

レストランレビューにおける 料理特有の評価表現の抽出と分析

立命館大学大学院 情報理工学研究科
鈴木 稜也 山西 良典 西原 陽子

研究背景

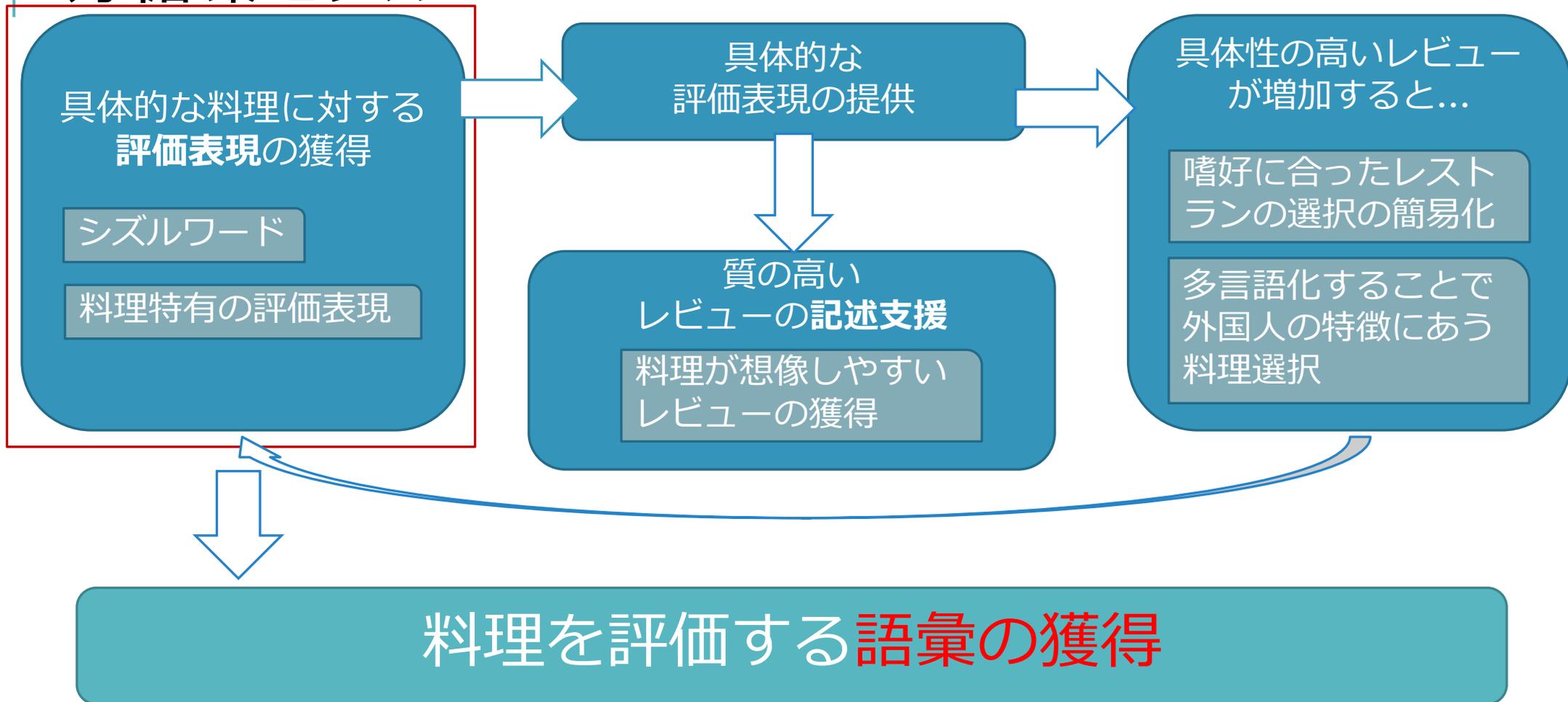
- 料理に対するレビューには**単純なもの**が多い
 - ☆☆☆☆☆:お肉が美味しい
 - ☆ :野菜がまずい...
⇒料理の様子の説明や評価に有用な**料理特有の表現が明らかではない**
- 情報量の少ないレビュー
 - レストランを検索するユーザ
 - レストラン側

