

レシピタイトルに頻出するシズルワードの多義性抽出手法と評価

高田 実央† 佐藤 哲司††

† 筑波大学 〒 305-8550 茨城県つくば市春日 1-2

†† 筑波大学図書館情報メディア研究科 〒 305-8550 茨城県つくば市春日 1-2

E-mail: †{takada18,satoh}@ce.slis.tsukuba.ac.jp

あらまし 近年、投稿型レシピでは「シズルワード」とよばれるおいしさを表現する修飾表現がよく用いられている。シズルワードには美味しさの感性が反映されているため、シズルワードは明確な意味の把握や解釈・定義が難しい場合がある。その1つにシズルワードの多義性が挙げられる。例えば、シズルワードの1つに「からい」という形容詞がある。「からい」には「Hot」「Salty」「Spicy」という少なくとも3つの意味がある。本研究では、「からい」と同様に、シズルワードには複数の意味があると仮定し、シズルワードの多義性を明らかにする手法を提案する。そのために、レシピの「概要」を多次元尺度構成法と階層的クラスタリングを用いて分析し、シズルワードの多義属性数と多義属性のキーワードを明らかにした。さらにアンケート調査を用いて、キーワードを絞り込み、それを多義属性として抽出した。また、アンケート調査の結果から各クラスターの重み付き平均経路長を算出し、抽出した多義性の妥当性を評価した。今回は「からい」と、cookpadで最もよく用いられている「たっぷり」「和風」「ヘルシー」を対象とした。キーワード シズルワード、多義性、cookpad

1 はじめに

近年、ユーザー発信型の情報が増えている。加藤ら[1]によると、ユーザー発信型の情報には「シズルワード」とよばれるおいしさを表現する修飾表現がよく用いられている。特に、投稿型レシピのcookpadでは、Twitterやwebと比較して多くのシズルワードが用いられていることが明らかにされている。

「シズルワード」とは、米国の経営コンサルタントE.Wheelによって提唱された概念である。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明確に表現する用語である。

おいしいを感じる言葉[2]によると、図1に示すように、味覚系表現93語、食感系表現119語、情報系表現119語の331語がシズルワードとして定義されている。シズルワードの詳細は図1に示す。

シズルワードに関する研究は盛んに行われており、その研究方法はアンケート調査による定量的な解析とマイクロブログなどのテキストデータの解析の2種類に大別される。どちらも、シズルワードのニュアンスや意味構造を明らかにすることを目的としている研究が多い。しかし、シズルワードの多義性に注目し、その多義性の数と意味を明らかにしようとする研究は充分にはなされていない。例えば、シズルワードの1つに「からい」という形容詞がある。デジタル大辞林によると、「からい」には「舌やのどを強く刺激するような味である」Hot、「塩気が多い」Salty、「甘みが少なくさっぱりとしていて、ひきしまっている」Spicyという少なくとも3つの意味があると考えられる。

