. まとめ

聴覚障がい者と健聴者の両者が手話通訳者を介さずに、円滑なコミュニケーションを行うことが出来るようにするための一手法として開発してきた手話・日本語間相互翻訳システムの一部を紹介した。改良は進んできたが、現場に導入するには課題の多いこともわかった。紙幅の都合で記載できなかったが、このシステムを応用して、現在は3次元手話アニメーションや中国手話アニメーションの生成にも取り組んでいる。

なお、手話アニメーションの生成には手話をどのように表記するかが表現方法に影響する。このため手話アニメーション生成に適した手話表記統一フォーマットの提案もされている。また、手話表現に関しては、手話の使われる状況により表現の異なることはよく知られており、それをどのようにコード化して手話アニメーションに反映させればよいか今後の課題である。

謝辞

本研究は科学研究費（基盤研究(B)(1)14370122ならびに基盤研究(B)16300029）の補助を受けた。謝意を表する。

参考文献

- 川村, 森本, 黒川 (2005) 手話アニメーションの時空間構造と理解度の関係—胃部レントゲン検査指示用の場合—, ヒューマンインタフェースシンポジウム 2005 論文集, 671-676.
- 北原, 埜田, 西山 (2001) 聴覚障害者の受療に関する医療機関側の調査—医療機関を対象とした面接調査の分析—, 社会医学研究, 19, 45-56.
- 森本, 川村, 黒川 (2005) 胃部レントゲン検査の指示に用いる手話アニメーションの作成とその評価, 電子情報通信学会技術研究報告, 67, 37-42.
- K. Morimoto, T. Kurokawa and S. Kawamura (2006) Improvements and Evaluations in Sign Animation Used as Instructions for Stomach X-Ray Examination, ICCHP 2006, LNCS4061, 607-614.
- M. Takahashi, N. Kuwahara, K. Morimoto (2009) Design of Sign Animation for Mammographic Screening, Human Computer International 2009, Elsevier, 759.
- 埜田, 北原, 西山 (1996) 連続手話通訳作業の負担に関する実験的研究, 産衛誌 38, 59-69.

文末表現を考慮した文章の特徴量を用いた質問回答文の因子得点の推定 Estimation of factor scores using feature values of question and answer statements with consideration of sentence-end expressions

横山 友也*, 宝珍 輝尚*, 野宮 浩揮*, 佐藤 哲司**

Yuya Yokoyama, Teruhisa Hochin, Hiroki Nomiya, Tetsuji Satoh

* 京都工芸繊維大学 情報工学専攻, 京都市左京区松ヶ崎御所海道町

Kyoto Institute of Technology, Graduate School of Information Science

Goshokaidocho, Matsugasaki, Sakyo-ku, Kyoto

** 筑波大学大学院 図書館情報メディア研究科, 茨城県つくば市春日 1-2

Graduate School of Library, Information and Media Studies, University of Tsukuba

1-2 Kasuga, Tsukuba-shi, Ibaraki

あらまし: Q&A サイトにおける質問者と回答者のミスマッチの問題を解消するために, 50 語の印象語を用いて, Yahoo!知恵袋に投稿された 60 個の質問回答文の印象評価実験を行った. 実験を通して得られた結果に因子分析を適用したところ, 文章に関する因子が 9 個得られた. 重回帰分析を使用することによって, 他の文章の因子得点の推定を試みてきている. 品詞の数や記号の数といった文章の特徴量を加えて, 推定精度を向上させるために, 本論文では文末表現を考慮して重回帰分析を試みる. 分析の結果, 文末表現を考慮することによって, 因子得点の推定精度の向上が可能であることを示す.

Summary: In order to avoid the problem of mismatch between the questioner and the respondent at Q & A sites, we have experimentally evaluated the impression of 60 question and answer statements posted at Yahoo! Chiebukuro by using 50 impression words. Nine factors as to statements have been obtained by applying the factor analysis to the scores obtained through the experiment. Factor scores of any other statements have been tried to be estimated by using multiple regression analysis. To improve the estimation accuracy, this paper tries the multiple regression analysis considering sentence-end expressions by adding such feature values as word classes like noun and verb, and the number of appearance or the percentage of Chinese and alphanumeric characters. The result of the analysis shows that the estimation accuracy of factor scores could be improved with the considerations of sentence-end expressions.

キーワード: Q&A サイト, 印象評価実験, 因子得点, 重回帰分析, 二次の項, 文末表現

Keywords: Question & Answer site, impression evaluation, factor score, multiple regression analysis, quadratic term, sentence-end expression

1. はじめに

インターネット上において, 質問回答サイトの利用者が近年急増している. 質問回答サイトとは, インターネット上でユーザー同士が互いに質問と回答を投稿しあうコミュニティの一種であり, 様々な悩み事・相談事を解決する場であると同時に, 膨大な知識が蓄積されたデータベースとして活用されるようになってきている. あるユーザーが質問を投稿すると, 他のユーザーがその質問に対して回答を投稿する. 質問者は, 質問文

に対して最も適切と判断した回答文を「ベストアンサー」に選定し, その回答を行った回答者に謝礼として手持ちのポイントを贈与する. ここで, 「ベストアンサー」とは, 質問文に対する満足度が最も高いと質問者が主観的に判断した回答文である.

質問回答サイトの参加者が増え, また, 投稿される質問数が膨大になると, 回答者が自身の専門性や興味に合った適切な質問文を探し出すことが困難になるという問題が顕在化してくる. あるユーザーが質問文を

投稿しても、その質問文が必ずしも適切な回答者の目に留まり、回答を得られるわけではないという問題である。また、適切な回答者に巡り会えないミスマッチから、質問者にも不利益も生じる。つまり、質問回答サイトの課題は、日々投稿され続けている幾多の質問と、様々な興味・関心や専門性を有する回答者とを適切にマッチングすることであるが、質問者や回答者の努力に任せているのが現状である。そこで、ある質問文に適切な回答ができるユーザーをその質問文に引き合わせるための方法が研究されている[2-16]。

これまでの研究において、筆者らは、質問者に適切な回答者を引き合わせるために、質問者と回答者の相性を判断する手段として質問者と回答者の文章の印象評価を行ってきた[2,3]。Yahoo!知恵袋[1]に投稿された質問文と回答文の計60個の文章に対して、50個の印象語を使用して、印象評価を行った。その結果、文章内容に関する因子が9個得られた。また、得られた因子の因子得点を適宜利用することで、通常の見返り文に対し、「ベストアンサー」を特定できる可能性を示した[2,3]。

しかし、ここで得られた因子得点は、評価実験を行った結果得られた質問文と回答文の文章60個に対するもののみであって、他の多数の質問文と回答文に対する因子得点は得られていない。そこで、どのような見返り文に対しても「ベストアンサー」の推定を可能にすることを目的として、文章の特徴量からの文章の因子得点の推定を試みている。ここでは、重回帰分析を使用して、因子得点を推定する。文章の特徴量として、名詞や動詞などの品詞、ひらがな、カタカナ、英数字の出現回数あるいは比率など、形態素解析を使用して求められるものを採用している[4]。

また、NTT データベースシリーズ[5,6]には、人が主観的に評定を行ったデータと、14年間にわたる新聞記事に出現した単語や文字の出現回数を計測した客観的データも収録されている。これらのデータは、人間の言語処理過程に大きな影響を及ぼすものとして広く知られており、収録されている各特性値や特性値間の関係は、日本語自体の特性を示しているといえる[6]。これらのデータも文章の特徴量として有用であると考えたため、単語心像性も文章の特徴量に追加して、因子得点の推定を試みたところ、4因子に関してはやや良好の推定精度が得られたが、残り5因子に関しては満足な推定精度が得られなかった[7]。また、二次の項(説明変数同士の積)[8]を考慮して推定精度の向上を図ったところ、6因子に関してはやや良好の推定精度が得られたが、残り3因子に関しては満足な推定結果を獲得できなかった[7]。

そこで、本論文では、推定精度を更に向上させるために、文末表現[9]も考慮した上で、質問回答文の因子得点の推定精度の向上を試みる。分析の結果、どの因子も推定精度が向上し、文末表現も考慮することで推定精度が向上できることを示す。

以降、2.では関連研究について述べ、3.では質問回答文に対する印象評価実験とその結果について述べる。4.ではこれまでの文章の因子得点の推定結果について述べる。そして、5.で文末表現も考慮した推定結果について述べ、6.では分析結果を評価する。最後に7.でまとめる。

2. 関連研究

これまでに、「ベストアンサー」を推定する研究が行われてきている[9-16]。Bloomらには、非テキスト特徴量とテキスト特徴量を用いて、「ベストアンサー」の推定を試みている[10]。Agichteinらには、内容や語法の特徴量を使用することによって、質問文と回答文の質の評価を試みた[11]。また、類推による手法も提案されてきている[12]。この手法では、過去の知見における質問文と回答文のリンクを使用することによって、「ベストアンサー」を探索する。Kimらには、「ベストアンサー」の選択基準を提案している[13]。情報型の質問には、文章内容の特徴量が重要である。また、提案型の質問には有用性が重要であり、選択型の質問には社会的な感情が重要であるということである[13]。

西原らには、ある1つの質問文に対する回答文群より、「ベストアンサー」になりやすいものを検出する手法を提案している。質問者と回答者の文末表現の相性に着目し、質問文と「ベストアンサー」の組み合わせをクラスタリングすることで、一定の成果をあげている[9]。しかし、この研究では質問文ならびに回答文の文末表現に着目した手法をとっており、文章内容に着目した手法をとっていない。そこで、本研究では、西原らが使用した文末表現[9]も特徴量として考慮した上で、文章全体の文体および内容から受ける印象評価に着目した検討を行う。

一方、熊本は新聞記事を対象として印象評価を行っている[17]。被験者100人が新聞記事10記事を読んで、印象語42語のそれぞれについて5段階(強い—わりと強い—わりと弱い—弱い—なし)で評価するという印象評価実験(アンケート調査)を9度実施して、印象評価データ(印象語42語×9000件)を収集・分析することによって、新聞記事の印象を表現するのに適した印象軸を提案している[17]。本研究では、質問回答文を対象として、印象語を用いた印象評価実験を行う。

3. 印象評価実験

3.1. 印象語

印象語とは、文章を評価するにあたって、その文体や内容から受けた印象や評価を表すために使用する語である。ここでは、まず「日本語の語彙特性」[18]に掲載された全 21690 語のうち、印象語に用いられ得ると第一著者が判断した語を抜粋した。

最初に抜粋した語の数は全部で 806 語である。次に、以下の手順で第一著者が印象語を選出した。

(1)806 語の印象語候補を、以下に示す A 群・B 群・C 群に分類する。

- ・A 群: ポジティブな意味に用いられやすい語
- ・B 群: ネガティブな意味に用いられやすい語
- ・C 群: A 群にも B 群にも分類できない語

(2)それぞれの群において、類似した意味を持つ語のグループを生成する。

(3)各グループに分類された語について、文章評価にあたって最も一般的な語を考査・選定する。

(4)類似した意味の語を併合する。

以上により、50 語を選出した。選出した語は、文体に関する語と文章内容に関する語に大別できる。精選した印象語の一覧を表 1 に示す。

3.2. 実験

3.2.1. 実験の前準備

表 1 に示した 50 語の印象語を用いて、19～23 歳の被験者 41 名(男 33 名, 女 8 名)を対象に印象評価実験(アンケート調査)を行った。被験者は、熊本が施行した実験[17]と同じ指標による段階評価で評価する。評価素材は、2005 年 9 月に Yahoo!知恵袋に実際に投稿された質問回答のうち、Yahoo!オークション、パソコン・周辺機器、恋愛相談・人間関係、政治・社会問題の 4 大カテゴリから各 3 組、全 12 組を選出し、それぞれの質問回答文において質問文と、それに対する回答文を 4 個(「ベストアンサー」を含む)選出した計 60 個の文章である。

表 1 選定した印象語

文章表現に関する印象語(22語)		文章内容に関する印象語(28語)	
易しい	しつこい	涙ぐましい	憤慨した
巧みな	たどたどしい	素晴らしい	幻滅な
丁寧な	味気ない	好ましい	怖い
美しい	不十分な	感動的な	残念な
清々しい	大袈裟な	的確な	不当な
流暢な	細かい	妥当な	非常識な
特殊な	単純な	重要な	呆れる
説得力がある	堅い	心温まる	真実味
明瞭な	長い	独創的な	仕方が無い
曖昧な	複雑な	充実した	熱い
難しい	斬新な	楽しい	力強い
		不快な	予想外な
		怪しい	不思議な
		辛辣な	懐かしい

なお、文章を選抜した基準は次の通りである。

- ・バイアスが強すぎる文章は除外
- ・公序良俗に反した内容を含む文章は除外
- ・他の HP の URL を参照する必要のある質問は除外
- ・短い文章と長い文章を織り交ぜることで、文字の量をバランス良く選抜
- ・内容、論旨の類似した文章を重複しないよう選抜
- ・各ジャンルから均等に選抜

12 組の質問回答文について、全て異なるジャンルを均等に織り交ぜるように、各組に対する回答記入用紙を回答記入冊子 A, B, C に振り分けた。そして、冊子単位では冊子 A→冊子 B→冊子 C の順で回答を行ってもらい、各冊子においては以下の手順で回答を行ってもらう。

I. 冊子 A: § 1【Q1→A1-1→A1-2→A1-3→A1-4】

⇒ § 4【Q4→A4-1→A4-2→A4-3→A4-4】

⇒ § 7【Q7→A7-1→A7-2→A7-3→A7-4】

⇒ § 10【Q10→A10-1→A10-2→A10-3→A10-4】

II. 冊子 B: § 2【Q2→A2-1→A2-2→A2-3→A2-4】

⇒ § 5【Q5→A5-1→A5-2→A5-3→A5-4】

⇒ § 8【Q8→A8-1→A8-2→A8-3→A8-4】

⇒ § 11【Q11→A11-1→A11-2→A11-3→A11-4】

III. 冊子 C: § 3【Q3→A3-1→A3-2→A3-3→A3-4】

⇒ § 6【Q6→A6-1→A6-2→A6-3→A6-4】

⇒ § 9【Q9→A9-1→A9-2→A9-3→A9-4】

⇒ § 12【Q12→A12-1→A12-2→A12-3→A12-4】

ここで、§ 1～§ 3 は Yahoo!オークションに関する質問、§ 4～§ 6 はパソコン・周辺機器に関する質問、§ 7～§ 9 は恋愛相談・人間関係に関する質問、§ 10～§ 12 は政治・社会問題に関する質問を、それぞれ掲載している。また、Q はそのセクションで取り扱っている質問文、A はその質問文に対する回答文である。例えば、Q1 ならば § 1 の質問文であり、A1-1, A1-2, A1-3, A1-4 は Q1 に対する回答文である。ここで、質問文の後に回答文を読んでもらうのは、回答文によっては質問文を読まないとい何に対する回答かがわからない文章があるからである。なお、各回答文のうち、A1-1, A2-1, …, A12-1 は実際に Yahoo!知恵袋で「ベストアンサー」に選出されたものである。ただし、被験者にはこのことは伝えない。

使用した質問文と回答文の例として、典型的な例を付録に示す。

3.2.2. 実験結果

実験の結果、各印象語に対して 2460 件(41 人×60 個の文章)の印象評価データが得られた。ここで、性差が印象評価に影響を与えているか否かを検証する

ために、男性33名の回答結果(33人×60個の文章×50個の印象語)と女性8名の回答結果(8人×60個の文章×50個の印象語)との間で、1%の有意水準で有意差があるかを検定した。その結果を表2に示す。これより、男女間に有意差があることが判明したため、以降では、男性の回答結果のみを使用することとした。

また、被験者の疲労効果を測るために、冊子A, B, C間に有意差があるかを、1%の有意水準で検定を行った。その結果を表3に示す。p値が有意水準より大きいので、冊子間に有意差が無いことが判明した。従って、疲労による影響は無いと判断した。そこで、以降では、全ての冊子の印象評価データを使用することにする。

印象評価データに対して、因子分析を実施し、バリマックス回転を施した。ここで、選択因子数を打ち切る基準として「固有値が1.0以上」[19]を採用して、因子数を9個とした。各因子の固有値、寄与率、累積寄与率を表4に示す。因子分析の結果、的確性、不快性、独創性、容易性、執拗性、曖昧性、感動性、努力性、熱烈性という9個の因子が得られた。9つの因子とそれぞれに対応する印象語をまとめて表5に示す。

また、これらの因子の因子得点を利用することで、通常の実験文に対し、「ベストアンサー」を特定できる可能性を示してきた[2,3]。

4. これまでの因子得点の推定

4.1. 形態素解析により抽出した特徴量

文の長さや数、品詞の数を求めるために、Text Seer[20]を用いて形態素解析を行った。ここでは、表6に示す64個の特徴量を使用している[4]。

文章において、複数回出現する単語が存在する可能性が高いことを考慮して、語彙数と語数とは独立した特徴量として抽出する。ここで、語彙数とは、文章中に同じ単語が複数回出現した場合でも1個と数えることを表し、語数とは、単純に単語の出現回数を表す。例えば、「私は私の夢を叶える」という文章を例にすると、「私」という単語が2回出現しているため、「私」という1つの語彙に対して、「私」の語数は2である。

表2 性差の有意差検定の結果

性別	平均	分散	p値
男性	1.255	1.143	1.768E-33
女性	1.627	1.937	

表3 冊子間の有意差検定の結果

冊子	平均	分散	p値
冊子A	1.228	1.563	0.429
冊子B	1.240	1.660	
冊子C	1.238	1.623	

表4 因子の固有値、寄与率、累積寄与率

因子	固有値	寄与率[%]	累積寄与率[%]
1	11.098	14.5	14.5
2	6.777	11.4	25.8
3	2.686	6.1	31.9
4	2.238	3.7	35.6
5	1.575	3.6	39.2
6	1.500	3.5	42.8
7	1.429	3.1	45.9
8	1.184	2.1	48.0
9	1.098	2.0	50.0

表5 9つの因子と対応する印象語

因子	印象語		
第1因子(的確性)	説得力がある 素晴らしい 真実味がある 充実した 丁寧な	流暢な 好ましい 清々しい 美しい	重要な 巧みな 妥当な 的確な
第2因子(不快性)	不快な 残念な 幻滅した	憤慨した 不当な 怖い	非常識な 呆れる
第3因子(独創性)	独創的な 斬新な	予想外な 不思議な	特殊な
第4因子(容易性)	易しい	明瞭な	難しい
第5因子(執拗性)	細かい	しつこい	長い
第6因子(曖昧性)	曖昧な	不十分な	
第7因子(感動性)	心温まる	感動的な	
第8因子(努力性)	涙ぐましい		
第9因子(熱烈性)	熱い	力強い	

また、ひらがな、漢字、カタカナ、記号、英数字に関しては、出現の割合によって印象が変わると考えられる。そこで、それぞれの出現回数だけではなく、それらを含む文章そのものの長さ及び文章内における含有率(全文字数に対する当該字種の文字数の比率)も考慮する必要があると考え、ひらがな、漢字、カタカナなどの文章における含有率も特徴量とした。例えば、表6のf16の「ひらがな(語数)」は、文章内のひらがなの単語数を表し、f36の「ひらがな(%)」は文章におけるひらがなの含有率を表している。

f26の「未知語」とは、Text Seerをデフォルトの状態で使用し、「未知語」と判定された語数を表している。なお、未知語と判定された語は、名詞または記号として辞書に登録した上で、改めて形態素解析を行った。f40のTTR(Type Token Ratio)は、語数に対する語彙数の比率を表している。

4.2. 単語心像性

NTT データベースシリーズ[5]には、人が主観的に評価を行ったデータと、14年間にわたる新聞に単語や文字が出現した回数を数えた客観的データが収録されている。これらのデータは、人間の言語処理過程に

表6 文章の特徴量(64個)

f	特徴量	f	特徴量
f1	文字数	f33	感動詞(語数)
f2	名詞(語彙数)	f34	助動詞(語数)
f3	動詞(語彙数)	f35	助詞(語数)
f4	形容詞(語彙数)	f36	ひらがな(%)
f5	副詞(語彙数)	f37	漢字(%)
f6	連体詞(語彙数)	f38	カタカナ(%)
f7	接続詞(語彙数)	f39	記号(%)
f8	感動詞(語彙数)	f40	TTR
f9	助動詞(語彙数)	f41	全角記号(%)
f10	助詞(語彙数)	f42	英数字(%)
f11	接頭詞	f43	全角英数字(%)
f12	記号(語彙数)	f44	半角英数字(%)
f13	文数	f45	名詞(%)
f14	文の長さ平均(語数)	f46	動詞(%)
f15	文の長さ平均(字数)	f47	形容詞(%)
f16	ひらがな(語数)	f48	副詞(%)
f17	漢字(語数)	f49	連体詞(%)
f18	カタカナ(語数)	f50	接続詞(%)
f19	記号(語数)	f51	感動詞(%)
f20	全角記号(語数)	f52	助動詞(%)
f21	英数字(語数)	f53	助詞(%)
f22	全角英数字(語数)	f54	「!」の数
f23	半角英数字(語数)	f55	「?」の数
f24	語数	f56	句点の数
f25	語彙数	f57	読点の数
f26	未知語	f58	中点の数
f27	名詞(語数)	f59	3点リーダーの数
f28	動詞(語数)	f60	鍵括弧の数
f29	形容詞(語数)	f61	鍵括弧閉の数
f30	副詞(語数)	f62	括弧の数
f31	連体詞(語数)	f63	括弧閉の数
f32	接続詞(語数)	f64	「/」の数

大きな影響を及ぼすものとして広く知られており、収録されている各特性値や特性値間の関係は、日本語自体の特性を示しているといえる[6]。これらのデータも文章の特徴量として有用であると考えられる。

これらのデータの中でも、単語心像性を文章の特徴量に追加する。単語心像性とは、単語から喚起される様々なイメージが、どの程度思い浮かべやすいかを示す主観的特性である。例えば、「りんご」という言葉を聞くと、赤・黄・緑の丸い形の果物、甘みずみずしい味・匂い、サクッとした音や歯ざわり、持った時の感触を思い浮かべることができる。一方、「世界」「経済」は、「りんご」に比べると具体的なイメージを思い浮かべにくいと思われる[5]。

ここで、単語心像性の特性値は、「単語の非言語的感覚イメージの喚起力」に関して、「1:イメージを非常に思い浮かべにくい(または思い浮かべない)~7:イメージを非常に思い浮かべやすい」の7段階尺度で評定させた値である。新聞記事を対象としたデータ[5]と、質問回答文の文章を形態素解析したデータとを比較して、収録データに合致する単語が形態素解析したデータに存在するならば、その単語の単語心像性の値を特性値として使用する。なお、形態素解析したデ

ータに収録データと単語が合致しない場合、その単語の単語心像性の値は考慮しないものとして処理する。

また、単語の同じ表記でも、意味または読みが異なる場合がある。例えば、意味が異なる例としては、「アース」という単語は、「電気を逃がすために接地すること」、「地球」、「殺虫剤(メーカー)」の意味がある。読みが異なる例としては、「間」という言葉は、「あいだ」、「ま」の読みがある。このような単語が形態素解析したデータに存在する場合は、文脈から判断しながら手動で意味または読みを決定する。

このようにして、単語心像性の特徴量を抽出した。これを表7に示す。特徴量としては、単語心像性に該当した単語の数や該当した単語の割合や、単語心像性の値が1点台、2点台……のように、1点間隔で特徴量をとったものや、1.0以上1.5未満、1.5以上2.0未満、……のように、0.5点間隔で特徴量をとったもの、を採用した[8]。

4.3. 多重共線性の考慮

重回帰分析を実施する際は、複数の説明変数同士は無相関であるという前提が必要となり、説明変数は以下の条件を考慮して選択しなければならない。

- a) 目的変数との相関係数が高い説明変数の選択
- b) 高い相関を示す説明変数の組のうち、一方を説明変数から除外

ここで、b)の事項に反した場合、偏回帰係数が正しく求まらないことがあり、この状態を多重共線性という。

表7 単語心像性の特徴量

該当単語(語彙数)	1点台(語数)
該当単語(語数)	1.0~1.5未満(語数)
該当単語率(語数)	1.5~2.0未満(語数)
1点台(語彙数)	2点台(語数)
1.0~1.5未満(語彙数)	2.0~2.5未満(語数)
1.5~2.0未満(語彙数)	2.5~3.0未満(語数)
2点台(語彙数)	3点台(語数)
2.0~2.5未満(語彙数)	3.0~3.5未満(語数)
2.5~3.0未満(語彙数)	3.5~4.0未満(語数)
3点台(語彙数)	4点台(語数)
3.0~3.5未満(語彙数)	4.0~4.5未満(語数)
3.5~4.0未満(語彙数)	4.5~5.0未満(語数)
4点台(語彙数)	5点台(語数)
4.0~4.5未満(語彙数)	5.0~5.5未満(語数)
4.5~5.0未満(語彙数)	5.5~6.0未満(語数)
5点台(語彙数)	6点台(語数)
5.0~5.5未満(語彙数)	6.0~6.5未満(語数)
5.5~6.0未満(語彙数)	6.5~7.0未満(語数)
6点台(語彙数)	
6.0~6.5未満(語彙数)	
6.5~7.0未満(語彙数)	

多重共線性を確認するには、「目的変数との相関係数」と「回帰係数」との符号が逆転している説明変数を調べる方法がある[21]. 符号が一致しない原因は、説明変数の組の中に高い相関のある説明変数が含まれているからである.

多重共線性を回避するために、表6と表7に示す説明変数に関して、説明変数同士の相関係数の値を調べ、0.7以上である組に関しては、一方を説明変数から除外した. その結果、説明変数は38個となった. これらを表6、表7に網掛けを施して示す.

4.4. 単項のみを考慮した推定結果

3.2.2.で述べた9つの因子の因子得点を、それぞれ y_1, y_2, \dots, y_9 とする. ここでは、3.の印象評価実験で使用した60個の質問回答文に対して、多重共線性を考慮した結果採用した38個の特徴量を説明変数とし、因子得点を目的変数として、ステップワイズ選択法[22]による重回帰分析を行った.

この結果、重回帰式(1)が得られた. ただし、第9因子に関しては、重回帰式が得られなかった.

重相関係数と、選ばれた説明変数を、それぞれ表8、表9に示す. 重相関係数は、その値が0.9以上ならば、分析精度が非常に良好であるとされ、0.7以上ならば、分析精度がやや良好であるとされ、0.7未満ならば、分析精度が不良であるとされている[23].

表8の結果から、第5因子(執拗性)は、0.9以上の値であるので、分析精度が非常に良好であるといえる. また、第1因子(的確性)は、0.7以上の値であるから、分析精度はやや良好であるといえる. 一方、その他の7因子は0.7未満の値であり、分析精度は良好とは言えない. また、第9因子は、該当する説明変数が得られなかった[7].

$$\left\{ \begin{array}{l} y_1 = 0.0287f_{65} + 0.195f_{22} + 0.0541f_{41} + 0.0118f_{20} \\ \quad - 0.0127f_{38} - 0.155f_{55} + 0.0794f_{29} - 0.426 \\ y_2 = -0.0938f_9 + 0.369 \\ y_3 = -0.0845f_9 + 0.0444f_{22} + 0.245 \\ y_4 = -0.0503f_{65} + 0.0828f_{29} - 0.161f_{55} + 0.0583f_9 \\ \quad + 0.214f_{54} + 0.103 \\ y_5 = 0.0898f_9 + 0.0157f_{18} + 0.0974f_{55} + 0.0130f_{37} \\ \quad + 0.00304f_{15} + 0.0268f_{13} + 0.403f_{59} - 1.14 \\ y_6 = -0.0624f_{57} - 0.0967f_9 - 0.0368f_{22} + 0.659 \\ y_7 = 0.456f_{33} + 0.0836f_{30} + 0.102f_{47} - 0.0330f_{56} \\ \quad - 0.193 \\ y_8 = 0.0914f_9 - 0.00605f_{42} + 0.0944f_{47} - 0.00907f_{20} \\ \quad - 0.336 \end{array} \right. \quad (1)$$

表8 重相関係数(単項)

因子	重相関係数
第1因子(的確性)	0.809
第2因子(不快性)	0.350
第3因子(独創性)	0.475
第4因子(容易性)	0.737
第5因子(執拗性)	0.881
第6因子(曖昧性)	0.734
第7因子(感動性)	0.562
第8因子(努力性)	0.590

表9 選択された各因子の説明変数(単項)

第1因子(的確性)	単語心像性4点台(語数) 接続詞(語数) 全角英数字(%) 全角記号(語数) カタカナ(%) 「?」の数 形容詞(語数)
第2因子(不快性)	助動詞(語彙数)
第3因子(独創性)	助詞(語彙数) 全角英数字(語数)
第4因子(容易性)	単語心像性4点台(語数) 形容詞(語数) 「?」の数 助動詞(語彙数) 「!」の数
第5因子(執拗性)	助動詞(語彙数) カタカナ(語数) 「?」の数 漢字(%) 文の長さ平均(字数) 文数 3点リーダの数
第6因子(曖昧性)	読点の数 助動詞(語彙数) 全角英数字(語数)
第7因子(感動性)	感動詞(語数) 副詞(語数) 形容詞(%) 句点の数
第8因子(努力性)	助動詞(語彙数) 英数字(%) 形容詞(%) 全角記号(語数)

$$\left. \begin{aligned}
 y_1 &= 0.0423f_{65} + 0.00649f_{12}f_{20} + 0.0365f_{43} - 0.00986f_{18}f_{30} - 0.422f_{11}f_{50} - 0.0201f_{49}f_{65} + 0.00655f_{30}f_{38} \\
 &\quad + 0.00510f_{42}f_{57} - 0.508f_{30}f_{51} + 0.142f_{31}f_{47} - 0.0139f_{38} - 0.00201f_{15}f_{43} + 0.00307f_{38}f_{39} - 0.0199f_{11}f_{38} \\
 &\quad - 0.00870f_{38}f_{55} + 0.00127f_{38}f_{65} + 0.00448f_{22}f_{56} + 0.00957f_{43}f_{57} - 0.543 \\
 y_2 &= -0.00282f_9f_{36} + 0.0249f_{43}f_{57} - 0.0141f_{18}f_{55} - 1.42f_{40}f_{40} - 0.00630f_{45}f_{47} + 0.0195f_{37}f_{40} + 1.10 \\
 y_3 &= 0.524f_{62}f_{64} + 1.30f_{40}f_{40} - 0.0132f_{36}f_{40} + 0.00253f_{37}f_{43} + 0.0103f_{18}f_{31} - 0.280 \\
 y_4 &= -0.0798f_{65} - 0.0997f_{48}f_{55} + 0.0910f_{29} - 0.0183f_{37}f_{40} + 0.486f_{31} + 0.00668f_{38}f_{47} + 0.00124f_{37}f_{65} \\
 &\quad + 0.109f_{49} + 0.518 \\
 y_5 &= 0.00509f_{13}f_{15} + 0.0693f_{48}f_{55} - 0.341f_{62}f_{64} - 0.199f_{43}f_{50} + 0.923f_{49}f_{59} - 0.0149f_{41} - 0.00625f_{22}f_{56} - 0.554 \\
 y_6 &= 1.21f_{40}f_{40} + 0.0165f_{37}f_{40} - 0.857 \\
 y_7 &= 0.439f_{33} + 0.131f_{30} - 0.530f_{31}f_{51} + 0.102f_{47} - 0.0242f_{48}f_{57} - 0.0118f_{56}f_{58} - 0.218 \\
 y_8 &= 0.00158f_9f_{36} - 0.00269f_{12}f_{20} - 0.00450f_{42} + 0.681f_{49}f_{59} - 0.231
 \end{aligned} \right\} (2)$$

4.5. 二次の項を考慮した場合の推定結果

単項のみの分析で使用した38個の説明変数に関して、二次の項(説明変数同士)を考慮する。ここでも二次の項同士の多重共線性を考慮した結果、説明変数は129個となった。単項の場合と同様に、3.の印象評価実験で使用した60個の質問回答文に対し、129個の特徴量を説明変数とし、因子得点を目的変数として、ステップワイズ選択法による重回帰分析を行った[22]。

この結果、重回帰式(2)が得られた。ここでも、第9因子に関しては、重回帰式が得られなかった。

また、この時の重相関係数を表10に示す。第1因子(的確性)は、0.9以上の値をとっているため、分析精度が非常に良好であるといえる。また、第2因子(不快性)、第4因子(容易性)、第5因子(執拗性)の3因子は、0.7以上の値であり、分析精度がやや良好であるといえる。一方、第3因子(独創性)、第6因子(曖昧性)、第7因子(感動性)、第8因子(努力性)の4因子は0.7未満で、第9因子(熱烈性)は重回帰式が得られなかったため、分析精度は不良であるといえる[7]。

