

# ウェブ上の自然言語解析で描く都市のイメージ

The Image of the City Using Natural Language Analysis on Web Resources

手塚 太郎<sup>▽</sup> 李 龍<sup>◇</sup>  
高倉 弘喜 上林 弥彦<sup>▲</sup>

Taro TEZUKA Ryong LEE  
Hiroki TAKAKURA Yahiko KAMBAYASHI

人間の主観的な地理空間認知を反映させた「イメージマップ」は、地域情報検索に用いるインターフェースとして最適である。従来のイメージマップは地図作成者の主観に基づくか、聞き取り調査によって作成されてきた。本研究ではウェブ上の和文テキストにおける位置説明文を解析し、イメージマップ描画に利用する手法を示す。

日本語の位置説明文は、格助詞・相対名詞・形容詞を中心として構成される。これらの品詞と地名との共起関係を分析した結果、地名のカテゴリと位置説明に用いられる品詞との間に相関関係があることが示された。

分析の結果は人間が思い浮かべる都市の姿をあらわしており、都市のイメージ、すなわち「イメージマップ」の描画に利用することが出来る。

Image maps based on subjective understandings of geographical space are suited for user interface in local information search. Usually, image maps are either drawn from subjective understanding of map designer or from a questionnaire survey on local residents. The aim of this paper is to present a way to create image map using natural language analysis on web resources.

In Japanese, case particles, relative nouns, and adjectives play important role in place descriptions. By obtaining co-occurrence between place names with those parts of speech, it will be shown that categories to which place names belong have correlation with how they are used in the context.

The result shows that it is possible to characterize place names and draw image maps using web resources.

## 1. はじめに

地域情報検索に用いられるユーザインタフェースのひとつ

<sup>▽</sup> 学生会員 京都大学大学院情報学研究科修士課程

[tezuka@db.soc.i.kyoto-u.ac.jp](mailto:tezuka@db.soc.i.kyoto-u.ac.jp)

<sup>◇</sup> 京都大学大学院情報学研究科博士後期課程

[ryong@db.soc.i.kyoto-u.ac.jp](mailto:ryong@db.soc.i.kyoto-u.ac.jp)

京都大学大型計算機センター

[takakura@i.kyoto-u.ac.jp](mailto:takakura@i.kyoto-u.ac.jp)

<sup>▲</sup> 正会員 京都大学大学院情報学研究科

[yahiko@i.kyoto-u.ac.jp](mailto:yahiko@i.kyoto-u.ac.jp)

つとして、主観的な空間認識を反映した「イメージマップ」がある。イメージマップにおいては幾何学的形状の正確さは求められず、認知上の重要性やオブジェクト間の意味的連関によって大きく歪められた形で描かれる。たとえば都道府県を選択するインデックス画像として正しい縮尺の日本地図が用いられることはほとんどなく、小さな県も選びやすいよう、歪められた地図を使用するのが一般的である。都道府県のサイトなどでは市町村を選択する際にもイメージマップを表示させることもあるが、それより大きな縮尺になると、テキストで住所を入力させる方式が多い。これは、イメージマップ作成にかかるコストが大きいと考えられる。

イメージマップはこれまで地図作成者の直観によって描かれるか、聞き取り調査の結果などに基づいて作成されてきた。本研究ではウェブ上の位置説明文をイメージマップの描画に利用する方法を示す。これによって描画のコストが大幅に削減される。

日本国内の都市に対する地域情報検索システムの実現を目的としているため、日本語位置説明文からの情報抽出を行った。日本語における位置説明は格助詞・相対名詞・形容詞を中心として構成される。これらの品詞と地名の共起関係を分析して地名集合の中から重要なイメージマップ要素を抽出し、特性付けを行なった。

## 2. 日本語における位置説明文の特徴

自然言語テキストからの情報抽出の手法として、単語の出現頻度が広く用いられている。ところが地域情報の抽出においては出現頻度の情報だけでは不十分である。たとえば京都市の地名リストの中には「京都大学」が含まれるが、ウェブ上での出現頻度は非常に高い。だが、地名としてではなく機関名として使われていることがほとんどである。「京都大学を卒業」「京都大学教授」などの形で使われることも多い。そこで、真に地名としての使用だけを抽出するためには、それが位置説明文の中で使われている事例のみを選び出す必要がある。