本研究では、「からい」のように1つのシズルワードで複数の意味が表現されている場合を「多義属性がある」状態である

味覚系表現：93語	食感系表現：119語	情報系表現：119語
1. 酸っぱい	1. ぷるぷる	1. 簡単
2. 酸っぱい	2. ぷるぷる	2. 簡単
3. 酸っぱい	3. ぷるぷる	3. 簡単
4. 酸っぱい	4. ぷるぷる	4. 簡単
5. 酸っぱい	5. ぷるぷる	5. 簡単
6. 酸っぱい	6. ぷるぷる	6. 簡単
7. 酸っぱい	7. ぷるぷる	7. 簡単
8. 酸っぱい	8. ぷるぷる	8. 簡単
9. 酸っぱい	9. ぷるぷる	9. 簡単
10. 酸っぱい	10. ぷるぷる	10. 簡単
11. 酸っぱい	11. ぷるぷる	11. 簡単
12. 酸っぱい	12. ぷるぷる	12. 簡単
13. 酸っぱい	13. ぷるぷる	13. 簡単
14. 酸っぱい	14. ぷるぷる	14. 簡単
15. 酸っぱい	15. ぷるぷる	15. 簡単
16. 酸っぱい	16. ぷるぷる	16. 簡単
17. 酸っぱい	17. ぷるぷる	17. 簡単
18. 酸っぱい	18. ぷるぷる	18. 簡単
19. 酸っぱい	19. ぷるぷる	19. 簡単
20. 酸っぱい	20. ぷるぷる	20. 簡単
21. 酸っぱい	21. ぷるぷる	21. 簡単
22. 酸っぱい	22. ぷるぷる	22. 簡単
23. 酸っぱい	23. ぷるぷる	23. 簡単
24. 酸っぱい	24. ぷるぷる	24. 簡単
25. 酸っぱい	25. ぷるぷる	25. 簡単
26. 酸っぱい	26. ぷるぷる	26. 簡単
27. 酸っぱい	27. ぷるぷる	27. 簡単
28. 酸っぱい	28. ぷるぷる	28. 簡単
29. 酸っぱい	29. ぷるぷる	29. 簡単
30. 酸っぱい	30. ぷるぷる	30. 簡単
31. 酸っぱい	31. ぷるぷる	31. 簡単
32. 酸っぱい	32. ぷるぷる	32. 簡単
33. 酸っぱい	33. ぷるぷる	33. 簡単
34. 酸っぱい	34. ぷるぷる	34. 簡単
35. 酸っぱい	35. ぷるぷる	35. 簡単
36. 酸っぱい	36. ぷるぷる	36. 簡単
37. 酸っぱい	37. ぷるぷる	37. 簡単
38. 酸っぱい	38. ぷるぷる	38. 簡単
39. 酸っぱい	39. ぷるぷる	39. 簡単
40. 酸っぱい	40. ぷるぷる	40. 簡単
41. 酸っぱい	41. ぷるぷる	41. 簡単
42. 酸っぱい	42. ぷるぷる	42. 簡単
43. 酸っぱい	43. ぷるぷる	43. 簡単
44. 酸っぱい	44. ぷるぷる	44. 簡単
45. 酸っぱい	45. ぷるぷる	45. 簡単
46. 酸っぱい	46. ぷるぷる	46. 簡単
47. 酸っぱい	47. ぷるぷる	47. 簡単
48. 酸っぱい	48. ぷるぷる	48. 簡単
49. 酸っぱい	49. ぷるぷる	49. 簡単
50. 酸っぱい	50. ぷるぷる	50. 簡単
51. 酸っぱい	51. ぷるぷる	51. 簡単
52. 酸っぱい	52. ぷるぷる	52. 簡単
53. 酸っぱい	53. ぷるぷる	53. 簡単
54. 酸っぱい	54. ぷるぷる	54. 簡単
55. 酸っぱい	55. ぷるぷる	55. 簡単
56. 酸っぱい	56. ぷるぷる	56. 簡単
57. 酸っぱい	57. ぷるぷる	57. 簡単
58. 酸っぱい	58. ぷるぷる	58. 簡単
59. 酸っぱい	59. ぷるぷる	59. 簡単
60. 酸っぱい	60. ぷるぷる	60. 簡単
61. 酸っぱい	61. ぷるぷる	61. 簡単
62. 酸っぱい	62. ぷるぷる	62. 簡単
63. 酸っぱい	63. ぷるぷる	63. 簡単
64. 酸っぱい	64. ぷるぷる	64. 簡単
65. 酸っぱい	65. ぷるぷる	65. 簡単
66. 酸っぱい	66. ぷるぷる	66. 簡単
67. 酸っぱい	67. ぷるぷる	67. 簡単
68. 酸っぱい	68. ぷるぷる	68. 簡単
69. 酸っぱい	69. ぷるぷる	69. 簡単
70. 酸っぱい	70. ぷるぷる	70. 簡単
71. 酸っぱい	71. ぷるぷる	71. 簡単
72. 酸っぱい	72. ぷるぷる	72. 簡単
73. 酸っぱい	73. ぷるぷる	73. 簡単
74. 酸っぱい	74. ぷるぷる	74. 簡単
75. 酸っぱい	75. ぷるぷる	75. 簡単
76. 酸っぱい	76. ぷるぷる	76. 簡単
77. 酸っぱい	77. ぷるぷる	77. 簡単
78. 酸っぱい	78. ぷるぷる	78. 簡単
79. 酸っぱい	79. ぷるぷる	79. 簡単
80. 酸っぱい	80. ぷるぷる	80. 簡単
81. 酸っぱい	81. ぷるぷる	81. 簡単
82. 酸っぱい	82. ぷるぷる	82. 簡単
83. 酸っぱい	83. ぷるぷる	83. 簡単
84. 酸っぱい	84. ぷるぷる	84. 簡単
85. 酸っぱい	85. ぷるぷる	85. 簡単
86. 酸っぱい	86. ぷるぷる	86. 簡単
87. 酸っぱい	87. ぷるぷる	87. 簡単
88. 酸っぱい	88. ぷるぷる	88. 簡単
89. 酸っぱい	89. ぷるぷる	89. 簡単
90. 酸っぱい	90. ぷるぷる	90. 簡単
91. 酸っぱい	91. ぷるぷる	91. 簡単
92. 酸っぱい	92. ぷるぷる	92. 簡単
93. 酸っぱい	93. ぷるぷる	93. 簡単
94. 酸っぱい	94. ぷるぷる	94. 簡単
95. 酸っぱい	95. ぷるぷる	95. 簡単
96. 酸っぱい	96. ぷるぷる	96. 簡単
97. 酸っぱい	97. ぷるぷる	97. 簡単
98. 酸っぱい	98. ぷるぷる	98. 簡単
99. 酸っぱい	99. ぷるぷる	99. 簡単
100. 酸っぱい	100. ぷるぷる	100. 簡単
101. 酸っぱい	101. ぷるぷる	101. 簡単
102. 酸っぱい	102. ぷるぷる	102. 簡単
103. 酸っぱい	103. ぷるぷる	103. 簡単
104. 酸っぱい	104. ぷるぷる	104. 簡単
105. 酸っぱい	105. ぷるぷる	105. 簡単
106. 酸っぱい	106. ぷるぷる	106. 簡単
107. 酸っぱい	107. ぷるぷる	107. 簡単
108. 酸っぱい	108. ぷるぷる	108. 簡単
109. 酸っぱい	109. ぷるぷる	109. 簡単
110. 酸っぱい	110. ぷるぷる	110. 簡単
111. 酸っぱい	111. ぷるぷる	111. 簡単
112. 酸っぱい	112. ぷるぷる	112. 簡単
113. 酸っぱい	113. ぷるぷる	113. 簡単
114. 酸っぱい	114. ぷるぷる	114. 簡単
115. 酸っぱい	115. ぷるぷる	115. 簡単
116. 酸っぱい	116. ぷるぷる	116. 簡単
117. 酸っぱい	117. ぷるぷる	117. 簡単
118. 酸っぱい	118. ぷるぷる	118. 簡単
119. 酸っぱい	119. ぷるぷる	119. 簡単

図1 シズルワードの詳細

と定義し、シズルワードには複数の意味があると仮定する。その仮定のもと、テキストデータの解析によるシズルワードの多義性を明らかにする手法を提案する。シズルワードの多義性が明らかになれば、ユーザーが感じている微妙なニュアンスの違いが明らかになるため、商品パッケージやネーミングにシズルワードを用いる際に参考にすることが可能になる。

2 関連研究

関連研究は定量的な解析とテキストデータの解析の2種類がある。

アンケートをもちいた手動的な解析の研究として、以下の2つが挙げられる。大橋[3]は、15歳から59歳までの男女1,500人にアンケート調査を行っている。305個のシズルワードに対し、「おいしい」「食べたい」と感じる言葉にチェックをしてもらうという方法で、人気のあるシズルワードを明らかにしている。加えて、2003年から行っている同様の調査と結果を比

較することで、シズルワードの人気の時系列変化を明らかにしている。これによると、2015年に最も人気のあったシズルワードは「もちもち」であった。石橋[4]は、消費者の食感とシズルワードの関係に着目し、米菓の堅さを中心に分析調査を行っている。20代の学生と社会人50名を対象に、無造作に置かれた4種類の米菓を一枚ずつ食し、米菓に適している33個のシズルワードがその米菓にあてはまるか否か、5段階の評定尺度で判定させている。これによると、「歯ごたえのある」「さっくり」「サクサク」「軽い」といったシズルワードがよくあてはまるとされている。