5.2. 単項のみを考慮した重回帰分析

3.で使用した60個の質問回答文に対し、表12に示す64個の特徴量を説明変数とし、9因子の因子得点を目的変数として、ステップワイズ選択法による重回帰分析を行った[22]。この結果、重回帰式(3)が得られた。

また、重相関係数と、選ばれた説明変数を、それぞれ表13、表14に示す。重相関係数に関して、表8と表13とを比較すると、第4因子(容易性)と第5因子(執拗性)の分析精度がやや低下しているが、その他全ての因子は推定精度が向上している。

表10 重相関係数 (二次の項)

因子	重相関係数
第1因子(的確性)	0.972
第2因子(不快性)	0.703
第3因子(独創性)	0.697
第4因子(容易性)	0.862
第5因子(執拗性)	0.864
第6因子(曖昧性)	0.690
第7因子(感動性)	0.678
第8因子(努力性)	0.614

5. 文末表現を考慮した推定結果

5.1. 文末表現

4.までで因子得点の推定結果を示したが、第9因子(熱烈性)は重回帰式がここまで全く得られていない。そこで、第9因子の重回帰式の獲得と、因子得点の推定精度を全体的に向上させるために、文末表現を考慮して重回帰分析を行うことにした。

ここでは、西原らが使用している文末表現[9]を特徴量に採用した。西原らは、文末表現のパターンを「質問とベストアンサーにのみ含まれる文末表現」と定義している[9]。これに対し、本研究では「全ての質問文と回答文に含まれる文末表現」と定義する。文末表現を表11に示す。多重共線性を考慮して、実際に採用したものに網掛けを施している。表6、表7、表11で網掛けを施した特徴量64個をまとめて表12に示す。以降、これらをg1, g2, ...と表記する。

表11 文末表現

f	特徴量	f	特徴量
f67	ぞ(語数)	f91	よ(%)
f68	だ(語数)	f92	ね(%)
f69	よ(語数)	f93	あ(%)
f70	ね(語数)	f94	な(%)
f71	か(語数)	f95	し(%)
f72	な(語数)	f96	です(%)
f73	し(語数)	f97	ます(%)
f74	です(語数)	f98	たい(%)
f75	ます(語数)	f99	ない(%)
f76	たい(語数)	f100	ぞ(文末%)
f77	ない(語数)	f101	だ(文末%)
f78	ぞ(文末語数)	f102	よ(文末%)
f79	だ(文末語数)	f103	ね(文末%)
f80	よ(文末語数)	f104	か(文末%)
f81	ね(文末語数)	f105	な(文末%)
f82	か(文末語数)	f106	し(文末%)
f83	な(文末語数)	f107	です(文末%)
f84	し(文末語数)	f108	ます(文末%)
f85	です(文末語数)	f109	たい(文末%)
f86	ます(文末語数)	f110	ない(文末%)
f87	たい(文末語数)	f111	ですか(語数)
f88	ない(文末語数)	f112	ないです(語数)
f89	ぞ(%)	f113	ますか(語数)
f90	だ(%)	f114	ました(語数)

表12 多重共線性を考慮して採用した説明変数

g	特徴量
g1	助動詞(語彙数)
g2	接頭詞
g3	記号(語彙数)
g4	文数
g5	文の長さ平均(字数)
g6	カタカナ(語数)
g7	全角記号(語数)
g8	全角英数字(語数)
g9	形容詞(語数)
g10	副詞(語数)
g11	連体詞(語数)
g12	接続詞(語数)
g13	感動詞(語数)
g14	ひらがな(%)
g15	漢字(%)
g16	カタカナ(%)
g17	記号(%)
g18	TTR
g19	全角記号(%)
g20	英数字(%)
g21	全角英数字(%)
g22	名詞(%)
g23	形容詞(%)
g24	副詞(%)
g25	連体詞(%)
g26	接続詞(%)
g27	感動詞(%)
g28	「!」の数
g29	「?」の数
g30	句点の数
g31	読点の数
g32	中点の数
g33	3点リーダーの数
g34	鍵括弧の数
g35	括弧の数
g36	「/」の数
g37	単語心像性4点台(語数)
g38	単語心像性6.5以上7.0未満(語数)
g39	か(語数)
g40	な(語数)
g41	し(語数)
g42	たい(語数)
g43	ない(語数)
g44	だ(文末語数)
g45	か(文末語数)
g46	な(文末語数)
g47	し(文末語数)
g48	です(文末語数)
g49	ます(文末語数)
g50	たい(文末語数)
g51	ない(文末語数)
g52	ぞ(%)
g53	だ(%)
g54	よ(%)
g55	ね(%)
g56	か(%)
g57	です(%)
g58	ます(%)
g59	ない(%)
g60	か(文末%)
g61	ですか(語数)
g62	不是吗(語数)
g63	ますか(語数)
g64	ました(語数)

$$\left. \begin{aligned}
 y_1 &= 0.00295g_{37} + 0.150g_{12} + 0.0609g_{59} + 0.0533g_{21} \\
 &\quad + 0.0105g_7 - 0.113g_{29} + 0.0896g_9 + 0.0151g_{19} \\
 &\quad - 0.750 \\
 y_2 &= 1.73g_{44} + 1.79g_{50} - 0.115g_1 + 0.110g_{43} \\
 &\quad + 0.0341g_8 - 0.172g_{12} + 0.261 \\
 y_3 &= 0.183g_{52} + 0.0585g_8 - 0.104g_{43} + 0.776g_{50} \\
 &\quad + 0.0829g_{60} - 0.0808 \\
 y_4 &= -0.0468g_{37} - 0.0842g_{52} + 0.0915g_9 - 0.178g_{45} \\
 &\quad + 0.305 \\
 y_5 &= 0.0794g_1 + 0.0152g_6 + 0.107g_{29} + 0.0132g_{15} \\
 &\quad + 0.00310g_5 + 0.0302g_4 + 0.415g_{49} + 0.563g_{50} \\
 &\quad - 1.14 \\
 y_6 &= -0.0759g_{31} - 0.11163g_{43} - 0.0618g_{56} - 0.0342g_8 \\
 &\quad - 0.126g_{55} - 0.772g_{44} - 0.691g_{50} - 0.125g_{25} \\
 &\quad + 0.698 \\
 y_7 &= 0.391g_{13} + 0.0747g_{10} - 0.920g_{62} + 0.106g_{23} \\
 &\quad - 0.265 \\
 y_8 &= 0.0871g_1 + 0.0788g_{60} + 0.104g_{23} - 0.00560g_{20} \\
 &\quad - 0.00905g_8 - 0.388 \\
 y_9 &= 0.207g_{52} - 0.0169g_{16} + 0.864g_{50} + 0.373g_{51} \\
 &\quad + 0.802g_{44} + 0.0406
 \end{aligned} \right\} (3)$$

表13 文末表現を考慮した場合の重相関係数(単項)

因子	重相関係数
第1因子(的確性)	0.832
第2因子(不快性)	0.774
第3因子(独創性)	0.744
第4因子(容易性)	0.728
第5因子(執拗性)	0.893
第6因子(曖昧性)	0.872
第7因子(感動性)	0.581
第8因子(努力性)	0.650
第9因子(熱烈性)	0.683

5.3. 二次の項を考慮した重回帰分析

次に、単項のみの分析で使用した64個の説明変数に関して、二次の項を考慮する。二次の項同士の多重共線性を考慮した結果、説明変数は218個となった。単項の場合と同様に、3.の印象評価実験で使用した60個の質問回答文に対し、218個の特徴量を説明変数とし、因子得点を目的変数として、ステップワイズ選択法による重回帰分析を行った[22]。この結果、重回帰式(4)が得られた。

また、この時の重相関係数を表15に示す。表10と表15とを比較すると、全ての因子において重相関係数の値が向上している。第3因子(独創性)と第6因子(曖昧性)だけは値が0.9に少し及ばなかったが、それでも推定精度がやや良好であるといえる。

表 14 文末表現を考慮した重回帰分析で
選択された説明変数 (単項)

第1因子(的確性)	単語心像性4点台(語数) 接続詞(語数) ない(%) 全角英数字(%) 全角記号(語数) 「?」の数 形容詞(語数) 全角記号(%)
第2因子(不快性)	だ(文末語数) たい(文末語数) 助動詞(語彙数) ない(語数) 全角英数字(語数) 接続詞(語数)
第3因子(独創性)	ぞ(%) 全角英数字(語数) ない(語数) たい(文末語数) か(文末%)
第4因子(容易性)	単語心像性4点台(語数) ぞ(%) 形容詞(語数) か(文末語数)
第5因子(執拗性)	助動詞(語彙数) カタカナ(語数) 「?」の数 漢字(%) 文の長さ平均(字数) 文数 3点リーダの数 たい(文末語数)
第6因子(曖昧性)	読点の数 ない(語数) か(%) 全角英数字(語数) ね(%) だ(文末語数) たい(文末語数) 連体詞(%)
第7因子(感動性)	感動詞(語数) 副詞(語数) ないです(語数) 形容詞(%)
第8因子(努力性)	助動詞(語彙数) か(文末%) 形容詞(%) 英数字(%) 全角記号(語数)
第9因子(熱烈性)	ぞ(%) カタカナ(%) たい(文末語数) ない(文末語数) だ(文末語数)

その他の7因子に関しては、全ての値が0.9以上をとっているため、推定精度が良好であるといえる。

5.4. 考察

単項のみを考慮した場合、各因子の重相関係数に関して、第4因子(容易性)の分析精度がやや低くしているが、その他全ての因子は推定精度が向上した。特に、第2因子(不快性)、第3因子(独創性)の値はかなり向上している。

また、二次の項を考慮した場合、各因子の重相関係数に関して、第3因子(独創性)、第6因子(曖昧性)は値が0.9に少し及ばなかったが、推定精度はやや良好であるといえる。残り7因子に関しては、いずれも0.9以上の値をとっているため、推定精度が良好であるといえる。また、どの因子も値が向上している。

単項のみを考慮した場合も、二次の項を考慮した場合も、これまで得られなかった第9因子(熱烈性)の重回帰式を、文末表現を考慮したことにより、得ることに成功している。

6. 評価

6.1. 推定的良好性

推定的良好性を、推定誤差により評価する。実験の因子得点とその推定値の平均誤差の絶対値を求めたものを表16に示す。全体の平均誤差は非常に小さく、どの因子も因子得点に近い推定値が求まっていると考えられる。その中でも、第1因子(的確性)、第5因子(執拗性)、第7因子(感動性)、第9因子(熱烈性)は、残差平均の絶対値が特に小さくなっており、観測値に非常に近い予測結果が出ていると考えられる。

6.2. 文末表現の効果

推定における文末表現の効果を調べるために、文末表現の特徴量(g39~g64)のみを説明変数とし、9因子の因子得点を目的変数として、重回帰分析を行った。この時の重相関係数を、4.での分析結果(g1~g38)と、5.での分析結果(g1~g64)とあわせて表17に示す。

表 15 文末表現を考慮した時の重相関係数(二次項)

因子	重相関係数
第1因子(的確性)	1.000
第2因子(不快性)	0.947
第3因子(独創性)	0.877
第4因子(容易性)	0.908
第5因子(執拗性)	0.966
第6因子(曖昧性)	0.899
第7因子(感動性)	0.997
第8因子(努力性)	0.904
第9因子(熱烈性)	0.954

$$\begin{aligned}
y_1 &= 0.00180g_4g_5 + 0.0372g_{58}g_{59} - 0.00428g_4g_8 - 0.0968g_{16}g_{50} + 0.0245g_6g_{54} - 0.297g_{54}g_{64} - 0.00276g_3g_{29} \\
&\quad - 0.0065g_{16}g_{56} + 0.0264g_{59}g_{59} + 0.0149g_{19}g_{24} + 0.0232g_9g_{57} - 0.194g_{45}g_{54} - 0.282g_{12}g_{25} - 0.0102g_{52}g_{52} \\
&\quad - 0.192g_2 + 0.00456g_3g_7 + 0.00112g_{16}g_{20} - 0.0173g_9g_{16} - 0.0308g_{57}g_{59} - 0.0246g_{31}g_{54} - 0.449g_{11}g_{26} \\
&\quad - 0.00534g_5g_{63} + 0.0298g_1g_9 + 0.00553g_3g_{31} - 0.0131g_{23}g_{57} - 0.0106g_{49}g_{58} + 0.000135g_5g_{16} + 0.00220g_4g_{12} \\
&\quad - 0.0544g_{55}g_{58} + 0.0898g_{42}g_{55} + 0.393g_{26}g_{27} + 0.00683g_{16}g_{48} + 0.00828g_4g_{53} + 0.00209g_{17}g_{56} - 0.0118g_{32}g_{61} \\
&\quad + 0.000270g_5g_{57} + 0.0469g_{25}g_{41} - 0.0000448g_{15}g_{20} - 0.00372g_{12}g_{16} - 0.00120g_{56}g_{56} + 0.0145g_2g_{11} \\
&\quad - 0.00408g_{16}g_{29} + 0.00229g_4g_{11} + 0.0201g_{41}g_{60} - 0.0136g_{48}g_{54} + 0.00360g_{57}g_{60} + 0.00151g_{16}g_{41} + 0.00260g_{11}g_{59} \\
&\quad + 0.000791g_{57}g_{58} + 0.00288g_{21}g_{55} - 0.0000171g_4g_7 + 0.0000232g_{22}g_{23} - 0.000171g_{34}g_{56} - 0.489 \\
y_2 &= 1.58g_{33} + 0.187g_{16}g_{50} - 0.0338g_{16}g_{49} - 0.00856g_{37}g_{56} + 0.00397g_8g_{20} - 0.132g_{25}g_{60} + 0.00209g_4g_8 \\
&\quad + 0.184g_{21}g_{55} - 0.0406g_9g_{21} - 0.113g_{48}g_{54} + 0.00207g_{22}g_{32} + 0.209g_{27}g_{58} - 0.0350g_{40}g_{49} + 0.00694g_{22}g_{26} \\
&\quad + 0.0820g_{34}g_{55} - 0.640g_{27}g_{27} + 0.00436g_9g_{16} + 0.193 \\
y_3 &= 0.0197g_{52}g_{52} + 0.109g_4g_{13} + 0.0993g_{44} - 0.0288g_{58}g_{59} - 0.188g_{34}g_{48} + 0.0153g_{16}g_{55} + 0.0646g_{16}g_{50} \\
&\quad + 0.0265g_6g_{25} - 0.00406g_{16}g_{57} - 0.105g_{11} + 0.163g_{54}g_{64} - 0.0412 \\
y_4 &= -0.0698g_{10}g_{64} - 0.00742g_7g_{29} + 0.00663g_{16}g_{57} - 0.000414g_{14}g_{16} - 0.915g_{11}g_{53} + 1.40g_{33} - 0.0117g_{15} \\
&\quad - 0.0584g_{10}g_{42} + 0.125g_{20}g_{64} - 0.00478g_5g_{63} - 0.00530g_{57}g_{57} - 0.0122g_{12}g_{16} + 0.00796g_{16}g_{49} + 0.538 \\
y_5 &= 0.00540g_4g_5 + 0.0495g_{57}g_{60} - 0.0540g_4g_{13} + 0.0690g_{16}g_{50} + 0.0435g_{31}g_{46} + 0.397g_2 + 0.0142g_{15} \\
&\quad - 0.00364g_4g_{30} + 0.0237g_{57}g_{58} - 0.0400g_{48}g_{58} - 0.00410g_7g_{43} - 0.434g_{11}g_{35} - 0.449g_{33} + 0.00494g_{52}g_{52} \\
&\quad + 0.000159g_{15}g_{20} - 0.0249g_{34}g_{56} + 0.0497g_{44} - 1.12 \\
y_6 &= -0.00333g_4g_5 - 0.0305g_{31}g_{54} - 0.0133g_{16}g_{55} - 0.0337g_{58}g_{59} - 0.00811g_{56}g_{56} - 0.0677g_{16}g_{50} - 0.548g_{33} \\
&\quad - 0.0572g_{21}g_{60} + 0.000282g_3g_5 + 0.740 \\
y_7 &= 0.117g_4g_{13} + 0.190g_{23}g_{42} + 0.0584g_1g_0 - 0.00420g_4g_5 - 0.544g_{62} + 0.0223g_1g_{10} - 0.290g_{21}g_{55} - 0.000282g_{15}g_{20} \\
&\quad + 0.00533g_{52}g_{52} + 0.0803g_{21}g_{54} - 0.00301g_{21}g_{22} + 0.225g_{54}g_{64} - 0.00224g_3g_{15} - 0.0229g_{31}g_{57} + 0.0234g_8g_{23} \\
&\quad - 0.0187g_{21}g_{56} + 0.0100g_{12}g_{16} + 0.0508g_{16}g_{50} + 0.0892g_{43}g_{55} + 0.305g_{29}g_{64} + 0.0165g_{48}g_{58} + 0.00856g_{20}g_{36} \\
&\quad - 0.0748g_{32}g_{61} - 0.0796g_{12}g_{40} - 0.0111g_{58}g_{59} + 0.00213g_{16}g_{32} + 0.0719g_2 + 0.00884g_8g_{57} - 0.0323g_{10}g_{42} \\
&\quad - 0.0775g_{12}g_{25} - 0.169g_{11}g_{51} - 0.0000662g_{14}g_{16} - 0.101g_{60} + 0.117 \\
y_8 &= 0.0673g_{57}g_{60} + 0.201g_{23}g_{42} + 0.0312g_{31}g_{46} + 0.0392g_{10}g_{64} - 0.101g_{42}g_{48} - 0.00817g_{12}g_{16} + 0.0807g_{21}g_{60} \\
&\quad - 0.00202g_7g_8 + 0.0437g_{24}g_{40} - 0.210g_{60} - 0.000574g_{16}g_{20} + 0.00781g_{57}g_{58} - 0.186 \\
y_9 &= 0.232g_{42}g_{55} + 0.0163g_{52}g_{52} + 0.638g_{33} + 0.0683g_{16}g_{50} + 0.0195g_{43}g_{59} - 0.00413g_{16}g_{30} - 0.0434g_4g_{25} \\
&\quad + 0.135g_{44} - 0.0479g_{10}g_{60} + 0.0200g_{10}g_{24} + 0.0441g_{43}g_{54} - 0.0286g_{56}g_{59} + 0.00671g_{17}g_{56} + 0.328g_{26}g_{54} \\
&\quad - 0.00233g_{19}g_{57} + 0.0410g_{12}g_{40} - 0.00119g_3g_7 - 0.0782
\end{aligned}
\tag{4}$$

表 16 各因子の残差平均の絶対値

因子	残差の絶対値
第1因子(的確性)	0.0000420
第2因子(不快性)	0.114
第3因子(独創性)	0.164
第4因子(容易性)	0.124
第5因子(執拗性)	0.0874
第6因子(曖昧性)	0.147
第7因子(感動性)	0.0202
第8因子(努力性)	0.112
第9因子(熱烈性)	0.086
残差平均	0.0949

ここで、g1~g38 とg39~g64 とを、重相関係数の点から比較する。g1~g38の結果に関して、第1因子(的確性)、第4因子(容易性)、第5因子(執拗性)の因子得点の推定精度は、g39~g64を上回っている。一方、g39~g64の結果に関して、第2因子(不快性)、第6因子(曖昧性)、第7因子(感動性)、第9因子(熱烈性)の因子得点の推定精度は、g1~g38を上回っている。従って、文末表現に特化した特徴量(g39~g64)と、それ以外の要素に着目した特徴量(g1~g38)とは、それぞれ長けている点と不足している点とがあることが分かる。

表 17 文末表現だけを説明変数として重回帰分析を行った場合の重相関係数

因子	g1～g38		g39～g64		g1～g64	
	単項のみ	二次項も考慮	単項のみ	二次項も考慮	単項のみ	二次項も考慮
第1因子(的確性)	0.809	0.972	0.681	0.795	0.832	1.000
第2因子(不快性)	0.350	0.703	0.695	0.834	0.774	0.947
第3因子(独創性)	0.475	0.697	0.590	0.637	0.744	0.877
第4因子(容易性)	0.737	0.862	0.697	0.797	0.728	0.908
第5因子(執拗性)	0.881	0.864	0.787	0.705	0.893	0.966
第6因子(曖昧性)	0.734	0.690	0.813	0.780	0.872	0.899
第7因子(感動性)	0.562	0.678	-----	0.895	0.581	0.997
第8因子(努力性)	0.590	0.614	0.548	0.695	0.650	0.904
第9因子(熱烈性)	-----	-----	0.593	0.859	0.683	0.954

一方、これら両方を特徴量として考慮すると、推定精度が非常に良くなる。これは、互いの不足している点を補い合うことにより、推定精度が非常に良くなるものではないかと考えられる。従って、西原らの使用した文末表現[9]にも、文章の因子得点の推定精度に一定の効果があるが、それだけでなく、文章の文の数や品詞の数など、文章全体の文体や内容にも着目することにより、文章の因子得点の推定精度が向上するものと考えられる。

7. まとめ

本論文では、質問者と回答者の相性を判定することを目的として、文章の因子得点の推定精度の向上をめざして検討を行った。ここでは、文末表現[9]も文章の特徴量に追加した上で、質問回答文の因子得点の推定精度の向上を試みた。その結果、全ての因子において推定精度が向上し、文末表現も考慮することで分析精度を大きく向上させることができた。また、文末表現だけでなく、文章全体の文体や内容にも着目することにより、文章の因子得点の推定精度が向上するのではないかと考えられる。

今後の課題としては、今回得られた重回帰式を用いて、まだ求まっていない質問回答文の因子得点を求め、「ベストアンサー」の推定を行う予定である。また、すでに特徴量に追加している単語心像性以外にも、単語親密度、表記妥当性、単語複雑度など[5]、文章の印象に影響すると考えられる特徴量が知られていることから、今後、これらの特徴量に加えた上で、因子得点の推定精度の向上を図る予定である。

謝辞

本研究は一部、科研費(21500091)の助成を受けて行われたものである。また、実装・評価に際し、大学共同利用機関法人国立情報学研究所から提供を受けた、

Yahoo!知恵袋のデータを利用している。ここに記して謝意を示す。

参考文献

- [1] Yahoo!知恵袋
<http://chiebukuro.yahoo.co.jp/>
- [2] 横山友也, 宝珍輝尚, 野宮浩揮, 佐藤哲司: 質問回答サイトの質問文と回答文の印象評価, 第2回データ工学と情報マネジメントに関するフォーラム(DEIM2010), C4-2, 2010.
- [3] 横山友也, 宝珍輝尚, 野宮浩揮, 佐藤哲司: 質問回答サイトの質問文と回答文の印象評価とベストアンサーの推定, 日本感性工学会論文誌, Vol.10, No.2, pp.221-230, 2011.
- [4] 横山友也, 宝珍輝尚, 野宮浩揮, 佐藤哲司: 文章の特徴量を用いた質問回答文の因子得点の推定, 第6回日本感性工学春季大会, 22D-2, 2011.
- [5] 佐久間尚子, 伊集院睦雄, 伏見貴夫, 辰巳格, 田中止之, 天野成昭, 近藤公久: 単語心像性①, NTTデータベースシリーズ日本語の語彙特性 第3期(第8巻), (社)三省堂, 2005.
- [6] NTTデータベースシリーズ,
<http://www.kecl.ntt.co.jp/mtg/goitokusei/>
- [7] 横山友也, 宝珍輝尚, 野宮浩揮, 佐藤哲司: 単語心像性を用いた質問回答文の因子得点の推定精度の向上, 平成23年度情報処理学会関西支部支部大会, C-13, 2011.
- [8] 横山友也, 宝珍輝尚, 野宮浩揮, 佐藤哲司: 文章の特徴量を用いた質問回答文の因子得点の推定精度の向上, 日本感性工学会関西支部大会 2011.
- [9] 西原陽子, 松村真宏, 谷内田正彦: Q&A コミュニティでの質疑応答パターン理解, 第22回人工知能学会全国大会, 1H2-7, 2008.
- [10] Blooma, M.J. and Chua, A.Y.K. and Goh, D.H.L.: A Predictive Framework for Retrieving the Best

- Answer, Proc. of 2008 ACM Symposium on Applied Computing (SAC08), pp.1107-1111, 2008.
- [11] Agichtein, E., Castillo, C., Donato, D., Gionis, A. and Mishne, G.: Finding High-Quality Content in Social Media, Proc. of the Int'l Conf. on Web Search and Web Data Mining (WSDM08), pp.183-194, 2008.
- [12] Wang, X. J., Tu, X., Feng, D. and Zhang, L.: Ranking Community Answers by Modeling Question-Answer Relationships via Analogical Reasoning, Proc. of 32nd Int'l ACM SIGIR Conf., pp.179-186, 2009.
- [13] Kim, S., Oh, J. S. and Oh, S.: Best-Answer Selection Criteria in a Social Q&A site from the User-Oriented Relevance Perspective, Proc. of American Society for Information Science and Technology (ASIS&T) 2007 Annual Meeting, 2007.
- [14] Adamic, L. A., Zhang, J., Bakshy, E. and Ackerman, M. S.: Knowledge Sharing and Yahoo Answers: Everyone Knows Something, Proc. of 17th Int'l Conf. on World Wide Web (WWW2008), 2008.
- [15] Jurczyk, P. and Agichtein, E.: Discovering Authorities in Question Answer Communities by Using Link Analysis, Proc. of 16th ACM Conf. on Inf. and Know. Management (CIKM2007), pp.919-922, 2007.
- [16] Hovy, E., Gerber, L., Hermjakob, U., Junk, M. and Lin, C.-Y.: Question Answering in Webclopedia, Proc. of 9th Text Retrieval Conf., pp. 655-664, 2000.
- [17] 熊本忠彦: 新聞記事を対象とする印象空間の構築, 第12回Webインテリジェンスとインタラクション研究, WI2-20075, 2008.
- [18] 佐久間尚子, 伊集院睦雄, 伏見貴夫, 辰巳格, 田中正之, 天野成昭, 近藤公久: 単語心像性①, NTTデータベースシリーズ日本語の語彙特性 第3期(第8巻), (社)三省堂, 2005.
- [19] 朝野熙彦: 入門多変量解析の実際 第2版, p.64, (社)講談社, 2000.
- [20] Text Seer マニュアル
http://www.valdes.titech.ac.jp/~t_kawa/ts/manual.htm
- [21] 菅民郎, 初心者がらくらく読める多変量解析の実践上, pp.37-41, (社)現代数学社, 1993.
- [22] 菅民郎, 初心者がらくらく読める多変量解析の実践上, pp.42-45, (社)現代数学社, 1993.
- [23] 重回帰分析(ステップワイズ変数選択),

<http://aoki2.si.gunma-u.ac.jp/R/sreg.html>

付録

印象評価実験で使用した質問文と回答文の例として、パソコン・周辺機器に関する質問とその回答文を、評価者に提示した形式で以下に示す。

•Q4(質問文)

パソコン初心者です。

デジカメで撮った画像をプリントアウトしたところ画像が暗いのですが、明るくする方法をご存知の方回答をお願いします。

•A4-1(回答文:「ベストアンサー」)

Microsoft Officeをお持ちなら、Microsoft Photo Editorというソフトが付いています。

スタート→すべてのプログラム→Microsoft Office ツールというフォルダにたいていあります。インストールされていない場合もありますから、その場合は Office CD より追加でインストールしてください。

このソフトで修正したい画像を読み込み、イメージ→自動調整を実行すると、かなりきれいになります。

また、手動でも各種の調整が可能です。

•A4-2(回答文)

デジカメに付いていた編集ソフトを、インストールして明るさを変更する。

•A4-3(回答文)

フォトタッチソフトで、明るさを調整できます。

デジカメに付属されていませんか？

なければ、フリーソフトで。

例えば、PictBear

<http://www.vector.co.jp/magazine/softnews/001004/n0010041.html>

•A4-4(回答文)

おすすめは「デジカメの達人」で補正する
 しかし、あれは有料ソフトなので、強要はしない

デジカメで撮った写真は全部補正掛けてそれを印刷する。

ページに載せるときも補正した後カットしてリサイズしたものを載せる

感性のあいまい性を考慮した印象評価に向けて

Towards an impression evaluation considering the vagueness of *Kansei*

赤井 俊介、宝珍 輝尚、野宮 浩揮

Shunsuke Akai, Teruhisa Hochin, Hiroki Nomiya

京都工芸繊維大学 情報工課程, 京都市左京区松ヶ崎御所海道町

Kyoto Institute of Technology, Goshokaido-cho, Matsugasaki, Sakyo-ku, Kyoto

あらまし: 現在、感性の評価手法として SD 法がよく使われている。しかしながら、SD 法は感性というあいまいな事柄を評価するにもかかわらず、決められた範囲での評価を強制している。その結果、感性のあいまいな部分の評価が困難である可能性がある。そのため、評価者にとってあいまいな評価ができるような、柔軟な評価手法が必要であると考えられる。そこで本論文では、対象の印象を表す言葉である感性ワードが散りばめられた空間を用意し、そこから対象の印象に当てはまる部分を領域として囲んでもらうことで評価するという手法を提案する。また、その手法の実現に向けて、印象の似ている感性ワードほど近くに配置されている平面の導出を試みる。ここでは、アンケート調査によって得られた感性ワード同士の非類似度から、多次元尺度構成法を用いて導出を行った結果を報告する。

Summary: The Semantic Differential method is often used as an evaluation method of *Kansei*. Although the Semantic Differential method evaluates a vague thing called *Kansei*, it forces the evaluation in a predefined range for statistical processing. As the result, it is hard to evaluate a vague part of *Kansei*. Therefore, it is thought that the flexible evaluation method that enables vague evaluation is required. This paper proposes such a method. This method uses the space containing impression words. The impression of an object is specified by surrounding the impression words fitting the impression. This paper tries to construct the space. In the space, the more similar the impressions of impression words are, the nearer the impression words are. The space is constructed by applying the Multi Dimensional Scaling to the scores of dissimilarity between impression words obtained through a subjective experiment.

