

## 共起語ネットワークに基づいたシズルワードの特性分析

## An analysis of sizzle words based on co-occurrence networks

笠井信吾<sup>\*1</sup>  
Shingo KASAI齊藤史哲<sup>\*1</sup>  
Fumiaki SAITOH石津昌平<sup>\*1</sup>  
Shohei ISHIZU<sup>\*1</sup> 青山学院大学・理工学部・経営システム工学科  
Department of Industrial and Systems Engineering, Aoyama Gakuin University

In recent years, Sizzle Words have been attracted attention as it is important in marketing. In this study, we analyze Sizzle Words based on the characteristics of co-occurring word networks, as a novel approach for conventional semantic-based research. We attempt to visualize the properties of sizzle word which could not be grasped by conventional analysis by cluster analysis using co-occurrence word network features such as mediocentricity and order centrality. It can be possible to grasp the contents by detecting the Sizzle word which is used with similar properties, despite the meaning and nuance being completely different.

## 1. はじめに

近年、食におけるマーケティングにおいてシズル感に注目が集まっている。シズル感とは、米国の経営コンサルタント E.Wheeler によって提唱された概念であり、消費者の五感を刺激することによって食欲や購買意欲を刺激する感覚である。これは商品のデザインや広告、製品コンセプトの策定などの要因と深く結びついており、マーケティング周辺分野を中心に重要視され始めている。

これに対して、シズル感に直結する曖昧な質感のニュアンスを端的かつ的確に表現する用語としてシズルワードがある。これは、日常会話のもとよりツイートやレビューなどといった、製品・サービスやレシピなどを対象とした評価に関わる Web 上への書き込みにおいても広く利用されている。シズルワードはオノマトペなどの感覚的な表現とも密接に結びついており、特に日本語はオノマトペの種類が豊富であることから、これらに関する知識は製品のパッケージや広告においても重要な役割を果たすことが期待されている。

このような背景から、シズルワードは様々な角度から分析がなれており、盛んに研究がなされている。シズルワードに関する既往の研究に目を向けると、シズルワードの用法や印象などに関するアンケート調査をベースとした研究[石橋 15]と、実際の用例であるテキストデータの解析をベースとした研究[渡辺 15, 森田 17, 長谷川 18]に大別できる。前者は実際に複数の被験者を募り、アンケートを通じたシズルワード(あるいはオノマトペなどの感覚的な表現)に関する結果を定量的に解析することで、その意味構造の解明を目指している。対して後者は、用例に基づくテキストデータを解析することによって、それぞれのシズルワードが持つ意味構造の解明を目指している。

これらはいずれも、アプローチや目的が異なるものの、利用者が抱く印象や感覚、利用対象などといったあいまいな言語に対する意味の類似性が単語間の類似性に影響するという前提に基づいた解析やツールの開発である。よって、これまでの研究の多くは、アンケートベース、文書解析ベースいずれにおいても、シズルワードのニュアンスや意味構造を解析に軸足を置いた研究であったといえる。曖昧さを含む言語に対する意味・ニ

ュアンスの解析は最重要項目の一つであるが、関連研究の大多数がこのアプローチに関するもので、たとえば、共起した単語間の関連性やその傾向といった、単語の使われ方の特徴やその可能性に関する議論は充分になされていない。

そこで、我々はこれまでのアプローチとは異なる新たな視点として、各シズルワードの使われ方に着目した。ひとえにシズルワードといっても、利用範囲が多岐にわたるものや限定的なもの、など様々な切り口に基づいて分類できると考えられることから、これらに関する新たな気づきの獲得を目指し、実際の用例の文書データを対象とした解析を行った。分析においては、共起語ネットワークの特性に着目することで、各シズルワードの性質の解析を行っている。各単語をノードとしたネットワークに対して計算した中心性指標を変数とした分析により、意味解析のみでは把握しきれなかった性質について報告する。

## 2. 基礎事項の準備

ここでは、本研究で行う分析において基礎をなす共起語ネットワークおよびその中心性指標について概説する。

## 2.1 共起語ネットワーク

共起語ネットワークとは、同一文書内において共起した単語をリンクで結んだグラフである。ここでは、各単語がノードであり、リンクが同時出現の関係(共起)を表している。共起語ネットワークを用いた研究に目を向けると、文書データからのキーワード抽出に共起語ネットワークの指標を SVM の特徴量として扱う研究[山本 09]が挙げられる。本研究では、シズルワードの性質の解析においても同様に、共起語ネットワークの特徴量を素性として解析することで新たな知見が得られると考え、共起ネットワークを解析対象とした。

