

「景観文字調査」のための調査結果分類・可視化ツールの試作

田島 孝治[†] 米田 純子[‡] 高田 智和[‡]
[†]東京農工大学 工学府 [‡]国立国語研究所

景観の中に現れる文字の調査は、地理学や言語学などの研究において様々な観点から考察が行われてきた。しかし、先行研究では、サンプルの定義、取得範囲、分類方法などが不明瞭なものが多く、過去の調査結果の再利用性・再現性に乏しい点が課題であった。本稿ではこの課題を解決するために、景観文字の調査結果を記録するツールを試作した。また、愛知県岡崎市において景観文字調査を行い、試作ツールの有効性を検証した。

A Classifying and Visualization Tool for the Survey of Linguistic and Notational Landscape

TAJIMA Koji[†] YONEDA Junko[‡] TAKADA Tomokazu[‡]
[†]Tokyo University of Agriculture and Technology
[‡]The National Institute for Japanese Language

The survey of linguistic and notational landscape has the problems of little reproducibility and low reusability. We proposed classifying and visualization tool for the survey of linguistic and notational landscape. And we also show the validity of our tool using the linguistic and notational landscape survey in Okazaki City.

1. まえがき

景観の中に表れる文字・言語の調査は地理学・日本語学・社会言語学などの各方面から、それぞれの研究目的に応じて行われている。これらの調査では、対象地域の店名や看板を調査することで、文化面での地域差や、外国人の滞在や来訪に対応するための社会の多言語化を考察している。また、正井泰夫(1983)の調査のように経年調査によって特定の地域における社会の変化を考察しようという試みもある。

先行研究の調査は、複数の研究領域において、それぞれの関心からサンプルや地域を決定し行われたものである。このため、サンプルの定義やサンプル取得地域の設定において、曖昧な点が残る調査もある。その結果、同一地域の調査であっても、調査範囲や分類基準が一致しているか判断できず、調査結果の相互比較を行うことが難しい。経年変化や地域差を考察する上で、調査結果の比較を適切に行えないことは重大な問題であり、採集結果の再利用性・再現性に乏しいという課題を解決する必要がある。

調査結果の有効性・利便性を高めるためには、最終的な分類・統計結果だけでなく、採集したサンプルデータ(写真)と、記録結果の分類に用いる表記言語や使用文字種の情報、採集位置や日時など(メタデータ)を同時に保存しておく必要がある。さらに、記録結果を必要に応じて何度でも再分類し、過去の記録結果を有効に再利用するためのツールが必要である。

2. 目的

本稿では、景観文字調査のための調査結果分類・可視化ツールを提案し、試作を行う。

本ツールで扱うデータは、デジタル化された写真と GPS 機器を用いて収集した時系列に基づく調査員の位置情報、デジタル化された地図画像である。これらのデータ形式は、景観文字調査の結果を電子的に扱う上で、極めて有効である。調査用紙を用いて人間が記録し、後に電子化する方式では、電子化のための作業の負担が大きく、そのためにサンプル数に制限が生じたり、十分な考察が行えなかったりという問題が生じる可能性がある。電子機器の利用はこれらの問題の解決に有効だが、機器の設定ミスや調査環境により、記録された調査結果に大きな誤差が発生している可能性に留意しなければならない。高田・田島・米田(2008)により、GPS、デジタルカメラ、デジタル地図を使う上での問題点が議論されている。

ただしここで提案するツールはデジタル画像の自動認識や分類、位置情報の自動補正などの自動化を行うものではなく、研究者による調査結果の分類や集計、過去のデータの再集計を補助するためのものである。