キーワード: 感性, あいまい性, 印象評価

Keywords: *Kansei*, vagueness, impression evaluation

1. はじめに

近年、消費者の要求は機能や利便性といった物理的なものだけでなく、見た目の美しさなど消費者がどのように感じるかという感性による部分が重要になってきている。しかし、感性というものは個人によって異なり、あいまいで、定量化が困難であるという特徴があり、そのような感性を表現することは難しい。

そのような感性の評価手法として、現在 SD 法[1]がよく使われている。SD 法は印象を数値化することで統計処理を可能にし、様々な分析を行うことができる。しか

し、統計処理を行うために決められた範囲での評価が強制されており、実験者にとって都合の良い項目しかないといった、恣意的要素が介在してしまう恐れがある。そのため、感性というあいまいな事柄を評価する際に、評価しにくい、あるいは評価できないような部分が生じてしまう可能性がある。そこで、評価者にとって、あいまいな評価ができるような、柔軟な評価手法が必要であると考えられる。

そこで本論文では、感性のあいまい性を考慮した評価手法を提案する。提案する手法は、対象の印象を

表す言葉(以下、感性ワードという)が散りばめられた感性空間から、対象の印象として当てはまる部分を領域として囲み、その領域内でも当てはまる割合を評価してもらう方法である。また、その前段階として、感性空間の作成に関する実験を行った。その結果について報告する。

以降、2. では従来手法について概説する。次に、3. で感性のあいまい性を考慮した印象評価手法を提案する。そして、4. で感性空間の作成実験とその結果について述べ、5. で考察を行う。最後に6. でまとめる。

2. 従来手法

2.1 SD 法

SD 法(Semantic Differential method)とは、社会・政治心理学者である Osgood により開発された、対象の持つ含意的な意味、すなわちその対象が人に感じさせる(あいまいな)意味を測定する方法である[1]。つまり、対象に対して人が抱く印象やイメージを明らかにするために用いられる手法である。

SD 法では、対象物を多数の評価語に基づき評価してもらう。具体的には「明るいー暗い」のように、対象のイメージを表す言葉(感性ワード)を、それと反対の意味の語とペアにして感性ワード対を作り、それに5段階や7段階といった尺度を持たせ評価してもらうことになる。そこから得られたデータを分析することで、類似している、すなわち相関が強い感性ワードを、一般的な因子としてまとめることができる。そして、その相関構造から意味空間を定義することができる。

しかしながら、感性というあいまいな事柄を評価することを考えると、評価しにくい、あるいは評価できない部分が生じてしまうと考えられる。そのような SD 法の問題点を以下に示す。

① 点数が離散値

感性というあいまいな事柄を評価するのに明確な数値で答えなければならない。何となくこのくらいという、あいまいな評価ができない。

② 評価する部分

対象によっては、部分的な印象と全体の印象が異なっていたり、観点を変えると違う印象になるものもある。そのような場合、表現することができない可能性がある。例えば、星空の画像を対象とすると、全体的には「暗い」と感じるが、その中の星を見ると「明るい」と感じる人もいるであろう。つまり、感性ワード対で評価するため、「明るい」と「暗い」を同時に感じることを、表現することができない。

③ 感性ワードの解釈

感性ワードの解釈が人によって、あるいは場面によって異なる可能性がある。そのため、複数の意味が考えられる感性ワードの場合、どの意味で評価されているのかが分からない。例えば、「重いー軽い」という項目があるとすると、力士、葬式、こってりしたラーメンという画像に対して、それぞれ「重い」と感じる人が多いのではないのだろうか。しかし、これらの「重い」の意味は微妙に違うはずである。

④ 感性ワードの選び方

SD 法では、評価に用いる感性ワードを実験者が自由に決めることができる。そのため、実験者にとって都合の良い感性ワードしか選ばれていないなど、恣意的に選ばれている可能性がある。そのため、対象の印象に適した感性ワードで評価できない場合がある。反対に、対象の印象に適していない感性ワードについても評価が強制されている場合もある。

⑤ 感性ワードが離散的

SD 法の項目となっている感性ワード対に対して独立に評価しなくてはならない。感性というあいまいな事柄を評価することを考えると、明確な言葉では表せない部分、例えば「かわいい」と「きれい」の間といった感じ方もあるのではないかと考えられる。

2.2 ファジィ

あいまいな概念を表す方法にファジィ[2]があり、その基本的な考え方としてファジィ集合がある。通常の集合は、ある範囲に対して属するか属さないかのみを示すもので、クリスプ集合と呼ぶ。それに対して、ファジィ集合はある範囲にどの程度属するか属さないかを示す集合のことである。

例えば、気温について考えると、「暑い」と「寒い」を明確に決めることはできない。これを無理やりクリスプ集合で表そうとすると図1のようになる。つまり、0か1かで表さなければならない。この場合、29度は「普通」で、1度だけ高い30度は「暑い」になってしまう。このような表現は、人の感覚からすると違和感を感じるものになっている。しかしながら、これをファジィ集合で表すと図2のように表すことができる。つまり、0か1だけでなく、0.5といった中間値であることを許す。この場合、30度は「普通」と「暑い」のどちらにも当てはまる可能性があることを表現することができる。このように、ファジィ集合はクリスプ集合に比べて、人間の感性に近い表現ができる。

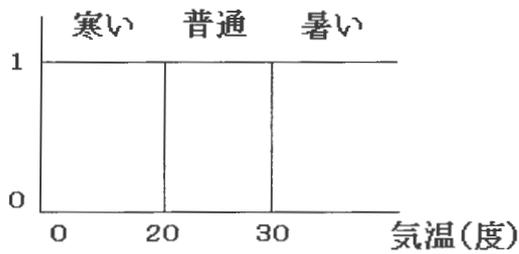


図1 クリस्प集合

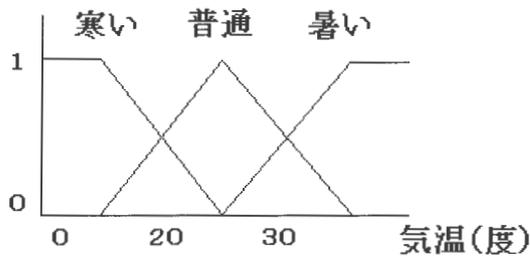


図2 ファジィ集合

そのファジィを使って感性的あいまい性を表現しようとしている研究もある。中森ら[3]は、SD法で得られたデータに対して、ファジィを用いてばらつきを表現することであいまいな部分を表現しようとしている。そして、そのばらつきのあるデータを用いて因子分析などの多変量解析を行う方法を提案している。

このように、ファジィを用いることであいまいでないデータに対して、少し幅を持たせ、あいまいなデータとして表現することができる。

しかし、使うデータは結局あいまいでないデータで、それに対して人為的に幅を持たせているだけに過ぎず、評価者にとってあいまいな評価ができるわけではない。

2.3 ニューラルネットワーク

ニューラルネットワーク[4]とは、脳の機能の一部をモデル化したシステムである。そして、それを用いて人間の感性を表そうとする研究もある。つまり、人間のあいまいな感性を、人間の脳を模倣したニューラルネットワークを用いることで表現しようということである。

人間の脳には多数のニューロンと呼ばれる脳細胞が存在し、互いにつながっており、それらが信号のやり取りを行っている。そのニューロンをコンピュータ上で扱うために単純化したモデルが図3である。つまり、それぞれの入力信号に対して重みをつけて、それを足し合わせた値が、閾値を超えるかどうかで出力が決まることになる。このニューロンをつなぎ合わせることでできるものが図4のようなニューラルネットワークである。

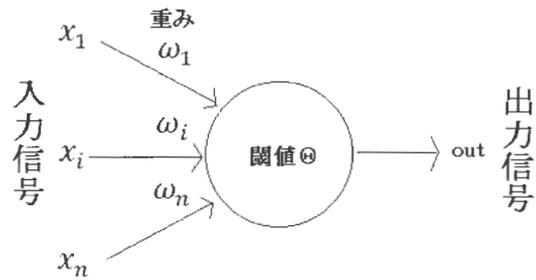


図3 ニューロンのモデル図

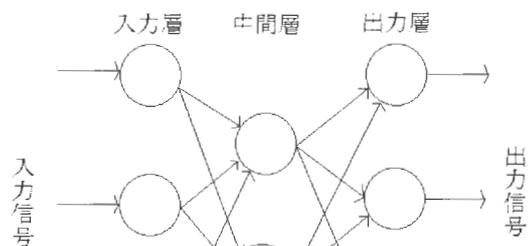


図4 ニューラルネットワーク

関谷ら[5]は、このニューラルネットワークを用いて、ユーザの感性を模倣する評価モデルを提案している。そこでは、入力部は人が受けた刺激であり、出力部は人が受けた感情に対応すると仮定している。そして、ニューラルネットワークの重みを調節して、入出力関係を、ユーザが受けた刺激とその刺激に対する感情との関係に最適化する。これにより、人に何らかの印象を与える刺激情報を入力することで、ユーザの感性を表現できるとしている。

しかし、これは複雑であいまいな感性プロセスをニューラルネットワークで表すものであり、本論文の目標である、評価者のあいまいな評価には適用できないと考えられる。

3. 感性空間と感性評価法

3.1 感性空間について

ここでは、我々が対象の印象を評価する際に思い浮かべる空間(以下、感性空間という)について考察する。

例えば、風景写真をみせてどのような印象を受けるかを問うと、「爽やか」や「美しい」といった言葉で印象を表現できる。また、SD法でのように、提示された印象語に合致している程度を示すこともできる。従って、心の中(感性空間)では、印象を表す項目(以下、印象表現項目という)が存在し、対象物がどの印象表現項目にどの程度合致するかを判断していると考えられる。しかし、どのような印象表現項目がどのように配置されているかは明らかではない。

印象表現項目として最もよく使用され、また、最も重要であるのは語であると考えられる。そこで、今回は印象を表現する語(感性ワード)について考えていくことにする。

次に感性空間の特性について考察する。

① 感性ワード間の非類似度

2つの感性ワードを提示すると、それらの感性ワードがどの程度似ているか・似ていないかを答えることができる。従って、感性ワードの類似度・非類似度があると考えられる。

② 成長性

感性ワードは体験や学習により成長し続けると考えられる。

③ ダイナミック性

対象の印象は、その時の気分によって変化する。また、それまでの経緯によって印象評価は変化するが、日々の定常的な生活を進めていくうちに、元の定常状態に戻る。このように、印象空間は静的なものではなく、常に変化している。

④ パーソナル性

②や③から明らかであるが、感性空間は個人個人で異なる。例えば、20歳前後の工科系大学の日本人男子学生というように共通した属性を持つ者は、類似の感性空間を持つ傾向があるが、完全に同一ということはない。

また、感性空間中の感性ワードも個人によって異なる。例えば、ある者は「エレガントな」という語を用いて印象を表現するが、ある者はそうではないといった具合である。

3.2 感性評価法

ここでは、SD法での問題点を解決できるような、感性のあいまい性を考慮した評価手法を提案する。これは、対象の印象を表す感性ワードが散りばめられた平面

(図5)を感性空間としてあらかじめ用意し、その感性空間において対象の印象として当てはまる部分を領域として表し、その領域内でも当てはまる度合を表現してもらう方法である。例えば、「静かな」と「暗い」という印象を感じ、逆に「地味な」という印象は感じない場合を表現したものが図6である。

しかしながら、感性ワード間の関係は2次元で表せるものではないと考えられ、3.1で述べたように個人によっても差が生じるものである。そこで、対象の印象としてよく当てはまる感性ワードが感性空間に存在しなければ、評価者が自由に追加することができ、また、感性ワードの配置に違和感を感じる場合は、感性ワードの位置も移動できるものとする。例えば、「静かな」、「ゆったり」、「のんびり」の間に「のどかな」が配置されていると感じる場合、図7のように表現できるものとする。

このように自由度の高い感性空間であるが、何もないう状態から利用者に感性空間を構築してもらうのは非常に大変である。そこで、通常の利用者に共通の感性空間をベースライン感性空間として提供し、これをもとに必要に応じて修正を加えていくこととする。

なお、領域内での当てはまる度合いの表現方法については、現在検討中である。

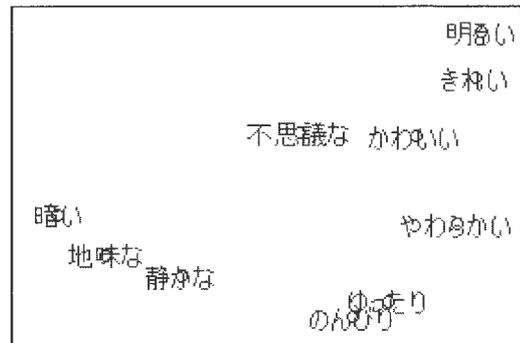


図5 感性空間の例

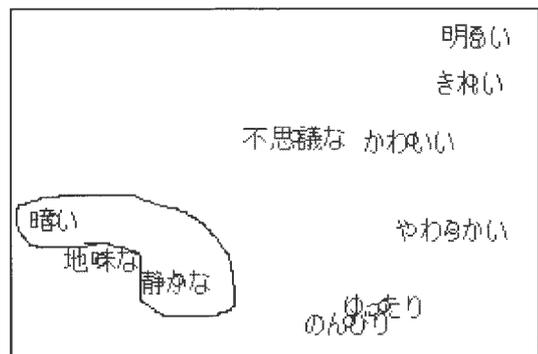


図6 評価方法1

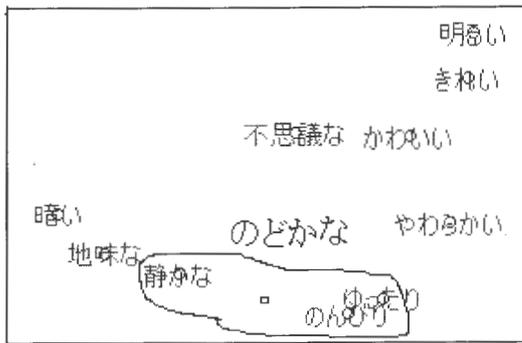


図7 評価方法2

この方法では、対象に対するイメージを領域で表すことで、SD法に比べて自由で柔軟な評価が可能になる。また、SD法では評価できなかった、反対の意味の感性ワードを同時に評価することもできる。さらに、感性ワード間も評価することができ、評価者の思う感性ワードの関係や、言葉として思いつかない場合でも、似ている感性ワードとの関係から評価することができる。

4. ベースライン感性空間の導出実験

4.1 実験方法

ベースライン感性空間の導出について述べる。ベースライン感性空間を求めるには、実際に被験者に自由に感性ワードを配置してもらう方法も考えられるが、矛盾が無いように配置するのは、感性ワードが多くなるに

つれて難しくなる。そこで、感性ワード間の非類似度のアンケート調査をし、その結果を多次元尺度構成法[6]を用いて2次元平面にプロットする方法を採用する。

4.1.1 感性ワードの選定

ベースライン感性空間には、評価に必要な感性ワード全てが含まれていることが望ましい。しかしながら、感性ワードは無数に存在する上、図に多数の感性ワードが含まれている場合、評価が難しく時間がかかる可能性も考えられる。また、感性ワード間の非類似度についてはアンケート調査を行うため、用いる感性ワードが多すぎるのは望ましくないと考えられる。つまり、感性の評価によく使用される感性ワードを選び出す必要がある。

小澤ら[7, 8]は、表1に示す階層構造からなる感性プロセスを提案しており、そこで感性評価によく使用されている100語(一部は言語対)を選定し、それを用いて階層構造の妥当性を検証している。その結果、それらの言葉が表1の6つの階層に分けられていた。今回はその結果を基に、各階層からイメージしやすいと考えられる言葉を選び、合計37個の言葉をベースラインに用いる感性ワードとして選定した。その37個の感性ワードを表2に示す。なお、上位の階層(数字が大きい)ほど対応する言葉が少なく、また抽象的でイメージしにくいと考えられる言葉が多かったため、選んだ感性ワードも少なくなっている。

表1 提案された感性プロセスの階層構造

(6)絶対感性	「家族愛など、人種・宗教を超えた絶対的価値観に関連して良い」と感ずるレベル。
(5)芸術感性	「それに有用性・快適性・嗜好との合致好感情のいずれも認められないが、それでも良い」と感ずるレベル。
(4)情緒感性	「それは自分の嗜好に合致していないが、実際に使ったら〇〇であった」と感ずるレベル。
(3)嗜好感性	「一般的にはそれがベターかもしれないが、私にとってはこれがベスト」と感ずるレベル。
(2)快適感性	「どちらも使えるならば、こちらの方がベター」と感ずるレベル。
(1)基礎感性	「これは使える」と感ずるレベル。あるいは、与えられた課題に対して生理学的反応として回答するレベル。

表2 選出した感性ワード(37語)

(6)絶対感性	不思議な	自然な	純粹な		
(5)芸術感性	うっとりしい	さわやかな	かつこいい	おしゃれな	素晴らしい
(4)情緒感性	かわいい	あっさり	のんびり	悲しい	ゆったり
(3)嗜好感性	派手な	地味な	さっぱり	こってり	鋭い
	なめらかな	強い	弱い		
(2)快適感性	うるさい	静かな	きれい	汚い	やわらかい
	硬い	広い	狭い		
(1)基礎感性	明るい	暗い	暑い	寒い	重い
	軽い	大きい	小さい		

Form5

2つの言葉のイメージがどのくらい近いと思うかを直観的に答えてください

似ているものは小さく、関係ないあるいは反対の意味の場合は大きく選んでください 例:うるさい - 騒がしい 1 大きい - 小さい 10

地味な	groupBox1	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 10	弱い
のんびり	groupBox2	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 10	強い
重い	groupBox3	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 10	やわらかい
静かな	groupBox4	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 10	うるさい
かわいい	groupBox5	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 10	純粹な
広い	groupBox6	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 10	きれいな
うっとりしい	groupBox7	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 10	素晴らしい
強い	groupBox8	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 10	なめらかな
のんびり	groupBox9	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 10	カッコいい
不思議な	groupBox10	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	<input type="radio"/> 8	<input type="radio"/> 9	<input type="radio"/> 10	きれいな

戻る 1/67 中断 次へ

図8 実験に用いたアプリケーション

4.1.2 アンケート調査

選んだ感性ワード同士のイメージについて、どのくらい似ているかを、図8のようなアプリケーションを用いて1～10の10段階で評価してもらった。なお、似ているほど値を小さくするようにしており、感性ワードのすべての組み合わせ666通りをランダムな順番で回答してもらっている。

被験者は21～25歳の大学生、大学院生11名(男性10名、女性1名)である。

4.1.3 多次元尺度構成法によるプロット

アンケートによって得られたデータから、感性ワード同士の非類似度を表す行列を作成し、それを使用して多次元尺度構成法により感性ワードを2次元空間にプロットした。

なお、今回は感性ワード間の非類似度をデータとして用いているので、非計量多次元尺度構成法を適用した。

4.2 実験結果

被験者個人のデータ、および被験者全員のデータを平均したものの計12個の感性空間を導出した。

個人のデータから導出した感性空間の中から特徴的なものを図9～図11に示す。また、平均から導出したものを図12に示す。

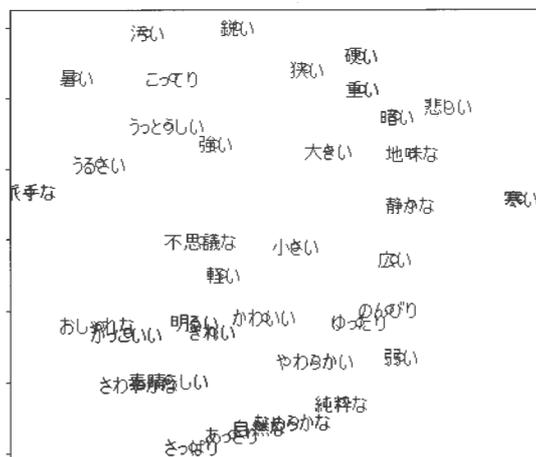


図9 被験者 A の感性空間

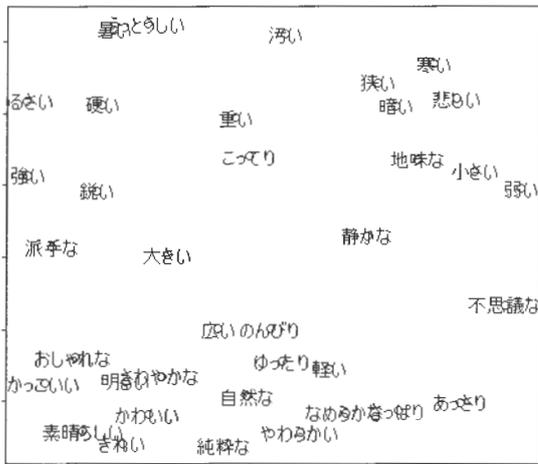


図10 被験者 B の感性空間

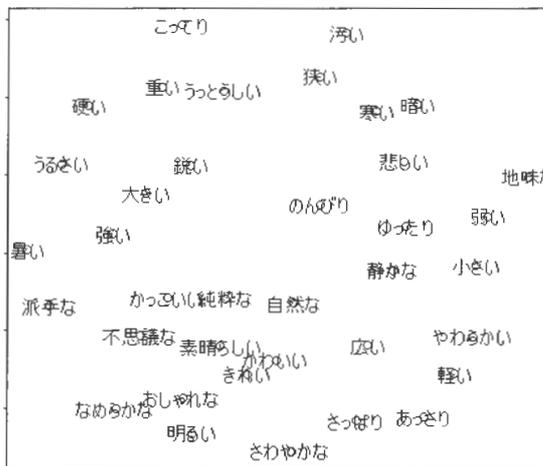


図11 被験者 C の感性空間

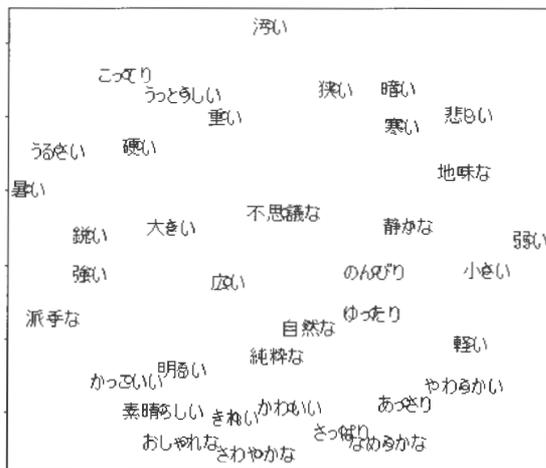


図12 平均から導出した感性空間

なお、比較しやすいように、回転と反転の補正を施している。具体的には、「悲しい」が右上になるように回転させ、その後「うっとうしい」や「うるさい」が左上になるように、反転が必要なものには反転を行った。

イメージが近い感性ワードは近くに固まっており、うまく配置されていると考えられる。

5. 考察

個人の結果を比べてみると、細かな配置には違いがあるものの、ある程度似ている部分もある。例えば、ほとんどの人の結果で、「素晴らしい」、「きれい」、「かっこいい」、「明るい」といった感性ワードが近くに集まるように配置されていることが分かった。つまり、これらの周りの感性ワードは、その人がポジティブに感じている感性ワードであると考えられる。

逆に、「暗い」、「地味な」、「汚い」、「悲しい」といった感性ワードも近くに集まるように配置されている。これらの周りの感性ワードは、その人がネガティブに感じている感性ワードであると考えられる。この結果から、似ているイメージの感性ワードが近くに配置されていることが分かる。

また、平均の結果である図12を見てみると、「不思議な」などの抽象度が高い感性ワードや、「大きい」などの観点によって、複数の意味が考えられる感性ワードは、個人差が大きいと考えられ、その結果真ん中付近に配置されているのではないかと考えられる。確かに図9～図11を見比べてみても、「不思議な」の配置はバラバラである。また、実際に得られたデータを見ても、「不思議な」に近いイメージの感性ワードは人によって違い、平均すると全ての感性ワードとの値がほぼ同じになるという結果になり、個人差が大きいことが分かる。

また、「大きい」に関しては、「重い」、「広い」、「強い」、「鋭い」といった感性ワードの間に配置されている。このようにイメージに近い感性ワードがばらけて配置されている場合は、それらの間に配置される。

ただ、個人差が大きい感性ワードや意味が複数考えられる感性ワードは、人によって同じ領域に囲む別の感性ワードに違いが出ると考えられるため、真ん中付近に配置されているという結果は、ベースライン感性空間として用いることを考えると妥当ではないかと考えている。

しかしながら、多次元尺度構成法を用いて、多次元のデータを2次元にしているため、少なからず誤差は生じている。例えば、ある感性ワード間の実際のデータが10(まったく似ていない)であっても、実際の配置ではかなり近い位置にあるという場合も見られた。多次

元尺度構成法は、全体を通しての誤差が最小になるような配置になるために起こる現象であると考えられる。この場合、イメージの似ていない感性ワードが近くに配置されることがある。しかし、ベースライン感性空間はあくまでも一つの指標、目安になるものであり、評価の際に違和感のある配置の場合は、感性ワードの配置を自由に変更できるため問題ないと考えている。ただし、ベースライン感性空間として用いようと考えている平均の結果を見る限り、あまりにも不自然な配置になっている感性ワードは無いのではないかと考えている。

以上の結果から、個人の結果を反映しながらも、似ている感性ワードは近くに、似ていない感性ワードは遠くに配置されており、妥当な結果になったのではないかと考えられる。よって、図12に示す感性空間をベースライン感性空間とする予定である。

6. まとめ

本論文では、感性のあいまい性を考慮した評価手法を提案し、その評価に向けて、感性空間の導出を試みた。多次元尺度構成法を使用することにより、感性ワードの非類似度のみで、比較的妥当な2次元の配置にすることができた。印象の類似した感性ワードが近くに配置されていることで、領域が囲みやすくなっていると考えられる。

今後は、提案した手法について細部を検討し、感性のあいまい性を考慮した評価手法を実現させ、その手法を用いた評価実験を行い、SD法との比較を行いたいと考えている。

参考文献

- [1] 長町 三生 編:商品開発と感性、海文堂出版 (2000)
- [2] 田中 一男:応用をめざす人のためのファジ理論入門ーファジ集合からファジ制御までー、ラッセルブックス (1991)
- [3] 中森 義輝:感性データ解析ー感性情報処理のためのファジ数量分析手法ー、森北出版 (2000)
- [4] 村上・泉田研究室 ニューラルネットワーク <http://ipr20.cs.ehime-u.ac.jp/column/neural/index.html> (2011)
- [5] 関谷 拓郎、徳丸 正孝、村中 徳明:ニューラルネットワークを用いた感性検索エージェントの最適化に関する研究、第26回ファジシステムシンポジウム、pp.51-56 (2010)
- [6] 金 明哲:R と多次元尺度法、ESTRELA10月号、pp.56-61 (2010)

[7] 小澤 賢司、松本 恵:感性プロセスにおける階層構造の検討ー(1) モデルの提案と語彙調査、第5回日本感性工学会予稿集、pg.44 (2003)

[8] 松本 恵、小澤 賢司:感性プロセスにおける階層構造の検討ー(2) 階層構造の検証実験、第5回日本感性工学会予稿集、pg.45 (2003)

個人の好みに関する感性のモデル化に関する研究

A Study on *KANSEI* Modeling of Preference about Items

荻野 晃大

Akihiro Ogino

京都産業大学 コンピュータ理工学部, 京都市 北区 上賀茂本山

Kyoto Sangyo University, Motoyama Kamigamo, Kita-Ku, Kyoto

あらまし: 本論文では、個人適応型のショッピングサイトを実現するために、商品に関するユーザの好みをモデル化する方法を提案する。本論文で提案するモデルは、ユーザが商品に対する好みを決定する過程を、(1) 商品の各属性を認知する、(2) 商品の各属性について好みの態度を決める、(3) 各属性に関する好みの態度を連合して、最終的な好みの態度を決定すると捉え、それに基づいてユーザの注目している商品の属性・属性値の抽出、抽出した属性・属性値に対する好みの態度の数量化を行い、ユーザが商品の好みを判断する基準となるモデルを作成した。

Summary: This paper proposes a model of user profile for a personalized e-commerce site. The model of user profile estimates attributes and their values of products that are related to user preference. The model of user profile also calculates the degree of relation with the attribute values of product and the user preference. The result shows that we can detect the attribute value related to user preference.

キーワード: 感性モデル, 好み, 個人化, Eコマース

Keywords: *KANSEI* Modeling, Preference, Personalization, E-Commerce

1. はじめに

現在のショッピングサイトには、ユーザが一見ただけでは全ての商品を見られない程の商品が登録されている。そのため、商品名などの商品に関する明確な情報を持ち得ないユーザは、自分の好みに適する商品をショッピングサイトから手早く見つられなくなっている。そのような状況の中で、ユーザの好みに基づいて商品を選別したり、推薦したりすることで、ユーザの商品検索を支援する個人適応型の情報サービス(以後、個人化情報サービス)が注目を浴びている。個人化情報サービスは、ショッピングサイトにおけるユーザの検索・購買履歴やユーザへのアンケートなどの情報から、商品に関するユーザの好みのモデル化を行う。そして、各ユーザのモデルを用いてユーザの好みに適する商品の推定を行う。

好みのモデル化手法には、大きく分けてベクトルを用いる方法、ネットワークを用いる方法、ルールを用いる方法がある。ベクトルを用いる方法は、商品の特徴をベクトルで表現し、ユーザの好きな商品と嫌いな商品の特徴を表すベクトルの違いを SVM(Support Vector

Machine)等により機械学習することで、ユーザの好みに適する商品を推定するモデルを生成する。ネットワークの方法は、ユーザの好き・嫌いと商品の特徴との間の関係をページアンネットワークなどによりネットワークとして形成し、確率的にユーザの好みに適する商品を推定するモデルを生成する。ルールを用いる方法は、アソシエーションルールやラフ集合などを用いてユーザの購買履歴等をデータマイニングし、そこから抽出したルールを用いて推薦する方法である。

ベクトルとネットワークを用いる方法は、ユーザの好みを推定する方法として有効であるが、モデルの構造が複雑になる。そのためユーザは、モデルによって推定された結果に対して善し悪しを判断する事はできるが、「モデル自体がユーザの好みをどのように解釈しているのか?」ということを直感的に理解することは難しい。一方、ルールを用いる方法は、IF-THEN形式としてモデルを表現する。そのためユーザは、モデルの内容を直感的に理解できる。しかし、複数のルールが生成された場合には、ユーザが全てのルールの内

容を網羅して、モデルがどのように自分を理解しているのかを理解することは難しい。

本論文では、ショッピングサイトを利用する各ユーザが、個人化情報サービスにより推定された商品の理由と、その元となっている好みのモデルの内容を直感的に理解できる好みのモデル化手法を提案する。

2. 好みのモデル化

本研究では、ユーザがある商品に関する好みの態度を決定する過程を、図1に示すような過程として捉えている。

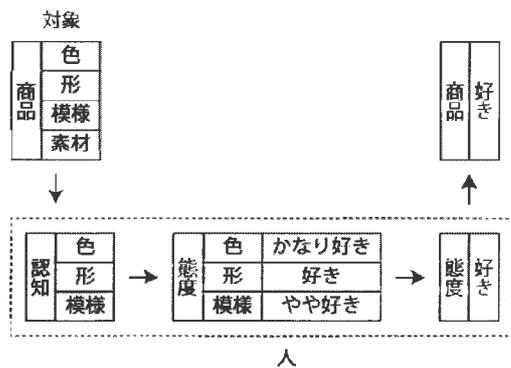


図1 商品に対する好みの態度を決定する過程

1. 商品の各属性を認知する。

ユーザは、各自のもつ動機(例:青いTシャツがほしい)にマッチする商品(例:Tシャツ)を見て、物の中から認知する。また、認知した商品においても、各自の動機に関連のある商品の属性(例:青)を注目して認知する。

2. 商品の各属性について好みの態度を決める。

認知した商品の各属性(例:Tシャツの色や首のデザインなど)について、各自の好みの基準により好みの態度をそれぞれ決定する。例えば、「Tシャツの色は好きだけど、首のデザインは、好きではない。」のように態度を決定する。

3. 各属性に関する好みの態度を連合して、その商品に関する最終的な態度を決定する。

認識した商品の各属性に関する好みの態度を連合させ、総合的に判断して商品に対する態度を決定する。例えば、あるTシャツについて、「色がとても気に入っている」と「胸に施されているロゴのデザインも悪くはない」という2つ属性に関する好みの態度から、この商品は好みであるという態度を決定する。

また本研究では、各ユーザの好みの基準は、そのユーザの重要視する属性に関連していると考えている。

例えば、商品A(色が青で、丸首のTシャツ)と商品B(色が青で、V首のTシャツ)に関する好みの態度の決定について考える。色に関心の高いユーザは、両商品に対して好意的な態度を示す可能性が高いと考えられる。一方、首のデザインに関心の高いユーザは、好みに適する商品にしか好意的な態度を示さないと考えられる。このようにユーザは、好みの基準として商品の属性の重要度を保持しており、その重要度がユーザの最終的な好みの態度に影響を与えていると考えられる。

本研究では、上記に示した、「ある商品に関する人の好みの態度の決め方」と「好みの基準」をベースとして、以下の3つのステップにより好みのモデル化を行う。

1. 商品属性の認知処理

ユーザが商品に対する好みの態度を決定するために不可欠である商品の属性・属性値をラフ集合[1,2]の縮約を用いて抽出する。ラフ集合の縮約とは、ある分類(例:好き)に属する対象を確実に他の分類(例:嫌い)の対象から識別するのに十分な極小の属性の集合のことである。

2. 属性(属性値)に関する好みの決定処理

ラフ集合の縮約により抽出した好みに関連する属性とその属性値に関するユーザの好み(好き・嫌い)の態度を度合いとして算出する。本研究では、この度合いを重要度と呼ぶ。属性に関する重要度は、商品の属性の内、どの属性に注目しているのかを指し示す。また属性値に関する重要度は、特定の属性において重要視している属性値を指し示す。

3. 商品に関する好みの決定処理

属性に重要度をユーザの好みの基準と捉え、商品の特徴を示す属性値に重みとして掛け合わせた度合いを嗜好度として算出する。本研究では、この嗜好度の集合を嗜好モデルと呼ぶ。この嗜好モデルは、ユーザが商品の特徴の内、どの属性値をどの程度好んでいるのかを表す。

以下の節において、嗜好モデルを算出する各ステップにおける情報処理の仕方について述べる。

2. 1 商品属性の認知処理

本研究では、ユーザが商品に関する好みの態度を決定するときに注視する属性・属性値を注視する過程をラフ集合によりモデル化する。

ラフ集合では、対象を識別する条件となる対象の属性(条件属性)の集合と、識別の目的となる属性(決定属性)の集合により表現した表を決定表と呼ぶ。また、決定属性の属性値に基づいて分類される対象の集合を決定クラスと呼ぶ。