英語においては前置詞によって位置関係を表現するが、日本語の位置説明文は主として以下の二種の形式をとる。

1) 名詞句 + 格助詞 + 相対名詞 + 格助詞 + 文節

例： 京都駅 + の + 前 + で + 写真を撮った。

2) 名詞句 + 格助詞 + 形容詞 + 格助詞 + 文節

例： 銀閣寺 + の + 近く + を + 散策した。

格助詞はそれに先行する名詞が文全体の中で示す役割、すなわち位置・起点・着点・経由地などの「格」を与える。しかし格助詞は種類が少なく、位置関係を十分に表現出来ないため、「上・右・前・中・近く」などの相対名詞/形容詞を補うことで詳細な位置指定を可能としている。

## 3. 相対名詞と形容詞を用いた特性付け

地名の既知の特徴が、相対名詞/形容詞との共起にどのような影響を及ぼすのかを明らかにすることで、共起頻度から逆に地名を特性付け出来るようになる。そこで、地名をカテゴリに分割し、それぞれのカテゴリに属する地名と相対名詞/形容詞との共起頻度を調べた。

実験では「寺社」「駅」「交差点」「公共施設」の4つのカテゴリについて、「付近/近辺/の近く」(三種をまとめて「近傍表現」と名づける)、「周辺」, および「の前」との共起数を調べた。地名データとして176の寺社, 857の交差点, 54の駅, 1458の公共施設を用いた。分析の結果を図1に示した。

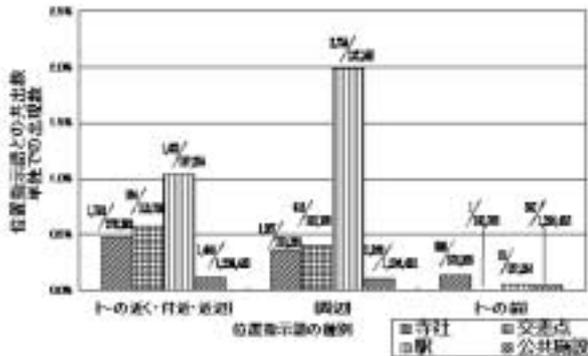


図1 地名カテゴリと相対名詞・形容詞の共起頻度

Fig.1 Co-occurrences of Place Names with Relative Nouns and Adjectives

得られた結果から、以下の特徴を読みとることが出来る。

- 1) 寺社・交差点・公共施設で近傍表現が多い。狭い領域を指示する上で重要な目標物として使われているためと考えられる。
- 2) 「駅」で「周辺」が多い。駅は広範囲の移動における結節点としての役割を担っているためと考えられる。

#### 4. 格助詞を用いた特性付け

位置説明に用いられる地名はほとんどの場合、格助詞を伴って現れる。そこでまず、位置説明に用いられる格助詞についてその意味を明らかにし、それらが実際に地名と共起する頻度を調べた。

日本語の格助詞において「デ/ニ」および「ヘ/ニ」は意味が類似するが、厳密には差がある。その違いが小さいため、両者の頻度に差現れた場合、原因の推定が比較的容易である。ゆえに本研究では「デ/ニ」「ヘ/ニ」の二対を分析し、地名の特性付けを行なった。

##### 4.1 「デ/ニ」格の相違とその利用

格助詞「デ/ニ」はいずれも位置を示す格として用いられるが、使い分けが存在する。二格は『状態』の位置を表すのに対し、デ格は『過程』および『行為』の位置を表す。動詞は状態/過程/行為の三種に分けることが出来るが、二格は状態動詞に、デ格は過程/行動動詞に、それぞれ対応する。  
**状態:** 阪急デパートにたくさんの客がいる。  
**過程:** 阪急デパートで安売りが始まる。  
**行為:** 阪急デパートで買い物をする。

一方、過程/行為動詞において二格が用いられた場合、それは着点を意味する。ひとつの文の中で、位置を示すデ格と着点を示す二格とが併用されることも可能である。例として、「四条通りでリプトンの喫茶店に入った」が挙げられる。デ格は全体的な行為の位置、二格は行為の着点を示す。

格助詞との共起数が地名の特性とどのように関係しているかをウェブ上の位置説明文に基づいて調べた。

表3は京都における代表的な地名等について、ウェブ上での格助詞との共起数を集計したものである。

最右列の「比」はデ格としての出現数を二格としての出現数で割った値である。自然地形に比べて駅や大学でデ格の頻度が高いのは、人間が活動する場所となりやすいためと考え

られる。

地名	で	に	比
桂川	723	1540	0.469
宇治川	523	1070	0.489
賀茂川	165	384	0.430
宝ヶ池	180	280	0.634
深泥池	113	264	0.428
大沢池	30	80	0.375
比叡山	1610	4180	0.385
愛宕山	317	1220	0.260
鞍馬山	316	590	0.536
金閣寺	880	2290	0.384
銀閣寺	257	1260	0.204
清水寺	1810	3870	0.468
京都駅	6940	9620	0.721
大阪駅	5490	6360	0.863
東京駅	21000	25100	0.837
京都大学	2700	2900	0.931
同志社大学	946	976	0.969
立命館大学	1190	1330	0.895

