

Web の内容解析に基づく地理的領域の特性付けと 地域情報検索への応用

手塚 太郎[†] 李 龍[†] 高倉 弘喜[‡] 上林 弥彦[†]

[†] 京都大学情報学研究科 〒606-8501 京都府京都市左京区吉田本町

[‡] 京都大学学術情報メディアセンター 〒606-8501 京都府京都市左京区吉田本町

E-mail: [†] {tezuka, ryong, yahiko}@db.soc.i.kyoto-u.ac.jp, [‡] takakura@media.kyoto-u.ac.jp

あらまし World Wide Web (以下、WWW) は日常的な情報収集の手段として広く用いられているが、地域関連情報の検索はその重要な一角を占める。本研究では WWW 上のテキストデータに対する内容解析によって人間の地理空間認知の構造を明らかにし、地域情報検索の効率化に役立てる。自然言語において地名は格助詞を伴って現れることが多い。そこで格助詞の意味分析を行ない、共起しやすい格助詞の種類によって地名を特性付けした。地名には狭義と広義のふたつの意味を持つものが多く、解釈において困難を生じさせているが、その判定に格助詞との共起パターンを用いられることを示した。

キーワード 地理空間認知, WWW, 地域情報検索, 自然言語表現

Characterization of Geographic Regions Using the Web and an Application to Local Information Search

Taro TEZUKA[†] Lee RYONG[†] Hiroki TAKAKURA[‡] Yahiko KAMBAYASHI[†]

[†] Graduate School of Informatics, Kyoto University Yoshida-Honmachi, Sakyo-ku, Kyoto shi, Kyoto,

606-8501 Japan [‡] Academic Center for Computing and Media Studies, Kyoto University

Yoshida-Honmachi, Sakyo-ku, Kyoto shi, Kyoto, 606-8501 Japan

E-mail: [†] {tezuka, ryong, yahiko}@db.soc.i.kyoto-u.ac.jp, [‡] takakura@media.kyoto-u.ac.jp

Abstract World Wide Web is now widely used as a tool for daily information search. One important usage of WWW is to search local information. Our paper discusses human cognition on geographic space. In Japanese text data, most place names appear with case intensifying particles. We performed analysis on case intensifying particles and characterized place names based on which particles it is likely to co-occur with. There are many place names that have more than one meaning: the original meaning and the extended meaning, which surrounds the original. We showed that the characterization based on case intensifying particles could be used to determine such duplication of the meaning.

Keyword Geographic space cognition, WWW, local information search, natural language expression

1. はじめに

World Wide Web (以下、WWW) は日常的な情報収集の手段として広く定着しているが、その重要な用途のひとつとして地域情報の検索がある。

効率的な地域情報検索を可能とさせるため

には、人間が地理空間をいかに認識しているかの把握が必須である。たとえば自然発生的な地名には明確な境界線がなく、周辺に行くに従って使用頻度が減少するという構造を持つ。また、「近く」「そば」等の言葉で表される地理的近傍の概念は、地理的クエリにおいて重要な役割を果たしているが、日本語においては省略されて

しまうことも多い。

そこで我々の研究では WWW 上のテキストを解析し、地理空間認知の特徴を調べた。自然言語の文において地名は格助詞を伴って現れることが多い。それぞれの地名について、格助詞と共起する割合によって特性付けを行なった。

2. 地名の意味範囲の内部構造

地名が意味する範囲は固定的ではなく、通時的な変化を受ける。ここでは特に、言語学において用いられる「メトニミー」の概念を用いて地名の意味変化の過程を追う。

2.1. メトニミーによる地名の意味範囲の拡張

形状の類似性に基づいて単語の意味が拡張される現象を、メタファーと呼ぶ。たとえば、「京都市の財政はパンク寸前だ」という表現は、「パンク」という単語をメタファーによって拡張して用いている。

これに対しメトニミーとは、対象間の近接性に基づいて語彙の意味が拡張することを指す [11]。たとえば、「ドンブリを平らげた」は、器によってその中身を指示するメトニミーである。

日常的に用いられるメトニミーの例として、表 1 のようなものが挙げられる。

- | |
|--|
| (1) 目にクマをつくる
(2) 机に座って手紙を書く。
(3) コップに花びらが浮いている |
|--|

表 1. メトニミーの例

(1)は、目に直接クマが出来たのではなく、目の周辺にクマが出来たことを意味する。また(2)では机に座っているのではなく、机の前に座っている。(3)はコップではなく、コップの中の水に花びらが浮いている。

地名が意味する範囲の拡張においても、メトニミーは重要な役割を果たす。

道路 両側に拡張 「四条の古本屋」 「烏丸の郵便局」
交差点 周囲に拡張 「河原町三条の喫茶店」 「四条木屋町のラーメン屋」
自然地形 周囲に拡張 「鴨川で昼寝をする」 「桂川でバーベキューをする」 「宝ヶ池の国際会議場」 「深泥池の病院」
ランドマーク 周囲に拡張 「嵐山の土産物屋」 「三条大橋のスターバックスコーヒー」