テキストデータの解析の研究として、以下の3種類がある。加藤ら[1]はインターネット上のメディアのうち、Twitter、Web、cookpad上でのおいしさを表す言葉であるシズルワードと食品との関係の比較分析を行っている。Twitter上にはユーザが実際に食べた「経験」に基づくシズルワードが多いと考えられるのに対し、一般のWeb上には「秘伝の」「手作りの」「自然の」といった宣伝文句と考えられる情報系のシズルワードが多く、CookPadでは「Twitterや一般のWebと比べて多くのシズルワードが共起しており、特に味覚系、食感系シズルワード」が「Twitterや一般のWebの場合と比べ多く使用」されている特徴があると考えられると述べている。橘ら[5][6]は「簡単」と「子どもが喜ぶ」という修飾表現に対して、料理名における修飾表現の根拠をネーミングコンセプトと定義し、「料理名における典型的な要素と比較することで差異を抽出し、差異間の組み合わせによる特徴パターンで分類することによりネーミングコンセプトを抽出する方法」を提案している。本研究では、対象とする修飾表現をシズルワードに拡張している点に新規性があることに加え、橘らが材料と調理器具からネーミングコンセプトの抽出を試みているのに対し、本研究ではレシピの概要から多義性の抽出を試みる点で新規性がある。長谷川ら[7]は楽天レシピのユーザからのレビューとそれに対するレシピ投稿者からの返事を対象に、シズルワードの特性に関する要因抽出を、NMFを適用することで行っている。NMFの因子における単語の重みに着目することで、「同様な印象を受けるシズルワードから構成される食における感性の表現に関する共通性を抽出することができた」と述べているが、分析対象をレビューに限定していることから、シズルワードの多義性について考慮されていないという課題がある。本研究ではCookPadの概要を分析対象として点に加え、シズルワードの多義性を明らかにすることを目的としている点で新規性がある。

3 シズルワードの多義性を抽出する手法の提案

3.1 シズルワードの出現分布

本研究では、投稿型レシピサイトの中でも、掲載レシピが最も多く、先行研究が豊富なCookPadを利用することとし、CookPadで高頻度で用いられるシズルワードを分析対象とする。まず、おいしいを感じる言葉[2]によると、味覚系表現93語、食感系表現119語、情報系表現119語の331語がシズルワードとして定義されている。これらを辞書登録し、タイトル

に用いられているシズルワードをカウントする。

その結果、「たっぷり」が26,302回で最も多く、次いで「ヘルシー」が24,738回、「和風」が24,686回であった。この3ワードでタイトルに使用されているシズルワードの約25%をしめることが分かった。図2はシズルワードのパレート図を表す。左軸がシズルワードの出現回数の軸であり、出現回数は青色で示されている。右軸はシズルワードの累積%の軸であり、累積%はオレンジ色で示されている。

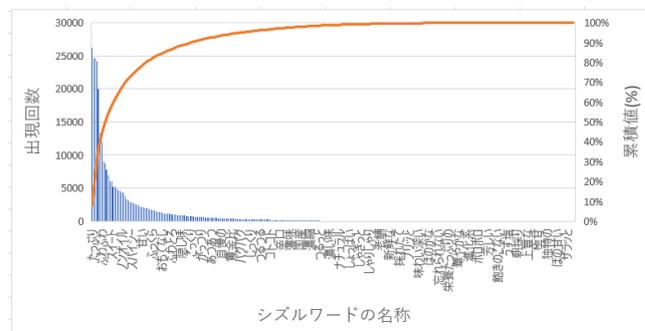


図2 シズルワードの出現頻度と累積値

図3にシズルワードの散布図を示す。横軸は登録したシズルワードの数の対数、縦軸はシズルワードの出現回数の対数を表す。極端な右下がりになっていることから、レシピで用いられているシズルワードには偏りがあり、一部のシズルワードのみが非常によく使われていることが分かる。

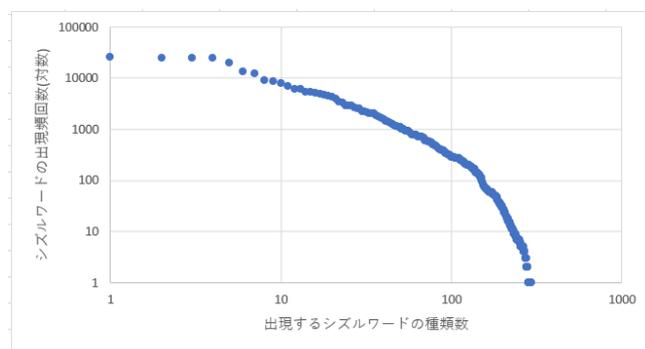


図3 シズルワードの両対数グラフ

3.2 属性タグ数の推定

本研究では対象とするシズルワードをタイトルに含む全レシピを分析の対象とする。レシピにはタイトルと概要、材料一覧、料理手順が記載されている。本研究では、レシピの特徴が簡潔に記載されている概要を用いてシズルワードの多義属性を明らかにする。

具体的には、レシピの「概要」部分をKHcoderを用いて多次元尺度構成法で分析する。多次元尺度構成法は分類対象物の関係性を低次元空間における点の配置で表現する分析手法であり、分析対象の類似性を可視化することができる。レシピの「概要」には、シズルワードを用いずにシズルワードが意味することが説明されている。例えば、「野菜たっぷり!もやし入り春

巻き」というレシピの概要には、「春巻きなのにあっさりしていて、たくさん食べられちゃいます」と書いてあり、「食べたえがある」ことを「たっぷり」を使わずに表現している。このため、レシピの概要を多次元尺度構成法で分析することで、シズルワードの多義性に関係のある言葉の類似性を可視化することができる。

今回はプロットの解釈の手助けとしてクラスター分けも行う。クラスター数の決定は、レシピの「概要」を階層的クラスタリングにかけ、併合水準で傾きが0に近くなる9以下のクラスター数を目視で確認し決定する。クラスターを数えることで、シズルワードの多義性数を推定できる。

3.3 多義性の抽出

3.2で得られた各クラスターから属性タグを抽出するために、アンケート調査を行う。クラスターごとに頻出する10単語を取り出した質問紙を作成し、クラスターを代表すると思われる3単語を選んでもらう。本研究では選ばれた単語をキーワードと呼ぶ。

このようにして得られたキーワードの出現回数に関する累積グラフを作成し、累積単語数の80%を占めるキーワードを導く。累積単語数の80%を占めるキーワードの出現分布と、キーワードの組み合わせを参照しながら、クラスターの特性を表す文章を作成し、得られた文章を多義性として抽出する。クラスターの特性を表す文章が作成できない場合は、そのクラスターからは多義属性は抽出不可と判断する。

