

料理とスイーツにおけるオノマトペの違いの分析

Comparison of Cuisine and Sweets using Onomatopoeia

廣川佐千男*1

Sachio Hirokawa

中藤哲也*1

Tetsuya Nakatoh

殷成久*1

Yin Chengjiu

鈴木孝彦*1

Takahiko Suzuki

橋本喜代太*2

Kiyota Hashimoto

*1九州大学情報基盤研究開発センター

Research Institute for Information Technology, Kyushu University

*2大阪府立大学現代システム科学域

College of Sustainable System Sciences, Osaka Prefecture University

Since evaluation of food is subjective and the feeling is expressed in Japanese, onomatopoeia is used in many cases. In cooking and sweets, the onomatopoeia used is different. In order to grasp the meaning of onomatopoeia, we need to read and compare the situation where each onomatopoeia is used. In this report, the "Double Rank" method which the first author proposed is applied to Blog documents to compare and to analyze Blogs.

1. はじめに

インターネットの発展により多くの人々が個人的経験や意見を Blog として発信できるようになった。企業や公的機関からの発信される情報は、事実を多く含み、一定の客観性が担保されていると考えられるのに比べ、個人の Blog 記事は、その人の経験や体験に基く主観的で感性的な意見が多いという特徴がある。しかし、そのような多数の個人的意見を分析することで、公的情報からは推定が難しい世の中の変化を読むことが期待されている。本稿では、そのような典型的な事例として観光情報に関連する Blog 記事に着目し、特に、食べ物についての感性的な意見を分析する方法を提案する。

Blog 記事に含まれる評判情報の分析についての研究が多くなされている。例えば、[鍛冶 2007] では、HTML 文書集合から自動構築した評価文コーパスから評価表現辞書を構築している。このような辞書を用いることで、記事の極性判定が可能となる。[小林 2003] では、商品についての評価と極性を合せた分析を行っている。

ところが、感性的な記事の中には、かならずしもポジティブとネガティブという極性だけではとらえない事柄もある。食べ物についての意見はその典型といえる。好きか嫌いかを表している、その表現には多様性がある。また、同じ表現でも、文脈によってポジティブとネガティブの極性が違うこともある。日本語の特色の一つでもあるオノマトペはそのような感性を表す典型的な単語である [橋本 2011]。

筆者らは、[Hirokawa 2011] において、1つの検索入力に対して、二つの観点から検索して結果をマトリクス状で示す Double Rank 法を提案している。特徴語を選択する観点として、オノマトペを使うことでポジティブとネガティブでは捉えられない微妙な違いの分析が可能であること示した。具体的には [Hirokawa 2011] のシステムでも、マトリクス上に表示されたキーワードを見ることで、料理とスイーツの違いが現れるオノマトペや、類似したオノマトペでも微妙に関連する料理が異なることが分る、というような分析のヒントが得られた。しかし、二種類の食べ物で検索してそれぞれ異なるオノマトペがマトリクスに表示されても、それらの単語が具体的にどの

ような状況で使われているかを確認することはできなかった。筆者らは、これとは別に、[Yin 2010] において検索結果の順位も着目する観点で異なるり、観点を表す方法として単語集合を限定する検索方式を提案した。検索結果のスニペットを表示するとき、その着目観点の単語を強調表示することで、それぞれの文書がなぜ検索結果の上位に現れるかが納得してもらえらる。本論文では、この二つの手法を融合することで、二つの観点に応じて検索結果の概観がわかり、かつ、その詳細な比較を具体的な文章で確認する方式を提案する。

2. ダブルランク法によるマトリクス表示

[Hirokawa 2011] で提案したダブルランク法では、二つの観点を検索語 u 、 v としてして表し、検索語 q について qu ならびに qv の AND 検索を行う。得られる検索結果の文書集合に現れる単語 w について、一つ目の観点での順位が $r_1(w)$ 、二つ目の観点での順位が $r_2(w)$ のとき、マトリクスの $(r_1(w), r_2(w))$ の位置に、 w を含む文書数の組 $(n_1(w), n_2(w))$ を表示した。具体的に表示されるのは、上位 K 個の特徴語に限定した。殆どの単語がどちらの検索結果でも同じ順位にあれば、マトリクスの対角線上に文書数の組は並ぶ。ところが、一方で上位だが他方では下位だと、対角線から離れたマトリクスの端に表示される。このように、二つの観点について、何が同じで何が違いかを概観することができる。

3. 検索事例の紹介と考察

本稿では、[Hirokawa 2011] で使った 1303 件のブログ記事を対象とする。これらはすべて九州地域の観光に関するもので、九州観光推進機構^{*1}により人手で集められたものである。

3.1 「ラーメン」と「うどん」

図 1 は「ラーメン」と「うどん」をオノマトペを特徴語として分析した結果のマトリクスである。

「しっかり」「シャキシャキ」「まるやか」「カリカリ」などはほぼ対角線上にあり、どちらにも共通の特徴語となっていることが分る。一方、「こってり」は対角線上から離れた上の部分にあり、「こってり」はラーメンの上位語だがうどんに関連