表 1 商品の属性・属性値と商品に関する個人 X の好みの関係を示した決定表

対象	条件属性						決定属性
	Color (Co)	Texture (T)	Category (Ca)	Neck (N)	Sleeve (S)	Chest (Ch)	Preference
ID_1	Dark-Brown (DBr)	Point (Po)	Hood (Ho)	Crew (Cr)	Long (L)	Zip-up (Z)	Positive (P)
ID_2	Blue (Blu)	Point (Po)	Hood (Ho)	Crew (Cr)	Long (L)	Line (Li)	Positive (P)
ID_3	Black (Bla)	Uniform (U)	Cut Sewn (CS)	Henley (He)	Long (L)	Non (No)	Positive (P)
ID_4	Deep-Green (DGr)	Uniform (U)	Cut Sewn (CS)	Crew (Cr)	Long (L)	Non (No)	Negative (N)
ID_5	Pink (Pi)	Point (Po)	Hood (Ho)	Crew (Cr)	Long (L)	Mark (M)	Negative (N)
ID_6	Deep-Green (DGr)	Uniform (U)	Shirts (S)	Collar (Col)	Long (L)	Button (Bu)	Negative (N)

表 2 決定クラス D_p (好き) に関する決定行列

	ID_1	ID_5	ID_6
ID_1	{(Co, DBr), (T, Po), (Ca, Ho), (Ch, Z)}	{(Co, DBr), (Ch, Z)}	{(Co, DBr), (T, Po), (Ca, Ho), (Ne, Cr), (Ch, Z)}
ID_2	{(Co, Blu), (T, Po), (Ca, Ho), (Ch, Li)}	{(Co, Blu), (Ch, Li)}	{(Co, Blu), (T, Po), (Ca, Ho), (Ne, Cr), (Ch, Li)}
ID_3	{(Co, Bla), (Ne, He)}	{(Co, Bla), (T, U), (Ca, CS), (Ne, He), (Ch, No)}	{(Co, Bla), (Ca, CS), (Ne, He), (Ch, No)}

表 1 は、洋服を商品とし、洋服の色(Color)などの 6 つの条件属性とその属性値、決定属性の個人 P の好み(Preference)に関する評価とその値を示した決定表の例である。

表 1 の商品 $ID_1 \sim ID_6$ は、決定属性の属性値 (好き (Positive) と嫌い (Negative)) により 2 つの決定クラス D_p (好き) と D_n (嫌い) に分類される。

本論文では、ある分類に帰属する対象を確実に含めた確実性集合を表現する上で最小限必要な属性の集合を縮約と呼ぶ。本研究では、決定表から決定クラス D_p と D_n に属する商品を確実に類別する縮約を抽出する。例えば、 $ID_1 \in D_p$ と $ID_5 \in D_n$ は、属性 Color, または属性 Chest により確実に区別できる。このとき、属性 Color, または属性 Chest を、決定クラス D_p を識別するための縮約と呼ぶ。

この縮約を用いると、決定クラス D_k を識別する最小限の属性とその属性値の組み合わせを抽出できる。その組み合わせは、決定クラス D_k を導き出す IF-THEN ルールとして捉えることができる。

上記の例では、決定クラス D_p を導き出す属性とその属性値は、「IF Color=Dark Brown THEN D_p 」という IF-THEN ルールで記述できる。ラフ集合では、決定クラス D_k を識別する最小限の属性と属性値の組み合わせを極小決定ルールと呼ぶ。

縮約に基づく決定表内の全ての対象を確実に好きと嫌いに識別する極小決定ルールは複数存在する。そのため本研究では、決定行列を用いる方法 [3] を用いて、全ての極小決定ルールを抽出する。ここで、決定クラス D_p の決定行列を $M(D_p)$ とし、その成分 (i, j) を $M_{ij}(D_p)$ とする。決定行列 $M(D_k)$ から抽出する決定クラス D_k を識別する極小決定ルールの集合を $F(D_k)$ とする。表 2 は表 1 から作成した決定クラス D_p (好き) に関する決定行列 $M(D_p)$ である。

$M(D_p)$ の各成分は、 D_p (好き) を D_n (嫌い) と識別するために最小限必要な属性・属性値の組み合わせを示している。

表 3 好き (Positive) という態度に関する属性・属性値の決定ルール： $F(D_p)$

$F_i(D_p)$	属性=属性値	$F_i(D_p)$	属性=属性値
$F_1(D_p)$	Co=DBr	$F_4(D_p)$	Ch=Z
$F_2(D_p)$	Co=Blu	$F_5(D_p)$	Ch=Li
$F_3(D_p)$	Co=Blu	$F_6(D_p)$	Ne=He

表 4 好き (Positive) に関連する属性 (a)・属性値 (v_a) の重要度

a	重要度： $ID_{D_p}(a)$	重要度の比率
Ne	0.33	0.16
Co	1.00	0.50
Ch	0.66	0.33

v_a	重要度： $ID_{D_p}(v_a)$	重要度の比率
Bl	0.33	0.16
He	0.33	0.16
Blu	0.33	0.16
DBr	0.33	0.16
Z	0.33	0.16
L	0.33	0.16

表 5 洋服に関する個人 X の嗜好度

属性値	嗜好度	属性値	嗜好度
Blu	0.27	No	-0.01
DBr	0.27	U	-0.02
Z	0.23	Cr	-0.04
Li	0.22	Col	-0.13
He	0.13	M	-0.23
CS	0.00	Bu	-0.23
Sh	0.00	Pi	-0.27
		DGr	-0.27

*注：値は小数点 3 位で四捨五入

例えば、表 2 に示した決定行列 $M(D_p)$ の成分 $M_{12}(D_p) : \{(Co, DBr), (Ch, Z)\}$ は、 $ID1 \in D_p$ と $ID5 \in D_n$ とを属性:Color (Co) の属性値:Dark-Brown (DBr), または属性:Chest (Ch) の属性値:Zip-up (Z) により識別できることを示している。

決定行列 $M(D_p)$ に内在する極小決定ルール $F(D_p)$ は、以下の通りとなる。

$$F(D_p) = ((Co, DBr) \vee (Ch, Z)) \vee ((Co, Blu) \vee (Ch, Li)) \vee ((Co, Bla) \vee (Ne, He))$$

上記の論理式より、好き (Positive) という態度を決定するために不可欠な属性・属性値の組み合わせは、表 3 まとめられる。同様に嫌い (Negative) という態度を決定するために不可欠な属性・属性値について求めることができる。

2.2. 属性 (属性値) に関する好みの決定処理

2.1 節で極小決定ルールとして抽出した属性とその属性値について、重要度を算出する。

ここで、極小決定ルールに含まれる属性を a としたとき、その取りうる属性値 v_a とする。そのとき、決定クラス (D_k に属する全ての商品 $O(D_k)$ の内において、決定ルール $F_i(D_k)$ を含む商品の割合 (被覆度) : $CI(F_i(D_k))$ を式 (1) とする。

式 (1) :

$$CI(F_i(D_k)) = F_i(D_k) \cap O(D_k) / O(D_k)$$

例えば、表 3 に示した決定クラス D_p (好き) の極小決定ルール $F_1(D_p) = [Co = DBr]$ の被覆度は、 $CI(F_1(D_p)) = 1/3$ となる。

また、極小決定ルールを構成する属性・属性値の個数 (ルールの長さ) を $Len(F_i(D_k))$ とする。ここで、決定ルールの集合 $F(D_k)$ の内で、属性 a を含む極小決定ルールの数を n とすると、決定クラス D_k の属性 a の重要度 : $ID_{D_k}(a)$ を式 (2) と定義する。

式 (2) :

$$ID_{D_k}(a) = \sum_{i=1}^n \left(\frac{CI(F_i(D_k))}{Len(F_i(D_k))} \times \frac{1}{(Len(F_i(D_k)))^e} \right)$$

ただし、ここで用いる $CI(F_i(D_k))$ は、属性 a を含んだ $F_i(D_k)$ に関する $CI(F_i(D_k))$ の値とする。

同様に、 $F(D_k)$ の内で属性値 v_a を含む極小決定ルールの数を m とすると、決定クラス D_k の属性 v_a の重要度 : $ID_{D_k}(v_a)$ を式 (3) と定義する。

式 (3) :

$$ID_{D_k}(v_a) = \sum_{j=1}^m \left(\frac{CI(F_j(D_k))}{Len(F_j(D_k))} \times \frac{1}{(Len(F_j(D_k)))^e} \right)$$

ただし、ここで用いる $CI(F_i(D_k))$ は、属性値を含んだ $F_i(D_k)$ に関する $CI(F_i(D_k))$ の値とする。

例えば、表 3 より決定クラス D_p (好き) の属性 Color の重要度は、 $n = 3$, $CI(F_1(D_p)) = CI(F_2(D_p)) = CI(F_3(D_p)) = 1/3$, $Len(F_1(D_p)) = Len(F_2(D_p)) =$

$Len(F_3(D_p)) = 1$ より、以下の通りとなる。

$$ID_{D_p}(color) = \left(\frac{1}{3} \times \frac{1}{1^e}\right) + \left(\frac{1}{3} \times \frac{1}{1^e}\right) + \left(\frac{1}{3} \times \frac{1}{1^e}\right) = 1.0$$

決定クラス D_p (好き)の属性Chestと属性Neckの重要度を計算すると、それぞれ $ID_{D_p}(chest) = 0.66$ 、 $ID_{D_p}(neck) = 0.33$ となる。3つの属性の重要度を比較することで、個人 X においては属性Colorが最も好きに関連している属性であることがわかる。表4は、好きに関する極小決定ルールに出現する属性と属性値の重要度とその比率を示した表である。決定クラス D_n (嫌い)についても、 D_p (好き)と同様に各属性・属性値に関する重要度を算出する。

各属性・属性値に関して好きと嫌いの重要度を統合し、総合的に属性・属性値の好みへの重要度として、統合重要度を算出する。式(4)と(5)に、属性の統合重要度： $UD(a_i)$ と属性値の統合重要度： $UD(v_{a_i})$ の算出式を示す。

式(4)：

$UD(a_i)$

$$= \begin{cases} ID_P(a_i) \times ID_N(a_i), & ID_P(a_i), ID_N(a_i) \neq 0 \\ 0 \times ID_N(a_i), & ID_P(a_i) = 0 \\ ID_P(a_i) \times 0, & ID_N(a_i) = 0 \end{cases}$$

式(5)：

$UD(v_{a_i})$

$$= \begin{cases} ID_P(v_{a_i}) - ID_N(v_{a_i}), & ID_P(v_{a_i}), ID_N(v_{a_i}) \neq 0 \\ 0 - ID_N(v_{a_i}), & ID_P(v_{a_i}) = 0 \\ ID_P(v_{a_i}) - 0, & ID_N(v_{a_i}) = 0 \end{cases}$$

好きと嫌いの両方に関して重要度の高い属性は、個人 X により注視されおり、商品の好きと嫌いを評価に寄与していると捉えることができる。そのため本研究では、属性に関する統合重要度を好き・嫌いの属性の重要度を乗算することで算出する。これにより、個人の注目している属性の重要度はより強調し、注目されていない属性の重要度を低下させる。

また、好きと嫌いの両者に属性値が現れるということは、その個人にとってその属性値に対する評価は定まっていないことを表している。したがって本研究では、属性値に関する統合重要度を好きから嫌いの属性値の重要度を減算することで算出する嫌いにのみ出現する属性値の重要度や両者に出現する属性の重要

度も低下させる。

2.3. 興味度の算出

好きと嫌いに関する属性・属性値の統合重要度から、個人 X の嗜好モデル $PM(X)$ を作成する。嗜好モデル $PM(X)$ の成分を $PM_i(X)$ とすると、 $PM_i(X)$ は以下のように定義できる。

表6 洋服を表現する属性とその属性値

属性	属性値
Color (Co)	Black, Blue, Brown, Dark-Brown, Deep-Blue, Deep-Green, Gray, Green, Pink, Purple, Red, Red-Purple, Yellow
Texture (T)	Check, Horizontal-Stripes, Point, Uniform
Category (Ca)	Cut-Sewn, Hood Polo-Shirts, Shirts, Knit
Neck (N)	Collar, Crew, Henley, High, Hood Turtle, V
Sleeve (S)	Half, Long
Chest (Ch)	Button, Mark, Zip-Up, None

表7 各セットに対する解答属性値

セット	好きと判断する基準
1	属性値=色：青 (Blue, Deep-Blue)
2	属性値=首：ヘンリーネック (Henley)

表8 評価実験データより算出した嗜好度

セット番号 (解答属性値)	各属性値の嗜好度 (上位5位)
データセット：1 (Blue と Deep-Blue)	Deep-Blue +0.14
	Blue +0.11
	Gray +0.04
	Uniform +0.01
	Half +0.01
データセット：2 (Henley)	Henley +0.26
	Gray +0.10
	Black ± 0.00
	Uniform ± 0.00
	None ± 0.00

$PM_i(X)$ を個人 X の属性値 i に関する嗜好度と呼ぶ。式(6)：

$$PM_i(X) = \{UD(v_{a_i}) \times UD(a_i)\}$$

本研究では、式(6)に基づいて算出される嗜好度

$PM_i(X)$ の値が正 (+) 値の場合、その個人の好きに起因している属性値と考え、負 (-) 値は嫌いに起因している属性値と捉える。表 5 は、個人 X の嗜好モデルの各成分である嗜好度 $PM_i(X)$ を降順にまとめた例である。嗜好度 $PM_i(X)$ の値より対象の色が青や黒の場合、個人 X は好むと推定できる。

3. 嗜好モデルの評価

洋服 Z に関する個人 X の好みの態度が「青色」に係している場合、青色の嗜好度の値が正 (+) 値で、かつ、他の属性値の嗜好度より相対的に高い値を示したならば、嗜好モデルは個人 X の嗜好を推定できていると見なせる。

上記の考えに基づいて、被験者に対象を好きと判断する基準とする属性値を与えた上で、洋服の好みを判断してもらい、嗜好モデルにより好みの判断基準とした属性値の推定する評価実験を行った。本評価実験では、嗜好モデルの評価を以下の手順で行った。

1. 男性用上着 520 着からランダムに選択した 50 着を 1 セットとし、2 セットの決定表データを用意した。表 6 は、本実験において対象を表現するのに用いた属性と属性値である。
2. 対象を好きと判断する基準とする属性値（以後、解答属性値）を被験者に与える。一人の被験者に対して、2 セットの実験を行った。表 7 は、各セットに対して設定した解答属性値を示している。
3. 被験者は、解答属性値を含む洋服を全て好きと判断する。それ以外の洋服は、被験者の判断に任せる。例えばデータセット 1 の場合、解答属性値である青 (Blue と Deep-Blue) を含む対象は、無条件で好きと答え、それ以外は被験者の判断に任せる。
4. 評価データから嗜好度を算出する。解答属性値とした属性値の嗜好度が正 (+) 値であり、かつ、全属性値における相対的な割合が他の属性値と比べて高ければ、嗜好モデルは好みを推定できたとする。

表 8 は、被験者 X がデータセット 1 と 2 に対して行った評価実験データから嗜好度を算出した結果を示している。

データセット 1 に関しては、解答属性値に設定した Blue と Deep-Blue の嗜好度の値が 1, 2 位となっている。データセット 2 も同様に、解答属性値に設定した Henley の嗜好度の値が 1 位となっている。この結果から、本論文で提案した嗜好度により人の好みを推定できていることが示された。

4. まとめ

本論文では、個人適応型の情報検索システムを実現するために、商品に関して人が好みを決定する過程を以下の 3 つの過程と捉え、その過程をシステム上で模倣するための嗜好モデルを提案した。

1. 商品の各属性を認知する。
2. 商品の各属性について好みの態度を決める。
3. 各属性に関する好みの態度を連合して、その商品に関する最終的な態度を決定する。

嗜好モデルを作成する方法として、ラフ集合の縮約を用いて対象に関する個人の好みに関連する対象の属性・属性値を抽出し、かつ、対象の属性・属性値と対象の関する好みの関連の度合い（嗜好度）として算出した。人が対象の好みを判断するときに重要視している属性値を嗜好モデルにより推定できるかを主観評価実験のデータから嗜好モデルを生成して評価を行った。その結果、主観評価実験において被験者が好みの判断を行う上で重要視していた対象の属性値を、本論文で提案した嗜好モデルにより推定できた。

今後の課題は、この嗜好モデルを情報検索システムに適用し、情報検索システムが利用者に提示した結果の理由が利用者に直感的にわかる情報検索システムを開発する。

参考文献

- 1) 森典彦, 田中英夫, 井上勝雄: “ラフ集合と感性—データからの知識獲得と推論”, 海文堂出版, 2004
- 2) Pawlak, Z.: “Rough Sets”, *International Journal of Computer & Information Sciences*, vol.11, issue 5, pp.341-356, 1982
- 3) Shan, N., Ziarko, W.: “Data-based acquisition and incremental modification of classification rules”, *Computational Intelligence*, vol.11, pp.357-370, 1995.

介護場面の日本語学習に向けた例文データベースの構築支援技術 Support techniques for developing an example sentence database in the nursing-care situations

竹内 和広

Kazuhiro Takeuchi

大阪電気通信大学 情報通信工学部, 寝屋川市初町 18-8

Osaka Electro-Communication University, 18-8 Hatsucho, Neyagawa, Osaka

あらまし:近年, 医療福祉の分野に就業する外国人従事者が増えている. そのような従事者がサービスの受け手と会話をする場面では, 既存の教科書では経験のしたことのない言語表現に遭遇するため, 場面に対応した例文集の作成が有益である. 本稿では, そのような例文を収集し, データベース化する上での支援技術を紹介する. 具体的には, 登録する例文が教科書的な例文と相違する点を多面的に分析する技術と, その前提となる発話者の心理的・身体的な状態や状況の中間表現に関する研究について報告する.

Summary: Recently, it is said in Japan that more and more foreign workers are needed in the field of health and welfare. These workers on the job face various Japanese expressions that they didn't learn in the ordinary text books for Japanese language learners. In this paper, we discuss an efficient way to construct sentence database for the learners in specific purpose. In particular, we introduce two support techniques for constructing an intermediate semantic representation used in various methods to evaluate sentences from several points of view.

キーワード: 日本語学習, 学習用例文, 意味表現

Keywords: Japanese Language Learning, Example Sentences of Language Learning, Intermediate expression of meaning

1. はじめに

本研究では, コンピュータ支援により医療福祉分野の日本語を学習・運用する際に役立つ対話例集の作成・収集について検討する. 具体的には, 登録しようとする対話例文の適切性, 非文法性を計算文法, および, その他の意味処理機構に基づいて評価し, 外国人医療福祉従事者が日本語を学習する上で役立つ, かつ実際の医療福祉コミュニケーションを支援するシステム構築にも役立つような言語資源の構築を支援する技術を考える.

医療福祉分野に従事する外国人労働者は, 既に看護師や介護士の資格を取得し, 一般的な日本語能力や専門能力については一定の水準に達してから現場に入るとされている. しかし, このような施設や拠点で定常的に正規の日本語教員を十分な人数採用するこ

とは難しく, 日本語教育のボランティアや現場の関係者が, 上記のような外国人労働者(以下, 単に学習者と呼ぶ)への日常的な日本語教育の任を担うことが多い. そのため, 現在, 様々な機関や施設が, 学習者向け医療福祉分野日本語例文集を提供するようになって来たが, そのような例文に関して, 適切な日本語教師のアドバイスが得がたいことが問題となる.

そのような背景から, 介護場面において想定される, 未知の例文を日本語学習者に提示する上で必要となる日本語学習・運用向けの補足情報を提示する図1のような概要のシステムを構築している. このようなシステムは専門日本語教育の例文データベースを構築する上で有益であることが期待されるが, 既存の言語処理技術を単純に応用するだけでなく, 当該の専門分野のドメイン知識や辞書を整備することが必須である.

また、言語教育の専門知識も取り入れる必要がある。本稿では、そのようなシステムを構築する上で、言語教育支援システムの一般性を視座に、我々がシステム構築に際して使用、あるいは新規に研究開発した計算機技術について報告する。

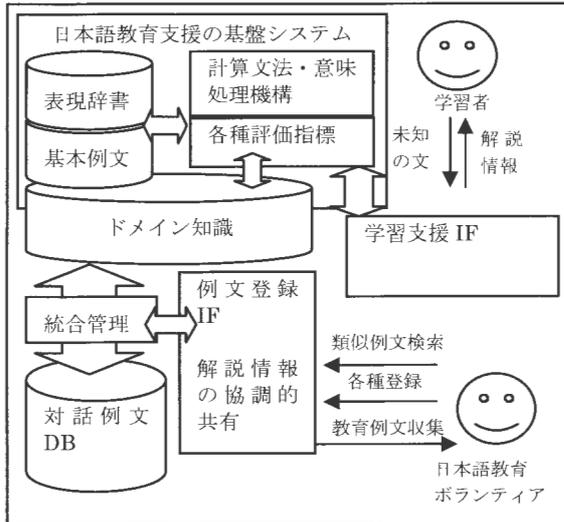


図1. 専門日本語学習支援システムの概要図

2. 計算機による言語学習支援

2.1 専門日本語教育支援の方向性

本研究では専門日本語教育をとりまく状況に関して、主に図1中にあるような2つの方向性の支援を考えている。

1つは現場で学習者の日本語教育に携わる関係者をサポートする方向性である。そのような関係者は必ずしも日本語教育を生業としているわけではないため、教材作成や日本語教育法の熟達に十分な時間をかけることはできない。この環境に対して、教材作成や日本語の教育上のポイントを低コストで実現でき、他の日本語教育ボランティアと情報を協調的に蓄積し共有する環境を提供する。

他方は、学習者の自習支援である。具体的には、未知の日本語をどのような位置づけで学べばよいかを、日本語教師のアドバイスを理想とした、近似的なアドバイスを、例文の解説情報として計算機により提供する方向性である。

さらに、コミュニケーションや協調の拠点として、計算機を使う方向性もある。そのような方向性においても、図1の基盤システムにおける処理機構は応用技術の開発において重要な役割を果たしうると考えているが、本稿の範囲を超える。

2.2 日本語文解説知識の構造化

日本語学習支援に対して有益なのは、現場に即した文に対しその解説情報を提供することである。しかし、言語の性質の本質として、日夜に生成される発話や文が多様であり、そのバリエーションは無限に存在する。従って、言語学習の際の無限に多様な文に対して、それぞれに個別な解説情報を詳細に提供することは不可能である。つまり、学習に即した解説情報として、当該の専門分野における日本語学習に適した限られた例文を用意し、その例文を用いて学習者に日本語の重要な性質、特徴、発話機能等を解説することが現実的である。

実際、言語教育において主流となる考え方の一つは、一般的な言語の性質をまず例文を用いて学習し、その上で専門的な語彙を積み上げていく方法である。具体的には、学習例文に基づく日本語学習においては、次のような知識階層から伝統的な教材が構成されているように見受けられる。

基礎知識: 初級の教科書に出現する例文、語彙(旧日本語能力試験で3級、4級レベルの出題範囲とされる文型、語彙)の列挙・解説。主に動詞の活用を基準として文型を規定する。表現の形式と目的が一致することを前提としている。文体は丁寧体「です・ます」を基本とする。

専門知識: 専門分野の特徴的な場面や行動、知識に関連した例文が示され、場面や専門性に応じた個別の語彙の列挙・解説。話者の態度や、聞き手への配慮を示す表現等も発展として解説が与えられる。

ここで、専門分野での言語運用では、語彙の列挙

中心の知識記述だけで、現場で接する日本語に対応することは容易ではなく、現場の日本語運用が、既に学習した基礎知識とどのような相関にあるかを解説情報として与えることが求められる。すなわち、協調的な例文データベースの構築・蓄積においては、現場で出現しそうな文を、解説情報を考慮せずに「むやみやたら」に登録しても教材として価値は低いと言える。

他方、計算機によって言語学習支援を行うためには、言語処理技術が必須となるが、言語処理技術においても言語の多様性に対応することは重要課題である。そのため、言語表現の多様性を評価する解析モジュールが提案されてきた。例えば、実際の言語運用で使われる日本語は、母語話者であっても誤用が存在し、そのような日本語もある程度は計算機によって解析が可能である必要がある。我々の研究グループも、既にそのような誤用例のデータベースを開発し、誤用や文体に即した不適切性を認識するための評価指標の比較[1]や学習時に気づきを促す上での効果的な情報提示方法、編集方法[2][3]を研究してきた。

また、日本語における誤用だけではなく、文体が多様性を生じさせる要因になる。例えば、丁寧語・敬語分析については、岩下ら[4]、白土ら[5]をはじめとする研究や学習支援システム構築の蓄積がある。それらの研究では、例えば、「食事」が「お食事」、「言う」が「おっしゃる」というように語彙レベルの言い替えが基本データとなる。

以上のような文の多様性を扱うためには、それを扱うための体系が必要なことを、誤用と敬語・丁寧表現の語彙知識を例に示した。このように、学習教材となる文は多面的な角度から言い換えを分析することが必要となる。

そのような背景から、我々は、計算機による学習者支援において、現場の学習者が知っている基礎知識を基本例文とその解説の形でデータベース化しておき、現場に即した文がどのような点で基本例文と類似し、また、どのような特徴を持つのかを多面的に評価し、

解説情報として学習者に提供する技術が基盤になると考えている。

2.3 解説情報基盤としての基本例文集

本研究では、日本語学習用に作成された一般的な基本例文集を用いて、学習者向けの日本語文の解説情報を提供することを考えている。この例文集は、学習者の基礎知識を考慮し、従来の文型積み上げ型の初級日本語教材と整合性を保ちつつ、実際の言語運用を重視して、次のような特徴を持つ。

- 1) 例文の提示:対話対
- 2) 例文の選択:日常的な行動, 状況が選択基準
(文法, 動詞の活用が主たる基準ではない)
- 3) 例文の情報:
語用論的バリエーションについて解説
基本的な応対関係の崩れについて例示

上記3)の語用論的バリエーションの例として、「評価を求める「どう」質問-返答」となる対話対を示す。

Q:「食事はどうでしたか」

A(基本):「おいしかったです」

A(発展):「たくさん食べられました」

このように、評価を求める質問に対しては形容詞などの評価表現を含む文で返答することが基本的な対話対と考えられる。しかし、形式上は評価表現を用いずに、内容は「評価」として返答することも可能であるため、発話対として成立する。このような言語運用についても、既存の教科書を用いた学習による対話の練習や暗記にとどまらず、運用の場での学習支援を目指すために、バリエーションの提示や解説も視野に入れた。

2.4 基本例文からの言い替え例文集

すでに2.2節で述べたように、基本例文から多様な文の解説情報を提供するためには、基本例文から当

該の文がどのように言い替えられたかを評価することが必要となる。また、そのような評価指標は多面的であることが望ましい。

そのような評価指標の実現には、文の特徴を捉える上での前提となるドメイン知識を整備する必要がある。ドメイン知識の整備には、次の節で説明するが、知識を体系化することが必要であり、文の言い換え、誤用分析、敬語等をより適切に整理するために文を計算機で扱うための意味表現を設計することが必要となる。それは、MedSLT[6]等の特定分野の機械翻訳や異文化コラボレーションツールの内部意味表現として、目的に応じた適切な機械翻訳用の意味判断のために中間言語と呼ばれる意味表現モデルが採用されていることに対応する。

以上の方針に基づき、基本例文とドメイン依存の対話例との対応付けを基本とし、対話例の言い換え可能性について、以下の視座から言い換え例を収集している。

1) 発話主体の属性や人間関係による言い換え

敬語や丁寧語に関わる。

2) 基本例文からの発話意図による言い換え

基本例文:「ここに名前を書いてくださいーはい」

言い替え:「ここに名前を書きましょうーはい」

↓対象ドメインでの対における言い替え

対話例文:「今から着替えてくださいーはい」

言い換え:「今から着替えましょうーはい」

3) 感性表現に関わる言い換え

発話者の多様な身体・心情の状況の提示についても対応が困難である。例えば、医療福祉分野ではオノマトペを中心とした、患者の心理、身体状況に関する表出の気づきについて資料が不足している。

頭が痛い → 頭がズキズキする

目に違和感がある → 目がゴロゴロする

以上のような言い替えは、分野に独自の運用にも対応するため、語彙表現の収集だけではなく、ドメイン知識の効率的な収集に計算機技術を援用することが期待される。

3. 介護ドメインの知識表現

3.1 対話に関するドメイン知識

我々は本研究の開始段階において、医療福祉分野での外国人労働者の就労状態や動向を現場関係者や関連学会・団体から調査し、彼ら彼女らの多くが、介護・看護の場面で活躍を期待されており、さらに、そのような現場において、被介護者とのコミュニケーションについて難しさを感じていることを認識した。

その知見に基づいて、既存の日本人向けの介護看護の実践テキストを対象に、テキストマイニングの手法を利用して調査・分析・整理し、介護場面における対話を計算機で扱う上で必要な知識を体系的に整備し、対話例を整理・処理する上での基盤の構築を目指した。このような調査は、特定の分野の専門日本語教育での対話を計算機で扱う上では、対象となる専門分野独特の用語や表現の収集は必須となり、本稿はそこにテキストマイニング技法を援用した事例紹介となる。

本稿が支援として想定する、例文提示型の教育において、対話例がどのような要素から体系的に整理づけるかを検討した。具体的には、我々が設計した対話例の整理体系は、対話例及び、それらに出現しうる語彙・表現を保存する体系であると同時に、次の特徴を持つ。

特徴の一つは、「場面」という整理の体系を導入し、該当分野の特徴を適切に把握・制限し、計算文法や意味処理を定義することである。ただし、「場面」という抽象的なラベルは、内部的な意味表現としては曖昧性が高いため、学習者が関わる行動を基準に、そこに関連して発生する対話例を保存する。このような整理をする利点は単に対話例の検索キーワードを場面よりも詳細化できるだけではない。例えば、介護、看護者に対して気遣いのある発話するために、食事をする時

に「おいしいですよ」と発話することが、行動に関して「提案」という機能的な役割をもつことがある。そういった現象の説明にも、この整理体系は役立つ。それは、発話が敬語であるかどうかを自動判断し、学習者に意識させる支援に整理体系を発展的に利用できる可能性を持つ。

もう一つの特徴は、行動に関わる語彙を対話例とは別に予め豊富に用意しておくことである。行動には、それに参加する主体の評価がつきものである。そのため、行動に関係する対話においては評価的側面があることを常に留意する必要がある。例えば、被介護者が着替えを行った後に「チクチクします」といった発言をした場合は、着替えという行動に関連し、被介護者からマイナスの評価が出ていることを、介護者は受け取らなければならない。後述するように、専門家ではないサービスの受け手は、上の「チクチク」のような直感的で感覚的な表現[8,9,10]を用いがちであるが、そのようなドメイン依存性の高い表現の辞書の整備はなかなか進んでおらず、それらの表現が使われた対話例をすべて登録されることを期待することは難しい。また、不完全な文体をとることが多いため、計算機支援においても技術的な課題が多く、解決のためには不完全な文体に対応できる語彙知識をドメインに即した適切に整備する必要がある。この方向へのテキストマイニング技術の応用は、4節に述べる。

3.2 テキストマイニングを利用したドメイン知識整理

以上に説明したように、患者や被介護者が日常生活の中で、看護および介護に携わる学習者がいかなる形で接しているかを体系的に整理するために、テキストマイニングを利用する。

学習者が被介護者と関与する行動は、日本人のための介護・看護の教科書が日本語で多数出版されている。このような教科書から、学習者と被介護者が関係する日常生活上の行動に関連する語を収集し、それらの行動を体系付けする。その際の観点は、教科書における「洗髪の介助技術」「足浴の介助技術」といった

学習項目である。ただし、それらの学習項目に対して、対話例の実例を整理する上で適切と考えられる介護の日常生活における「場面」は、一対一では対応しない。そのため、以下のような階層を導入して関連語彙群を整理し、ドメイン知識の体系として利用する。

場面—協調活動—行動—関連語彙群

具体的な例で説明する。介護分野の人材養成教科書には、例えば、「着替えの介助」の技術解説として以下のような記述例[7]を考える。

「健側の手を持ちながら、腕を抜いてもらう。」

この文の係り受け解析を行い、その結果を利用した、「着替えの介助」という協調活動に関する行動および、関連語句を整理する。上の例ならば、解説文内に出現する行動に相当する「持つ」「抜く」という語と、それらの行動の対象となる「手」「腕」という語を抽出し、行動とその対象をグラフ構造で保存する。このような作業を当該の学習項目の解説文集合に適用すると、当該項目に関係する語の関連性がグラフ構造に表現ができる。これをドメインの知識表現として利用する。

グラフ構造を用いた知識表現は、当該の協調活動に関する語の関連性を表現することに役立つ。例えば、協調活動の「着替えの介助」において、「手」に関する行動は図2のように抽出することができる。図2は「着替えの介助」に関して、特定条件で抽出した部分グラフである。