4 実験

本節では、提案手法の有効性について、「からい」「たっぷり」「ヘルシー」「和風」の4つのシズルワードを用いて検証する。

4.1 データセット

本研究では、Niiが研究用に公開しているCookPadのレビューデータ¹を使用した。クックパッドで2014年9月30日までに公開されたレシピに関するデータで、レシピのタイトルや概要、手順、つくれば（「作りましたフォトレポート」の略）、カテゴリなどのデータが含まれている。

4.2 「からい」の多義属性数と抽出した属性タグ

4.1のデータから、レシピタイトルに「たっぷり」を含むレシピをすべて抽出し、そのレシピの概要を多次元尺度構成法で分析した。その結果は図5である。図4から、併合水準の傾きが0に近くなるクラスター数は5であると判断した。これより、「からい」の多義属性数は最大5であると推定した。

3.3に基づいて、図6の質問紙を作成した。

選ばれた単語を数字に置き換えて集計すると、クラスター1では図7のようになった。単語と数字の対応表は表1に示す。これに基づいて累積累積グラフを作成した。作成した累積グラフは図8であり、クラスター1は濃い青色で示されている。こ

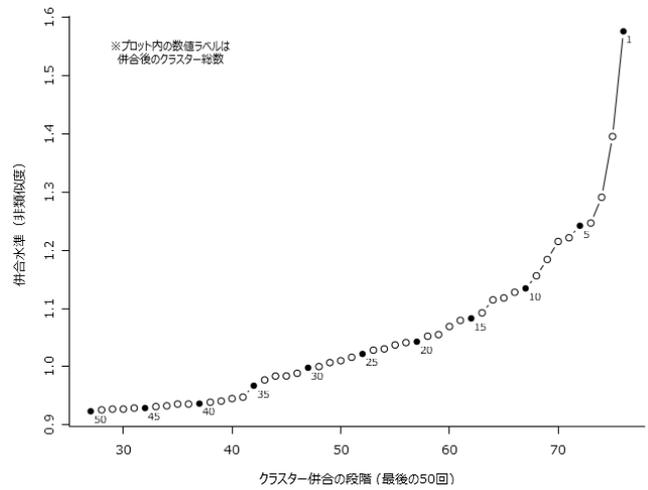


図4 「からい」の併合水準

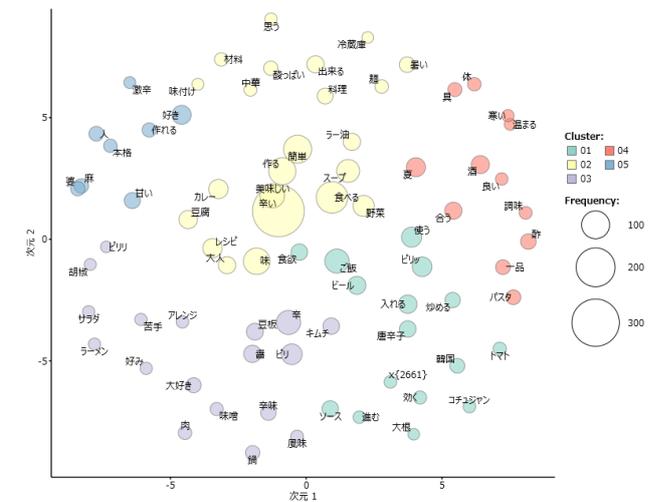


図5 「からい」の多次元尺度構成法

【例】	からい				
	クラスター1	クラスター2	クラスター3	クラスター4	クラスター5
キーワード	ご飯	カレー	キムチ	パスタ	甘い
	ソース	スープ	サラダ	一品	激辛
	ビール	ラー油	ピリ	夏	好き
	ピリッ	レシピ	胡椒	寒い	作れる
	韓国	簡単	醬	具	人
	使う	作る	辛味	合う	妻
	食欲	食べる	大好き	酒	本格
	唐辛子	豆腐	豆板	酢	麻
	入れる	美味しい	鍋	調味	
	炒める	野菜	風味	良い	
写真					
文章					

図6 作成した質問紙（からい）

れより、累積単語数の80%を満たす単語数は4であり、そのキーワードは図7よりキーワード4・8・7・3であることが分かる。クラスター1のキーワードの出現組み合わせは図9のようであることから、クラスター1から「ビールにあう、食欲をそる唐辛子のからさ」という多義属性を抽出することができる。