ツールは以下の点に留意して設計した。

- (1) 画像・位置情報・メタデータの保存は将来的に利用可能な汎用的な形式を用いる
- (2) 任意のメタデータを用いて容易にクロス集計を行うことを可能とする

(3) 最終結果を随時確認しながら、メタデータの内容を変更することが可能であり、何度でも、必要に応じたフォーマット（図・表）で出力することができる。

3. 調査結果の処理手順

図1に提案したツールを用いた景観文字調査の記録の流れを示す。

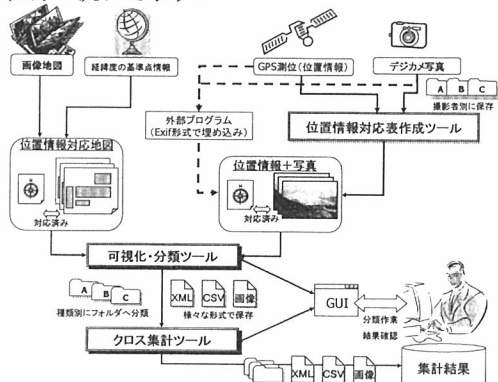


図1 景観文字の調査結果記録過程

提案ツールは機能ごとに分割し、「位置情報対応表作成ツール」、「可視化・分類ツール」、「クロス集計ツール」の三種類から構成する。

3.1 位置情報対応表作成ツール

位置情報対応表作成ツールは、調査において利用した地図画像（スキャナにより取り込んだ地図やデジタルデータとして発行された地図）、調査時のGPS測位により得られた位置情報からなる移動履歴、撮影されたデジタル写真の三点に対し、分類作業のための事前処理を行う。

地図画像には、国土地理院のWebサイトや実測値などを用いて経緯度の関連付けを行い、位置情報を付与する。この処理は位置情報対応表作成ツールの機能とは分離し、他の二つのツールからも実行可能とする。これにより、外部プログラムで位置情報を付与した写真に対しても、分類作業を可能とする。

移動履歴とデジタル写真は、記録時刻と撮影時刻を比較して、一致もしくは誤差が設定値以内になる組み合わせを検索する。そして、撮影位置付与済み写真として表1の列を持つCSV形式で出力する。ファイル名は絶対パスで記録するが、分類ツールは、事前処理後のファイルが手動でサブフォルダへ移動された場合でも、自動的に検索しパスの変化に対応する。測地系は日本測地系（TOKYO）と世界測地系（WGS-84）に対応し、出力は世界測地系に統一する。

加えて、デジタル写真はファイル名のリネームを行う。ファイル名のリネームは、特定のエリアを複数人で調査した結果の統合処理に必要である。同一の機器を利用してデータ収集を行うと写真などのファイル名に重複が起こる可能性がある。リネーム処理により、撮影者別に分けられたファイルを、連番のファイル名に変換し、各撮影者の行動履歴と関連付けることが可能となる。

表1 写真へ位置情報を関連付けるCSVのカラム構成

列	カラム名	内容例
1	ファイル名	C:\hogehoge\test.jpg
2	日付	2008-12-20
3	時刻	9:45:00
4	緯度	36.085879
5	経度	140.108085
6	高度	60.0
7	測地系	WGS-84 または TOKYO

3.2 可視化・分類ツール

景観調査の結果は、研究者の目的に合わせ、表記言語や使用文字種、出現地域、さらには素材や表記文字の大きさなどの分類軸にしたがい、考察を行うことが一般的である。

可視化・分類ツールは、デジタル写真を分類する作業の補助ツールである。分類結果は、デジタル写真をファイルシステム上のフォルダに分ける方式で保存する。この方式は人間にとって直感的であり、分類結果が同一であれば、どのような方法で分類作業を進めてもよい。このため、撮影地域や撮影時刻など画像を確認しなくても分類できる写真に対しては、機械的な分類が可能となる。

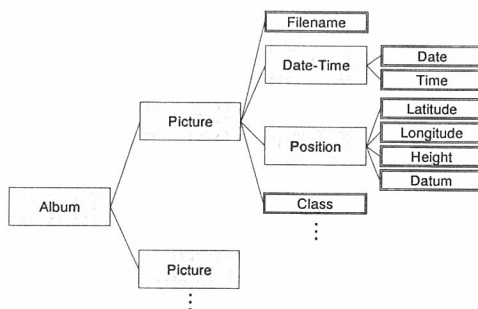


図2 保存用XMLのデータ構造

本分類ツールを利用すると、各写真のサムネイルを確認しながら、分類作業を行うことが可能である。また、結果は地図上に記号として表示され、分類結果に合わせ自動的に更新される。