このような知識表現は、例えば例文が、どの協調活動と強く関連するかを算出する上で役立つ。例えば、「手を持ち上げてください」という対話例の一部が与えられたときに、この知識表現から対話例に関連する協調活動と行動の対応を検討することができる。また、協調活動に「起床」、「食事」、「排泄」、「入浴」といった日常的な介護の場面に対応付けておき、当該の協調活

動が行われる場所、時間等の情報を付与すれば、文からの場面对応付けも可能となる。

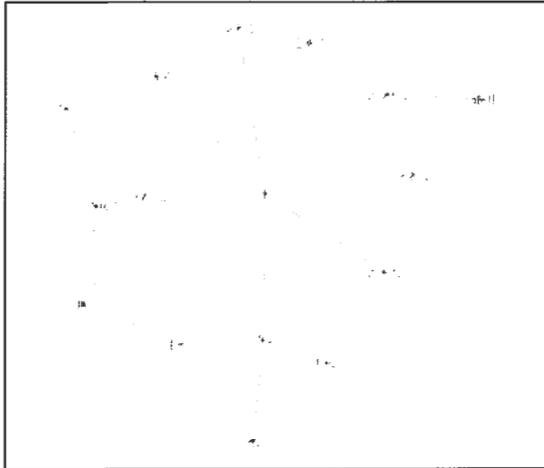


図 2. 協調活動に関係した手を対象とする行動群の例

上記では、行動とその対象の結びつけ手順のみを例示したが、通常、行動の対象語となる名詞の分類や関連性の定義が必要となるが、言い換えを踏まえて、次の評価対象となりうる名詞は、身体の部位、協調行動に関連した動作、「患者の身体、心情的な状態の対象となりうるもの」を関連資料から同様に半自動的な手段で整備している。

4. 多様な言い替えへの対応

4.1 評価表現のドメイン依存性

3 節では、医療福祉分野の日本人向けの教科書データからテキストマイニングすることにより、当該分野のドメイン知識を効率的に抽出整理する例を示した。それに対して、医療福祉サービスの受容者が自然に発する、それぞれの行動に関連付けるける直感的で感覚的な発話に対して、そこで使用される語彙をあらかじめ人手で網羅的に用意することには限界がある。そこで、インターネット上に大量に存在する直感的な記述をその使用文脈とともにあらかじめ収集しておき、当該の表現がシステムに入力された際に、その使用文

脈が一般的であるか否かを検討することを考えた。これは日本語教育ボランティアの教材作成時には、データ追加時の一般性を検討することに役立つ。また、学習者が自習支援の際には、当該の表現が特異であるか否かを検討する上で役立つ。ここでは、ドメイン依存性が高く、発話者の身体や心情の状況を表現する上で重要となる表現を自動収集する技術を紹介する。

ある表現が肯定的／否定的という二値の文脈においてどちらで出現しやすいかを検討した Turney らの研究[11]がある。Turney らの研究では、例えば、「良い」、「美しい」などの形容詞を代表とする評価語は肯定的な極性、「悪い」、「汚い」などの表現は否定的な極性を持つといったように、肯定的・否定的な文脈に特定の表現が出現する度合いをその表現の感性極性として捉える。

インターネット上には、商品についてユーザがレビューを行い、例えば 5 段階評価で 5 なら肯定的評価であり、1 ならば否定的評価であるといった、テキストデータとその評価値が組になったデータが多く存在する。これを教師データとして使用し、評価語に関して Turney の手法を適応すれば、良い評価に関連する語と、悪い評価に関連する語を数値化することが可能となる。

なお、このような評価語の極性値が評価対象となるジャンルごとに異なることは、既に知られている。我々は数多くのジャンルでの評価的意味を体系的に扱うために、意味的な中間表現として感性評価軸を導入した。

感性評価軸は、特定の評価対象について、対立する評価語を見出し語として提示し、当該の性質がどちらの語に似ているかを表現する図 3 のような形式のアンケート等でよく利用される意味表現方法である[12]。我々は、この評価軸をアンケート分析に関わる書籍を網羅的に調査し、できるだけ多くの評価的意味を表現する約 400 軸を用意した。

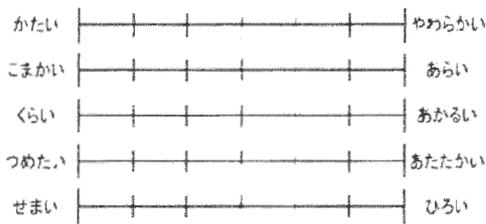


図3. 感性評価軸の例

この約400軸は直感的には肯定的評価を表す評価語と否定的評価を表す評価語の対であり、 i 番目の感性評価軸を形式上、以下のように示す。

$$P_i \leftrightarrow N_i$$

例として、Turneyの感性極性を応用した実験例を簡潔に示す。ここでは、食べ物の評価軸に関わる「スイーツ」ジャンルのテキストデータ、肌に対する評価軸を得ることが期待できる「化粧品」というテキストデータをインターネット上の評価サイトの書き込みから収集し、2つのコーパスとした。それぞれの分野の上位の特色となる感性評価軸の上位3軸は次のようになった。

- 化粧品ジャンル
細かい \leftrightarrow 粗い, かたい \leftrightarrow やわらかい, 引き締まった \leftrightarrow たるんだ
- スイーツジャンル
おいしい \leftrightarrow おいしくない, かたい \leftrightarrow やわらかい, 好き \leftrightarrow 嫌い

以上のように、2ジャンルごとに共通する軸があることを含め、ジャンルの特徴となる感性評価軸が存在する特性を端的に示している。

4.2 機械学習を用いたドメイン依存言い替え表現の収集

前節では、既を用意されている感性評価軸に対して、評価を行ったものであるが、用意した約400の感性評価軸には存在しない評価語が、意味的な中間表現である感性評価軸のうち、どの使用文脈と近いかを類似度の算出により、評価軸の評価語と入れ替えが可能かを検討した。手順は以下の通りである。この手

順は Turney の感性極性算出法を機械学習手法の代表的な手法の一つである Support Vector Machines(SVM)を使って改良した研究[13,14]に基づいて、感性評価軸のドメイン適合性判定に応用したものである。

- 1) ジャンル j のコーパス $C^{(j)}$ を教師付学習コーパスとして使い、約400軸の感性評価軸の評価語それぞれの感性極性を算出する。軸 i の評価語 P_i, N_i の極性値をそれぞれ $Sp_i^{(j)}, Sn_i^{(j)}$ とする。
- 2) $Sp_i^{(j)}$ と $Sn_i^{(j)}$ のそれぞれに閾値で足りしをした上で、 $|Sp_i^{(j)} - Sn_i^{(j)}|$ の値上位30位以上の感性評価軸を、ジャンル j を代表する感性評価軸の集合 $S^{(j)}$ とする。
- 3) $S^{(j)}$ に含まれる評価語それぞれの出現周辺文脈をSVMによりモデル化する。
- 4) コーパス $C^{(j)}$ において語 w が出現する周辺文脈が $S^{(j)}$ に含まれる評価語のそれぞれのSVMモデルにおいて最も類似する評価語を選択し、その類似度が閾値以上なら当該の評価語と w を言い替え可能と見なす。

上記の手順で、各コーパスに出現するオノマトペ表現上位100語に関して実験を行った。LIBSVM[15]を用い、カーネルにはRBFを選択した。

実験で、オノマトペ表現を対象にしている理由は、言語的には、口語体で身体や心理的な状態を表現する上で、直感的、感覚的に用いられ、意味のドメイン依存性も高いことが知られているからである。また、応用的には上記のような理由から、オノマトペ表現の辞書記述は難しく、また、生成的であるため3節のような方法で予め意味内容を記述しておくことが困難な表現の代表だからである。

実験結果を表2,3に示す。表内では、言い替え可能であったオノマトペ表現を矢印で示してある。このように評価する対象の側面や属性の異なるジャンルごとに、辞書登録がなされていないオノマトペ表現が、意

味的な中間表現である感性評価軸の評価語に関して言い替え例を提案することが可能であることを示している。

現在、医療福祉ドメインでの評価対象と評価表現を評価する上での基盤資源となる、評価サイトのデータを蓄積中であり、オノマトペ表現だけでなく、特定の表現が感性評価軸上と相関をもつ表現かを評価できるように整備を進めている。

表2 化粧品ジャンルでのオノマトペ言い換え

かたい	↔	柔らかい	→	しっとり
重い	↔	軽い	→	すっきり
乾いた	→	さらさら	↔	湿った
多い	→	たっぷり	↔	少ない
清潔な	→	さっぱり	↔	汚い

表3 スイーツジャンルでのオノマトペ言い換え

かたい	↔	柔らかい	→	ふわふわ
重い	↔	軽い	→	さっぱり
乾いた	→	ぱさぱさ	↔	湿った
多い	→	ずっしり	↔	少ない
強い	↔	弱い	→	あっさり

5. まとめ

医療福祉ドメインに限らず、本研究の基盤部分は様々な専門分野の日本語教育に適応できる可能性がある。その際には、専門ドメインの知識を効率的に作成することが必要となる。そのような状況において期待できる例文の構築方法、関連知識の半自動抽出方法について論じた。ここで論じた基盤技術や知識表現を、工場などの技術研修向けの専門日本語学習支援等に、より一般的に適用できるよう、言語グリッドのサービスとして実装することを検討していきたい。

謝辞

本研究は科研費(基盤研究(C) 研究課題番号: 21500113)の支援を受けて実施した。

参考文献

- [1] 安藤秀明, 秋吉信吾, 竹内和広: 作文指導に向けた表現データベースの構築と表現評価指標の検討, 第15回公開シンポジウム「人文科学とデータベース」, pp.37-42, 2009.
- [2] So NODA, et al.: Awareness Promoting Visualization of Style Inconsistencies in Texts,

- IEEE 2nd International Symposium on Aware Computing, pp.291-297, 2011.
- [3] Kazuki SHIMAMURA, et al.: An effective visualization of style inconsistencies for interactive text editing, 16th International Conference on Artificial Life and Robotics, pp. pp.154-157, 2011.
- [4] 岩下志乃, 岩切智希: 状況に応じた対話による敬語学習システム, 知能と情報, Vol.20, No.5, pp.709-719, 2008.
- [5] 白土保, 丸元聡子, 村田真樹, 神崎享子, 井佐原均, 日本語の敬語誤用判定システム, 計量国語学会, 計量国語学 28 巻 1 号, pp.1-20, 2011.
- [6] M. Rayner et al.: The MedSLT System, Coling 2008 Workshop on Speech Processing for Safety Critical Translation and Pervasive Applications, 2008.
- [7] 下正宗, 深見悦司: 『はじめての介護』, 成美堂出版, 2006.
- [8] 田守育啓, ローレンス・スコウラップ: 『オノマトペ - 形式と意味-』, くろしお出版, 1999.
- [9] Yumi NISHIMURA et al.: A Consideration on the Support for Japanese Onomatopoeia Learning in Japanese for Specific Purposes, 2nd Asian Conference on Education, pp.1164-1180, 2010.
- [10] 西村由美, 竹内和広: 目的別日本語教育におけるオノマトペ表現の重要性, 言語処理学会第17回年次大会, 2011.
- [11] Peter D. Turney: Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews, 40th Annual Meeting of ACL, pp.417-424, 2002.
- [12] 岩下豊彦: 『SD 法によるイメージの測定』, 川島書店, 1983.
- [13] 北中佑樹, 秋吉信吾, 竹内和広: 日本語教育に向けたオノマトペ表現データベースの構築, 日本知能情報ファジィ学会 第20回ソフトサイエンスワークショップ講演論文集, pp.49-52, 2010.
- [14] 北中佑樹, 竹内和広: 印象調査アンケートのためのオノマトペ表現による評価尺度の拡張, 電子情報通信学会技術研究報告「思考と言語」, 2010.
- [15] LIBSVM - A Library for Support Vector Machines.
<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

文化遺産のバーチャルアーカイブのデジタルコンテンツ化に関する研究

Study on virtual archives of cultural heritage content in digital

渡邊 俊祐

Shunsuke Watanabe

同志社大学 文化情報学研究科

Doshisha University Culture and Information Science, 1-3, Miyakodani, Tatara, Kyotanabe, Kyoto,
Japan

あらまし: 本研究では、文化遺産のバーチャルアーカイブを文化遺産自体のリアリティーの保持と有意性を鍵として、それを構成する景観構成要素である遺物(オブジェクト)と、それに関連して、本来遺物が保持している遺物が遺物であるがための景観構成要素を併せてアーカイブする方法の検討を行う。その為にイラン国立博物館所蔵粘土板資料の保存を目的としたプロジェクトの一環として、3Dアーカイブによる粘土板のデジタルコンテンツ化に際し、スキャニングから表現法までの一連の作業を通して、その有用性と可能性について議論する。

Summary: In this paper, I argue about the usefulness and possibility from the document investigations such as the 3D archives of the clay tablet of the Iran National Museum.

キーワード:バーチャル、アーカイブ、3次元スキャン、文化遺産

Keywords: virtual, archive, 3DScanning, cultural heritage

1.はじめに

今回、調査に参加したプロジェクトは、イラン国立博物館所蔵粘土板資料の保存と活用を目的としたプロジェクトである。その中で、3次元アーカイブを行い、その技術指導とデータ公開に関わる部分について実務を担当している。

さて、文化遺産に関する調査データは、近年まで2次元のアナログデータ(遺構図や遺物のトレース図)や現場でのメモ写真などとして膨大な調査数に比例して取得されてきた。しかし、調査により判明した遺構をそのまま保存されるケースは稀であり、調査終了後に文化遺産自体へのアプローチは難しいのが現状である。最近では、保存場

所や資料の劣化などの問題からデジタル保存のためのアナログ資料からデジタルデータでの保存への変換が行われている。だが、現在GIS¹などに代表されるデジタル情報での文化遺産データの統合をするための絶対的な情報の欠如により、過去に調査されたデータを活用することが難しいのが現状である。今後の方針として、文化遺産自体のリアリティーを併せ持ったデジタルコンテンツ化を目標に、それを構成すると考えられる景観構成要素をも併せた3次元スキャンの為に、バーチャルアーカイブ手法の検討を行った。イラン国立博物館

¹ Geographic Information System 地理情報システム

所蔵の粘土板資料のデジタルコンテンツ化に従事し、アーカイブデータだけではなくアーカイブ課程についても検討できるよう調査環境のデータ²についての検討を行い、現状での有用性と可能性について考察する。

2. 景観とは

まず、文化遺産のリアリティーの保持のために欠かせない景観構成要素から検討を行う。

景観構成要素とは読んだとおりであるが、受け取る側の人によって意味が異なるのではないだろうか。「とりわけすぐれた眺めを指す言葉として景観が挙げられよう」(柴田、2006)や「景観は、一般には風景に比してより狭域で用いられる概念とされている」(後藤、2011)という風に風景に含まれているという考えがある。一方、風景は主観的な審美的意識が含まれるが、景観は客観的で普遍的³な価値判断が含まれるという様な別物として捉えるという考えも存在する(中川、2008)。意味論的な見地だけでも大きく分けると風景に含まれるか含まれないという考えに分かれているため、総合的な景観という言葉は様々な捉えられ方をしていて当然である。そこで、はじめに本論での景観とはどのようなイメージなのかを検討する。

景観を考えると常に比較対象として登場するのは風景である。そして、風景もまたイメージとしては固まり切っていないように感じる。「風景とは、風景内の個別的な事物は同一の空間と一定の時間に所属して、その事物は、多数の意味によって関係し、構造の中で場所や方向や状況を獲得する。その意味や関係は偶然にゆだねられている」(柴田2006)。つまり、風景内の事物は、同一(主観的な統一された共同体)の時間的・空間的要因に所属していると考えられる。従って、概念として客観

² 3次元スキャナの設定条件を中心とした情報である。詳しくは4.3次元スキャンを参照

³ 一般的な対象を定義するもの(プラトン)

性を除外している風景ではなく、個人の主観を排除した客観的で普遍的なものとして「景観」が存在している。しかし、個人の主観を排除した結果として文化的要素の欠如を招くことになった(例建物の外観だけの保存による内部の歴史の抹消)⁴。つまり、自然を人間のまなざしで主観的に統一体にした「風景」でも、主観を排除した「景観」でも文化的事象の保存についてはそのまま適用するには足りない部分が存在するのである。本論ではイエスかノーか極端な選択である「風景」と「景観」を選択するのではなく、中間的な文化という統一されたものを客観的に見るために文化的景観⁵(後藤 2011)を今後は使用することにする。次に、今回の調査対象となった粘土板の文化的景観とは何かを考える。

3. 粘土板の文化的景観

「粘土板」と呼ばれている資料は、遺物としての分類上の名称であり、内容の仕様、つまり、記載言語が統一されているわけではない。そのため、出土地域や作成年代により楔形文字ではあるが、複数の言語⁶で書かれており、取引記録や支出管理や王の年代記などの内容が記載されている(Fig1)。楔形文字はラテン語の「楔」を意味するcuneusが語源であり、実際に読み書きしていた人以外が触れたのは、ヨーロッパ人が楔形文字を書き写した17世紀になってからである。文字の羅列方向は左から右に上から下に書かれており、記載内容の楔の方向によって判別することが基本である。また、Fig1のように粘土板に書かれているのが主流であるが、一部日干しレンガなどの部材にも刻ま

⁴ 中川 2008

⁵ 人がかかわった、すなわち生活の営みによって、自然と人口が影響し合って形成される景観のこと(後藤 2011)

⁶ シュメール語 ヒッタイト語 アッカド語 古代ペルシア語など

れている場合が存在する。メソポタミア地方においてはティグリスユーフラテス川によってもたらされる良質の粘土が手に入ったことで、家屋建築にも日干しレンガとして利用されていたが時が経つと共にくずれるので今日では巨大な遺丘（テペと呼ばれる）になっている。粘土板は、形を整えて、文字を書いて、乾燥させて保存するという工程で作成されている。また、熱を加えると石のように硬くなる。東はインダス川流域、西は東地中海（クレタ島やギリシア）、北はアルメニア、南はペルシア湾までエジプトを除く古代オリエント世界で使われていた。

粘土板の記載内容から、どのような用途、状況で使われたのかなどを知ることが出来る。例えば、取引記録では、都市や国家の人間同士の関係性を示すオブジェクトになる。また、支出管理の記録では、穀物や家畜の動きを知ることが出来、動的な関係性を示すオブジェクトになる。さらに、王の年代記では政治体制やその時代の事象を知ることが出来る。もちろん、どの言語が使用されているかということから、時代や地域、国などの情報を得ることが可能である。これらのことから、楔形文字文書をアーカイブすることは3次元的な記録としてアーカイブでき、かつ、何が刻まれているのか、どのような内容かを読み取れる状態でアーカイブすることが、文化的景観と結びついてアーカイブできる状況となると考えられる。



Fig1 粘土板

さらに、文字情報だけでは十分ではなく、それがどこに書かれているか、どのように書かれているかの情報が技法によるオブジェクトの差異につながることを期待される。将来的に記載内容からのアプローチだけではなく、形状という面からのアプローチという可能性も広がる。

4.3 次元スキャン

上記のような認識のもとに粘土板の3次元スキャンを行った。今回調査で使用した機材はNext Engine, Inc.のNext Engine (Fig2)という3Dスキャナで、Multi Stripe Laser Triangulationという4本のレーザー（クラス1）を2組使用し（Fig3）、また、3つのモードを駆使して最小0.126mmピッチのトライアングルで点群データを取得できるものである（Fig4）。

先に述べたように、この段階での作業内容を検証できるように複数の項目を設定し、情報を取得している（Table1）。この項目は、遺物の大きさ⁷、Fig2からもわかるように、機材と遺物を乗せるターンテーブルの距離と高さは任意であるためその距離、スキャン時の機械設定（解像度、明度、焦点距離、スキャン時間など）や遺物の状態を基本としたスキャン時における正の方向）など、本来ならば情報として残らず、スキャン後のオブ



Fig2 Next Engine

⁷調査時間の関係から大体の大きさ。詳細なデータは管理台帳に記載済み。

ジェクトからではアプローチしにくかった情報を意識的に情報化して様々なデータを保存した。



Fig3 レーザー発射口 (マクロモード)



Fig4 スキャン作業時画面イメージ

5. 3次元データ

今回、3D スキャナを使用した粘土板のデジタルデータ化は、イラン国立考古博物館碑文部門の一室を専用の部屋として利用させていただき作業を行った。

調査においてデジタルデータ化したオブジェクトは、54点(碑文部門52点、考古学部門2点)で85.6GBのデータ規模になり、作業日数19日で1日

約4時間、総作業時間は約76時間に及んだ⁸。ただし、現地博物館職員への3次元スキャンニングのレクチャーも上記の作業日数に含まれているので、オブジェクトのスキャンニングの実作業時間は数日分減少している。参加させていただいたこのプロジェクトでは、写真の補完的な位置づけ⁹で3Dデータを活用するという目的であったため、楔形文字自体を明瞭にスキャンすることを主眼とした。しかし、それだけでは将来的なデータの活用は望めない。つまり、現在大量に保管されているアナログデータと同じよう情報として保存されているが、活用することが難しい情報と将来的な価値が変わらなく恐れがあった。そこでアーカイブ可能な最大限の最小ピッチのデータを文字の形が鮮明にスキャンできることを最優先とし、次にオブジェクトの形も最低限表現できるように作成し、先に述べたようにスキャンニングの情報をすべてデータとして残すこととした。そして、データ処理ソフトを使用してノイズデータやホールやマージを行う作業が進行中である。今回、スキャンニングで使用した制御ソフトは機械とセットであるScanStudioHDを使用した。このソフトはスキャンからデータの成型まで一貫して作業することができるようになっており、初心者でも3次元のコンテンツ化が容易になっている。しかし、今回の調査におけるデータの成型は様々な諸事情からRapidWorksを使用することにした。このソフトは、RapidFormという3次元モデル加工用ソフトの廉価版であり、機能が少々制限されているが、Next Engineでスキャン可能なサイズのオブジェクトの加工には問題ないものである(Fig5)。

⁸ 調査時期が、イスラム教の宗教行事の断食月と重なっていたため、1日最大でも5時間しか作業時間を確保できなかった。

⁹ 写真では平面部の文字については明瞭に判読が可能であったが、局面部については1枚の写真では条件が厳しく、複数の写真にわたって確認する必要があったため、自由に角度を変えて確認可能な3次元データでのアーカイブが採用された。



Fig5 スキャニング作業時画面イメージ

このソフトを用い、基本的には3つの加工を行う(名称はソフト内の機能名に準拠する)。1つ目がTrim、2つ目がAlign、3つ目がFile holeである。Trimは、ノイズデータとなるオブジェクトの構成には不必要となるデータを削除する機能。さ

らに、誤差の大きいデータなど形状が大きく実物からかけ離れている部分に関しても、細かく指定して取り去ることが可能である。Alignは、複数のスキャンデータ同士を自動で位置あわせを行う。どうしても自動でできない場合は、手動で位置あわせを行うこともできる。ただし、その場合には、任意の移動を行ってしまうとオブジェクトの持つ究極的なオリジナル性が失われる危険も含んでいるということ念頭に、極力は計算による位置あわせを使用することが望ましいと考えられ、慎重に作業を行う必要がある。そして、File holeはその名の通りデータの欠損部分の穴を埋める機能である。元々点群にテクスチャーの張り付けやポリゴンとして形状を作成する際は点同士を線で結び、

Object Number	Date	File name (Rapidworks)	Trim	Align (Auto)	Much	File hole	other				
M547	8月8日	7*8*2	18	8	MACRO HD2	Light	4	360	wide		
M552		7*8*2	18	8	MACRO SD3	Light	4	360	long		
M551		8*8*1	18	8	MACRO HD2	Light	12	360	long		
M305		8*3.5*2	18	8	MACRO SD3	Light	4	360	long		
M558		8*8*2	18	8	MACRO HD2	Light	12	BRACKET	long		
M559		8*3.5*1.5	18	8	MACRO HD2	Light	8	360			24
M234		8*8*2	18	8	MACRO HD2	Light	8	360	long		23
M560		4.8*2*1.5	18	8	MACRO HD2	Light	8	360	long		23
M561		8*5*2	18	8.5	MACRO HD2	Light	8	360	wide		26
M368		6.3*4*5.5	18	8.5	MACRO HD2	Light	17	ONESHOT	wide		3.3
M341	7*8*2	18	8.5	MACRO HD2	Light	8	360	long		24	
M414	5*4.2*1.3	18.5	8	MACRO HD2	Light	8	BRACKET	long		9.1	
M315	8*3*1.8	18.5	8	MACRO HD2	Light	8	360	wide		23	
M1786	3*3.5*2	17.5	10	MACRO HD2	Light	8	360	long		23	
M1757	7*4.5*2	17.5	10	MACRO HD2	Light	8	360	wide		23	
M125	none	18.5	9.5	MACRO HD2	Light	4	360	long		2.5	
TEST 01642	none	18.5	10.5	MACRO HD2	Light		ONESHOT	long		2	

Table1 スキャン設定メモ情報

Object Number	Date	File name (Rapidworks)	Trim	Align (Auto)	Much	File hole	other
M547	8月8日	Test					
M172	8月8日	Default					
M202	8月7日	Br3994 Trim					
M542	8月7日	Br4005 Align					
M2917 Br3994	8月7日	Br4005					Request Transfer
M450	8月7日	Br4005 Align					
M547	8月8日	Br4005 Align					
M262	8月8日	M552 Trim					
M551	8月8日	M551 Trim					
M305	8月8日	Br4005 M559					
M558	8月8日	Br4005 M559					
M234	8月8日	M559 Trim					
M559	8月8日	M559 Trim					
M305	8月8日	M305 Trim					
M341	8月8日	M341 Trim					
M414	8月8日	M414 Align					
M315	8月10日	M315 Trim					
M1786	8月10日	M1786 Trim					
M1757	8月10日	M1757 Trim					
M525	8月10日	M525 Trim					

Table2 データ処理進行状況表

三角形の面を多数作った上で面やポリゴンが作られているが、境界となってしまう部分＝(穴が開いている部分を自動的に認識し、穴埋めを行うことが可能である。さらにオプションとして、計算により周辺部と同じような曲線の面での穴埋めも行うことが可能である。

上記3つの作業の進行状況を把握することを中心とした加工の進行状況の表を作成している。ではなぜ、一つのオブジェクトに複数のファイルを存在させる必要があるのか、それは、データ管理の面からすると、スキニング終了時点において膨大なデータ量が存在しているうえに、ファイル単体ごとのデータ量も決して軽いとは言えないものである。しかし、ステップごとにファイルを保存しておくことは将来的にメリットとなる可能性があることによる。その可能性とは、Rapid worksで1つの加工をしてしまった場合、前の段階に戻ることができないステップが存在するため、加工終了時のごとのファイルを保存している。この事は、加工終了後は問題とされなかった事象が、時間がたった段階で問題となった場合に完成形とされたファイルから元に戻して加工を再度行うということが困難なためである。将来的に起こりうる可能性のある先のような問題は、再度加工しなおすなどの場合、スキニング終了時のデータからやり直すことは大幅な時間のロスにつながるため必要な情報であると考えた。Fig6は上記の作業をする前のデータである。ターンテーブルなどのわからない情報や複数回によるスキニングによるスキアンデータ間での位置の誤差などが見受けられる。Fig7, 8は、上記の作業が終了したデータである。Fig7は点群データ上にスキャン時に同時に取得した写真データをテクスチャーとして張っているもので、Fig8はテクスチャーを張っていない点群だけのものである。そのため、Fig8の方が任意の色に変更が可能となっており、視覚的に知覚しやすいようになど、目的によって意図的に変更が

可能となっている。例えば、Fig7, 8を見比べた場合に、また、意図的な変更を加えるのは色だけではなく、光源の変更もソフト上で可能となっている。人間が立体感のある楔形文字の文字を認識するには陰影によるコントラストが必要不可欠であり、写真(Fig9)では撮影の際に様々な方向と強弱をつけての撮影が必要になってくるが、ソフト上で最適な位置と強弱を選択可能になっている。つまり、最適な条件を求めて複数回オブジェクトに接触することは保存の観点からするとリスクが高まることになり避けなければいけない。特に粘土板のように劣化が進みやすい文化遺産に対してはなおさらである。文化遺産への接触が最少回数で済む可能性の高い3次元スキャナは、文化遺産保存の観点からも有用であるが、長所と短所が存在することを理解する必要がある。



Fig6 スキニング終了時データ



Fig7 加工処理終了コンテンツ



Fig9 写真によるオブジェクトデータ保存



Fig8 色変更コンテンツ

6. 3次元データの表現法

加工が終了すると、3次元のデジタルコンテンツ化のすべての作業が終了するわけではない。完成したデータをどのようにして、パソコンの外へアウトプットできるのかということを考えなければならぬ。もちろん、パソコン内部において、加工ソフトに付属している見るためだけのビューワソフトを使用すれば、簡単に360度回転させ、自由に3次元のデジタルコンテンツを見ることは可能である。しかし、パソコンにインストールされているソフトを使うとなると、どこでもだれでも簡単というわけにはいかない。これが、現状で最も苦慮している点がここである。パソコン環境で万人が観察することができるという条件ならば、ハードルは下がる。例えば、一番簡単な手法は、ビデオ形式での保存である。この場合、3次元データを2次元の写真にし、映像にするという本来の趣旨からすれば、少々本末転倒な感が否めな

いが、想定されうる回転角度をあらかじめビデオにすることで直観的ではないが、ある程度の任意での角度から観察することは可能となる。また、写真データにすることで、本来はCADに分類される加工ソフトでは表現できないカラーリングも専門ソフトを介すことで可能となる。

現状では、3次元のままカラー情報を意図するままに加工することは難しく、想定されたものではなかったとしてもいい意味の副産物である。

7. まとめ

ここまで、粘土板の3次元データスキャンについて述べてきたが、当初に設定した目標である景観構成要素を含むバーチャルアーカイブにまで現状では到達していないのは明らかである。

ここ数年によるパソコン機器やソフトによる技術的進歩により、デジタルコンテンツの作成は容易になっていることは確かである。しかしながら、多くの場合、先にも述べたように、デジタルコンテンツの表現として、パソコンなどの画面上から抜け出すことに苦労している感が否めない。もちろん、3次元プリンターやプロジェクターなど3次元の情報を3次元のままアウトプットすることも一部は一般的にも可能とはなっている。しかし、あくまで一部であり、それが3次元データの活用の解であるかという絶対であるとは言い切れず、選択肢のうちの一つのように考えられる。次のステップとしては、どのようなアウトプットをすれば、どのようなニーズに対応でき、さらに、どのシチュエーションに最適であり、より簡単に3次元であることの長所を示すことが可能になるのかの検討が必要であると考えている。

謝辞

本研究は、基盤研究A・海外学術調査「イラン国立博物館所蔵粘土板文献の調査・研究」代表者前川和也氏（国士館大学・教授）23-27年度の共同

研究成果の一部である。上記プロジェクトに関係し、日頃お世話になっている全ての方に記して感謝申し上げる。また、3次元データのスキニング後のデータ処理について、多数の助言を頂いている株式会社相互技研並びに同志社大学時空間情報科学研究センター研究員の田中良明氏、内山幹夫氏にも感謝申し上げる。

引用、参考文献

- 杉 勇 著 楔形文字入門 講談社 2006
- 飯島 紀 著 楔形文字の初歩：歴史と文法：シュメール語・ハッチ語・アッカド語・新アッシリア語・ウガリト語・ベルシャ語 泰流社 1994
- E. キエラ 著 板倉勝正訳 粘土に書かれた歴史：メソポタミア文明の話 岩波書店 1958
- 中川 理著 風景学 風景と景観をめぐる歴史と現在 共立出版株式会社 2008
- 景観学への道 藤沢 和 著 日本経済評論社 2009
- 木岡 伸夫 著 風景の倫理 沈黙から語りへ 世界思想社 2007
- 柴田 陽弘 編著 風景の研究 慶應義塾大学出版会株式会社 2006
- 都市へのテキスト/ディスクールの地図 ポストグローバル化社会の都市と空間 後藤伸一 著 株式会社建築資料研究者 2011
- エドワード・ホール著 日高 敏隆 佐藤 信行訳 かくれた次元 みすず書房 2000
- 荒山 正彦 大城 直樹編 空間から場所へ 地理学的想像力の探求 古今書院 1998
- 時間と空間 エルンスト マッハ著 野家 啓一 訳 法政大学出版局 2008

奈良文化財研究所におけるデータベース Database in the Nara National Research Institute for Cultural Properties

森本 晋

Susumu Morimoto

独立行政法人国立文化財機構 奈良文化財研究所, 奈良市二条町 2-9-1
Independent Administrative Agencies National Institutes for Cultural Heritage
Nara National Research Institute for Cultural Properties
2-9-1 Nijocho, Nara

あらまし:奈良文化財研究所は長年に渡り、調査研究に基づいて文化財関係のデータベースを作成してきており、その多くを公開している。データベースそれぞれの特徴について、入力内容の工夫とともにデータベース構造の特色についても紹介する。また合わせて考古資料の総合的な情報化と活用についても述べる。

Summary:Nara National Research Institute for Cultural Properties has been continuing to develop the databases about the cultural heritage for a long time and disclosed them. It is mentioned about the characteristics of databases and the device of data entry. I also express the total digitization, informatization and application of archaeological data.