同様に他のクラスターでも多義属性の抽出を行うと、クラス

1: <https://www.nii.ac.jp/dsc/idr/cookpad/cookpad.html>

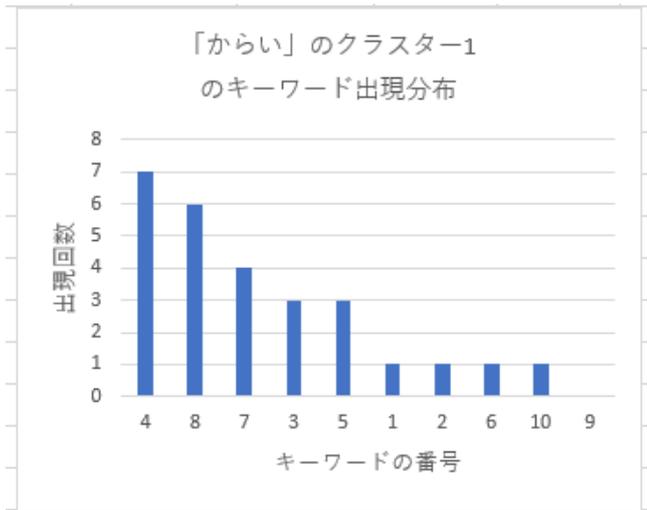


図7 「からい」のクラスター1のキーワードの出現分布

表1 クラスター1のキーワードと数字の対応表

クラスター1	数字
ご飯	1
ソース	2
ビール	3
ピリッ	4
韓国	5
使う	6
食欲	7
唐辛子	8
入れる	9
炒める	10

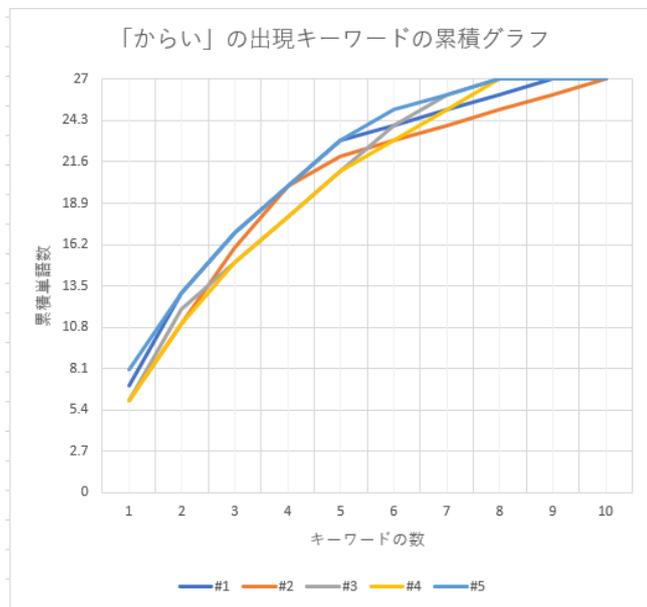


図8 「からい」のキーワードの累積グラフ

ター4以外の4つのクラスターで多義属性を抽出できた。詳細は以下のようなものである。

クラスター1 ビールにあう食欲をそそる唐辛子のピリッとしたからさ

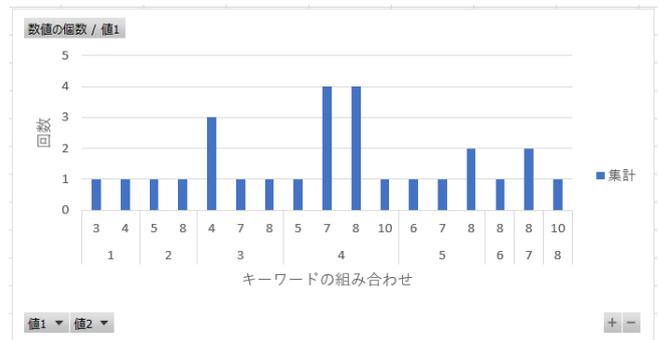


図9 「からい」のクラスター1のキーワード組み合わせ

クラスター2 簡単に美味しいラー油を使ったスープのからさ
 クラスター3 ピリ辛のキムチサラダのからさ
 クラスター5 本格的で甘味もある麻婆豆腐のからさ

4.3 「たっぷり」の多義属性数と抽出した属性タグ

4.1のデータから、レシピタイトルに「たっぷり」を含むレシピをすべて抽出し、そのレシピの概要を多次元尺度構成法で分析した。その結果は図10である。階層的クラスタリングの併合水準の傾きが0近くなるクラスター数は5であると判断した。これより、「たっぷり」のクラスター数は5であり、多義属性数は最大5であると推定した。

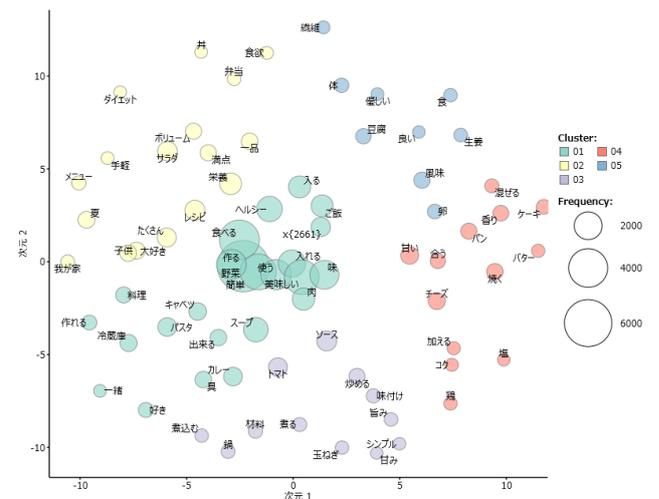


図10 「たっぷり」の多次元尺度構成法

3.3に基づいて、図11の質問紙を作成した。

選ばれた単語を集計し、累積グラフを作成した。作成した累積グラフは図12である。各クラスターで累積単語数の80%を満たす単語数を求め、それに含まれるキーワードを導いた。キーワードの出現組み合わせを参照しながら、各クラスターを特徴づける文章を作成したところ、クラスター5以外のクラスターから以下の多義属性を抽出することができた。

クラスター1 簡単にできて、美味しくヘルシーな野菜たっぷり

クラスター2 子供が大好きでボリュームと栄養がたっぷり

クラスター3 玉ねぎのソースで旨みたっぷり

クラスター4 焼くと香りがたつ甘さたっぷりのケーキ

【例】	たっぶり				
	クラスター1	クラスター2	クラスター3	クラスター4	クラスター5
キーワード	味	一品	味付け	合う	良い
	入れる	栄養	炒める	甘い	体
	美味しい	子供	旨み	香り	生姜
	簡単	サラダ	材料	ケーキ	食
	食べる	大好き	ソース	チーズ	繊維
	使う	たくさん	玉ねぎ	鶏	卵
	作る	夏	トマト	バター	豆腐
	肉	ボリューム	鍋	パン	風味
	ヘルシー	満点	煮込む	混ぜる	優しい
	野菜	レシピ	煮る	焼く	
写真					
	文章	簡単ヘルシーで野菜たっぷりの子供が大好きボリューム満点のたっぶり たっぶりのクラスターなので、何がたっぶりなのかを説明する文章をつくってください			

図 11 作成した質問紙 (たっぶり)

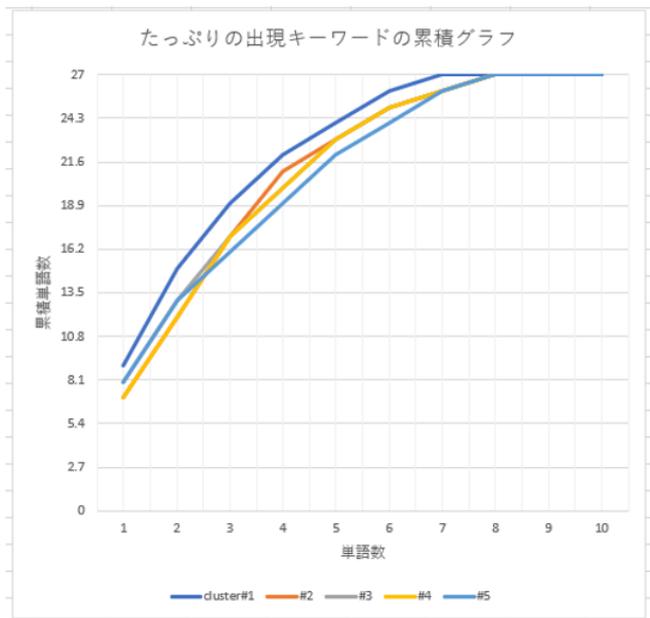


図 12 「たっぶり」のキーワードの累積グラフ

4.4 「ヘルシー」の多義属性数と抽出した属性タグ

4.1 のデータから、レシピタイトルに「ヘルシー」を含むレシピをすべて抽出し、そのレシピの概要を多次元尺度構成法で分析した。その結果は図 13 である。階層的クラスタリングの併合水準の傾きが 0 になるクラスター数は 6 であると判断した。これより、「ヘルシー」のクラスター数は 6 であり、「ヘルシー」の多義属性数は最大 6 であると推定した。