以下本稿中ではノード  $i$  とノード  $j$  の間のリンクに関する情報を記号  $a_{ij}$  で表し、リンクが存在する際は  $a_{ij} = 1$ 、リンクが存在しない際は  $a_{ij} = 0$  と表示する。なお、一般的には共起語ネットワークは無向グラフとして扱われるため、 $a_{ij} = a_{ji}$  となり、 $a_{ij}$  を  $i$  行  $j$  列の要素とした対角行列として隣接行列  $A$  として表現される。 $n$  はネットワーク内に含まれる全ての頂点の数を表している。

## 2.2 ネットワークの中心性指標

ネットワークやその構成要素(ノード)が持つ特徴を表す指標として、中心性指標が挙げられる。中心性指標を解析することに

連絡先: 齊藤史哲, 青山学院大学, 神奈川県相模原市中央区  
淵野辺 1-5-10

よってネットワークの特徴を定量的に把握することが可能になる。ここでは、本研究にて用いる特徴量について整理する。

#### 《次数中心性》

次数中心性は他のノードと繋がるリンク数が多いほど大きな値を持つ指標である。ノード  $i$  に対する次数中心性  $C_d(i)$  はネットワーク内でより多くの他ノードと繋がりを持つノードを高く評価されるように、次式によって定義されている。

$$C_d(i) = \sum_{j=1}^n a_{ij} = \sum_{j=1}^n a_{ji} \quad (1)$$

これは、リンクの多さを評価しているもので、共起語ネットワークにおいては、共起する単語のバリエーションの多さを表している。すなわち、対象のシズルワードの利用範囲の大きさと解釈することができる。

#### 《媒介中心性》

媒介中心性はネットワーク内において他の頂点間の最単経路上に位置する程度を評価する指標である。ノード  $i$  に対する媒介中心性  $C_b(i)$  は次式のように定義されている。

$$C_b(i) = \sum_{j \neq k} \frac{g_{jk}(i)}{g_{jk}} \quad (2)$$

ここでは、 $g_{jk}$  はノード  $j$  とノード  $k$  の間に存在する最短経路数を表しており、 $g_{jk}(i)$  はその内でノード  $i$  を介する経路数を表している。

これは、ノードのクラスタを介する媒介点を高く評価する傾向にあることから、レシピなどのジャンルや領域をまたいで利用することができる可能性を持つシズルワードを高く評価することができる指標と考えられる。

#### 《固有ベクトル中心性》

固有ベクトル中心性は次数中心性を拡張した中心性指標であり、ネットワーク内において他の(次数)中心性が高いノードと繋がりを持つノードを高く評価する指標である。ノード  $i$  に対する固有ベクトル中心性  $C_{ev}(i)$  は隣接行列  $A$  の固有ベクトルに基づいて、次式のように定義されている。

$$C_{ev}(i) = \frac{1}{\lambda} \sum_{j=1}^n a_{ij} C_{ev}(j) \quad (3)$$

中心性の高いノードと繋がりが多いノードを高く評価する中心性であることから、共起語ネットワーク内では、他の重要性の高いキーワードとの結びつきが強いシズルワードを評価していると解釈できる。

#### 《情報中心性》

情報中心性は媒介中心性が最短経路を対象としてノードを評価するのに対して、ノード間の最単経路以外の経路の重みも考慮した中心性指標である。ノード  $i$  に対する情報中心性  $C_{info}(i)$  は経路長に関する情報量  $I_{ij}$  を導入することによって、次式のように定義されている。

$$C_{info}(i) = \frac{n}{\sum_{j=1}^n \left( \frac{1}{I_{ij}} \right)} \quad (4)$$

他の頂点との距離が小さいノードを高く評価する傾向にあることから、直接的に関わるキーワードが多いシズルワードを評価できる可能性がある。

### 3. 分析方法について

本研究では、上記の共起語ネットワークにおける指標およびデータセット内における単語の総出現頻度(TF 値)を全単語に

対して算出し、その中からシズルワードのみを解析対象として選択する。これらの特徴量としてクラスタ分析を試みることで、シズルワードが文書表現において果たす役割の性質を調査する。このように各シズルワードのネットワーク指標を特徴量とすることで、意味内容に捕らわれずに、それら利用傾向や利用方法に関する特徴の把握が可能になる。