結果の保存は、フォルダ構造にデジタル写真をコピーするだけでなく、表形式の CSV と木構造の XML 形式にも対応している。

CSV 形式のカラム構成は表 1 のカラムに加え、8 カラム目に分類結果を記述したものである。XML 形式は、図 2 に示す構造で保存する。

3.3 クロス集計ツール

クロス集計ツールは、デジタル写真の分類結果を複数用いてクロス集計を行うためのツールである。可視化・分類ツールによる分類結果の CSV または、フォルダに分けられたデジタル写真を読み込むことで、最大三種類の分類軸を相互に重ね合わせ表示することができる。それぞれの分類軸は、地図上にプロットする記号の色・形・大きさを変化させることで表現する。割り当てる色や形については、地図を見ながら自由に設定することが可能である。

クロス集計ツールを利用することで、考察したい内容に合わせ記号を変化させた地図を、何種類も作成することができる。これに加えて集計結果を、CSV 形式または XML 形式で保存することも可能である。

4. 実装

4.1 実装環境

提案したツール群は、全て Java 言語を用いて実装した。Java のコンパイラおよび VM (Virtual Machine) は version 1.5.0_09 を利用した。Java 言語の特徴により Windows, Linux, Mac OS など多数のプラットフォームで動作することが可能である。

4.2 ツールが対応するファイル形式

本ツールで利用可能な移動履歴のファイル形式は NEMA, CSV, タブ区切りテキスト, gsd 形式の四種類である。NEMA は現在市販されている GPS 機器と、ナビゲーションシステムや地図ソフトなどの間で使われている通信プロトコルであり、最も一般的な形式である。テキスト形式であるため、受信したデータをそのままテキストファイルとして保存し、移動履歴として利用する。本実装では GPGGA および GPRMC メッセージを解析し位置情報を取り出す。CSV とタブ区切りテキストは、商用の表計算ソフトやユーザが自作したプログラムで利用される汎用性に優れた形式である。本実装では表 2 に示す形式とする。ID は任意の文字列を利用可能であり、速度および高度は空白でもかまわない。gsd 形式は実装時に利用したハードウェア (DG-100) に付属していたソフトウェアの独自形式である。一般的な形式ではないが、利用中のアプ

リケーションからそのまま出力できるため、利便性を高めるために対応した。

デジタル写真は JPEG 形式のみ対応とした。現在市販されている多くのデジタルカメラでは、写真にメタ情報を付与するための規格として Exif を採用している。BMP などの他の画像形式では Exif をサポートしていないため JPEG のみに対応することにした。

地図画像は読み込みに、Java の標準ライブラリである ImageIO クラスを用いた。このため、JPEG, PNG, BMP, WBMP, GIF に対応している。

表 2 位置情報ログとして入力する CSV, タブ区切りテキストのカラム構成

列	カラム名	内容例
1	ID	1
2	日付	2008-12-20
3	時刻	9:45:00
4	緯度	36.085879
5	経度	140.108085
6	速度	2.8
7	高度	60.0

4.3 ツールのモジュール構成

図 3 に各ツールのモジュール構成を示す。

位置情報対応表作成ツールは移動履歴とデジタル写真を読み込み、時刻で同期することで位置情報を付与した写真を生成する。ファイルの読み込みは FileReader パッケージに属する二つのクラスにより行う。移動履歴は GPSPositionReader クラスが読み込み、LogParser インタフェースによりファイル形式の違いを吸収する。デジタル写真は ExifReader クラスが読み込み、撮影時刻を Exif タグ中の DateTimeOriginal タグから取得する。二種類のファイルの読み込み完了後に、Data Mapper パッケージ中のクラスが、デジタル写真に位置情報を関連付け、位置情報付与済み写真を生成する。結果は CSV 形式の表として保存する。Exif における GPSInfo IFDPointer タグを利用し、位置情報を直接写真ファイルへ書き込む方式は、汎用性と誤差の修正の容易さを損なうため行わないことにした。ただし、他のソフトウェアにより位置情報を埋め込まれた写真の位置情報を読み込むことは可能である。

分類・可視化ツールは、位置情報付与済み写真と地図画像を用いて分類作業を行う。地図画像は三種類のツールで共通して利用可能な MapImage コンポーネントにより位置情報を付与し、背景画像として利用する。分類作業は GUI を用いて行い、ファイルの選択と移動は全てドラッグ&ドロップで行う。これによりスナップ写真を手で分ける感覚で作業が行える。