表1 代表的な地名における「デ/ニ」格出現数の比

Tab.1 Ratio of "de" Case and "ni" Case in Well-Known Place Names

図2は京都市の11の区、54の駅、176の寺院、25の大学、および京都の代表的な20の自然地形について、「デ/ニ」格それぞれの頻度を散布図で表したものである。表2はそのカテゴリ別平均値と分散を示す。

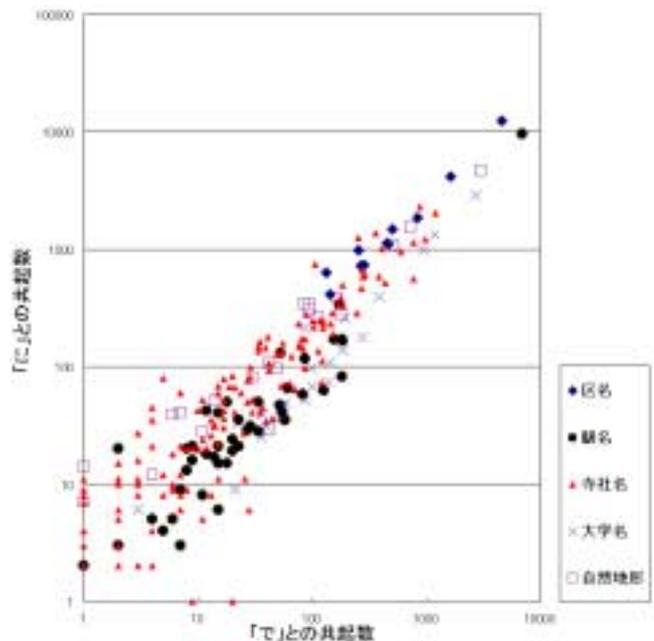


図2 「デ/ニ」との共起数の分布

Fig.2 Co-occurrences with "de" Case and "ni" Case

地名の カテゴリ	「デ／ニ」の、 カテゴリ内平均	分散
区名	0.360	0.004
駅名	0.915	0.307
寺社名	0.691	3.034
大学名	1.107	0.240
自然地形	0.408	0.071

表2 「デ／ニ」との共起数のデータ

Tab.2 Co-occurrences with “de” Case and “ni” Case

全体的な傾向として、区名・寺社名・自然地形において二格が多く、駅名・大学名においてデ格が多い。これは、駅および大学では人間の活動が頻繁に行なわれているため、過程や行為の動詞に用いられるデ格が多く現れるのだと考えられる。寺社は目的地としては重要であるが、そこで必ずしも行為が行なわれるとは限らない。「金閣寺に行く」ことは多くとも、「金閣寺で遊ぶ」「金閣寺で食事する」は少ない。また、自然地形はそれについて言及されることは多くとも、実際に行為の場となることは少ないため、デ格の頻度が低くなっていると考えられる。

この結果を、利用者に応じたイメージマップのカスタマイズに使用できる。図3にひとつの方法を示す。

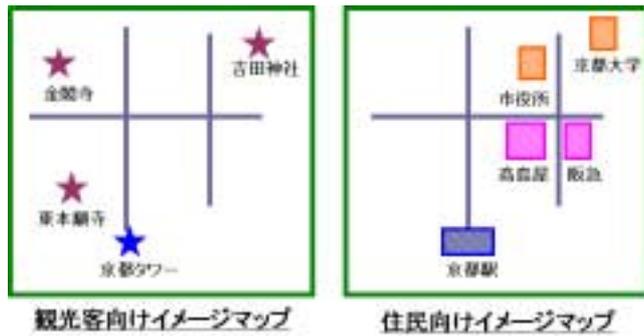


図3 「デ／ニ」格頻度を利用したイメージマップ

Fig.3 Using the Difference between “de” Case and “ni” Case in Drawing Image Maps

観光客を対象とするイメージマップにおいては地域の概略を示すことが主な目的であるため、二格と多く共起する地名、すなわち寺社や自然地形を表示させる。これらの地名には静的な情報が多く結び付けられ、観光客にとって重要である。一方、地元住民は新しく起きた出来事に関心を向けるため、デ格と多く共起する駅や大学を強調して表示させるべきである。