3.3 に基づいて、図 14 の質問紙を作成した。

選ばれた単語を集計し、累積グラフを作成した。作成した累積グラフは図 15 である。クラスター 6 は回答者の 1 人が回答不明と回答したため、全回答が 24 個となっている。各クラスターで累積単語数の 80% を満たす単語数を求め、それに含まれるキーワードを導いた。キーワードの出現組み合わせを参照しながら、各クラスターを特徴づける文章を作成したところ、クラスター 3 とクラスター 6 以外のクラスターから以下の多義属性を抽出することができた。

- クラスター 1 簡単でおいしいのにカロリーが低いのでヘルシー
- クラスター 2 栄養のある野菜をたくさん使った一品でヘルシー
- クラスター 4 材料を油で揚げないのでヘルシー

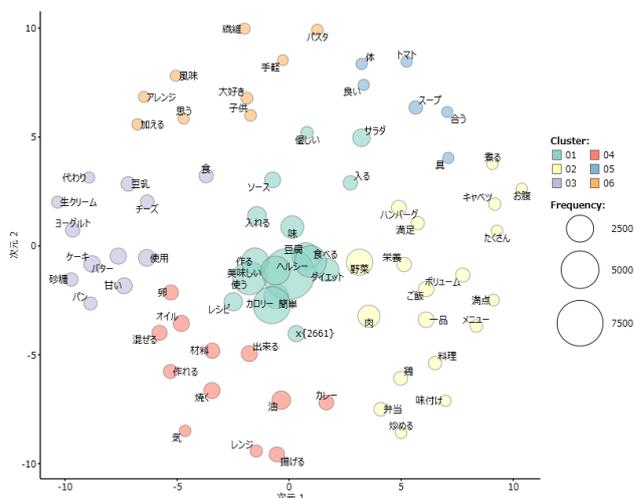


図 13 「ヘルシー」の多次元尺度構成法

【例】	ヘルシー					
	クラスター1	クラスター2	クラスター3	クラスター4	クラスター5	クラスター6
キーワード	カロリー	一品	甘い	揚げ	合う	アレンジ
	サラダ	栄養	ケーキ	油	良い	使う
	ダイエット	ご飯	砂糖	オイル	体	加える
	簡単	鶏	使用	カレー	具	子供
	作る	肉	食	材料	スープ	繊維
	食べる	ハンバーグ	チーズ	卵	トマト	大好き
	豆腐	ボリューム	豆乳	作れる		手軽
	入れる	満点	バター	出来る		バスタ
	美味しい	野菜	ヨーグルト	混ぜる		風味
	写真					
文章						

図 14 作成した質問紙 (ヘルシー)

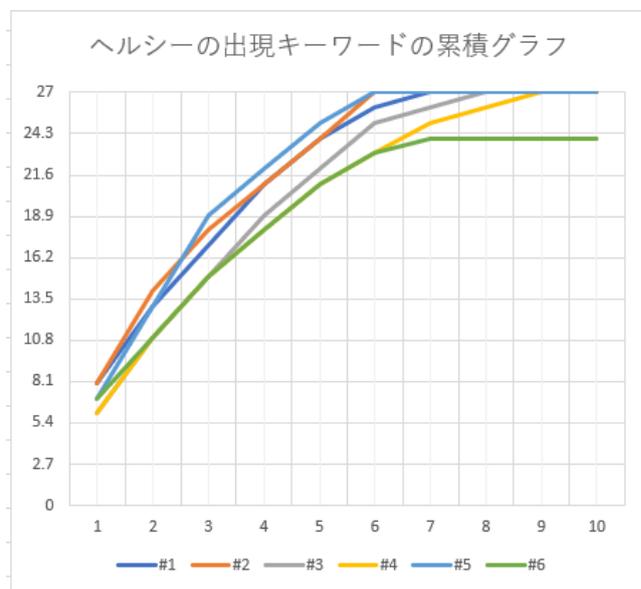


図 15 「ヘルシー」のキーワードの累積グラフ

クラスター 5 体にいい具材が使われているヘルシースープ

4.5 「和風」の多義属性数と抽出した属性タグ

4.1 のデータから、レシピタイトルに「和風」を含むレシピをすべて抽出し、そのレシピの概要を多次元尺度構成法で分析した。その結果は図 16 である。階層的クラスタリングの併合水準の傾きが 0 になるクラスター数は 4 であると判断した。これ