キーワード:文化財, データベース, 電子化, 情報化, 公開

Keywords:cultural properties, database, digitization, informatization, disclosure

1. はじめに

奈良文化財研究所(以下、奈文研)は長年に渡り、調査研究に基づいて文化財関係のデータベースを作成してきており、その多くを公開している。データベースのいくつかについて、内容を入力する時の工夫とデータベース構造の特色を紹介する。また合わせて、奈文研が取り組んでいる考古資料の総合的な情報化と活用についても述べる。

2. 公開データベース

奈文研がインターネット経由で公開しているデータベースには以下のものがある。

木簡関係のデータベース

- 木簡人名データベース
- 木簡画像データベース・木簡字典/電子くずし字典データベース連携検索
- 木簡データベース
- 木簡画像データベース[木簡字典]
- 全国木簡出土遺跡・報告書データベース

遺物関係のデータベース

- 軒瓦データベース
- 遺跡のデータベース
- 遺跡データベース
 - 地方官衙関係遺跡データベース
 - 古代寺院遺跡データベース

遺跡整備関係のデータベース

- 官衙関係遺跡整備データベース
- 遺跡の斜面保護データベース

庭園のデータベース

- 発掘庭園データベース
- Archaeologically Excavated Japanese Gardens
- Japanese Garden Dictionary

図書のデータベース

- 所蔵図書データベース
- 発掘調査報告書のデータベース
- 報告書抄録データベース
- 文書関係のデータベース

■薬師寺典籍文書データベース

■大宮家文書データベース

これらの中で、筆者が設計・入力・更新に直接関わっているデータベースのいくつかについて詳しく紹介する。他のデータベースについては、公開しているものなのでぜひ、実際にご利用いただき、ご意見をいただきたいと考えている。

<http://www.nabunken.go.jp>

遺跡データベース

遺跡データベースは、日本の遺跡全体を対象として構築を進めているデータベースである。

遺跡とひと口に言っても様々なものがあり、遺跡によっては、重層的な構造が見られる。例えば平城宮跡はひとつの遺跡としてとらえる場合もあれば、大極殿地区、朱雀門といった部分に分けて考えるべき場合もあろう。

遺跡データベースにおいては、情報の典拠となる文献の記述に従ってデータを入力している。よって、ある古墳群について「円墳5基」という情報しかなければ、古墳群としての情報しか入力できない。これに対し、群を構成している1基ずつの古墳それぞれについて情報が得られれば、個々の古墳ごとにレコードを作成して情報入力が可能である。

遺跡群とその構成要素である個々の遺跡といった重層的な構造をそのままデータベースの構成に持ち込むと、群と個別とが明確に分かれておらず、かつ十分な情報が得られないことが多い現状では、構成が複雑になりすぎる危険性がある。そこで、奈文研の遺跡データベースでは「種別」というフィールドを導入し、そこに集合・支群・個別・地区・調査といった重層構造中の想定位置を記述している。群としてとらえられる遺跡も単体としてとらえられる遺跡も同じようにフラットな構造の中で取り扱っている。

また、フィールドの設定において、その内容に関する情報が得られている遺跡数があまりに少ないものについては、当該フィールドを廃止して、「概要」の中に記述を移動した。例えば、「現状」に関する情報は「遺構概要」の中に「現状」という接頭語を付けて記述している。こうすることによって、検索結果の画面に空欄が並ぶという事態を回避している。

遺跡データベースには現在474,756レコードがあり、日々入力を続けている。典拠となる文献としては、都道府県単位の遺跡地名表・遺跡地図を優先して入力してきたが、最近では出版させているものをほぼ網羅し

たので、市町村単位の遺跡地名表・遺跡地図へも参照範囲を広げて入力を継続している。

また、遺物別の出土地名表の類も適宜参照してデータベースの充実を図っている。

発掘調査報告書抄録データベース

このデータベースは、遺跡の発掘調査報告書の内容に関するデータベースである。冊子体の報告書に添付されている抄録を電子化したものだが、抄録がない報告書についてもデータを入力するとともに、遡及入力を継続して行っている。

1レコードは、報告書内に記載されているひとつの遺跡に関する記述を単位としている。このため、1冊の本の中に複数の遺跡に関する記載があれば、複数レコードとして入力されている。

情報源は、都道府県教育委員会が、管下の市町村教育委員会を含めて取りまとめたデータと、全国埋蔵文化財法人連絡協議会が加盟機関のデータを取りまとめたものの提供を受け、奈文研が独自入力データを加えている。

外部から寄せられたデータについても、奈文研が用字の統一、用語の調整を行っている。これは検索によって確実に必要な情報が得られるようにするためである。

データ取りまとめの性質上、まとめた公開情報の更新は1年に1回であるが、1年分のデータの調整には10ヶ月ほどを要している。2011年度末には、1984年度と2010年度のデータが寄せられる予定となっており、カバーしている範囲が27年分となる。毎年の上及入力により、日本で出版された遺跡調査報告書に関する情報の全てをカバーできる日もそれほど遠くないと期待されている。

3. 業務用データベース

奈文研内部で業務用として使用しているデータベースは3種に分けることができる。

1. 公開しているデータベースと同じデータベースで利用しているレコードも一緒のもの。
2. 公開しているデータベースであるが、公開していないフィールドや公開していないレコードを含んでいたもの。
3. 全く外部には公開していないデータベースである。

1に属するものとしては、遺跡データベースなどがある。2に属するものとしては、木簡データベースなどがある。木簡データベースにおいては、木簡に記載された本文がたいへん重要な情報であるが、古代の遺物

に記された文字なので、木簡が破損していたり、保存状態が良くなかったりして文字の読み取りが困難な場合が多い。そこで、複数の研究者による検討によって積文が確定したもののみが、公開されており、確定していない木簡については業務用データベース内でのみ利用可能となっている。

第3のカテゴリーに属するものとしては、写真データベース、航空写真データベース、図面画像データベースなどがある。またその他に、奈文研の所内には、研究者個人が作成しているもの、研究室内でのみ共有しているデータベースがある。

写真データベース

奈文研が撮影したフィルムのひとコマを1レコードとするデータベース。電子化されたものを中心としており、362,605レコード。データ中に、奈文研所蔵ではない遺物の写真などを含むため非公開となっている。

奈文研では撮影されたフィルムは登録され、恒温恒湿の保管庫で保管されており、劣化は少ないが、それでも自家処理していない一部のカラーフィルムに退色が認められ、早急な電子化が望まれる。また、8×10インチといった大判のフィルムやガラス乾板のように取り扱いが困難な原板も、通常の出版物での使用といった用途には電子化したものを利用する方が楽である。

使用頻度の高いカラーキャビネ版から、奈文研内部での電子化を開始し、次いでカラーブローニュ版、カラー35mmスライドと範囲を拡大して継続している。また、大判やガラス乾板については、設備の関係で外注して電子化を進めている。

航空写真データベース

奈文研が航空測量の会社から預かっている航空写真の撮影位置を中心とする情報のデータベース。地図上に撮影位置を表示できる。

奈文研が預かっている航空写真には次の特徴がある。昭和29年から昭和63年に撮影されたフィルムで、発注者に納品されておらず、撮影機関自身が保管しているフィルムを除いたもの。測量や地図作製のために専用のカメラで撮影された垂直写真に限る。

データベース化の手順は次の通りである。原板は1コマが23cm角のモノクロネガフィルムであり、視認性が低い。長いものでは50m近い長尺であるのに対し、短く切られたものもあり、そのままでは扱いが困難である。そこで、4500コマ程度をひとまとまりとして、マイクロフィルム化を行っている。データベースへの入力は専用入力ソフトを使用し、地形図の画像をパソコンの画面に出し、空中写真の撮影位置をクリックすることに

よって位置情報をデータベースに取り込む。入力者の左側でマイクロフィルムをリーダーに掛けて表示しながら、右側でパソコンを操作するという作業である。

空中写真には、どこを撮影したというフィルムかというメタデータが、評定図という形で添付されているのが通常であるが、添付されていないこともあり、必要な情報が欠けている場合も多い。評定図が不備な写真については、撮影の件名と写真に写っている地表面の情報のみから、どこを撮影したフィルムであるかを特定しなければならない。20万分の1地勢図や5万分の1地形図を参照しながら作業を行うが、空中写真撮影後の地形変化が著しい場合もあり、同定はたいへん困難な作業である。

このデータベースは、内容となっている航空写真が奈文研の所有品ではないために公開することができない。ただ、ある場所を撮影した写真が奈文研にあるかどうかという所在情報については、個別の照会に対応可能である。現在までに、1,277,315レコードが登録されており、航空写真に関するデータベースとしては日本最大である。

図面画像データベース

60年に及ぶ奈文研の調査で得られた実測図などの図面をスキャンした画像を検索するためのシステムで、本年度に利用開始したばかりである。

大判用紙に対応したスキャナの導入により、発掘調査時の遺構実測図から順に電子化を行っている。スキャンはカラー24bit、400dpiで行っている。スキャン後の作業として、余白の切り取り、傾きの補正を行うとともに、複数図面を一度にスキャンした場合には、個々の図面への切り分けを行って、基礎となる画像を作成する。次いで、この基礎画像を元にインデックス用の画像を作成している。

メタデータとしての文字データを作成し、画像データとのリンク情報をつけたものをデータベース化している。発掘調査の現場で作成された原図は、大判であることもあって、遺構を検討する際にも参照するのが困難であったが、このデータベースにより細部の確認などが研究者各自のパソコンから可能となった。

文化財情報研究室では、研究所内の個々の研究室内でのみ共有されているデータを、研究所全体で共有できるようにする取り組みを続けている。専門性の高いデータであっても、異分野の研究者が参照することによって新たな視点からの研究が進むことも期待される。

全所的な共有や対外的な公開に関しては、各研究室のみの努力では実現できないので、文化財情報研究室が助言やコンピュータ資源の提供を行っている。

4. 総合的な電子化と情報化

奈文研内部においては、文章の電子化、写真の電子化、図面画像の作成という方向で電子化を進めている。長い調査・研究の蓄積がある機関だけに、紙媒体・フィルム媒体で保管している資料は膨大な量にのぼっている。

これらの資料の高度な利用を図るために電子化を進めてきているが、新規発生分を除く既存部分の電子化が終了するにも、まだかなりの年月を必要とする。ただ、写真資料についても電子的なデータによる貸し出しが行われるなど、日常業務として利用できる水準にまで電子化が進んできている。膨大な資料数に臆することなく作業を日々進めることが大切であると考えている。

文字列による検索を行うデータベースを統合的に活用する道筋も導入している。それは、データベースの横断検索である。

例えば、「朱雀門」という語句を検索したい時に、どのデータベースを検索すれば、望んでいる情報が得られるかよくわからない場合がある。こういう時に、横断検索を活用すると、木簡データベースには〇〇件、〇〇データベースには〇〇件というヒット数が表示されるので、それ以降はそれぞれのデータベースに入って詳細な情報を得ることができる。

写真情報に関しては、近年最初からデジタルカメラで撮影されたデータ、いわゆるボーン・デジタルなデータが増加している。これらのデータは、現像・焼き付けという過程を経ない手軽さもあって、大量に作成される傾向がある。このため、予備的な整理を行ってメタデータを付与しないと、写っているものの内容が不明となってしまうやすい。撮影日時などのメタデータが自動で取得できるという利点を活かしつつ、必要な情報を加えてから、上記の写真データベースへと情報の移行を図るための、デジタル写真用のシステムを別建てで運用している。

図面の情報に関しては、現在発掘調査時の遺構実測図に限られている対象を、順次建築調査時の図面、遺物の実測図などに広げていく予定である。

次の課題として、画像情報としてのみ電子化している図に意味を付加する作業がある。これは図の構造や

構成要素をどのようにデータベース化するかということと深くかかわっている。

すなわち、掘立柱建物として認識している遺構が、複数の柱穴と溝によって構成されている場合、遺構を構成する遺構素として穴・溝を認識している訳で、そういった遺構構造を記述しなくてはならない。ひとつの柱穴を取ってみても、図として表す場合には、柱穴の上端線、下端線、遺構ケバといった図の構成要素があり、それらは単独で存在しているのではなく、相互につながりがある要素となっている。よって、図として表現する時に、柱穴位置の変更ということが、上端線と下端線の位置を個別に変更することで解決されるのではおかしいことになる。柱穴という実体としての把握がデータの構造にも必要である。

こういった遺構構造、遺構図構造を適切に情報化することによって、蓄積された膨大な情報をよりよく活用することが可能となると考えられる。先行している文字情報、画像情報の電子化と統合も必要で、この分野に関する研究をさらに進めることが大きな課題となっている。

デジタルデータの脆弱性が指摘されて久しい。電子化を進めることで、貴重な文化財を記録した、多くは唯一で代替物のない記録が失われてしまわないような配慮が必要である。遺物や美術工芸品であれば、複数回の写真撮影も可能ではあるが、遺跡の発掘調査は再現不可能な一回きりの作業の記録である。

記録内容に対する適切な到達可能性を確保した上でのデータ・マイグレーションを継続して行っていくことは、発掘調査の主体である地方公共団体単独の努力では困難な面が多いと考えられる。奈良文化財研究所が行っている、情報の電子化・情報化のノウハウを参考にしてもらうことによって、人類の宝とも言うべき文化財情報を安全かつ完全に次の世代に引き継ぎ、活用してもらえることを強く望む。

西日本縄文・弥生時代集落の GIS データベース化と時空間動態評価 GIS Site Database and Multivariate Analysis for a Spatiotemporal Study of Jomon-Yayoi Settlements in Western Japan

山口 雄治

Yuji Yamaguchi

同志社大学 文化情報学部, 京田辺市多々羅都谷 1-3

Doshisha University, 1-3 Tataramiyakodani, Kyotanabe, Kyoto

あらまし: 西日本における縄文時代遺跡の特徴の一つとして、遺跡数や遺構の検出例が東日本のそれに比べて極めて少ないことが挙げられる。このような資料的制約によって、西日本各地の個別具体的な遺構論や集落論は遺存状態の良い、特定の遺跡情報のみを用いて定性的に展開せざるを得ない状況があった。そこで、遺跡の大局的な展開を定量的に評価するために、西日本における縄文・弥生時代遺跡のデータベースを構築した。これを用いて各種遺構を変数とした多次元定量解析(主に主成分分析)を行い、当該期における集落の動態とその背景について考察した。その結果、西日本の集落は縄文時代後期前葉に、家族的・長期的居住傾向を示す集落から共同体的・短期的居住傾向を示す集落へと変化することが明らかとなった。こうした変化は、縄文時代社会が家族的・短期的様相から共同体的・長期的様相への線形的に移行したのではなく、共同体的・短期的様相を経た非線形的な推移として展開したことを表していると考えられる。また、時期的に東日本縄文時代集落の解体と軌を一にしていることから、こうした動態が汎列島的な特徴であることが予測された。

Summary: We have constructed a database of sites from the Jomon to the Yayoi period in western Japan in order to develop a quantitative analysis of the settlement pattern changes. In this paper, we will discuss the dynamic state of settlement in these periods and their background by using a multivariate analysis (especially “PCA: principal component analysis”) of the sites database. As a result of our analysis, we have observed a characteristic change occurring in Jomon Late period in settlements of western Japan, from the “family-like-and-long-term-stay settlement” to “communal-and-short-term-stay settlement”. In other words, this is a non-linear process in historical change of settlement rather than a linear process. At the same time, in eastern Japan, Jomon settlement splits into small units although they also constructed communal monuments. This stream in eastern Japan can be also observed in western Japan. For this reason, we believe this transformation of Jomon settlement in this time has a generalized aspect in the Japanese archipelago.

キーワード: 縄文・弥生集落, 遺跡動態, 多変量解析, 縄文時代後期, 西日本

Keywords: Jomon-Yayoi settlement, sites movement, multivariate analysis, Jomon-Late period, western Japan

1. はじめに

西日本における縄文時代集落の現象面における特徴として、①集落が少数かつ小規模である、②時期によって立地が変化する、③環状集落が認められない、の3点を挙げる事ができる(大野2001)。これらの特徴は、おおよそ東日本縄文時代文化の集落とは対照的な姿であり、その評価をめぐって、これまでにさまざまな見解が提出されてきた。例えば、これが周辺環境

の生態的条件、とりわけ食料資源の量的な問題であると説いたサケ・マス論は有名である(山内1964)。相対的に遺跡検出数の少ない西日本においては、その資料的制約から当然個別具体的な遺構論や集落論は停滞気味であり、遺物論や遺跡動態論がもっぱらであった。具体的な集落論が展開し始めるのは資料が増加し始める1990年代後半以降であり、2000年以降、各地で集成や個別的な検討が行われるようになってきた(板倉

2010、稲田 2010、縄文時代文化研究会 2001、関西縄文文化研究会 1999、九州縄文文化研究会 2000、瀬口 2003、2010、富井 2000、2002、山田 2002、矢野 2001、山口 2008、2010 など)。しかしながら、各地の集落の様相が次第に明らかになってきた反面、地域間比較による西日本縄文時代集落の相対化と一般化はこれからの大きな課題である。

こうした現状を認識した上で、本論では、西日本における縄文時代遺跡の通時代的なイメージを手続き的な再現性を伴いながら再構築することを目的とする。特に、従来の特定の充実した資料情報に引きずられる定性的なイメージの醸成をより発展的に継承する意味も含め、ここでは、より広域で簡素ではあるが定量的評価が可能な資料情報を集積し、これを基盤として主成分分析などの多次元定量解析を実施する方法を採用する。また、遺跡立地については、近年一般的な分析ツールとしても具体事例で活用されつつある地理情報システム (GIS) の解析機能を適用する。本論では、特定のサンプルからの特定の解釈ではなく、多次元定量解析の手法を用いた通時代的な理解を通じて、縄文時代遺跡・集落イメージの構築を行う。

2. データベースの構築

GIS を基盤とした西日本地域における縄文～弥生時代 (前期) 遺跡データベースの作成を行った (図 1)。各遺跡のもつ変数としては、位置情報 (世界測地系準拠)、立地、標高、継続時期、遺構の種類を設定し、まずは報告遺書の記述どおりに集計した。それは、各遺跡報告書によって、遺構の認識にズレが存在し、必

ずしも遺構の名称が統一されていないからである。従って、一度データを入力した後、遺構の記述内容が同一とみなせるものは用語の最大公約数的な統一を図った。

また、遺跡立地の分類についても、報告書記載の遺跡立地表現の記述レベルに差異が認められるため、大まかな地形的分類を用いざるを得なかった。しかし、これについては GIS を用いた分析によって十分補完できるものと考えられる。データベースを入力する際には、例えば 50m メッシュデータに反映されない微地形をデータとして入力することがむしろ望ましいが、今回は行っていない。

なお、本データベースで採用した時期区分については表 1 の通りである。多少大雑把な時期区分ではあるが、西日本全体の動向を統一の基準で論じるためには、このような時期幅が限界であると考えられる。本来は

表 1 本論の時期区分

大別時期	時期	
0th	草創期	押型文以前
1th	早期	押型文
2th		(押型文～) 縄文土器～
3th		羽島下層式～
4th	前期	磯ノ森式～
5th		彦崎Ⅱ式～
6th		船元Ⅰ式～
7th	中期	里木Ⅱ式～
8th		中期末～
9th	後期	中津式～
10th		津雲A式～
11th		元住吉山Ⅱ式～
12th	晩期	滋賀里Ⅱ式～
13th		篠原式～
14th		谷尻式～
15th	晩期/早期	津島岡大式～
16th	弥生前期	板付Ⅱ式～

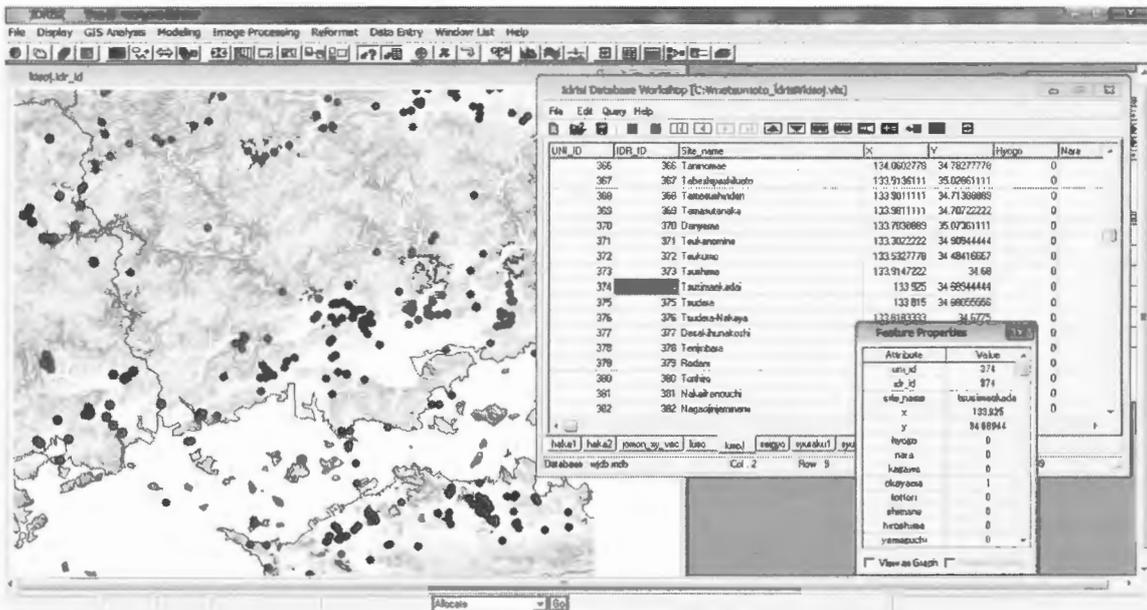


図 1 岡山県における遺跡分布と遺跡空間データベース

できるだけ細かな編年に依拠するべきであろうが、編年も細かくなれば見解の統一が困難な状況になるだろう。本論での分析には、この時期区分でも十分目的に適うと判断した。

分析に用いる試料数は延べ4656遺跡である。ただし、近畿地方は兵庫・奈良県域のみであり、四国地方は含めていない。前者については、片手落ちの感も否めないが、大局的な傾向はつかめるのではないかと考える。

3. 縄文・弥生時代の遺跡動態

(1) 各県別の遺跡動態

はじめに各県別の遺跡数変動を確認したい。しかし、各県において遺跡の絶対数は異なる。これでは標本平均(平均)や標準偏差(S.D.)も当然異なるため比較に適さない。従って、ここでは県毎にデータの「標準化」を行った。標準化とは、平均0、標準偏差1に統一する統計的操作である。これにより、すべて同じ土台で変動の幅を議論できるようになる。図2は、そうして作図した時系列のデータである。

0期から1・2期にかけて増加し、そのまま8期までほとんど遺跡数に大きな変化はみられない。1・2期に大きく増加しているのは南九州地方である。それ

を除けば、西日本各地の遺跡数は地域ごとの変化は認められない。

9期になると各県ごとに大きく異なる動態を示す。平均値も上昇するがS.D.の平均値が二倍以上の差を示していることから、全体としての動きが活発化しているのに加え、各地域の展開過程に大きな隔たりが存在していることがわかる。特に、近畿地方では8期に、中国地方では9期に、九州地方では9期に増加するものの10-11期にピークを形成する。

こうした9期以降の遺跡数の増加とそのピークの西漸については、東日本からの一定程度の人口移動や居住形態の変化があったことが想定されており、比較的知られている現象である(矢野2004、山口2008など)。しかし、それ以前の前・中期における低い変動幅や地域ごとの変化量などについては、未解明の部分が多く、今後の課題である。

(2) 遺跡立地の変動

次に、各時期の遺跡立地平均標高とそのS.D.、傾斜角などをSRTM3データ(DEM)からGISを用いて遺跡範囲の平均値として求めた(図3)。多様な地形を要する日本列島において、各時期の遺跡標高平均値は、考古

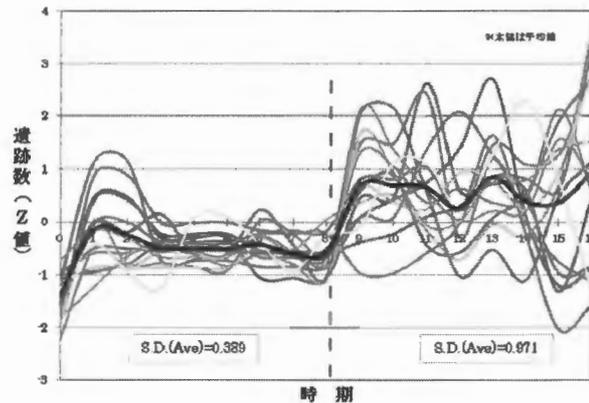


図2 遺跡数の変動

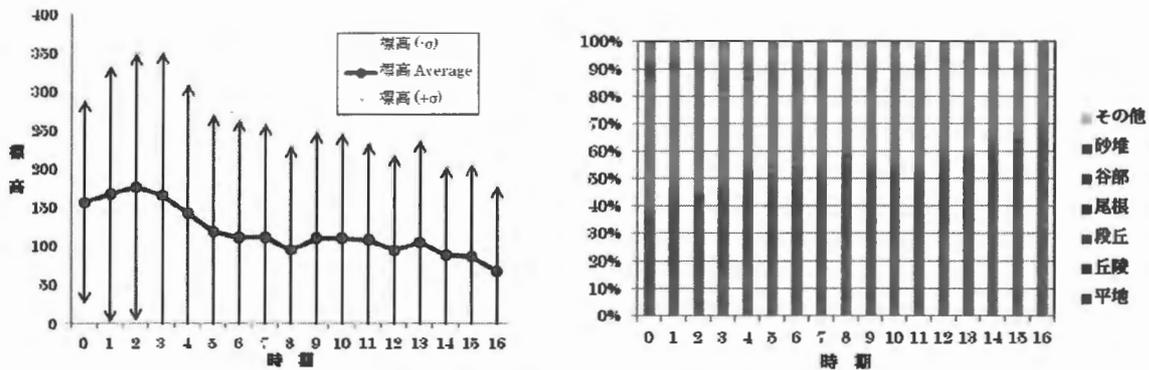


図3 遺跡立地標高と立地地形の変動

学的には無意味なものである。しかし、これらをS.D.も含めた時系列データとしてみれば、平均値の低下とS.D.の縮小を認めることができ、従って標高の低い場所とその周辺に遺跡が集まる全体としての動きを読み取ることができる。

また、大まかな立地データの時系列比率も、従来から沖積地への進出と形容されるように、これに適うようなあり方を示している。特に、平地とした沖積地と段丘上の立地の反比例は、立地平均標高の低下と極めて整合的なあり方を示していると考えられる。

遺跡立地場所の傾斜角はおおよそ4~6°であり、時間的な変化があまりみとめられなかったが、S.D.は時期が下るにつれて低くなる傾向が認められた(S.D. 7.54から3.77)。

(3) 小 結

以上、西日本における縄文~弥生時代の遺跡動態と立地の変動についてみてきた。これらをまとめると以下のようなになるだろう。すなわち、1期に遺跡が増加するもののそれらの利用地は標高の高い段丘上や丘陵上が主であった。ただし、S.D.も同時に高いことから、貝塚の形成に代表されるように海岸付近の平坦地や丘陵裾部などの利用も活発であったことが伺える。その後8期まで遺跡数に大きな変動はないが、立地標高とそのS.D.の低下、平地地利用の増大をみると、遺跡の立地は相対的に、高地部は減少傾向に、低地部は増加傾向を示すようなあり方へ変化しているといえよう。

大きな変化が認められるのは9期になってからである。西日本全体で遺跡数が増加するが、そのタイミングは地域ごとに異なるのは先に指摘した通りである。この時期になっても、遺跡立地標高とそのS.D.の低下、平地利用の増大は引き続き認められるので、遺跡数の爆発的増加は主に、低地部で起こったといえることができる。しかしながら、ピークは一時期のものであり、それが継続的に維持されているわけではない点には注意を要するが、それでも、8期以前の遺跡数にまで減少するわけではない。ピーク後に、緩やかに減少傾向を示しているとはいえ、確かな遺跡数増加が認められるのである。こうした増加現象の背後には、人口の(流入とは必ずしもいえないが)増加が想定されるが、一方で居住形態の問題とも関係する。この点については、次の分析を踏まえた上で後述したい。

4. 縄文・弥生時代遺跡の時空間動態解析

前節では、非常にマクロな動態を示した。では次に、遺跡レベル、すなわち、どのような遺構が時間的・空間的に展開し、遺跡を形成しているのか、についてみ

ていきたい。しかし、多種類の遺構で構成されている遺跡はその種類の数だけデータの特徴が異なるため、一般化することは容易ではない。ここでは、各遺構を変数として時期別にその検出数を集計し、主成分分析を行った。主成分分析とは、多次元データを情報の損失を最小限にする形で合成し低次元に要約する(新しい説明変数/主成分をつくる)多変量解析の一手法である。

ここで、この主成分が何を表現しているのかについて、明らかにしておきたい。縄文時代遺跡の遺構構成に関する多次元データの主成分は、遺跡全体の特徴を最もよく表現する合成変数となる。もとのデータはある時間的順序をもって集計された遺構(集落の構成要素)であるから、主成分は、各遺構の時間的動態を考慮したものとなる。つまり、ここで解釈される主成分とは、遺構構成の動態すなわち「集落」の展開を方向づける根本的な原理とでも呼ぶべきものと理解でき、主成分分析はそれを抽出する方法と理解することができる。

まずは西日本全体の遺構を変数とした主成分分析を行った。各変数の固有値をみると(表2下)、第2主成分までで87.1%を説明することが可能である。また、固有ベクトルをみると(表2上)、第1主成分はすべて一値であり、寄与率71%である。溝状遺構や住居、水田遺構や土坑、墓などが高い値を示し、共同体的要素をもつもので構成されている。従って第1主成分は「属共同体(-)⇔属家族(+)」と解釈した。第2主成分は、寄与率16%であり、+値に土器埋設や灰、集積遺構、一値に環濠や貯蔵穴、墓などをとる。従って、第2主成分は「長期的生活志向(-)⇔短期的生活志向(+)」と解釈した。これら主成分から得た主成分得点の散布図が図4である。

前~中期にかけて家族的・短期的から長期的へ変動し、後期へ至ると共同体的・短期的特徴を示すようになる。後期は、短期と長期の揺り戻しのような小変動をしつつ、晩期に至ると次第に共同体的性質を強める。突帯文期(15期)には長期化を志向するようになり、弥生前期(16期)には共同体的・長期的志向が強く顕在化しているようにみえる。こうした8~9期における画期は、主成分得点をクラスター分析にかけた場合でも、明瞭に分類される結果となっている。

また、第1、第2主成分の固有値の値は、前者が後者を大きく上回っている。これは、縄文~弥生時代遺跡の時間的変化について、第1主成分である属共同体⇔属家族の特徴つまり、親族集団や他者との関係を重視しながら展開していったことを示していると考えられる。これは、集団の住む場を解析した結果であるか

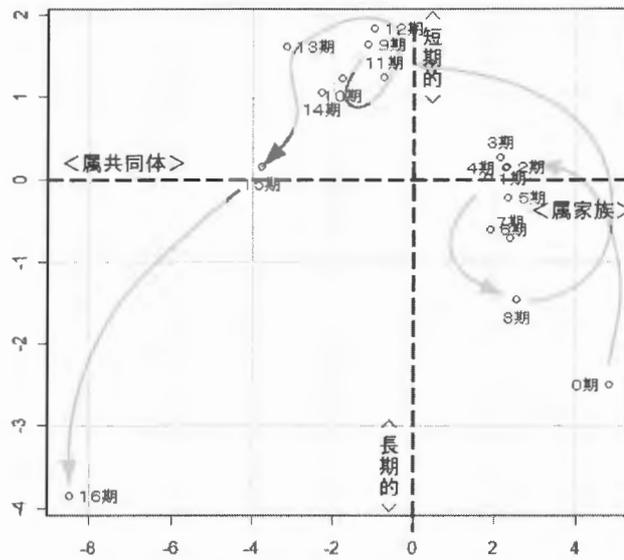


図4 主成分の解釈と得点のプロット (西日本全体)