これらのデータに対してクラスタ分析を用いることによって、各シズルワードの類似性について議論を行う。クラスタリング方法としてはウォード法を採用し、解析結果をデンドログラムとして表示することでシズルワード間の類似性を可視化する。併せて、特徴量間の類似性についてもデンドログラムとして表示することでクラスタリングの結果に対して解釈を与える。

## 4. 実データに基づいた分析

### 4.1 実験設定

本研究では、NII 情報学研究データリポジトリにおいて楽天株式会社より提供されている“楽天レシピ”のデータセットを解析対象としている。楽天レシピのデータセット内においては、レシピに対するレビューである“つくったよレポート(つくレポ)”のデータおよびレシピ提供者からそれに対する返答がテキストデータとして提供されている。本研究では、隣接行列のサイズを抑えるために提供された約 11 万件のレビューデータの中からランダムに選択した 5000 件を用いている。形態素解析によって抽出した品詞は名詞、形容詞、動詞、副詞である。

本研究にて解析対象としたシズルワードは、[大橋 15]にて紹介されているものから単語リストを作成し、その中でもデータセット中に存在したものを解析対象のシズルワードとしている。本研究は、レシピサイトのレビューデータを扱っていることから、シズルワードとレシピ内容の関連性が重要になる。このため、共起は同一レビュー内に出現した単語を共起関係とみなしている。クラスタリングにおいては、先述の各中心性をそれぞれ標準化し、無単位化したものを特徴量として解析を行う。

### 4.2 解析結果

解析結果の一部を図 1、図 2 に示す。図 1 は中心性指標と TF 値を特徴量としたシズルワードのクラスタリング結果をデンドログラムとして表示したものであり、図 2 はその結果に対する各特徴量の大きさを表している。図 2 クラスタリング結果に対する各特徴量の大きさをヒートマップとして表示したものである。ここでは、左の列から順に固有ベクトル中心性、次数中心性、媒介中心性、TF 値、情報中心性の順に表示されており、その上のデンドログラムは特徴量間の類似度を表している。

また、シズルワードは大橋らによって“食感系”、“味覚系”、“情報系”に類別されている[大橋 15]。表 1 は一部のクラスタとその関係をまとめたものを表している。

### 4.3 まとめ

クラスタリングの結果より、例えば、「甘い」と「たっぷり」は全く意味や用途は異なるものであるが、データセット内では類似した役割を持っている可能性がある。これは次数中心性や媒介中心性が高く、情報中心性が低いことから幅広い用途で利用できるキーワードであることが見て取れる。「さっぱり」や「香ばしい」、「コク」「香ばしい」などは、固有ベクトル中心性や次数中心性が高いが、媒介中心性が低いことから、対象領域が限定的ではあるが、対象領域内の単語の種類が多い単語と解釈できる。