連絡先: 廣川佐千男, 九州大学情報基盤研究開発センター, 福岡市東区箱崎 6-10-1, hirokawa@cc.kyushu-u.ac.jp

*1 <http://www.welcomekyushu.jp>

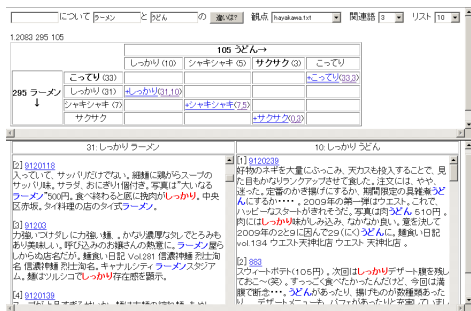


図 1: オノマトベによる「ラーメン」と「うどん」の比較

しては上位に現れないことがわかる。また、「サクサク」「ふんわり」「パリッ」は、下の部分にあり、うどんの特徴語と分る。マトリクス内に表示される結果は [Hirokawa 2011] では単に左上端を両者で上位に来る語、右ないし下に行くほど登場順位が落ちる語として表示するだけであったが、本研究においては検索結果の詳細を表示するスニペットにジャンプするインタフェースとする上で、それぞれでの具体的な頻度も「+こってり (33,3)」のように表示されるようにしている。詳細表示のスニペットでは、左右で対照的に詳細結果を閲覧でき、キーワードは色を変えて一瞥して分かるようにしている。また「(33,3)」の部分をクリックすることで、下のフレームに「スープは、割とこってりしていて濃い感じ」という文が表示される。この結果を観察することにより、同じ「しっかり」であっても何が「しっかり」と形容されているのかが分かる。

3.2 「シャキシャキ」と「サクサク」

3.1 の事例では「うどん」や「ラーメン」のように事物を検索語として、あらかじめ共起対象語として観点リストに登録してあるオノマトベを結果として表示しているが、観点リストとして逆に事物のリストを利用することで、どのようなオノマトベがその描写に利用されているかを調べることができる。図 2 は「シャキシャキ」と「サクサク」について検索を行なった結果例である。

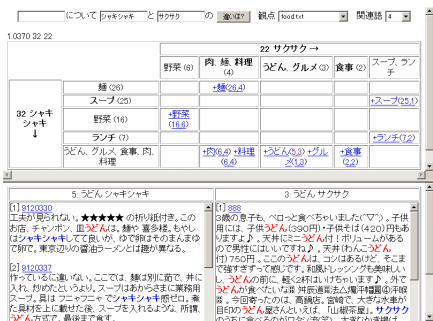


図 2: 料理語による「シャキシャキ」と「サクサク」の比較

3.3 考察

本研究は検索結果をどのように提示すれば検索者がより発見的な知見を得られるかを追及するものであり、オノマトベが具体的に何を描写しているかを言語処理的に分析することは現在行っていない。このため、図 1 に見られるように、「しっ

かり」が「ラーメン」と共起していると言っても、[2] の例では「挽肉がしっかり = たくさん存在する」であるのに対して、[3] の例では「麺がしっかり存在感を顕示 = 具材の中での顕著感が高い」となっているが、こうした違いを予め分析して、描写対象語別に色分けするなどには行っていない点は将来的な課題の一つである。

一方、特に外国人日本語学習者をターゲットに考えると、図 2 の「うどん」「シャキシャキ」の [1] と [2] は示唆的な結果を含んでいることが分かる。すなわち、[1] では「もやしはシャキシャキしてて」と、「うどん」において「シャキシャキ」が使われるのは「うどん」そのものではなく、「もやし」などであることが分かる。一方、[2] では「具は…シャキシャキ感ゼロ」とあるから、[1] と合わせれば、「うどんという料理は麺だけでなく、いろいろな具が入っていることがあり、その一つはもやしだ」、「うどんで言語表現としてのシャキシャキが使われるのは、もやしなどの具材に対してである」という 2 つの類推が容易に得られることになる。

このように本システムの一つの適用事例としては、日本語表現と日本の事物双方について未知が多い外国人日本語学習者に対して、オノマトベ表現の意味と日本の事物両方を同時並行的に学習するための素材を提供するというものが考えられる。言語である日本語と文化風俗である日本の事物は往々にして教育現場では分けてしまうが、学習者にとってはその両輪がともに学習できる方が興味の維持という点でも学習の効率性という点でも望ましいと考えられる。この応用事例については外国人日本語学習者に本システムを試用してもらうことで検証することを予定している。

4. まとめ

本稿では、二つの観点による違いをマトリクスマップとして表示し、さらに詳細な文を提示する検索手法を提案した。観光関連 Blog における食べ物についての分析を行い、特にオノマトベによる違い、あるいはオノマトベの違いが分析できることを示した。

なお、本研究は科研費 (24500176) の助成を受けたものである。

参考文献

[鍛冶 2007] 鍛冶伸裕, 喜連川優, 自動構築した評価文コーパスからの評価表現辞書の構築日本データベース学会論文誌, Vol.6, No.1, pp41-44, 2007.

[小林 2003] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, テキストマイニングによる評価表現の収集, 情報処理学会研究報告 NL154-12,77-84, 2003

[橋本 2011] 橋本喜代太, 竹内和弘, ジャンル, 文脈に応じたオノマトベ表現の感性極性値の推定手法の提案, 人工知能学会第 25 回年次大会, 1C2-OS4b-6, 2011

[Hirokawa 2011] S. Hirokawa, C. Yin, K. Hashimoto, K. Takeuchi, Search and Analysis of Gourmet Blogs with a Particular Reference to Onomatopoeia, ICIC Exp. Let., Vol. 5, No.8(B), pp.2971-2978, 2011

[Yin 2010] C. Yin, T. Nakatoh, S. Hirokawa, X. Wu, J. Zeng, A proposal of search engine XYZ for tourism events, Proc. JCAI, Vol.1, pp.178-181, 2010