より、「和風」のクラスター数は4であり、「和風」の多義属性数は最大4であると推定した。

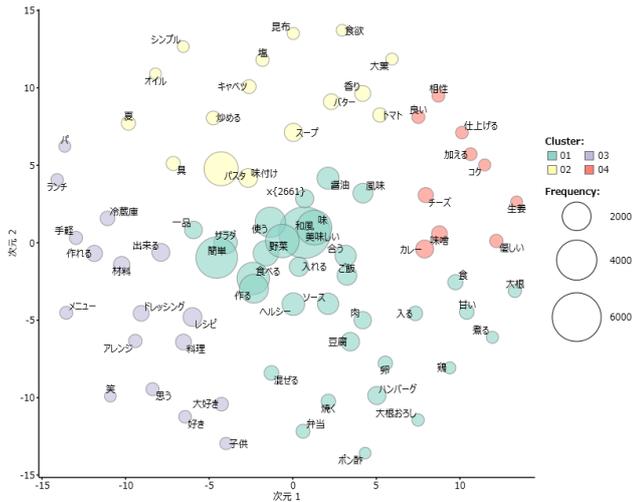


図 16 「和風」の多次元尺度構成法

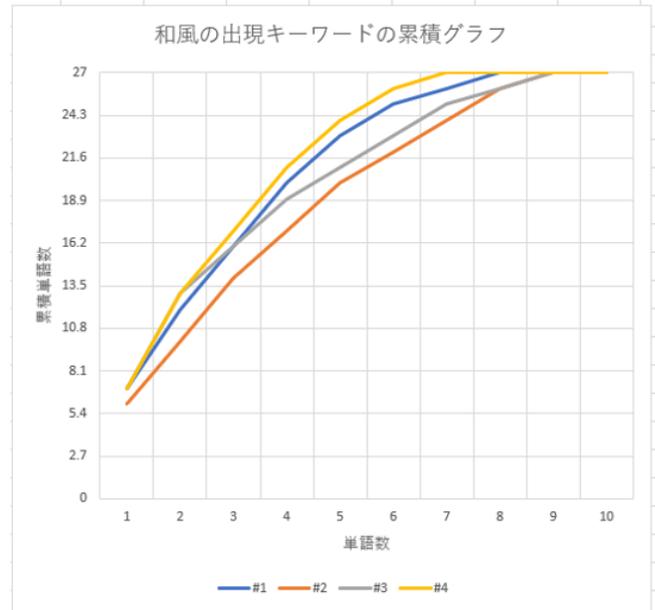


図 18 「和風」のキーワードの累積グラフ

3.3 に基づいて、図 17 の質問紙を作成した。

【例】	和風			
	クラスター1	クラスター2	クラスター3	クラスター4
キーワード	味	味付け	アレンジ	相性
	美味しい	炒める	ドレッシング	良い
	簡単	香り	レシビ	カレー
	サラダ	キャベツ	材料	加える
	醤油	昆布	作れる	コク
	食べる	スープ	手軽	仕上げる
	使う	トマト	出来る	生姜
	作る	夏	大好き	チーズ
	ヘルシー	パスタ	料理	味噌
	野菜	バター	冷蔵庫	優しい
写真				
文章				

図 17 作成した質問紙 (和風)

選ばれた単語を集計し、累積グラフを作成した。作成した累積グラフは図 18 である。各クラスターで累積単語数の 80% を満たす単語数を求め、それに含まれるキーワードを導いた。キーワードの出現組み合わせを参照しながら、各クラスターを特徴づける文章を作成したところ、クラスター 2 以外のクラスターから以下の多義属性を抽出することができた。

- クラスター 1 醤油味の和風野菜サラダ
- クラスター 3 手軽に作れてアレンジのできる和風ドレッシング
- クラスター 4 味噌を使った優しいコクのある和風カレー

5 評価

「からい」「たっぷり」「ヘルシー」「和風」の順に得られた多義属性の評価を行う。アンケート調査で得られた回答の組み合わせをネットワークとみなし、各クラスターの重み付き平均経路長を算出した。

4 で得られた多義性を基に、シズルワードが多義性を持つことのできる重み付き平均経路長を定める。

平均経路長はある回答の組み合わせから別の回答の組み合わせまで到達するのに必要なステップ数を表している。平均経路長を求めることでネットワークのおおよその大きさを表現することができる。今回は、回答の組み合わせによって選ばれた回数が異なることから、経路に重みをつけた重み付き平均経路長を算出する。

重み付き平均経路長は以下のように算出する。2 頂点 v_i と v_j の距離 $d(v_i, v_j)$ を、 v_i から v_j に行くために通らなければならない最小の枝数で定義する。重みを反映させるために、距離 $d(v_i, v_j)$ の逆数をとる。ネットワークの重み付き平均経路長 L は、距離 $d(v_i, v_j)$ の逆数の全ての頂点对にわたる平均であるので、頂点が N 個ならば、頂点对の選び方は $N(N-1)/2$ 通りあり、

$$L = \frac{2}{N(N-1)} \sum_{1 \leq i < j \leq N} \frac{1}{d(v_i, v_j)}$$

という式で求めることができる。

各シズルワードの各クラスターで重み付き平均経路長を算出すると図 19 のようになった。セルに色がついているクラスターは 4 で多義性が抽出できたクラスターである。

また、各クラスターで抽出された多義性の質的な考察についても行う。

重み付き平均パス長	からい	たっぷり	ヘルシー	和風
クラスター1	1.583333333	1.428571429	1.333333333	1.428571429
クラスター2	1.666666667	1.464285714	1.266666667	1.527777778
クラスター3	1.392857143	1.5	1.392857143	1.5
クラスター4	1.285714286	1.464285714	1.583333333	1.333333333
クラスター5	1.607142857	1.428571429	1.2	
クラスター6			1.428571429	

図 19 重み付き平均経路長

5.1 からの評価

図 19 から、からは重み付き平均経路長が約 1.393～約 1.607 であるときに多義性が抽出可能であると考えられる。重み付き平均経路長が約 1.286 であるクラスター 4 からは多義性が抽出できなかったことから、重み付き平均経路長が約 1.393 より小さいと、ネットワークが小さすぎて意味として成り立たないと考えられる。

また、クラスター 1 の「ビールにあう食欲をそそるピリッとした唐辛子のからさ」は Hot に該当すると考えられる。クラスター 3 の「ピリ辛のキムチサラダのからさ」は Spicy に該当すると考えられる。本研究では Salty に該当する多義属性は抽出できなかった。その原因として、関東では「しょっぱい」関西では「からい」というように、地域によって Salty を表す言葉が異なることが考えられる。Salty を抽出するためには、例えば、甘辛いといった別のシズルワードが適当かもしれない。

5.2 たっぶりの評価

図 19 から、たっぷりのクラスター 1 とクラスター 5 の重み付き平均経路長の値が同じことが分かる。クラスター 5 からは多義性が抽出できなかったことから、クラスター 1 の多義性は必ずしも適当であるとはいえないことが分かる。これより、たっぷりは重み付き平均経路長が約 1.464～約 1.5 であるときに多義性が抽出可能であると考えられる。

また、クラスター 2 の「子供が大好きでボリュームと栄養がたっぷり」は「ボリューム」と「栄養」という 2 つの単語がたっぷりにかかっている。これより、クラスタリングが不十分であった可能性が伺える。クラスター 4 の「焼くと香りがたつ甘さたっぷりのケーキ」のみ、具体的な料理名が多義属性についている。他の 3 つの多義属性と粒度が異なっているため、クラスター 4 を多義属性として抽出可能とするかどうかは検討の余地がある。

5.3 ヘルシーの評価

図 19 から、クラスター 5 の重み付き平均経路長 1.2 が最小で、クラスター 4 の重み付き平均経路長約 1.583 が最大であることが分かる。しかし、重み付き平均経路長が 1.2～1.583 の間にあるクラスター 3 とクラスター 6 の多義性は抽出されていないことから、クラスター 4 から抽出された多義性は必ずしも適当であるとはいえないことが考えられる。これより、ヘルシーは重み付き平均経路長が 1.2～約 1.333 であるときに多義性が抽出可能であると考えられる。