このように、共起語ネットワークの中心性指標に着目してデータ解析を行うことで、シズルワードの使われ方やその応用範囲

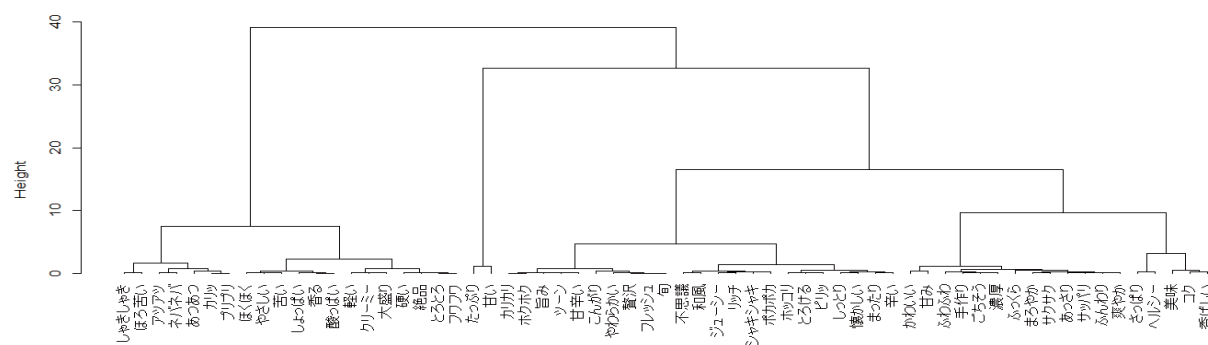


図 1: 共起語ネットワークの特徴量に基づいたシズルワードのクラスタリング結果

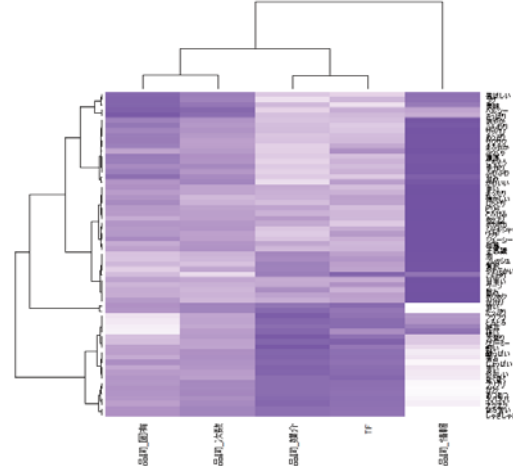


図 2: クラスタの特徴

に対して新たな解釈を与えられると考えている。また、表 1 のように一部のクラスタでは、“食感系”や“味覚系”に偏りがあるものも確認できたことから、この類別化と用法の間には部分的に関係性が認められる可能性がある。

表 1 クラスタリングの結果とシズルワードの関係性  
(食感系を青色, 味覚系を橙色, 情報系を緑色)

クラスタ A	クラスタ B	クラスタ C
しゃきしゃき	さっぱり	かわいい
ほろ苦い	ヘルシー	甘み
アツアツ	美味	ふわふわ
ネバネバ	コク	手作り
あつあつ	香ばしい	ごちそう
カリッ		濃厚
プリプリ		
ほくほく		

5. おわりに

本研究では、意味や用途の類似性に基づいた分析を前提としている一般的なシズルワードの研究に対して、シズルワードの用法の特徴に着目した新たな分析を試みた。方法としては、ネットワークの評価指標であるネットワーク中心性に着目し、共起語ネットワークにおける中心性を特徴量とした分析を行った。共起語ネットワーク内におけるシズルワードの特徴をクラスタ分析により比較することで、意味が全く異なるにもかかわらず、シズ

ル感を表現する上で類似した役割を果たす単語を類別化し、新たな気づきを与えた。

今後は、さらに大規模なデータを解析することやレシピ以外の食品の評価レビューなど、対象データの範囲を拡大する必要があると考えている。また、学習等のより高度なデータ解析を通じて、同様なデータからクラスタを構築する要因に関する知識の抽出を試みる必要がある。これらを通じて、意味が異なるにもかかわらず類似な共起の特徴を持つクラスタの原因や言語構造等を明らかにしていく予定である。

謝辞

実験用のデータセットをご提供いただきました楽天株式会社様に厚く御礼申し上げます。

参考文献

[石橋 15] 石橋賢, 深瀧創, 宮田一乗: 米菓を対象としたシズル語の印象評価, 人工知能学会論文誌, Vol. 30, No. 1, pp. 229-236, (2015)

[大橋 15] 大橋正房, 光岡祐子, 須藤正彦, 北村奈津子, 米山真知子, 猿渡一兵: sizzle word シズルワードの現在「おいしいを感じる言葉」調査報告, BMFT 出版部, (2015)

[長谷川 18] 長谷川永奈, 小宮香乃, 齊藤史哲, 石津昌平: 食における言語資源に基づいたシズル感に関する因子情報の抽出, 日本感性工学会論文誌, Vol. 17, No. 2, (2018)

[森田 17] 森田真季, 荒牧英治, 灘本明代, 宮部真衣: マイクロブログにおける“おいしさ”情報分類器の適用可能性の検証, 信学技報 (LOIS) Vol. 116 No. 488, pp. 229-234, (2017)

[山本 09] 山本優樹, 折原良平: 共起語ネットワーク特徴の言語・文書種非依存性に基づくキーワード抽出と見出し語の予測による性能評価, 人工知能学会論文誌, Vol. 24, No. 3, pp. 303-312, (2009)

[渡辺 15] 渡辺知恵美, 中村聡史: オノマトペロリ: 味覚や触感を表すオノマトペによる料理レシピのランキング, 人工知能学会論文誌, Vol. 30, No. 1, pp. 340-352, (2015)