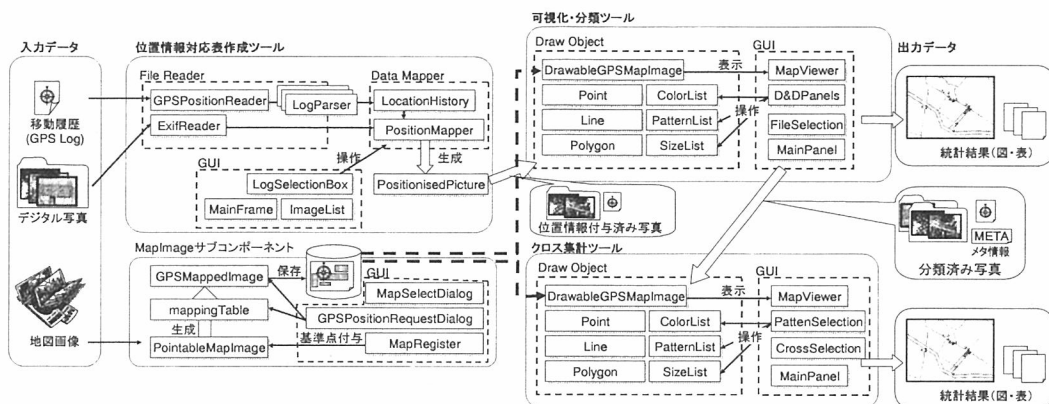


図3 ツールのモジュール構成

分類時には Draw Object パッケージ内のクラスが分類結果を地図画像の上に重ねた画像を生成する。画像は分類の変更に合わせて随時更新されるため、結果を確認しながら分類作業を進めることが出来る。最終結果は、フォルダ単位に分割したデジタル写真の形式で保存する。

クロス集計ツールは分類・可視化ツールの分類結果を利用し集計処理を行う。描画や基本的な GUI を実現するクラスは、分類・可視化ツールと共通である。分類軸に対し色や形を選択するための、CrossSelection クラスと Patten Selection クラスが新規クラスとなる。

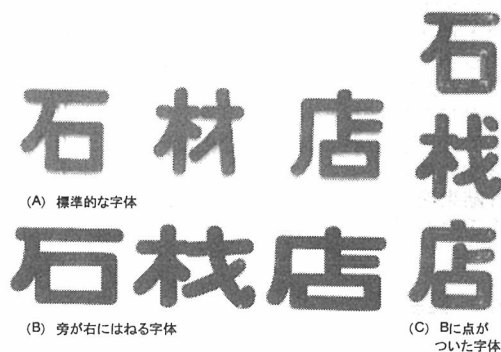


図4 調査フィールドに見える三種の「材」

5. 調査結果を利用した実証実験

5.1 検証実験のフィールドと内容

開発したツールの効果を検証するために、愛知県岡崎市をフィールドにした言語景観調査の結果を本ツールにより整理する。岡崎市は国立国語研究所が敬語敬意表現の経年調査を行っているフィールドであり、調査の第一回目は 55 年前の 1953 年、第二回目は 36 年前の 1972 年に行われている。2008 年中に第三回目の調査を行う。この調査に合わせて言語景観を調査することにより、実時間による経年変化を記録することが出来ることを期待している。

今回対象にした景観文字は、看板中に現れる「材」という漢字である。岡崎市は良質の花崗岩の産地であり、石製品の生産が盛んなことから石材店が多く存在している。今回の調査では岡崎市花崗町の石屋町通りに存在する石材店の看板を撮影した。

調査は 2008 年 10 月 7 日に調査員二名により行った。調査員は GPS ロガー (DG-100) とデジタルカメラ (CaplioR4) を利用し、毎秒の位置情報と 60 枚の写真を集めた。

図 4 に調査フィールドに見える三種の「材」の画像を示す。

写真に対して付与するメタデータの候補として次の五種類が考えられる。

- (1) 字体: 「材」の字体(A)~(C)
 - (2) 書体: ゴシック体か否か
 - (3) 文脈: ○○石材店, 石材, 石材一式など
 - (4) 縦書き/横書き: 表記の方向
 - (5) 表示の種類: 看板, ポスター, 暖簾など
- 今回のフィールドでは、「石材」以外のパターンが存在しないため文脈は利用しなかった。