表2 主成分分析 (西日本全体) の固有値と固有ベクトル

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
環濠	-0.255	-0.363	-0.024	0.266	0.038	0.373	0.330	0.130	-0.065	0.251
陥穴	-0.122	0.336	0.777	0.382	0.055	0.003	-0.127	0.202	0.168	0.035
建物	-0.296	-0.180	-0.083	0.121	0.344	0.302	-0.253	-0.292	0.295	0.388
溝状遺構	-0.310	-0.094	-0.046	-0.025	-0.045	-0.219	-0.415	-0.315	0.216	-0.126
集積遺構	-0.252	0.379	0.024	-0.155	0.356	-0.267	0.328	-0.349	-0.438	0.170
住居	-0.303	-0.016	-0.133	0.285	-0.031	-0.322	-0.383	0.096	-0.251	-0.234
水田遺構	-0.303	-0.095	-0.049	-0.281	0.204	-0.128	-0.124	0.095	0.121	-0.042
柱穴	-0.299	0.049	0.174	-0.168	-0.506	0.025	0.367	-0.423	0.365	-0.207
貯蔵穴	-0.280	-0.276	-0.021	0.246	-0.192	0.171	0.084	0.048	-0.365	-0.426
土器埋設	-0.059	0.500	-0.545	0.525	-0.008	-0.009	0.188	0.031	0.301	-0.053
土器溜まり	-0.286	0.161	-0.169	-0.346	-0.325	-0.064	-0.061	0.566	0.135	0.299
土坑	-0.310	0.025	0.061	0.108	-0.323	-0.117	0.025	-0.016	-0.318	0.459
墓	-0.299	-0.152	0.066	-0.084	0.440	-0.210	0.386	0.338	0.230	-0.285
炉	-0.225	0.429	-0.025	-0.284	0.094	0.662	-0.198	0.043	-0.188	-0.276

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
固有値	10.015	2.175	0.850	0.603	0.162	0.090	0.034	0.032	0.015	0.010
寄与率	0.715	0.155	0.061	0.043	0.012	0.006	0.002	0.002	0.001	0.001
累積寄与率	0.715	0.871	0.931	0.975	0.986	0.993	0.995	0.997	0.998	0.999

※Comp. 10までの表示にしてある

ら、ある種当然の結果であるといえるかもしれないが、それを拡大ないし強化していく歴史展開を抽出できたことは重要である。また、興味深いのは、共同体的性格の強化と居住期間の長期化が、正に相関しないという結果である。これは、縄文時代社会が、線形的ではなく、非線形的な歴史的展開として推移していたことを示すものであると考えられる。

では次に、西日本の中での地域差に目を向けてみたい。つまり、同様の手法を用いた場合、近畿・中国・九州地方が一様に推移したのか、もしくは地域差があったのか、についての検討である。ただし、先の主成分の解釈がそのまま適用できるわけではなく、再度データを見て解釈する必要がでてくる。

地域ごとに主成分分析を行ったのが表3・図5である。固有値をみると第2主成分までで約71%を説明す

ることが可能である(表3下)。各固有ベクトルをみると(表3上)、第1主成分はすべて一値であり、中でも土坑、水田遺構、土器溜まりなどが高い値を示すが、基本的にすべて高い値を示すのが特徴である。従って、ここでは第1主成分を「遺構構成の多様化(-)」と解釈する。第2主成分は、貯蔵穴、墓、住居が一値で高く、+値で陥し穴、炉、柱穴などが高い。これは一値は人口の規模を示す、+値は生業など一時的に利用されるもので構成されている。従って第2主成分は「相対的な人口密度(-)」と解釈した。

各地域において、この2軸上でやはり後期以降大きく方向性を異にしていることが読み取れよう(図4)。0~8期まではどの地域も大きな変化が見られず、遺構の構成や人口密度は同じような様相を呈している。しかし、9期以降、各地域ごとに異なる方向に進行し

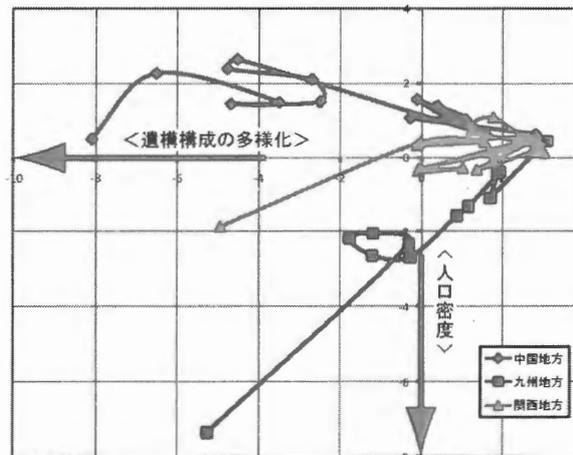


図5 主成分の解釈と得点のプロット (各地域)

表3 主成分分析 (各地域) の固有値と固有ベクトル

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
環濠	-0.257	-0.258	0.292	0.202	-0.345	-0.468	-0.064	-0.148	-0.011	0.442
陥穴	-0.213	0.337	-0.180	0.472	0.405	-0.003	0.202	-0.162	-0.365	0.087
建物	-0.316	0.121	-0.019	0.303	-0.344	-0.336	0.116	0.282	0.157	-0.241
溝状遺構	-0.317	-0.004	0.163	-0.472	-0.037	-0.116	0.241	0.069	0.054	-0.427
集積遺構	-0.224	0.265	-0.185	-0.006	-0.623	0.603	-0.004	-0.164	0.022	0.046
住居	-0.251	-0.394	-0.228	-0.009	-0.002	0.150	0.104	-0.239	0.120	0.300
水田遺構	-0.335	-0.065	0.051	-0.244	0.062	0.114	-0.065	0.566	-0.480	0.155
柱穴	-0.272	0.270	0.247	-0.285	0.294	-0.020	0.213	-0.335	0.412	0.138
貯蔵穴	-0.196	-0.422	0.259	0.290	0.053	0.188	-0.083	-0.339	-0.198	-0.586
土器埋設	-0.038	-0.257	-0.734	-0.260	0.008	-0.305	-0.110	-0.175	-0.033	-0.131
土器溜まり	-0.310	0.112	0.095	-0.100	0.178	0.043	-0.839	-0.085	0.048	0.047
土坑	-0.354	0.029	-0.065	-0.168	0.039	-0.036	0.251	-0.182	-0.345	0.108
墓	-0.225	-0.397	-0.080	0.204	0.281	0.321	0.134	0.385	0.429	0.106
炉	-0.278	0.297	-0.284	0.221	0.065	-0.143	-0.141	0.130	0.281	-0.189

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
固有値	7.268	2.706	1.281	0.831	0.572	0.473	0.315	0.231	0.112	0.082
寄与率	0.519	0.193	0.092	0.059	0.041	0.034	0.022	0.017	0.008	0.006
累積寄与率	0.519	0.712	0.804	0.863	0.904	0.938	0.960	0.977	0.985	0.991

※Comp. 10までの表示にしてある

ている様子が見て取れる。第2主成分とした人口密度は、九州が最も高く、ついで近畿・中国地方と、従来の認識との大きな齟齬はないものと考えられる。また第1主成分では、中国地方が弥生前期(16期)に最も多様化しているような結果がでたが、そのような状況は遺構論の立場からはあり得ないだろう。これはむしろ、縄文的要素の遺構と弥生的要素の遺構の混在状況を示しているものと考えられ、従って、他地域よりも「弥生化」しきれていない状況を反映しているのかもしれない。

近畿においては県単位のサンプルが少ないことによって縄文時代後期の様相が他地域と比べてはっきりとしないが、それでも、中国と九州の中間に位置してお

り、サンプル数が増えても、このあたりに分布するものと考えられる。

なお、こうした定量的な変動が、単位時間(絶対年代)当たりの変化量ではないことを批判する向きもあるかもしれない。各土器型式の時間幅は均一ではなく、従って本論で用いた大別時期も当然、均一ではありえない。例えば、後期の時間幅がそれ以前よりも長ければ、単位時間当たりの遺構検出数は少なくなるはずだから、よりスムーズな変化ないしは異なる解析結果がでるのではないかと、という批判である。しかし、各土器型式の継続時間は一部には公表されているものの、西日本においてはそのすべてがカバーされているわけではない。この問題を解決するには、なお一層の時間を要する。本論では従来の時間認識を用いて、定量的

な解析を行っているが、筆者は結果の基本的な傾向に変化はないものと考えている。なぜなら、縄文時代の時期区分は時期が古いほど、時間が長いことがC14年代法により明らかにされているからである。従って、縄文時代6期区分をさらにそれぞれ3時期区分したような多少荒い本論の時期区分を仮に単位時間当たりに変換したとしても、こうしたドラスティックな変化をより鮮明にこそすれ、スムーズな変化になることなどあり得ない。型式レベルの時間軸の設定を避けたのも、こうした理由が一部には存在しているからである。しかし、本論の時間軸が詳細な変動を一括したより概括的な時間軸であることに変わりはない。将来的には、細別型式レベルでの単位時間当たりの量的問題へと発展させなければならないことは論を待たない。

5. まとめ

本論では、西日本における縄文時代の時空間動態について、遺跡数や遺構構成におけるデータの振る舞い方からみてきた。すなわち、サンプルした標本の特定の側面についての議論ではなく、サンプルの全体的特徴についての議論である。従って、ここで述べてきた動態は、西日本縄文時代の集団全体の傾向として理解可能であろう。

ここで明らかにした動態は、縄文時代社会が家族的・短期的様相→共同体的・長期的様相へ線形的に移行するのではなく、共同体的・短期的様相を経た非線形的な推移として解釈できる。そしてこれと同時に、遺跡数と立地におおきな変化が認められた。この9期以降の状況は、従来の人口増の問題のみではなく、むしろ短期的生活志向によって見かけ上の大幅な遺跡数の増加や沖積地への進出が表出し、しかしそれらに伴ったであろう相対的人口増加、機能的な遺跡の顕在化などによる、相互依存ネットワーク（属共同体化）の強化によって特徴付けられる。その背後には生業や祭祀などの転換も起こったことが想定される。属共同体化については、後期（9期）以降に大規模祭祀空間の形成や集団狩猟の可能性が論じられるなど、遺構・遺物論からも認められるものである（稲田2007、岡田2005、山口2010など）。ただし、居住集団や世帯がどこまで独立していたのかについては別の問題であり、今後検討が必要であろう。

また、9期以降に西日本の各地域で異なる方向への遺跡展開が認められたが、この属共同体化の強化形態の差異が地域的多様性を促進させたことは間違いない。相互依存の強化から多様性の顕在化現象は、まさに生物進化における生態系の構築プロセスとほとんど同様のものである。すなわち、これはヒトの生存・文化シ

ステムの再構築に他ならない。こうした意味で、やはり後期以降は別世界といえよう。

そして、このようなデータの振る舞い方は、東日本のそれと比較的に通ったものとして理解することが可能なようにも感じる（津村2002）。また、中期における環状集落の盛行と後期における解体は、広く知られている現象である。縄文時代社会の東西問題を考える上で、個別としては相当に異なったものとして描けたとしても、全体としては通底している、ということは大変興味深い現象である。

本論で試みたような遺跡全体の特徴の平均的な状態を描くことを通じ、再び個別事象にフィードバックすることで、“どのように”異なるのか、をなるべく客観的に観測できるものとする。今後は、より時空間の幅を広げるとともに、ミクロな展開についても分析していくことが課題として挙げられる。

謝辞

本論は松本直子・岡山大学大学院准教授に交付された「科学研究費補助金（B）縄文・弥生社会の人口シミュレーションと文化変化モデル（課題番号：20320123）」の成果を含むものである。松本直子先生には、データの使用・分析を許可していただいた。また、同志社大学の津村宏臣先生からは、本論を執筆するに当たり多大なご教示を賜った。この場を借りて、深くお礼申し上げる次第である。

参考文献

- 板倉有太 2010 「九州地方の縄文集落」 『シンポジウム記録』7 考古学研究会
- 稲田陽介 2007 「山陰地方における縄文時代後期の石器製作技術構造」 『島根考古学会誌』24 島根考古学会
- 稲田陽介 2010 「山陰地方の縄文後・晩期集落」 『シンポジウム記録』7 考古学研究会
- 大野薫 2001 「近畿・中国・四国地方における集落変遷の面期と研究の現状」 『縄文時代集落研究の現段階』 縄文時代文化研究会
- 岡田憲一 2005 「大規模祭祀空間の形成—近畿地方における縄文時代後晩期集落のあり方—」 『関西縄文時代における石器・集落の諸様相 関西縄文論集2』 六一書房
- 関西縄文文化研究会 1999 『関西の縄文住居』
- 九州縄文研究会 2000 『九州の縄文住居』
- 縄文時代文化研究会 1999 『縄文時代』 10

- 縄文時代文化研究会 2001 『縄文時代集落研究の現段階』
- 瀬口眞司 2003 「関西縄文社会とその生業—生業=居住戦略の推移とそれに伴う諸変化—」『考古学研究』50-2 考古学研究会
- 瀬口眞司 2010 「関西地方の縄文集落」『シンポジウム記録』7 考古学研究会
- 津村宏臣 2002 『青森県縄文時代遺跡の遺跡空間データベースの構築と空間分析』（博士論文）総合研究大学院大学
- 富井 眞 2000 「近畿・中国・四国地域の縄文住居」『九州の縄文住居』九州縄文研究会
- 富井 眞 2002 「中四国地方縄文住居の先史学的研究に向けて」『四国とその周辺の考古学』犬飼徹夫先生古稀記念論文集刊行会
- 矢野健一 2001 「西日本の縄文集落」『立命館大学考古学論集』II 立命館大学考古学論集刊行会
- 矢野健一 2004 「西日本における縄文時代住居址数の増減」『文化の多様性と比較考古学』考古学研究会
- 山内清男 1964 「日本先史時代概説」『日本原始美術第1巻 縄文式土器』講談社
- 山口雄治 2008 「中国地方縄文時代中・後期の居住形態」『考古学研究』54-4 考古学研究会
- 山口雄治 2009 「中部瀬戸内における縄文時代後期の生業と集団関係に関する一試論」『日々の考古学2』六一書房
- 山口雄治 2010 「山陽地方における縄文時代後・晩期の集落」『シンポジウム記録』7 考古学研究会
- 山口雄治・津村宏臣・松本直子 2011 「西日本における縄文時代遺跡の時空間動態解析」『第77回日本考古学協会発表要旨集』日本考古学協会
- 山田康弘 2002 「中国地方の縄文集落」『島根考古学会誌』19 島根県考古学

箸墓古墳実測図に関する一考察

A Consideration on the Contour Map of Hashihaka Tomb Mound

小沢 一雅
Kazumasa Ozawa

大阪電気通信大学, 寝屋川市
Osaka Electro-Communication University, Neyagawa, Osaka

あらまし: 奈良県桜井市にある箸墓古墳は、最古級の前方後円墳であって、日本古代を考える上できわめて重要なモニュメントである。近年、邪馬台国畿内説の中で女王卑弥呼の墓とみる見解もあり、マスメディアの注目を集めてもいる。近年、宮内庁書陵部が管理する同古墳の墳丘実測図の精密な複製版が出版公開された。本稿では、この実測図を電子化した上で、等高線形状を逐一厳密に検討し、目視によって不規則な変動部を検出する。それが人為的な攪乱によると判断される場合には、経験則によって修復する。こうした修復処理を施した等高線形状を基礎にして、墳丘の規模（墳丘長・後円部径・くびれ部幅・前方部幅）を推測するとともに、墳丘体積の推計を行う。

Summary: Hashihaka Tomb, located in Nara Prefecture, is estimated as one of the oldest ancient Keyhole tomb mounds, which is playing a very important role in considering Japanese ancient regime. Recently a number of archaeologists argue that it should be identified with the tomb of Himiko who appeared in an ancient Chinese document as a leading Queen of Japanese ancient regime in the 3rd century. A high-definition contour map of Hashihaka tomb mound has recently been published, of which original one has been kept in the Household Agency since the tomb has officially been designated a mausoleum of an ancestor of the imperial family. In this paper, definitive dimensions of the tomb mound such as diameter of its round-back and so on have been estimated based on a restored contour map obtained by the following procedure: First, the published paper-printed contour map has been transformed into an electronic version given by a set of contours. Next, inspecting contours one by one, irregular parts due to artificial disturbance have been detected and restored from the sense of visual impression.

キーワード: 箸墓古墳, 等高線実測図, 復元, 墳丘規模

Keywords: Hashihaka tomb mound, Contour map, Restoration, Dimensions of the tomb mound

1. はじめに

前方後円墳の形態研究における基礎データは、墳丘の実測図である。筆者が現在公開中の前方後円墳データベースによれば、前方後円墳は、北海道と沖縄を除く全国で約5千基が遺存し

ている[1]。これらの古墳がすべて本格的に実測（実地測量）されているわけではないし、きちんと実測され実測図ができていない場合でも、その実測図が簡潔な利用環境の下で公開されているわけではない。推測であるが、研究者が簡潔な利用環境の下でアクセスできる良質な実測図は300枚を超えないと思われる（書店を通じての入手、もしくは大規模図書館における資料閲覧などを想定した場合）。

ねらう実測図に到達しえたとしても、形態研究の視点からすれば、どんな実測図にも共通する根源的な問題にまず直面する。実測図とは、遺存している墳丘の現状（現在形）を忠実に等高線形式で図化したものにすぎないからである。すなわち、形態研究において対象とすべきは築造時の原形であって、築造以来千数百年にわたる自然由来の経年変化、あるいは過去の人為的な攪乱による変化を含んだ現在形ではない。たしかに、保存状況が良好で実測図から原形が容易に推測できる古墳もあるにはあるが、多くはそうではない。にもかかわらず、多くの前方後円墳研究において、こうした根源的な問題を正面からとりあげようとする姿勢がきわめて希薄であったことは否定できない。

奈良県桜井市にある箸墓古墳は、邪馬台国問題を含め日本古代の実像を解明するにあたってのメルクマールとして近年とみにその重要性が高まっている。箸墓古墳は宮内庁が陵墓として管理していることもあって、墳丘への自由な立ち入りは禁止されている。発掘調査も行われていない。一方、墳丘の本格的な実測は第2次世界大戦前にすでに実施されており、精密な実測図も作製されていた。戦後になってから、この実測図の複製が書籍中の図版として公開されたが[2,3]、縮小率の関係で等高線の詳細を点検するにはやや難があった。しかし、近年になって縮小率が格段に改善された良質の複製版が出版されるに至った[4]。本稿では、この最新の複製実測図を基礎データとして用いている。

本稿の分析作業を大まかな流れとしてみれば、(1) 実測図等高線の電子化、(2) 等高線における不規則部の検出と修復、(3) 墳丘規模の推計、という3段階になっている。具体的な分析作業の概要と考察については、本稿の後半で述べる。

2. 箸墓古墳

畿内を圧倒的な中心として、前方後円墳が全国に分布している事実はだれの目にも明白である。前方後円墳の時代における中央政権が畿内にあったことはまちがいない。しかし、前方後円墳の時代がいつ頃にはじまったのかという、年代決定の問題は、厳密にはいまだに未解決である。昨今、前方後円墳の築造が3世紀中頃から開始されたにちがいないとする、考古学者主流の見解がマスコミをにぎわせている。いわゆる邪馬台国畿内説と連動する年代観である。

すなわち、3世紀中頃に卑弥呼が没した事実を魏志倭人伝が伝えているから、前方後円墳の時代が3世紀中頃にはじまったとするならば、邪馬台国は当然畿内でなければならない。魏に使者を派遣した卑弥呼は、中央政権（倭国）の女王だからである。

こうした年代観の根拠として注目されるのが箸墓古墳である。これまで、相対的な年代の順序関係の意味で箸墓古墳は“最古級”の前方後円墳として位置づけられてきた。そこで、もし箸墓古墳の絶対年代が確定できるならば、それは前方後円墳の時代の開始年代そのものにほかならない。箸墓古墳が注目される理由はここにある。

近年になって、①箸墓古墳の近隣地区の発掘で出土した土器の付着物の放射性炭素による年代測定（ ^{14}C 年代測定）によって、箸墓古墳が築造された時代の土器型式（布留0式）が“3世紀中頃”と判定されたとする新聞報道（2009年）、および②池上曾根遺跡出土のヒノキ柱根の年輪年代測定値（BC52年）から、従前AD100年頃とされてきた弥生時代後期の開始が100年余遡上し、これに連動して箸墓古墳の築造年代も3世紀中頃に位置づけられるとの新聞報道（1996年）など、科学的な年代「測定値」を根拠とする箸墓古墳3世紀説が喧伝されるに至った。

こうした一連の動きに対して、新たな批判が加えられつつある。たとえば、①については、考古学の専門学会において、データ操作が恣意的であるとの集中的な批判が加えられている。また、 ^{14}C 年代測定によってAD270年と判定された土器（布留1式）と共伴して木製の馬具（あぶみ）が箸墓近隣の調査区から出土した。東アジアにおける馬の使用は早くとも4世紀後半と考えられているから、もしこの年代測定値が正しければ、出土した馬具は中国を含めて東アジア最古の馬具ということになる。明らかな矛盾である。②については、過去に発表された年輪年代測定値の導出過程を綿密に追跡調査した結果、奈良時代以前の年輪年代「測定値」には、すべて100年の誤差があると指摘されている[5]。従前の年代観がむしろ正しいという結論である。

年代「測定値」の過誤がこのように明らかになりつつあっても、それを根拠にいったん喧伝された箸墓古墳3世紀説という、まさに結論だけがひとり歩きする状況がいまなお続いている。もちろん、こうした年代「測定値」によってではなく、三角縁神獣鏡やその他の物証についての旧来の解釈と推論だけに依拠して箸墓古墳3世紀説を唱える立場もあるが、新たな物証の出現などによっていまや説得力は希薄になったといわざるをえない。

近年の、突発的に盛り上がった年代錯誤の動向は一応しておくとしても、最古級の前方後円墳である箸墓古墳の年代決定は、日本古代の実像を解明する上できわめて重要な意味をもつことには変わりはない。以下では、箸墓古墳の年代を直接云々する前段階として、精密な実測図を基礎データとして箸墓古墳の規模（墳丘長・後円部径・くびれ部幅・前方部幅などの部位の大きさ、および墳丘体積）をできる限り正確に把握することを目標として行った基礎的な分析と考察について述べる。墳丘の規模は、築造にかかる労働量に比例することから、築造者（首長）の支配人口に関係すると考えられるからである[6]。大首長の支配人口は、古代人口論にしたがって推測可能と考えられるので、近い将来、この視点から箸墓古墳の年代を厳密に推定したいと考えている。

3. 箸墓古墳実測図の分析と考察

実測図の電子化

本稿で基礎データとする実測図は、前述のように、近年になって出版された宮内庁書陵部管理の実測図（原図）の良質な複製版である[4]。この複製実測図は、紙媒体（印刷体）の等高線図であるので、スキャナーを介してまず全体を画像形式（BMP）で収録する。これを等高線画像とよんでおく。

等高線画像から、おなじ高度の等高線1本ごとに座標点列として記述する。座標点列は、画

面上で1本の等高線を目視によって追跡しながら適当な間隔で曲線上の点をつぎつぎ選んでクリックしながら生成する(専用プログラムによる)。等高線は「種別」で分類される。種別は、後円部・前方部・基部の3種類である。「高度」によっても分類される。高度は、整数値(海拔メートル)である。1本の等高線は、(種別, 高度, {座標点列})なる組として定義することとし、全等高線をXMLファイル形式で一括保存する。すべての種別と高度について等高線を一齐に書き出したものを、電子化等高線図とよぶ。

図1に、等高線画像(a)と電子化等高線図(b)をしめしている。すべての等高線が適正に電子化されているかどうかについては、(a)と(b)の重ね合わせによって完全一致を確認している。なお、同図は左を北、上を東とした墳丘の等高線図である。同図(b)の電子化等高線図において、(a)にはみられない、ほぼ平行な3本の直線が斜めに走っている。原因は、基部(基底部)にあたる最下位3本の等高線が(a)で開曲線となっているからである。箸墓古墳を囲む現在の地面が、本来あった周壕を埋める形で形成されていった際に、墳丘に対して傾斜が生じたためであろう。

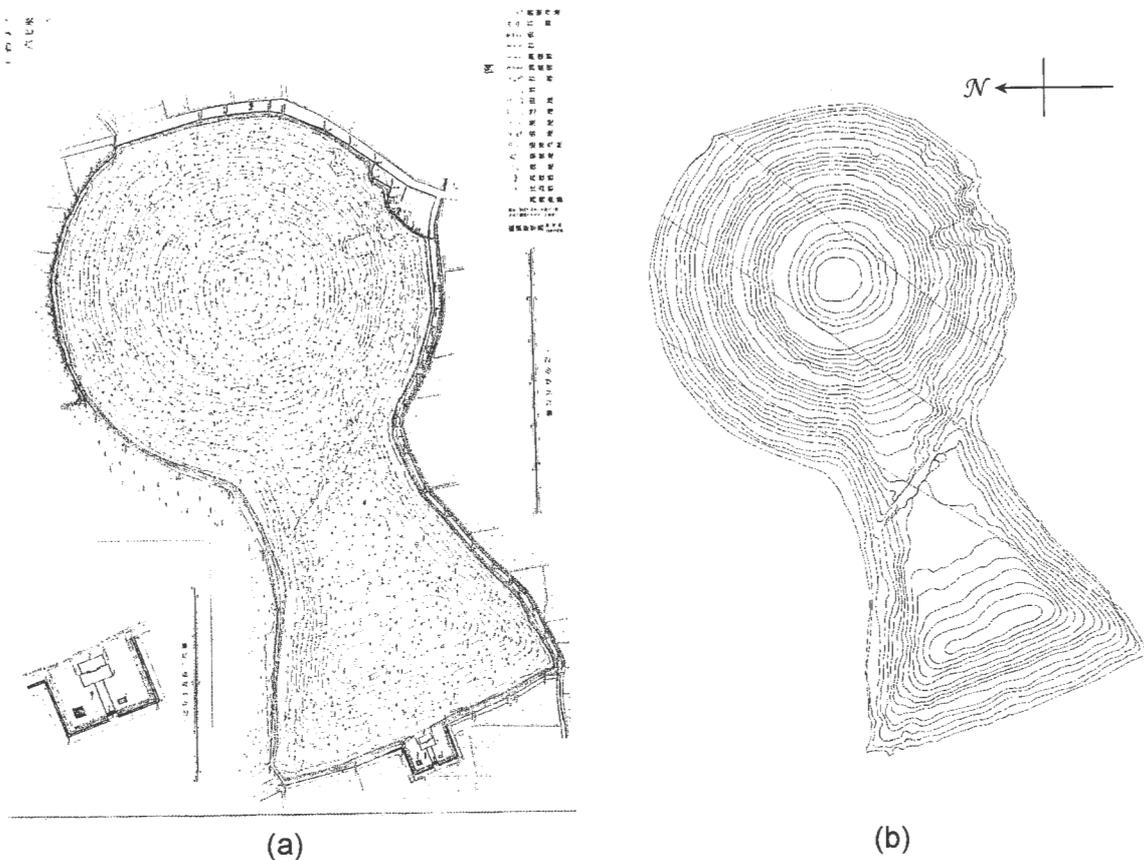


図1 実測図(a)と電子化等高線図(b)

等高線の点検と修復作業

図1(b)にしめされる墳丘等高線において、数個所に等高線の不規則な変動が目視できる。くびれ部付近にあるかなり激しい乱れは、箸墓古墳がまだ陵墓として管理されてはいなかった

江戸時代に人々が墳丘をまたいで通行するために日常的に利用していた通路の跡とみられている。また、後円部の東南部の基部にも乱れが目視できる。こうした明らかに過去の人為的な攪乱によって生じた乱れのほかに、築造後千数百年という長い時間の流れの中で草木の繁茂や枯死、あるいは無数にくり返された風雨の影響などによる封土のゆるやかな自然的変動も当然加わっているはずである。

こうしたすべての変動を修復し、築造時の原形にもどすことができれば理想的であるが、そのためには墳丘の全面発掘など、相当大がかりな対策によって得られる知見が必要になるであろう。発掘という行為は、情報の不可逆的な破壊や損傷という一面をあわせもっており、決して副作用のない「良薬」ではない。形態研究にとって必要なことは、墳丘の規模を数値的に推計することであって、必ずしも完全な復元をやらなければならないわけではない。本稿では、箸墓古墳の実測図で目視できる等高線の人為的な乱れだけに限定して修復し、上述のゆるやかな自然的変動は対象外とする“穏健な”レベルでの復元をめざすことにした。

等高線修復の対象として、図1 (b) をもとに以下の3箇所をとりあげる。

- (1) くびれ部付近で目視される近世までの“通路”の跡とみられる等高線の乱れ
- (2) 後円部東南部における墳丘部への人為的な掘り込みをしめす等高線の乱れ
- (3) 墳丘を囲む地面の傾斜に起因する最下位3本の等高線の不規則性

実測図の電子化に際して、上記の問題箇所を意識しつつ、等高線1本ずつ慎重に点検を行いながら修復作業を進めた。修復は原データにある座標点列の修正ということになるが、作業はすべてモニター画面上に等高線を表示し、目視によって試行錯誤的に進めた。

上記3箇所のうち、(1)と(2)については、平面企画の意味でほとんどが墳丘の内部における等高線の乱れであって、後述するように墳丘規模の推計への影響は比較的すくないが、(3)は墳丘を囲む地面に「埋もれた」等高線を復元することになり、墳丘規模の推計に直接影響を及ぼすことになる。したがって、その修復はとりわけ慎重に進める必要があった。すなわち、「開かれた」等高線の消失部分を復元してぐるっと周回する閉曲線へと修復することによって、後円部東南部端および東端周辺に埋もれて消失していた等高線が新たに出現することになり、墳丘長や後円部径などの墳丘規模にかかわる数値に直接的な変化をもたらす可能性が出てくるからである。

前述の手順にしたがって修復した等高線図(等高線の一括表示)を下に掲げる図2にしめす。同図の縮尺は、原図にある縮尺表記のうち、メートル法による部分を抽出して転記したものである。

墳丘規模の推計 (平面企画4部位)

図2にしめされる修復された等高線データにもとづいて、箸墓古墳の墳丘規模の推計を行う。まずは、形態研究で不可欠な平面企画における4部位(墳丘長・後円部径・くびれ部幅・前方部幅)の計測である。4部位の計測については、すでに開発済みの古墳計測支援システム(「古墳測りま専科」)を用いることとした[7]。計測結果を表1にしめす。同表には、筆者による従前の計測結果[6]も対比させている。後円部径以外は、従前の計測結果と比べて大きな差異はみられない。後円部径の推計においては、一般に後円部等高線と真円との仮想的なマッチング

が基本になるが、従前の推計においては、原図における後円部等高線の外周（東南部）にみられる不規則な等高線形状に影響され、やや小さめの後円部径を推計値として選んだことが原因と考えられる。