表 2. メトニミーによる意味範囲拡張の例

図 1 は、メトニミーによって意味範囲が拡張するプロセスを図示したものである。

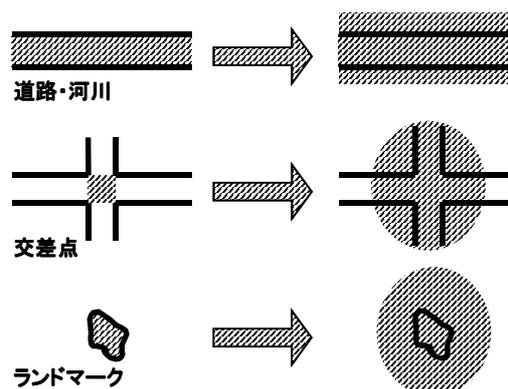


図 1. 意味範囲拡張のプロセス

これらの現象は次節で述べるように、地名の後に続く近傍表現が省略された結果と見られることも出来る。

2.2. 近傍関係の明示と暗示

メトニミーを含む表現は、品詞の後に「の近く」などの近傍表現を付け加えることによって、通常の表現に直すことが出来る。しかし、この省略はあまりにも一般的であって、「の近く」を付け加えると、逆に一般的でない表現になることもある。以下は、Google による検索結果である。

宝ヶ池の国際会議場	20 件
宝ヶ池の近くの国際会議場	0 件
鴨川のカップル	41 件
鴨川の河原のカップル	0 件

前節で挙げた例についても、近傍表現を補うことで本来の形に戻すことができる。

「四条の古本屋」 「四条に面した古本屋」
「鴨川で昼寝をする」
「鴨川の河原で昼寝をする」

メトニミーによる近傍表現の省略は、特に日本語において著しい。英語では near あるいは beside といった前置詞で表現されるところが、日本語ではメトニミーの結果、格助詞のみの出現となる。よって日本語において地名が指す領域を適切に扱うためには、メトニミーに対する理解が不可欠である。

2.3. 地名概念のコアモデル

メトニミーによる地名の意味範囲の拡大は、一回限りの出来事ではなく、漸進的に進むと考えられる。その一方で、地名が本来意味していた部分は最後まで、狭義の意味範囲として維持される。

「嵐山」：広義：京都市西京区に広がる観光地
狭義：大堰川右岸に位置する
標高381メートルの山
(広義の嵐山に含まれる)

メトニミーによる意味拡張が生じた後、狭義の意味として残っている中核部分を、「コア」と呼ぶことにする。そしてその周辺に、その地名を使うことの妥当性の度合いを表す等高線を描くことができる。

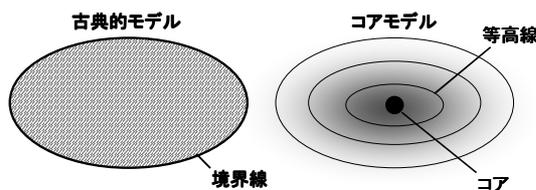


図2：古典的モデルとコアモデル

図2に描かれた等高線は便宜的なものであり、その線を区切りとして離散的な変化が起こるわけではない。妥当性の強さは、連続的に減衰していくものとしてモデル化される。

コアモデルは、地名の意味する範囲がメトニミーによって拡張していく歴史的背景に考慮したモデルであると言える。

2.4. 行政的地名と自然発生的地名

「京都市」や「左京区」など、行政的に定められた地名は境界線が確定しており、地表上の各点はそれに含まれるか含まれないかの二通りしかない。そのような領域は上記の古典的モデルに対応しよう。

一方、自然発生的な地名は明確な境界線を持たない。建造物や河川沼沢などを除き、境界線がひとつに定まっているという例は少ない。例を挙げるならば、「祇園」「西陣」と呼ばれる地域は、明確な境界線を持たない。

山や丘陵においても、境界線が等高線等によって確定されているわけでもなく、どこからどこまでが山なのかは不明瞭である。「富士山はどこから始まるのか」といった議論は、少なくとも日常的な用法に関する限り無意味である。

コアモデルは、このような状況を適切にモデル化するものと期待される。

3. 格助詞との共起を用いた地名の特性分析

WWW 上には多くの地名が出現するが、それらが文の中で現れる際、格助詞を伴うことが非常に多い。そこでまず、地理空間認知と関わりのある格助詞について分析し、またそれらが実際に WWW で地名と共に現れる頻度を調べた。