4.2 「ヘ／ニ」格の相違とその利用

格助詞「ヘ／ニ」は共に移動に関わり、意味も類似する。だが、ヘ格では移動の方向に焦点が置かれるのに対し、ニ格では目的地に焦点が置かれるという違いがある。これは二格が着点を表す格であることとも関連する。

表3は京都の代表的な地名等のウェブ上での「ヘ／ニ」との共起数およびその間の比を示す。寺院や駅においてヘ格の頻度が高く、自然地形や大学において低いことが読み取れる。

地名	へ	に	比
桂川	288	1540	0.187
宇治川	224	1070	0.209
賀茂川	41	384	0.107
宝ヶ池	68	280	0.243
深泥池	32	264	0.121
大沢池	11	80	0.138
比叡山	753	4180	0.180
愛宕山	392	1220	0.321
鞍馬山	161	590	0.273
金閣寺	749	2290	0.327
銀閣寺	759	1260	0.602
清水寺	2410	3870	0.623
京都駅	3280	9620	0.341
大阪駅	2270	6360	0.357
東京駅	9540	25100	0.380
京都大学	509	2900	0.176
同志社大学	227	976	0.233
立命館大学	366	1330	0.275

表3 代表的な地名における「ヘ／ニ」格出現数の比

Tab.3 Ratio of “e” Case and “ni” Case in Well-Known Place Names

図4では、京都市内の地名集合について「ヘ／ニ」格それぞれの出現頻度を散布図で示した。表4はその平均値と分散を示す。特徴として読み取れるのは、区名・寺社名・自然地形において二格が多く、駅名においてヘ格が多いことである。区名においてはヘ格の頻度が極端に小さい。

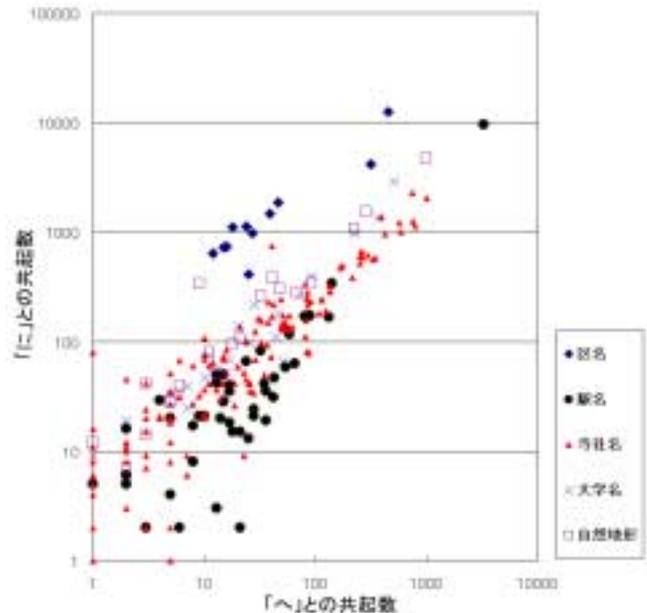


図4 「ヘ／ニ」との共起数の分布

Fig.4 Co-occurrences with “e” Case and “ni” Case

地名の カテゴリ	「へ/二」の、 カテゴリ内平均	分散
区名	0.032	0.000
駅名	1.013	2.560
寺社名	0.368	0.261
大学名	0.249	0.022
自然地形	0.184	0.005

表4 「へ/二」との共起数の傾向

Tab.4 Co-occurrences with “e” Case and “ni” Case

この理由として、駅は移動におけるランドマークとしての役割を果たしていることが挙げられる。一方、自然地形はその大きさゆえに移動の方向として用いられることは少ない。区名においてへ格が少ないのは、市町村や字と比べて文化的統一性に乏しく、そのため移動の方向として意識されることが少ないためと考えられる。

これらの結果は図5のような形でイメージマップに反映できる。へ格が多い駅は移動の方向として重要な役割を果たしているため、経路（パス）とセットにするなどして、強調描画する。一方、へ格が少ない区名はコミュニティとしてのまとまりが弱く、人間の地理的イメージの中での印象が薄い。そこで、区の境界線は弱く描くことが望ましい。