5.2 ツールの利用手順

この調査結果を本ツールにより分類し集計する。第一の手順として位置情報と写真の関連付けを行った。位置情報対応作成ツールにより、GPS ログを読み込み、写真中に埋め込まれた撮影時刻と同期した位置情報を取得する。その後ファイル名のルールを統一し、表 1 の形式により結果を出力した。

次に、背景として利用する地図画像を準備する。分類ツールは任意の画像を読み込み、位置情報を付与し、背景画像とすることができる。今回は国土地理院発行の数値地図 25000 (地図画像) および市販の地図ソフトから取り込んだ画像を利用した。本稿の画像は国土地理院の地図を背景として利用している。図 5 に地図画像

に対し位置情報を付与する GUI を示す。画像を読み込み適当なポイントを指定すると、経緯度の入力ダイアログが表示され位置情報を入力することが可能となる。入力した地図画像を位置情報付き地図画像として利用するには、二点以上の位置情報を入力する必要がある。



図5 地図画像への位置情報付与画面

続いて、設定するメタデータに従い写真を分類する。この際同じ看板を撮影した写真が複数ある場合や不要な画像などは、分類対象外として除去することが可能である。事前にファイルを別フォルダに移動しておく方法でも除去できる。分類時の GUI 画像を図6^(注1)に示す。写真をドラッグすることにより、分類の種類を変更することが可能である。また分類の種類数をボタンにより増減可能である。分類を変更すると地図画像上のマークが変更され、リアルタイムに分類結果を確認することができる。また、写真を右クリックすることで、原寸大のイメージを確認することも可能である。

分類作業はメタデータの種類に応じて行い、結果は CSV またはフォルダごとに分類した画像として保存する。

最後にメタデータを複数組み合わせることでクロス集計を行った。(1)字体と(2)書体の関係に対して分析した、GUI の画像を図7に示す。クロス集計ツールでは、分類結果を CSV もしくはフォルダ単位で読み込み、それぞれの要素に対して色・形・大きさの分類方法を選択する。その後、赤や青、丸や四角などを選択することで、クロス集計結果の画像が生成される。この画像は選択項目の変化によりリアルタイムに更新され、どのタイミングでも保存することが可能である。またクロス集計結果は、CSV または XML として出力することも可能である。

5.3 分析結果

字体と書体によるクロス結果を図8に示す。また、結果を表3に示す。集計は撮影した

1) 画像の分類は色を用いて行うが、本稿では印刷時にモノクロになる点を考慮し、形状を変更したものを利用した。また、拡大写真の一部はプライバシー保護の関係上加工を施した。

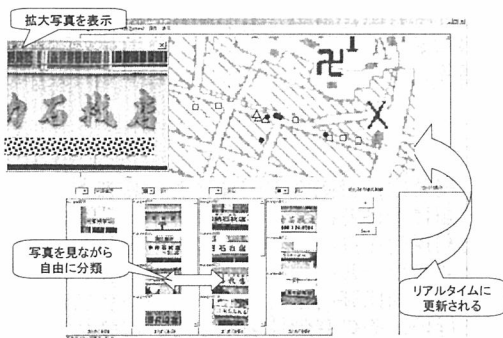


図6 景調査結果の分類画面



図7 景調査結果のクロス集計

60 枚の写真の中で、「店名を表示している大型看板」のみに注目し分類した。対象となった写真は 13 枚である。集計の結果から、この地域における「材」の中心となる字体は(B)であり、(A)はゴシック体に、(C)はゴシック体以外で表現されることがわかる。

表3 字体と書体によるクロス集計結果

	(A)材	(B)材	(C)材
ゴシック体	4	4	存在しない
ゴシック体以外	1	3	1

5.4 分類時の問題点

本ツールの検証実験において、元の写真を利用して分類を行うことの問題点が指摘された。調査時においては、同じ被写体を複数回撮影することが多くある。例えば建物の全体像と、文字が書かれた看板の拡大写真のような場合である。この場合、ツールで両方の写真が入ったフォルダを分別すると、属性を付与するたびに除去せねばならず手間が大きい。しかし、文字のみに特化して写真を切り出してしまうと、日付情報が書き換わるため自動的に位置情報を付与することが困難になってしまう。