なお、ここでの共起は地名と格助詞とが接触した形で現れることを意味する。

3.1. 格助詞による位置関係の表現

前置詞によって対象物間の位置関係を示す英語とは対照的に、日本語においては「から/まで/へ/に/で/を」などの格助詞によって位置関係が示される。ただし、「から/まだ/へ」の意味は確定し易いのに対し、「に/で/を」の意味は共起する単語の語彙情報に大きく依存する。前者に関しては以下のような意味付けが可能である。

から： 起点
 まで： 限界点
 へ： 方向

一方、格助詞「に／で／を」の役割は多岐に及ぶ。その働きには少なくとも以下のものが含まれる。

に： 着点・原因・方式・位置
 で： 道具・原因・位置
 を： 行為が作用する対象・位置

以下では特に「に／で／へ」を用いた空間的な位置指示に関して比較を行なう。

3.2. 「に」と「で」の相違

格助詞「に／で」は、いずれも位置を示すのに用いられるが、使い分けが存在する。すなわち、「に」は状態の位置を示すのに対して、「で」は過程あるいは行為の位置を示す。以下では、述語を状態／過程／行為の三類型に分ける[7]の分類に基づいて分析を進める。なお、印は、その選択が日本語として相応しくないことを意味する。

状態： 阪急デパート[で／に]沢山の客がいる。
 過程： 阪急デパート[で／ に]安売りが始まる。
 行為： 阪急デパート[で／ に]買い物をする。

過程あるいは行為において「に」が用いられた場合、それはもっぱら着点を意味する。

行為： 四条[で／に]行く。

さらに、位置を示す「で」と、着点を示す「に」とが併用されることも可能である。

過程： 河原町[で／ に]自動車が電柱[で／に]ぶつかった。

「に」と「で」のいずれも可能であるが、意味に違いが現れる例としては、以下のものが挙げられる。

嵯峨野[に／で]遊ぶ。

どちらも一般的な表現であるが、ニュアンスが異なる。後者は嵯峨野を「遊ぶ」行為が発生する位置としてしか見ていないが、前者における「嵯峨野」は着点／目的であり、文化的総合体としての嵯峨野が含意されている。たとえば嵯峨野のポーリング場で遊ぶことは「嵯峨野で遊ぶ」の例にはなっても、「嵯峨野に遊ぶ」の例とはならない。

以上の分析に基づき、「に／で」との共起頻度の差を、地名の特性付けの一手段として使用することが考えられる。すなわち「に」と頻繁に共起する地名は描写の対象となる場所、「で」と頻繁に共起する地名はイベントや事件が発生しやすい場所、といった意味付けが可能である。表3は京都における代表的な地名等について、以上の数値を集計したものである。

最右列の「比」は、「で」との共起数を「に」との共起数で割った値である。

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3. 代表的な地名での「で／に」との共起

自然地形に比べ、「駅」や「大学」は人間の活動の場となりやすいため、「で」の頻度が高くなっていることが読み取れる。本来はそれほど頻繁に行為の場所となりえない地名(たとえば水域である宝ヶ池など)が「で」と多く共起している場合、そこにメトニミーが生じていることが推察される。

3.3. 「へ」と「に」の相違

格助詞「へ／に」は共に移動に関わり、ある程度まで互換性がある。ただし、「へ」が移動の方向に焦点が置かれているのに対し、「に」は目的地に焦点が置かれるという違いがある。すなわち「へ」においては話者の心理的視点が出発点にあるのに対し、「に」においては到達点にある。

京都の代表的な地名等における「へ／に」との共起頻度を、表4に示す。

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4. 代表的な地名での「へ／に」との共起

3.4. WWWにおける格助詞の使用頻度の差

前節までの分析を実データに基づいて裏付けるため、WWW上のテキストにおいて様々な地名が格助詞とどのような頻度で共起しているのかを調べた。検索エンジンとしてはGoogleを用いた。

図3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の自然地形名(山、川、池の三分野より)について、「で」「に」それぞれとの共起数を散布図で表したものである。

明らかな特徴としては、区名・寺社名・自然地形において「に」が多く、駅名・大学名において「で」が多い。

これは駅および大学において人間の活動が頻繁に行なわれているためだと考えられる。寺社は目的地としては重要であるが、そこで

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」等、何らかの行為を行なうことは少ない。自然地形はそれについて言及されることは多くとも、実際に行う行為の場となることは少ないため、相対的に「で」の頻度が低くなっている。