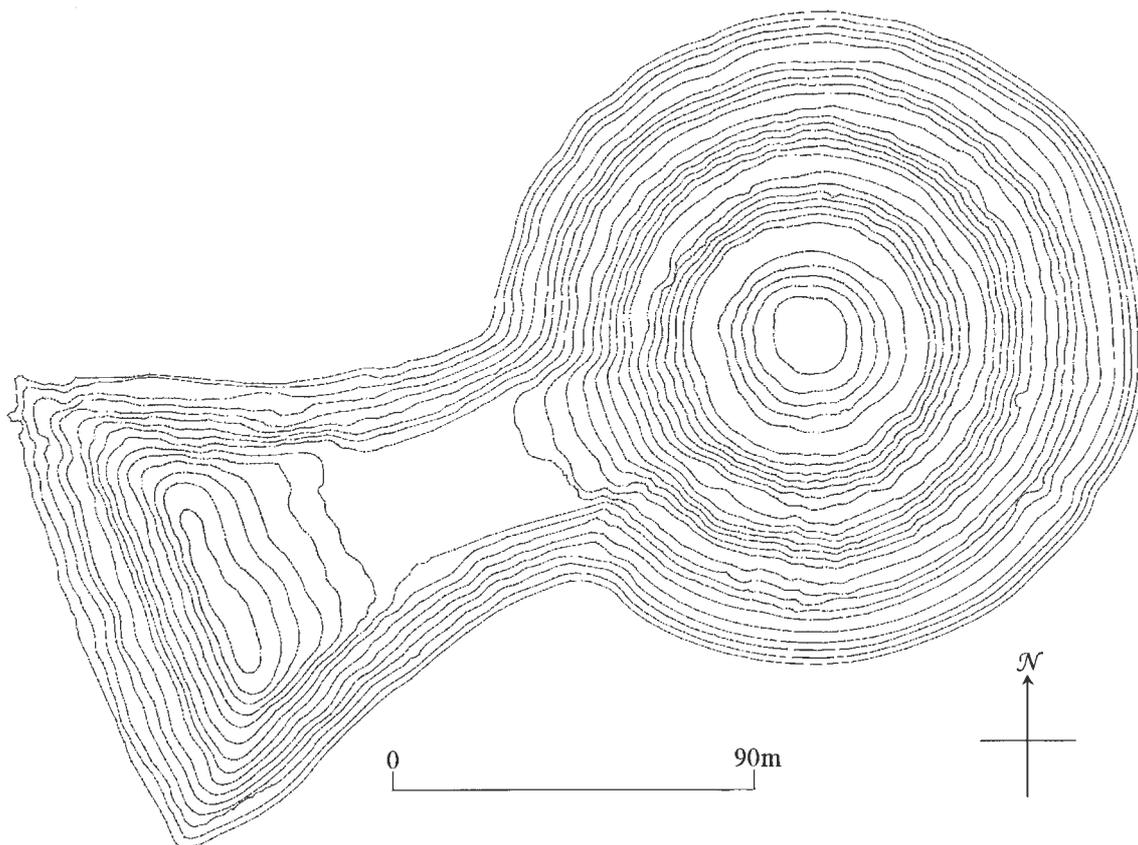


図2 修復等高線図（小沢一雅；2011）

表1 平面企画4部位の計測値

	墳丘長 (m)	後円部径 (m)	くびれ部幅 (m)	前方部幅 (m)
新推計値	281	163	64	132
旧推定値	280	161	64	132

墳丘規模の推計（墳丘体積の算定）

箸墓古墳の墳丘体積に関して、これまでに実際に推計作業を行った上で公開された推計値は筆者の知るかぎり、3例しかない。年代順に言えば、①筆者による古い推計値（1980）[8]、②大林組による推計値、③石川昇氏による推計値（1989）[9]、の3例である。ただし、②の推計値はNHK特集『巨大古墳の謎』（1987年）の番組中に紹介された可能性があるものの、数値の詳細は不明である。印刷体資料で確実に参照できる推計値は、①と③のみである。それぞれの体積値は、後にしめす表2に記載している。

筆者の古い推計値①は、形態研究の初期段階で試みた墳丘の3次元マッチングに用いるため

に作成した数十基のデジタル古墳データをもとに算出した数値であって、あくまで個々の墳丘体積の概数を与えたものにすぎない。当該論文にもこの旨の断り書きを入れている。石川昇氏の数値③は、氏の考案した簡易計算法にもとづいて手計算で算出されたものである。

さて、ここでは、図2にしめす修復等高線データにもとづく箸墓古墳の墳丘体積の厳密な計算手順の概要と計算結果について紹介する。

等高線にもとづく体積計算の基本は、まず各等高線が囲む領域の面積計算である。図2の等高線図では、最下位等高線の海拔高度は74mであり、そこから上位に向かって順次1mおきに等高線が描かれ、最高位等高線高度（後円部中央）は102mである。等高線の総数は29本である。そこで最下位の等高線が囲む面積を S_1 とし、つぎの等高線が囲む面積を S_2 、つぎを S_3 、・・・のように表していくと最高位は S_{29} となる。ただし、計算理論上、仮想の最高位（面積ゼロ） S_{30} をさらに追加設定している（ $S_{30}=0$ ）。

もっとも簡単な方法（台状計算法）は、各等高線が囲む面を底面とした高さ1mの「台」が積み上がって墳丘が形成されているとする近似計算法である。計算は簡単で、墳丘体積を V とすると、つぎの式で計算できる。

$$V = S_1 \times 1 + S_2 \times 1 + \dots + S_{29} \times 1 + S_{30} \times 1 \quad (\text{“} \times 1 \text{” は台の高さの乗算})$$

$$= S_1 + S_2 + \dots + S_{29} \quad (\text{立方メートル； m}^3)$$

等高線はたしかに1m間隔であるが、墳丘の実態は1m間隔で台状に積み上がった形状をしているわけではなく、自然的な経年変化もあってなめらかな斜面になっている。この点を考慮した計算法（直線補間法）の1つは、台状計算における「台状」を緩和するために直線的斜面に置き換える計算法である。おなじく、もっとなめらかな斜面を追求した計算法（シンプソン公式による補間計算法）もある。数式はやや複雑なので省略する。

表2に今回の計算結果（台状計算法・直線補間法・シンプソン法）および過去の2例の数値を対比的にしめしている（有効数字2桁として表示）。

表2 箸墓古墳の墳丘体積計算値一覧

台状計算法	直線補間法	シンプソン法	小沢一雅 (1980)	石川昇 (1989)
37万 m ³	35万 m ³	33万 m ³	37万 m ³	30万 m ³

誤差にかかわる一点について付言しておく。上述のように、29本の等高線が囲む面積が体積計算の基礎になっている。台状計算を例にとると、もっとも最上位の等高線がつくる「台」の体積（ $S_{29} \times 1$ ）は、わずか300m³（0.03万m³）である。その一段下位の台（ $S_{28} \times 1$ ）でも607m³にすぎない。墳丘体積という量の視点でいうと、最上位から数段の等高線は、あってもなくても大勢にほとんど影響がない存在ともいえるのである。

台状計算法による計算値37万m³は、筆者の古い数値と一致しているが、これは完全な偶然であって、両者はデータ形式もちがうし、計算法もまったく異なる。表2にある5つの数値

のうちどれがもっとも信頼度が高いかといえば、筆者は今回の計算におけるシンプソン法による計算値33万 m^3 を採用すべきと考えている。誤差にかかわる直感的な判断としていえば、仮に箸墓古墳が原形に完全復元されたとしても、その墳丘体積は33万 m^3 からほとんど動かないとみてよいだろう。

4. むすび

本稿では、箸墓古墳という、日本古代を考える上で重要な意味をもつ前方後円墳をとりあげた。とくに、その墳丘についての形態研究の一環として実測図等高線の数理的分析を行った。陵墓としての箸墓古墳に関する考古学的な学術情報は、かつて墳丘表面で検出された埴輪片、あるいは周壕と目される区域の発掘調査から検出された土器などの遺物群、および宮内庁所管の墳丘実測図である。考古学は「宝探しの発掘」からはじまった、とエドワード・ハリス[10]はいうが、たしかにどちらかといえば、新たな遺物発見をねらう発掘の方向に関心が集中する傾向がある。いかに精密であっても、印刷物としてすでに公開されてしまっている墳丘実測図にいまさら大きな関心をよせる空気は希薄、というのが本当のところであろう。

当該墳丘実測図は、箸墓古墳についてのきわめて品質の高い考古学データにほかならず、しっかりした分析をやれば、貴重な知見が得られる可能性が大きいはずである。そういう意図と動機の下に、ひとつの調査研究を企画した。本稿は、その第一歩をふみ出すための準備作業を述べたことにもなる。作業の過程で得られた墳丘規模に関する新しい知見を紹介したが、今後は、これをベースにさらに考察を進めていきたいと考えている。

最後に、等高線電子化作業を補助してくれた奥島健一君に感謝する。

【参考資料】

- [1]前方後円墳データベース (<http://www3.kcn.ne.jp/~yuka-o/kofun/>)
- [2]末永雅雄『日本の古墳』朝日新聞社 (1961)
- [3]末永雅雄『古墳の航空大観』学生社 (1972)
- [4]宮内庁書陵部陵墓課編『宮内庁書陵部陵墓地形図集成』学生社 (1999)
- [5]鷲崎弘朋「木材の年輪年代法の問題点—古代史との関連について」『東アジアの古代文化』136号 (2008)
- [6]小沢一雅『卑弥呼は前方後円墳に葬られたか—邪馬台国の数理』雄山閣 (2009)
- [7]上月誉士・小沢一雅「前方後円墳の墳形計測システム (第2報)」『情報処理学会研究報告』2004-CH-64, pp25-32 (2004)
- [8]小沢一雅「前方後円墳の形態研究とその計数的方法の試み」『考古学研究』25巻2号 (1980)
- [9]石川昇『前方後円墳築造の研究』六興出版 (1989)
- [10]エドワード・ハリス (小沢一雅訳)『考古学における層位学入門』雄山閣 (1995)

■第1回公開シンポジウム「人文科学とデータベース」

1995年12月25日(月曜日)、26日(火曜日) 大阪電気通信大学(寝屋川キャンパス)

特別講演

古地震データと活断層(pp.1-4)

寒川旭(通産省地質調査所)

一般講演

Intelligent Pad システムを用いた歴史学研究支援データベースの構築(pp.5-12)

赤石美奈・中谷広正・伊東幸宏・阿部圭一・田村貞雄(静岡大学)

4次元歴史空間システムにおける地理情報処理について(pp.13-18)

小林努・加藤常員・小沢一雅(大阪電気通信大学)

視点に依存する属性付け機構をもつ木簡研究支援システム

ー構造化型データベースの概念ー(pp.19-28)

森下淳也(姫路獨協大学)・上島紳一(関西大学)・大月一弘(神戸大学)

古典籍と JIS 漢字

ーテキストの本文校正とのかんけいについてー(pp.29-36)

當山日出夫(花園大学)

手書き文字時系列筆跡パターンの一解析と今後の計画(pp.37-42)

東山孝生・山中由紀子・澤田紳一・中川正樹(東京農工大学)

絵画 DB とイメージ検索

ー浮世絵の線画表現とデータ圧縮効果ー(pp.43-48)

濱裕光・志賀直人(大阪市立大学)

画像データベースの自然言語インターフェースについて(pp.49-54)

伊東幸宏・中谷広正(静岡大学)

多視点距離データを用いた3次元形状モデリング(pp.55-60)

横矢直和(奈良先端科学技術大学院大学)・増田健(電子技術総合研究所)

ハイパーメディア・コーパスの構築と言語教育への応用について(pp.61-66)

上村隆一(福岡工業大学)

「歌物語」語彙の数量的分析と研究(pp.67-74)

西端幸雄(大阪樟蔭女子大学)

高次辞書データベースのための語彙知識自動獲得システム(pp.75-82)

亀田弘之(東京工科大学)・藤崎博也(東京理科大学)

社会調査結果の視覚化データベース(pp.83-88)

吉田光雄(大阪大学)

「間」に関するデータベースの構築(pp.89-98)

中村敏枝(大阪大学)

方言音声データベースの作成と利用に関する研究(pp.99-104)

田原広史・江川清・杉藤美代子・板橋秀一(大阪樟蔭女子大学)

■第2回公開シンポジウム「人文科学とデータベース」プログラム
1996年12月21日(土曜日) 大阪電気通信大学(寝屋川キャンパス)

招待講演

三浦梅園の主著『玄語』のデータベース化と解読の試み
ー江戸時代のハイパーテキストを読み解くー(pp.1-10)
赤星哲也(日本文理大学)・北林達也(三浦梅園研究所)

江戸図データベースの作成と今後の課題(pp.11-22)
黒川隆夫(京都工芸繊維大学)

一般講演

Shape from motion を応用した什器類の立体データ作成(pp.23-34)
中島重義・岡本次郎・濱裕光・細川省一(大阪市立大学)

古地図に描かれた内容のデータベース化の試み(pp.35-44)
出田和久・正木久仁・小方登・山近博義(奈良女子大学)

考古学のためのデータベースシステム(pp.45-54)
宝珍輝尚・中田充・白井治彦・都司達夫(福井大学)

インターネット・イントラネットにまたがる分散型図書館目録データベースの構築と運用
(pp.55-56)
芝勝徳(神戸市外国語大学)

音楽における印象語検索システムの開発とその有用性(pp.57-66)
原田章・吉田光雄(大阪大学)

『方言認知地図』プログラムと統計処理地図(pp.67-78)
ダニエル・ロング(大阪樟蔭女子大学)

■第3回公開シンポジウム「人文科学とデータベース」プログラム
1997年12月20日(土曜日) 大阪電気通信大学(寝屋川キャンパス)

招待講演

「邪馬台国大和説を科学する」

倭人社会と卑弥呼の王権(pp.1-10)

吉田晶(岡山大学名誉教授)

前方後円墳成立の歴史的意義

ー『邪馬台国時代』における首長と農民層の共同幻想ー(pp.11-17)

広瀬和雄(奈良女子大学)

一般講演

歴史学研究支援システムの構築(pp.19-30)

三浦崇・伊東幸宏・小西達裕・田村貞雄(静岡大学)・赤石美奈(北海道大学)

中谷広正・阿部圭一(静岡大学)

地理情報を利用した遺物データベースシステムについてー(pp.31-38)

宝珍輝尚・都司達夫(福井大学)・河合秀夫(大阪電気通信大学)

木簡研究支援データベースとシステム

ー知見と仮説に基づく再構造化ー(pp.39-46)

森下淳也・大月一弘(神戸大学)・上島紳一(関西大学)・大庭脩(皇學館大学)

杉山武司(姫路獨協大学)

短編推理小説の論理構造の分析(pp.47-54)

西島恵介・神山文子・藤田米春(大分大学)

形状分析ツールの開発とその応用

ー浮世絵に描かれた役者の同定と分類ー(pp.55-64)

モハメド・アミラン・ブイヤン・阿古弥寿章・濱裕光(大阪市立大学)・松平進(甲南女子大学)

技術紹介

3次元形状入力へのおさそい

ー人文科学の道具としてー(pp.65-72)

濱裕光(大阪市立大学)

■第4回公開シンポジウム「人文科学とデータベース」プログラム
1998年12月9日(土曜日) 大阪電気通信大学(寝屋川キャンパス)

招待講演

邪馬台国は北九州甘木市付近にあった(pp.1-22)
安本美典(産業能率大学)

特別講演

選挙研究とデータベース(pp.23-24)
三宅一郎(関西大学)

一般講演

中期インド・アリアン聖典のデータベース(pp.25-34)
逢坂雄美・山崎守一(仙台電波工業高等専門学校)・宮尾正大(室蘭工業大学)

高地性集落遺跡データベースからみた弥生時代の情報通信(pp.35-42)
加藤常員(大阪電気通信大学)

データベース倫理について(pp.43-52)
江澤義典(関西大学)

階層構造グラフによるデータモデルの適用例：木簡データベース(pp.53-59)
杉山武司(姫路獨協大学)・森下淳也・大月一弘(神戸大学)・上島紳一(関西大学)

遺物破片の計測
－照度差ステレオ計測装置の製作と計測－(pp.61-70)
結城宏和・宝珍輝尚・都司達夫(福井大学)・河合秀夫(大阪電気通信大学)

技術紹介

地場工芸品立体展示システム
－人文科学の道具として－(pp.71-78)
橋本隆之・青木功介・釣裕美(インテックシステム研究所)

■第5回公開シンポジウム「人文科学とデータベース」プログラム
1999年12月18日(土曜日) 関西大学総合情報学部

招待講演

顔の認知研究における顔データベースの利用(pp.1-10)
加藤隆(関西大学)

Recent Research on Paleolithic Arts in Europe and the Multimedia Database (pp.11-22)
Cesar Gonzalez, Roberto Cacho Toca (University of Cantabria, Spain)

一般講演

Photo VR 考古資料データベース『北スペインの旧石器洞窟美術』
ー日本・スペイン産学共同プロジェクトの実現と諸問題ー(pp.23-34)
深沢武雄(テクネ)

バーチャルリアリティによる遺物探訪 3D Archaeo-Copter (pp.35-38)
中村健・小沢一雅(大阪電気通信大学)

照度差ステレオ法を用いた遺物の表裏形状の計測(pp.39-46)
結城宏和・宝珍輝尚・都司達夫(福井大学)

文字データベースのための文書の構造化と意味管理(pp.47-58)
横田一正・三宅忠明・国島丈生(岡山県立大学)
劉渤江(岡山理科大学)・田槇明子(リョービスシステムサービス)

相対インデックス法を使った文構造分析(pp.59-66)
雄山真弓・岡田孝・黒崎茂樹(関西学院大学)

仮想電子辞書システムの設計と構築(pp.67-78)
芳野学・都司達夫・宝珍輝尚(福井大学)

■第6回公開シンポジウム「人文科学とデータベース」プログラム
2000年12月20日(土曜日) 静岡大学情報学部

特別講演

「ええじゃないか」研究を振り返って(pp.1-2)
田村貞雄(静岡大学)

一般講演

文化資源情報における Topic Map の適用と評価
ー『源氏物語』画像 DB を使ったモデル作成の試みー(pp.3-10)
長瀬真理(静岡大学)

文字冗長度による日本文学分析(pp.11-16)
福田宏(静岡県立大学)・山下泰弘(科学技術振興事業団)・勝矢光昭(静岡県立大学)

広領域分野資料の横断的アーカイブ論に関する基礎研究(pp.17-26)
八重樫純樹(静岡大学)

ジャイナ教聖典のデータベース
ー特殊フォントで表現されたデータの PDFー(pp.27-34)
逢坂雄美(仙台電波工業高等専門学校)

パスカルデータベースシステム (1) (pp.35-42)
白石修二(福岡大学)

吾妻鏡データベースの構築(pp.43-52)
安道百合子(国文学研究資料館)

全国遺跡データベースの構築(pp.53-62)
森本晋(奈良国立文化財研究所)

地理情報システムを用いた城下町の復元的研究
ー彦根城下善利組足軽屋敷地図を中心としてー(pp.53-72)
生方美菜子・濱崎一志(滋賀県立大学)

考古学データベースにおける検索エンジンの研究(pp.73-80)
三浦宙明・小沢一雅(大阪電気通信大学)

■第7回公開シンポジウム「人文科学とデータベース」プログラム
2001年11月17日(土曜日) 関西学院大学情報メディア教育センター

特別講演

感性情報研究の動向とデータベース(pp.3-10)
井口征士(大阪大学基礎工学部)

一般講演

広重の版画と江漢作らしい油絵
—二つの「東海道五十三次」について—(pp.11-27)
荒木啓介(科学技術振興事業団) 資料提供、大畠洋一(東海道研究家)

古文書文字列に対するキャラクタスポッティング(pp.29-38)
橋本智広・梅田三千雄(大阪電気通信大学)

日商簿記検定問題の電子化(pp.39-46)
福田宏・小津稚加子(静岡県立大学)

米国におけるテレビニュースデータベース構築の歴史的・法的経緯と現状について
(pp.47-50)
魚住真司(関西外国語大学)

画像・音声の超高再現性 PC の開発(pp.51-63)
片岡裕(大谷大学)

全国遺跡データベースの構築 2001年度の動向(pp.65-68)
森本晋(奈良文化財研究所)

日本語キエルケゴール文献データベース(pp.69-72)
平林孝裕・橋本淳(関西学院大学)

前方後円墳データベース検索システムと地理情報処理(pp.73-80)
西上昌治・小澤一雅(大阪電気通信大学)

貝類のマルチメディアデータベースの構築(pp.81-88)
高田茂樹・雄山真弓(関西学院大学)

■第8回公開シンポジウム「人文科学とデータベース」プログラム
2002年12月21日(土曜日) 帝塚山大学

特別講演

- ・空間コンテンツの検索とプレゼンテーション
田中克巳(京都大学)

自由論題報告

- ・尾張藩士・朝日文左衛門の生活行動空間
—GISを用いた『鸚鵡籠中記』の分析—
村田祐介(名古屋大学研究生)
- ・3次元空間を共有するマルチユーザCSCW環境と
仮想考古遺跡ウォークスルーへの応用
坂田義則(立命館大学院生)・八村広三郎(立命館大学)

SYMPOSIUM “人文科学における空間情報の利用”

I. 古地図・衛星画像の分析と利用

- ・高精細絵図画像データの利活用 —阿波国絵図・徳島城下絵図を例に—
平井松午(徳島大学)
- ・分散型GIS「GLOBALBASE」の実装
森洋久(国際日本文化研究センター)

II. 考古学におけるGIS利用

- ・古環境復原のための考古学情報クリアリングハウスの構築
—Java Script と Dynamic HTML を使用して—
河野一隆(九州国立博物館(仮称)設立準備室)・塚本敏夫(元興寺文化財研究所)・
魚津知克(大手前大学)
- ・平安京における空間情報システムの整備と条坊復原
宮原健吾(京都市埋蔵文化財研究所)・内田賢二(ライカジオシステムズ)
- ・中世都市の景観と構造
鋤柄俊夫(同志社大学)

■第9回公開シンポジウム「人文科学とデータベース」プログラム
2003年12月20日(土曜日) 大阪工業大学 情報科学部

特別講演 テーマ：著作権

- ・他人の著作権を侵さないために — 法律的視点からデジタル著作権
岡村久道(英知法律事務所 弁護士)
- ・自分の著作権を守るために — 技術的視点から
佐野睦夫(大阪工業大学)

一般講演

- ・コンピュータグラフィックスを用いた花型学習システム
三原比呂美, 西尾孝治, 小堀研一(大阪工業大学)
- ・デジタルアーカイブとデータベース構築
清水宏一, 山口豊博(京都デジタルアーカイブ研究センター)
- ・絵巻物を利用した心理検査支援システム
池田瑞穂, 雄山真弓(関西学院大学)
- ・浮世絵を通してみた江戸時代女性の人体表現について
森下あおい(成安造形大学), 黒川隆夫(京都工芸繊維大学大学院)
- ・縄文語による地名語源の解釈—山名の例を中心に—
永田良茂
- ・奈文研航空写真検索システム(NARS)について
森本晋(独立行政法人文化財研究所奈良文化財研究所)
- ・兼永本古事記・出雲国風土記データベースの構築
松本智子(国文学研究資料館研究情報部)
- ・心理学入門のWeb自習教材として項目反応理論を適用した項目プール(データベース)の開発
田崎美弥子(東京理科大学)
- ・感性に基づく動画検索について
井田俊博(福井大学大学院), 宝珍輝尚(大阪府立大学), 都司達夫, 樋口健(福井大学)

■第10回公開シンポジウム「人文科学とデータベース」プログラム
2004年12月18日(土曜日) 大阪府立大学

特別講演

考古学データベースとXML
八重樫純樹(静岡大学)

招待講演

中世の国際都市 堺
續伸一郎(堺市立埋蔵文化財センター)

一般講演

百名山からの縄文地名解釈
永田良茂

遺跡と遺物・遺構表示の階層性に関する一考察
宝珍輝尚(大阪府立大学)

考古学情報のXML記述とデータベースシステム～データ統合と横断検索へのXSLTの利用
小笠原和慶, 八重樫純樹(静岡大学)

柔軟な構造を持つデータベース管理システムを用いた万葉集検索システムの構築法
中田充(山口大学)

前方後円墳の築造と方位観念・社会的背景の復元に関するデータベースの活用
北條芳隆(東海大学)

前方後円墳の墳形計測と築造企画
小澤一雅(大阪電気通信大学)

■第11回公開シンポジウム「人文科学とデータベース」プログラム
2005年12月3日(土曜日) 大阪樟蔭女子大学

特別講演

古典芸能研究におけるデジタルアーカイブの効用
赤間 亮(立命館大学)

一般講演

言語地図作成方法の変遷 -GIS ソフトの導入と展開-
鳥谷善史・田原広史(大阪樟蔭女子大学)

立地分析のための分布地図表現 -分布地図作成支援システムの構築-
中島高司・加藤常員(大阪電気通信大学)

人文科学のための地理情報・資源共有化システム・プロジェクト
石川正敏(東京農工大学), 川西陽一(京都大学), 奥村英史(ヒューマンオーク), 原正一郎(国文学研究資料館), 桶谷猪久夫(大阪国際大学), 貴志俊彦(島根県立大学), 村尾義和(島根県立大学), 柴山 守(京都大学)

4D-GIS による遺跡空間データベースの構築と先史文化生態解析
津村宏臣(同志社大学)

国書古典籍中の挿絵・絵本に描かれた実在キャラクター達の存在意義 -情報学から文学論へのエチュードとして-
相田 満(国文学研究資料館)

浮世絵と写真に基づく幕末から明治に至る女性の形態の定量的考察
森下あおい(滋賀県立大学), 黒川隆夫(京都工芸繊維大学)

嵯峨本『伊勢物語』の木活字及び組版分析モデルに関する報告
津田光弘(イパレット/奈良女子大学), 鈴木広光(奈良女子大学)

色彩のデジタル化の諸問題 -日本文学作品画像データベースと色彩-
當山日出夫(花園大学)

万葉集校本データベースから古字書データベースへ
西端幸雄(大阪樟蔭女子大学), 鈴木榮一(万葉情報システム調査会)

古今集データベースによる歌語の視覚化
山元啓史(オーストラリア国立大学)

PDF ファイルを用いた歴史的資料のデジタル化とデータベースの試作
竹内さおり・白川哲郎(大阪樟蔭女子大学)

階層に基づく遺物データベースシステム
宝珍輝尚(大阪府立大学)

■第12回公開シンポジウム「人文科学とデータベース」プログラム
2006年12月23日(土曜日) 大阪大学

特別講演

時と場：デザインにとっての歴史と環境
藤田 治彦 (大阪大学)

一般講演

祇園を祇園と書くのは誤字か -浮世絵アーカイブを資料とした検証-
當山日出夫 (花園大学)

中期インド・アリアン聖典のデータベース II
逢坂雄美 (仙台電波工業高等専門学校)

脈波の「ゆらぎ」を用いた精神健康度自己チェックの可能性
雄山 真弓 (関西学院大学)

唐代人物知識ベースについて ～漢字文献知識ベース構築へ向けて～
山本 一登 (京都大学)

『古事記』学術支援データベースの構築 -基本機能の検討-
柴田みゆき・斎藤晋・生田敦司 (大谷大学)

作文指導のための作文添削データベースの構築
竹内和広 (大阪電気通信大学)

言語年代学における日本語系統、縄文語を推定する
永田良茂

■第13回公開シンポジウム「人文科学とデータベース」プログラム
2007年12月22日(土曜日) 奈良女子大学

特別講演

キトラ・高松塚古墳のフォトマップ撮影と画像の保存活用
～埋蔵文化財写真でのデジタル画像とデータベース～
中村一郎(奈良文化財研究所)

自由論題報告 I

分類語彙表による歌ことばのシソーラスの開発
山元啓史(オーストラリア国立大学)

『古事記』学術支援データベースの構築 ～系譜の図像化とインターフェイスの検討～
生田敦司・齋藤晋(大谷大学), 杉山正治(立命館大学), 柴田みゆき・宮下晴輝(大谷大学)

人文系研究資料・データの構築・公開ツールと国文学資料事例の紹介
～オープンソースソフトウェア Greenstone を利用して～
北村啓子(国文学研究資料館)

研究者のためのマイ・データベース・システムの開発
及川昭文(総合研究大学院大学), 山元啓史(オーストラリア国立大学)

共通課題報告

ウォークスルーCG を利用した情報の可視化
木村寛之(株式会社イビソク), 宇田 晃(有限会社アシストコム)

WebGIS のアクセスログによる地理学的研究の可能性～バーチャル京都 3D マップを事例に～
瀬戸寿一・桐村喬・渡辺広織・矢野桂司(立命館大学)

近畿地方における前方後円墳の分布論的検討 ～GIS データベース利用の試み～
出田和久(奈良女子大学)

GIS・数理モデルによる集落分布の立地分析
石崎研二(奈良女子大学)

自由論題報告 II

透過性をもつ能装束のリアルタイム表現法
蔡 康穎・尹 新・田中弘美(立命館大学)

仏象写真の数理的色彩復元方の試論
坂田年男・竹之内和樹・能野謙介・前原一満(九州大学), 堀内隆彦(千葉大学)

文化財ローカル・ナレッジの集積と Web-GIS を基盤とした文化現象解析
津村宏臣(同志社大学)

顔画像のフラクタル次元と好感度の関係
雄山真弓(関西学院大学)

崇神天皇の崩年はいつ頃か ～崩年モデルによる数理的検討～
小澤一雅(大阪電気通信大学)

■第14回公開シンポジウム「人文科学とデータベース」プログラム
2008年12月13日(土曜日) 同志社大学

自由論題報告 I

同志社校地内歴史的建造物群のアーカイブ化に向けて
中谷正和 (NPO 法人 京都文化財建造物研究所)

古典籍と JIS 漢字についての再考察 ー何が変わったか、変わらないでいるかー
大田 靖 (大阪大学)

古典籍と JIS 漢字についての再考察 ー何が変わったか、変わらないでいるかー
當山日出夫 (立命館大学)

弾性波による構造物調査における数学的貢献について
大田 靖 (大阪大学)

特別講演

『浮世絵の計量分析』
村上征勝 (同志社大学)

自由論題報告 II

GIS を用いた遺跡のデジタル測量と遺跡空間データベースの構築
寺村裕史 (総合地球環境学研究所)

遺跡調査における文化的景観把握の試行ー文化財の継承のためにー
山口欧志 (国際日本文化研究センター)

先史人類学への数理モデルによるアプローチ
鎌倉快之 (大阪工業大学)

遺跡探査法による削平を受けた古墳の形状復元
岸田徹 (天理大学), 酒井英男 (富山大学)

■第15回公開シンポジウム「人文科学とデータベース」プログラム
2009年11月28日(土曜日) 神戸大学

招待講演

国立民族学博物館におけるデジタルアーカイブズ
久保正敏 (国立民族学博物館)

自由論題報告

継承語教育文献データベースの構築
田中順子・森下淳也(神戸大学), 中島和子(トロント大学)

日本語の起源、解明方法と数値評価
永田良茂

表現豊かな自然発話コーパスのアクセスについて
モクタリ明子 (神戸大学), ニック・キャンベル (ダブリン大学), 田畑安希子 (神戸大学)

和歌解析用 MeCab 辞書の開発 —八代集解析済みコーパスによる学習—
山元啓史(東京工業大学)

作文指導に向けた表現データベースの構築と表現評価指標の検討
安藤秀明・秋吉信吾・竹内和広 (大阪電気通信大学)

『古事記』学術支援データベースの構築 —系図表示システムにおけるデータ構造の検討—
生田敦司 (大谷大学), 杉山正治 (立命館大学), 柴田みゆき (大谷大学),
齋藤晋 (国際日本文化研究センター), 宮下晴輝 (大谷大学)

古事記崩年干支についての疑念
小沢一雅 (大阪電気通信大学)

■第16回公開シンポジウム「人文科学とデータベース」プログラム
2010年11月27日(土曜日) 花園大学

一般報告

中尾佐助資料公開のための画像データベースの構築と運用
小島篤博・青木茂樹・宮本貴朗 (大阪府立大学)

デジタル写真資料のWWWに基づく共有・管理手法について
守岡知彦 (京都大学人文科学研究所)

2010年代の考古データベースはどう展開するか？
近藤康久 (東京大学)

京町家の3次元CG復元をめぐる問題
師茂樹・明珍健二 (花園大学)

ブーリアン演算による歌ことばモデルの解析
山元啓史 (東京工業大学)

「現代日本語書き言葉均衡コーパス」のための形態論情報データベースについて
小木曾智信・中村壮範 (国立国語研究所)

『古今和歌六帖』の本文と表現—デジタル校本の試み—
福田智子 (同志社大学)・竹田正幸 (九州大学)・南里一郎 (立命館大学)

主催：第17回公開シンポジウム
実行委員会

後援：人文系データベース協議会

委員長：宝珍輝尚（京都工芸繊維大学）
委員：加藤常員（大阪電気通信大学）
小森政嗣（大阪電気通信大学）
野宮浩揮（京都工芸繊維大学）

議長：出田和久（奈良女子大学）
副議長：川口 洋（帝塚山大学）
副議長：加藤常員（大阪電気通信大学）
赤間 亮（立命館大学）
江澤義典（関西大学）
及川昭文（総合研究大学院大学）
桶谷猪久夫（大阪国際大学）
小沢一雅（大阪電気通信大学）
後藤 真（花園大学）
小森政嗣（大阪電気通信大学）
柴山 守（京都大学）
関野 樹（総合地球環境学研究所）
高橋晴子（大阪樟蔭女子大学）
中谷広正（静岡大学）
八村広三郎（立命館大学）
原正一郎（京都大学）
深海 悟（大阪工業大学）
宝珍輝尚（京都工芸繊維大学）
三宅真紀（大阪大学）
村上征勝（同志社大学）
森下淳也（神戸大学）
師 茂樹（花園大学）

人文系データベース協議会 第17回公開シンポジウム「人文科学とデータベース」

発行日 2012年1月7日

発行所 第17回公開シンポジウム実行委員会

〒606-8585 京都府京都市左京区松ヶ崎御所海道町

京都工芸繊維大学・工芸科学研究科・情報工学部門

宝珍輝尚（シンポジウム実行委員長） Email:hochin@kit.ac.jp