

E-43

格助詞解析に基づくランドマークの特性付け

Landmark Characterization Based on Intensifying Particles

手塚 太郎 李 龍 高倉 弘喜 上林 弥彦

Taro TEZUKA Lee RYONG Hiroki TAKAKURA Yahiko KAMBAYASHI

1. まえがき

地域情報の検索は World Wide Web (以下、WWW) の重要な用途のひとつである。効率的な地域情報検索を可能とさせるためには、人間が自分たちの地理空間をいかに認識しているかを明らかにする必要がある。我々の研究は WWW 上のテキストを解析し、地理空間の認知情報を抽出する手法を提案する。

2. 自然言語テキストからの地理情報抽出

2.1 近傍表現を用いたランドマークの特性付け

人間の地理空間認知においては、ランドマークが重要な役割を果たす。

たくさんの地名の中から、どれがランドマークとなっているのかを明らかにするためには、自然言語テキストにおける近傍表現の多寡を調べるという方法がまず考えられる。すなわち「Aの近く」という表現が頻繁に用いられるようなAは、認知上重要な役割を果たしていると考えられる。

「Aの近く」という表現は、英語における“near A”あるいは“by A”などと対応させるのが一般的である。

ところが以下の例に見られるように、日本語においては「の近く」が省略されることが極めて多い。

宝ヶ池の近くの国際会館 ⇒ 宝ヶ池の国際会館

(英文: Kyoto International Conference Hall
[near / by / beside] the Takaragaike Pond.)

日本語の場合、「の近く」を省略した表現の方がむしろ広く使われている。英語の場合、地名間の関係は前置詞によって明示されるが、日本語の場合、地名間の位置関係は明らかにされないことが多い。

このため、近傍表現の多寡を用いた特性付けは、日本語においては容易でない。

2.2 メトニミーによる意味範囲の拡張

前節での観察を一般化する。メトニミーとは、空間的近接性に基づく意味範囲の拡張を意味する言語学の用語である。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。「の近く」の省略もまた、メトニミーの一貫と捉えられる。「頭が良い」は、頭と近接する脳の働きについて述べている。地名における近傍表現の省略も、この一環と捉えることが出来る。

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」
「深泥池の病院」

ランドマーク→周囲に拡張

「清水寺の坂」
「三条大橋のスターバックスコーヒー」

表 1. メトニミーによる意味範囲拡張の例

さらに、メトニミーが慣用化されることによって、地名が二重の意味構造を持つようになることも多い。例として、京都の「嵐山」などが挙げられる。

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 381m の山

3. 格助詞との共起数を用いた特性付け

WWW のテキストには多くの地名が出現するが、それらは格助詞を伴うことが非常に多い。また前章で述べたように、日本語においては近傍表現が格助詞に吸収される。そこで地名に伴う格助詞を比較し、それぞれが共起する頻度について分析を行なった。

3.1 「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、「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。

状態: 銀閣寺にたくさんの観光客がいる。

過程: 銀閣寺で夜間拝観が始まる。

行為: 銀閣寺で写真を撮る。

表 3 は京都等における代表的な地名について、「で/に」の頻度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。

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。本来はそれほど頻繁に行為の場所となりえない地名（たとえば水城である宝ヶ池など）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。実際、

† 京都大学情報学研究科

‡ 京都大学学術情報メディアセンター

宝ヶ池は本来は池を指すが、現在ではそれを含む広い地域を指すように意味範囲が拡張している。

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
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 2. 代表的な地名における「で／に」との共起

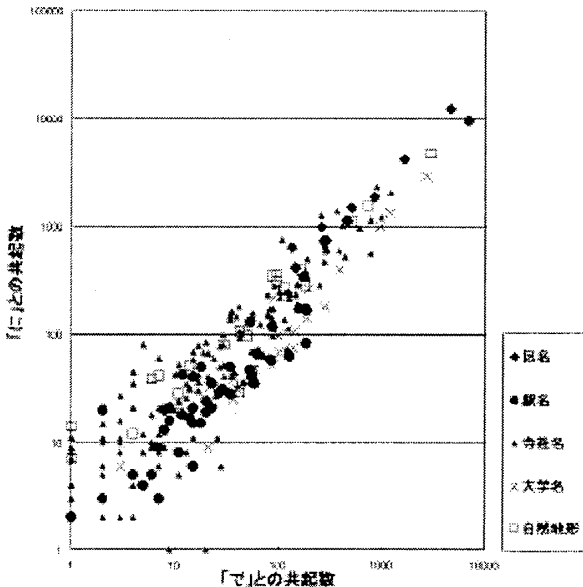


図 2: 「で／に」との共起数の分布

グループ	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 3: 「で／に」との共起数の傾向

3.2 「へ」と「に」の相違

格助詞「へ／に」は共に移動に関わるが、「へ」においては移動の方向に心理的注目が置かれているのに対し、「に」は目的地に注目が置かれるという違いがある。

図 3 と表 4 は「へ／に」における共起を表したものである。区名・寺社名・自然地形において「に」が多く、駅名において「へ」が多いことが読み取れる。

実際、駅は移動におけるランドマークとしての役割を果たす。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。ここから、「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、といった応用が考えられる。格助詞の分析は単にランドマークを抽出するだけでなく、その特性をも明らかにできる。

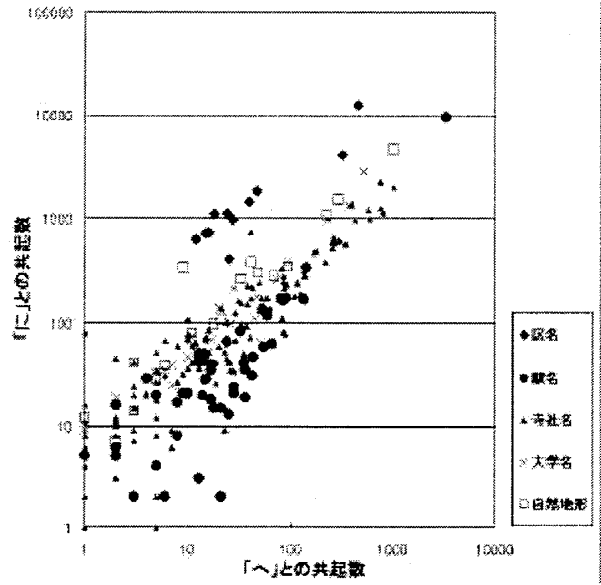


図 3: 「へ／に」との共起数の分布

グループ	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／に」との共起数の傾向

4. まとめと今後の展望

本研究では WWW 上の自然言語テキストを用いたランドマークの抽出および特性付けを目的として、地名が格助詞と共起する頻度を分析した。

今回の解析に用いられたのは地名の直後に現れる格助詞という局所的なものであり、今後は文中の離れた品詞および助詞をも分析の対象とする必要がある。

文 献

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)