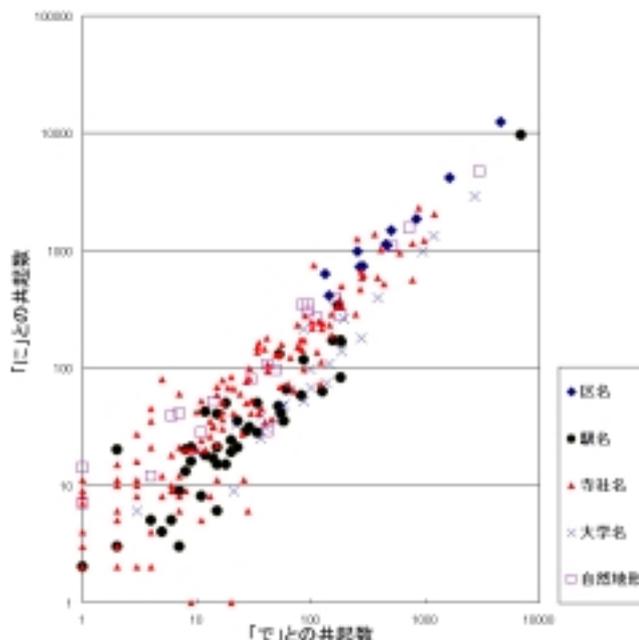


図3: 「で／に」との共起数の分布

地名のグループ	グループの平均	グループの分散
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5: 「で／に」との共起数の傾向

続いて図4では、上記と同じ地名集合について、「へ」「に」それぞれとの共起数を示している。まず、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「に」が多く、駅名において「へ」が多いことである。

駅は、移動におけるランドマークとしての役割を果たす。一方、自然地形は、それについて言及されることは多くても、移動の方向を表現する際に使われることは少ない。区名において「へ」が少ないのは、市などと異なり文化的一体性に乏しく、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。この結果のひとつの応用として、「へ」との共起数によって、それぞれの地名が持つコミュニティとしてのまとまりの強さを推定する、といった展開が考えられる。

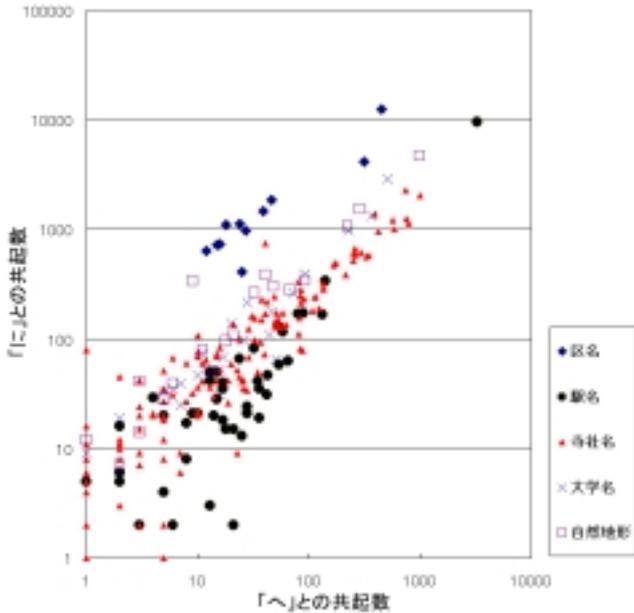


図 4: 「へ/に」との共起数の分布

地名のグループ	グループの平均	グループの分散
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 6: 「へ/に」との共起数の傾向

4. まとめと今後の展望

本研究では人間の地理空間認知の中でも特にメトニミーによる意味範囲の拡張について論じ、また、格助詞との共起数の差によって地名の特性付けを行なった。

ここで用いられたデータは地名直後の格助詞という非常に局所的なものであり、それゆえノイズの混入も最小限に抑えられたが、今後はさらに離れた品詞、あるいは助詞をも分析の対象とし、より多くの情報を取り込んでいく必要がある。

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, 研究者出版, 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, “『に』と『で』の棲み分け: 日英語の空間認識の型(1)” 英語青年, vol.140, no.10, pp.20-22, 1994
- [5] 中右実, “『に』と『で』の棲み分け: 日英語の空間認識の型(2)” 英語青年, vol.140, no.11, pp.22-24, 1994
- [6] 中右実, “『に』と『で』の棲み分け: 日英語の空間認識の型(3)” 英語青年, vol.140, no.11, pp.22-24, 1994
- [7] 中右実, 認知意味論の原理, 大修館書店, 1994
- [8] 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
- [9] 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, くるしお出版, 東京, 1993.
- [10] 牧野成一, ことばと空間, 東海大学出版会, 1978
- [11] 山梨正明, 認知文法論, ひつじ書房, 1995.
- [12] T. Bittner, “On Ontology and Epistemology of Rough Location”, Spatial Information theory - COSIT 99, Hamburg, Germany, 1999.
- [13] S. Borgo and N. Guarino, “An Ontological Theory of Physical Objects”, Proc. of 1997 Workshop on Qualitative Reasoning (QR 97). Cortona, Italy, 1997.
- [14] G. Lakoff, Woman, Fire, and Dangerous Things: What Categories Reveal about the Mind, University of Chicago Press, Chicago, 1987.
- [15] 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. (The Best Paper Award)
- [16] 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.