これらのことから、本ツールの利用においては事前処理がリネームのみという対応は不十分であり、写真同士の関係性を整理する必要がある。

6. まとめ

本稿では、景観文字調査における調査データの分類における問題点について述べ、調査結果の再利用性を高め、分類、可視化を効率的に行うためのツールを提案した。また、ツールを実装し、愛知県岡崎市における景観文字調査の分析を行い、その有効性を示した。この結果、ツールに求められる改良点として、写真の事前処理における同一内容の写真を整理する方法の実現と、景観文字により適応するために写真の一部のみを切り出した画像へ対応する必要性が明らかになった。

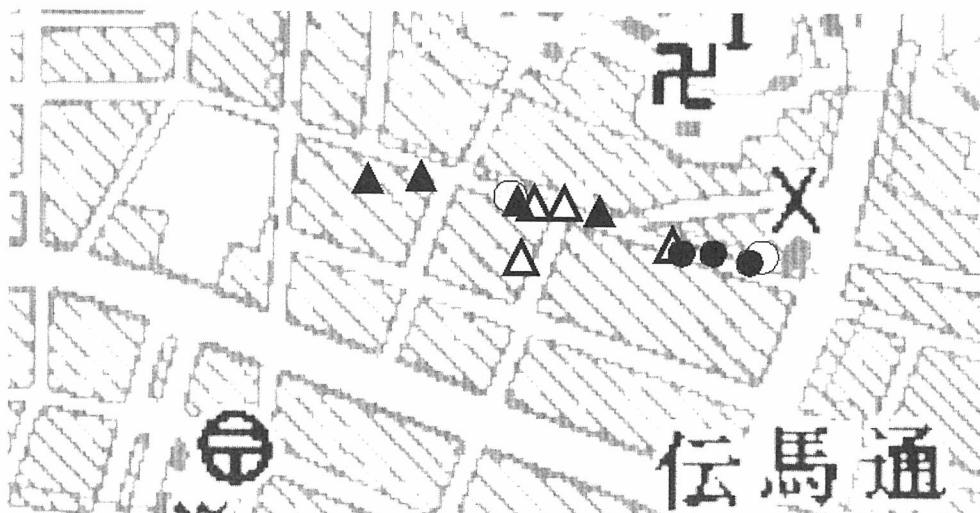
今後の展望としては、一つのサンプルが二種類の分類に属するなどの複雑な事例分析や、経年調査への適応による有効性の検証と、先に述べたツールの改良が考えられる。

参考文献

- [1] 正井泰夫: 新宿の喫茶店名一言語景観の文化地理一, 筑波大学地域研究 1, pp.49-61, 1983.
- [2] 宮島達夫: 多言語社会への対応—大阪: 1994年—, 阪大日本語研究, Vol.7, pp.1-21, 1995.
- [3] エツコ・オバタ・ライマン: 表記法から観察するビジネス・アイデンティティ—表参道商店街の店名 (I), 麗沢学際ジャーナル 13-1, pp.39-67, 2005.
- [4] 當山日出夫, 笹原宏之, 高田智和: 文字研究における GPS の利用, 情報処理学会研究報告, 2007-CH-73, pp.1-8, 2007.
- [5] 高田智和, 田島孝治, 米田純子: 文字生活研究における「景観文字調査」, 情報処理学会研究報告, 2008-CH-79, pp.1-8, 2008.
- [6] PETER BACKHAUS: *Sings of multilingualism in Tokyo — a diachronic look at the linguistic landscape*, INTERNATIONAL JOURNAL OF THE SOCIOLOGY OF LANGUAGE, 175/176, pp.103-121, 2005.

表 4 字体および書体の描画凡例

字体	色による分類	書体	形状による分類
(A) 材	白	ゴシック体	三角
(B) 枝	黒	ゴシック体以外	丸
(C) 枝	灰色		



※背景地図として国土地理院発行の「数値地図 25000 (地図画像)」を利用

図 8 愛知県岡崎市石屋町通りにおける材の字体と書体