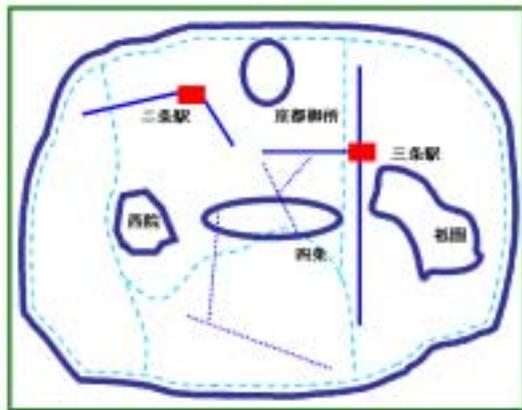


図5 「へ/二」格頻度を利用したイメージマップ  
Fig.2 Using the Difference between “e” Case and “ni” Case in Drawing Image Maps

## 5. まとめと今後の課題

本研究ではウェブ上の位置説明文から重要な地名を抽出し、特徴付けを行なう方法を示した。また、得られた特徴をイメージマップ描画に利用する手法を示した。格助詞・相対名詞・形容詞といった局所的なコンテキストを利用することでノイズの混入を最小限に抑えたが、今後は広い範囲での文脈を扱うことも求められる。また、他言語において同様な傾向が現れることを確認し、提案されたモデルを裏付けていくことが望まれる。

## [謝辞]

科学技術振興事業団の戦略的基礎研究推進事業(CREST)に深謝する。

## [文献]

- [1] 国立国語研究所, 日本語における表層格と深層格の対応関係(国立国語研究所報告 113), 国立国語研究所, 1997
- [2] 田中茂範, 松本曜, 空間と移動の表現(日英語比較選書), 中右実編, 研究社出版, 東京, 1997.
- [3] 中右実, “場所の『に』と『で』: 日英語空間の認知地図” 英語青年, vol.140, no.2, pp.28-30, 1994
- [4] 中右実, 認知意味論の原理, 大修館書店, 1994
- [5] 中右実, 西村義樹, 構文と事象構造(日英語比較選書), 中右実編, 研究社出版, 東京, 1998.
- [6] 中村豊, 岡本耕平, メンタルマップ入門(地理学選書), 古今書院, 1993
- [7] 仁田義雄, “日本語の格をめぐって”, 日本語の格をめぐって, 仁田義雄, くろしお出版, 東京, 1993.
- [8] 牧野成一, ことばと空間, 東海大学出版会, 1978
- [9] 山梨正明, 認知文法論, ひつじ書房, 1995.
- [10] T. Bittner, “On Ontology and Epistemology of Rough Location”, Spatial Information theory - COSIT 99, Hamburg, Germany, 1999.
- [11] S. Borgo and N. Guarino, “An Ontological Theory of Physical Objects”, Proc. of 1997 Workshop on Qualitative Reasoning (QR 97). Cortona, Italy, 1997.
- [12] G. Lakoff, Woman, Fire, and Dangerous Things: What Categories Reveal about the Mind, University of Chicago Press, Chicago, 1987.
- [13] T. Tezuka, R. Lee, H. Takakura, and Y. Kambayashi, “Web-Based Inference Rules for Processing Conceptual Geographical Relationships”, Proc. of the 2nd Int. Conf. on Web Information Systems Engineering, The 1st Int. Workshop on Web Geographical Information Systems, Kyoto, pp.375-384, Dec. 2001.
- [14] M. Worboys, “Metrics and Topologies for Geographic Space”, Advances in Geographic Information Systems Research II: Proceedings of the International Symposium on Spatial Data Handling, Delft, Kraak, M.J. and Molenaar, M. (eds.), International Geographical Union, pp. 7A.1-7A.11, 1996.

## 手塚 太郎 Taro TEZUKA

京都大学修士課程在学中。2001 京都大学総合人間学部卒業。地理情報システムの研究・開発に従事。情報処理学会学生会員。日本データベース学会学生会員。

## 李 龍 Ryong LEE

京都大学大学院博士後期課程在学中。2001 京都大学大学院情報学研究所修了。ウェブデータベースシステム・地理情報システムの研究・開発に従事。

## 高倉 弘喜 Hiroki TAKAKURA

京都大学大型計算機センター助教授。1995 京都大学大学院工学研究科情報工学専攻博士課程修了, 工学博士。高信頼データベースシステム・地理情報システムの研究に従事。情報処理学会会員。

## 上林 弥彦 Yahiko KAMBAYASHI

京都大学情報学研究所教授。1970 京都大学工学部電子工学科博士課程修了, 工学博士。データベースシステムの研究・開発に従事。情報処理学会フェロー。日本データベース学会会長。