また、クラスター 5 の「体にいい具材が使われているヘル

シースープ」のみ、具体的な料理名が多義属性についている。他の 3 つの多義属性と粒度が異なっているため、クラスター 5 を多義属性として抽出可能とするかどうかは検討の余地がある。

5.4 和風の評価

図 19 から、和風では重み付き平均経路長が約 1.333～1.5 であるときに多義性が抽出可能であると考えられる。

また、すべてのクラスターで具体的な料理名が多義属性についている。すべての多義属性で粒度が同じであるため、クラスターからの多義属性の抽出は妥当であると考えられる。

6 おわりに

本研究では、おいしいを感じる言葉 [2] に基づき、味覚系表現 93 語、食感系表現 119 語、情報系表現 119 語の 331 語を「シズルワード」として定義し、シズルワードには多義性があるという仮説に基づき、その多義性を明らかにすることを目的とする。そこで本稿では、「からい」と日本最大の投稿型レシピサイトである CookPad で出現回数の多い上位 3 件のシズルワード、「たっぷり」「ヘルシー」「和風」を対象に、それらのシズルワードの多義属性数と多義属性の抽出を行った。

提案手法では、まず、シズルワードを用いずにシズルワードの意味が説明されているレシピの「概要」を KHcoder を用いて、多次元尺度構成法と階層的クラスタリングで分析を行った。この分析により、対象とするシズルワードが持ちうる最大の多義属性数が推定できる。「からい」の多義属性数は最大 5 であると推定でき、「たっぷり」の多義属性数は最大 5 つ、「ヘルシー」の多義属性数は最大 6 つ、「和風」の多義属性数は最大 4 つであると推定できた。

次にアンケート調査を行い、多義属性の抽出を行った。その結果、「からい」からは「ビールにあう食欲をそそる唐辛子のピリッとしたからさ」「簡単に美味しいラー油を使ったスープのからさ」「ピリ辛のキムチサラダのからさ」「本格的で甘味もある麻婆豆腐のからさ」という 4 つの多義属性が抽出できた。

「たっぷり」からは「簡単にできて、美味しくヘルシーな野菜たっぷり」「子供が大好きでボリュームと栄養がたっぷり」「玉ねぎのソースで旨みたっぷり」「焼くと香りがたつ甘さたっぷりのケーキ」という 4 つの多義属性が抽出できた。

「ヘルシー」からは「簡単でおいしいのにカロリーが低いのでヘルシー」「栄養のある野菜をたくさん使った一品でヘルシー」「材料を油で揚げないのでヘルシー」「体にいい具材が使われているヘルシースープ」という 4 つの多義属性が抽出できた。

「和風」からは「醤油味の和風野菜サラダ」「手軽に作れてアレンジのできる和風ドレッシング」「味噌を使った優しいコクのある和風カレー」という 3 つの多義属性が抽出できた。

最後に、抽出した多義性を重み付き平均経路長を用いて評価した。「からい」では重み付き平均経路長が約 1.393～約 1.607 であるときに多義性が抽出可能であると考えられ、「簡単に美味しいラー油を使ったスープのからさ」「ピリ辛のキムチサラダ

のからさ」「本格的で甘味もある麻婆豆腐のからさ」という3つの多義属性が妥当であると考えられた。

「たっぷり」ではたっぷりは重み付き平均経路長が約1.464～約1.5であるときに多義性が抽出可能であると考えられ、「子供が大好きでボリュームと栄養がたっぷり」「玉ねぎのソースで旨みたっぷり」「焼くと香りがたつ甘さたっぷりのケーキ」という3つの多義属性が妥当であると考えられた。

「ヘルシー」では重み付き平均経路長が1.2～約1.333であるときに多義性が抽出可能であると考えられ、「簡単でおいしいのにカロリーが低いのでヘルシー」「栄養のある野菜をたくさん使った一品でヘルシー」「体にいい具材が使われているヘルシースープ」という3つの多義属性が妥当であると考えられた。

「和風」では和風では重み付き平均経路長が約1.333～1.5であるときに多義性が抽出可能であると考えられ、「醤油味の和風野菜サラダ」「手軽に作れてアレンジのできる和風ドレッシング」「味噌を使った優しいコクのある和風カレー」という3つの多義属性が妥当であると考えられた。

今後の課題として、2つのことが挙げられる。1つ目の課題として、キーワード抽出の精査が挙げられる。今回の実験では、質問紙に掲載する単語を、出現回数の多い上位10件とし、単語をすべて同じように扱った。しかし、単語によってレシピの概要での出現回数が大きく異なることから、出現回数に重みをつけた集計を行う必要があると考えられる。今後は重みをつけた集計によってクラスターを代表するキーワードが変化するか、変化するのであればどのように変化するかを検証する。

2つ目の課題として、アンケート調査の改善が挙げられる。今回の実験では料理経験の有無を問わずにアンケート調査を実施したが、cookpadは料理をする人向けに書かれているので、料理経験の豊富な人に限定してアンケート調査を行った場合、シズルワードを代表するキーワードに変化が見られると考えられる。

謝辞 本研究では、クックパッド株式会社と国立情報学研究所が提供する「クックパッドデータ」を利用しました。本研究はJSPS科研費JP16H02904の助成を受けたものです。ここに記して謝意を示します。

文 献

- [1] 加藤大介, 宮部真衣, 荒牧英治, and 灘本明代. インターネット上のメディア毎の「おいしさ」表現比較分析. *DEIM Forum*2015, 2015.
- [2] B・M・FT. おいしいを感じる言葉 *Sizzle Word 2017*. BMFT出版部, 2017.
- [3] 大橋 正房. シズルワードの現在と変化. 人工知能学会全国大会論文集, JSAI2016:3M4OS20b2-3M4OS20b2, 2016.
- [4] 石橋 賢, 深瀧 創, and 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*, 30(1):229-236, 2015.
- [5] 橘明穂, 若宮翔子, 難波英嗣, and 角谷和俊. 料理名のネーミングコンセプト抽出によるレシピ推薦システム. 第2回 *ARG Web* インテリジェンスとインタラクション研究会, pages 31-32, 2013.
- [6] 橘明穂, 若宮翔子, 難波英嗣, and 角谷和俊. 料理名の修飾表現の関係性に基づくレシピのネーミングコンセプト抽出. *信学技報*, *DE2013-36*, Sep, 2013.
- [7] 長谷川永奈, 小宮香乃, 齊藤史哲, and 石津昌平. 食における言