上記2点にとって有益なレビューになり得ないことが考えられる



ユーザが求めるものは食事をすることで得られる**料理の具体的な説明**

レストランレビューの記述支援で得られる 好循環モデル



具体性のあるレビューとは？

単調なレビュー

いわゆる洋食屋さんのハンバーグ。

500円ハンバーグ。

料理の様子が想像できないレビュー

具体性のあるレビュー

大き目のハンバーグは柔らかくて
ジューシーで、ソースがピリ辛で
美味しいです。

シズルワードや
料理特有の評価表現を用いることで
料理の様子が想像しやすいレビュー

料理の評価表現

一般的な味表現 = シズルワード

あっさり

カリカリ

揚げたて

マイルド

冷え冷え

ヘルシー

味覚系
食感系
情報系

計362語

料理特有の評価表現 = シズルワード以外の主に名詞

デミグラスソース

和風ソース

照り焼きソース

料理の味付けに
対する評価表現

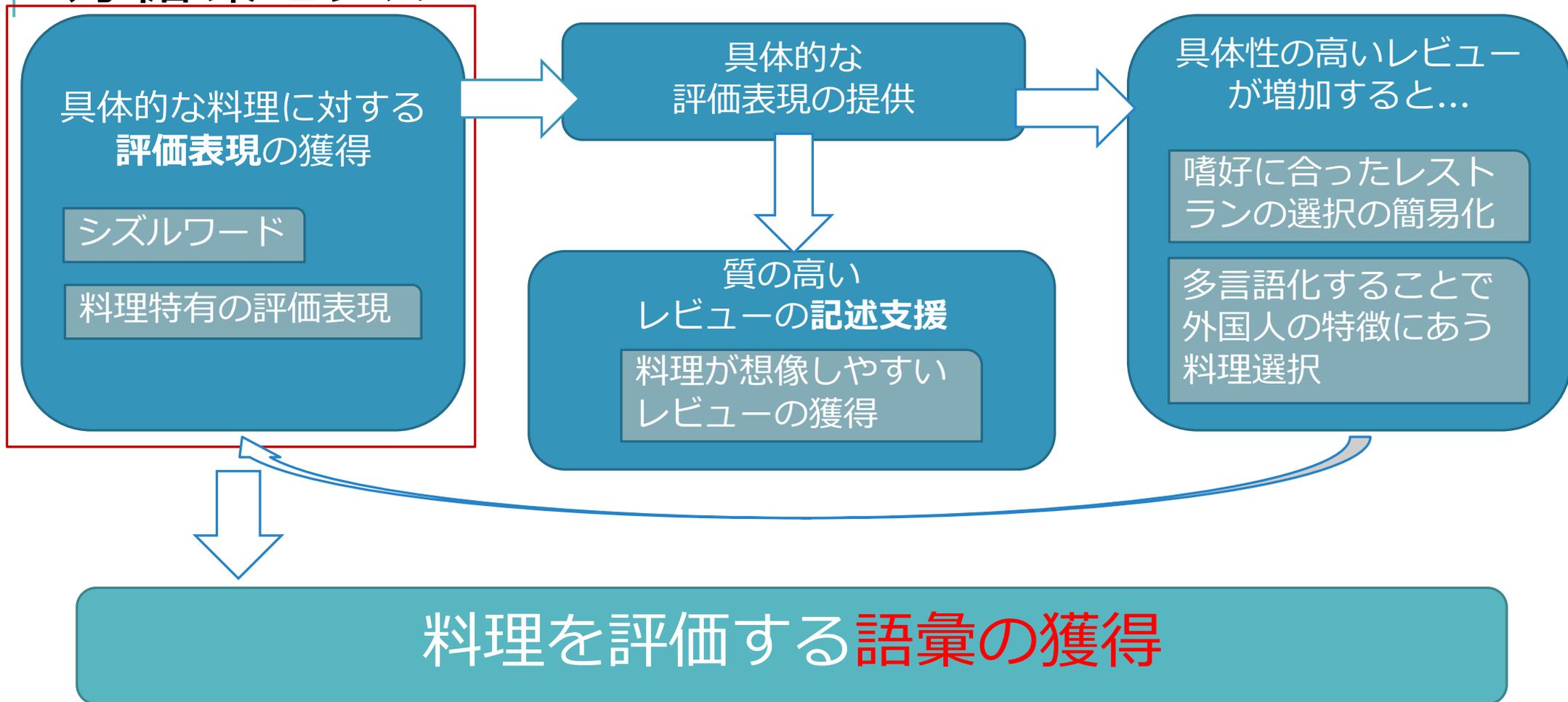
調理方法

値段

店の雰囲気

味に関すること以外の
料理に対する評価表現

レストランレビューの記述支援で得られる 好循環モデル

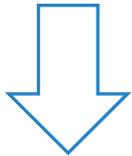


語彙獲得のための手法

具体的な料理に対する
評価表現の獲得

シズルワード

料理特有の評価表現



レビュー中での**評価表現**を獲得するためには...

データ準備

- ・ gurunaviのAPIより大量のレビューの取得
- ・ レビューを分析するために必要な部分を抽出

データ分析

- ・ 全文係り受け解析による分析
- ・ 単語単位での分析, TF-IDF値の計算

料理を評価する**語彙**の獲得

データ準備

1. gurunaviのAPIより料理ごとに**最大1000件**のレビューを取得
2. 取得したレビューを「, 」 「. 」, 記号で分割
3. 分析対象の料理名を含む文と, その前後3文の合計7文を取得

データ準備:レビューの取得

本研究で取り扱った50料理の一覧を示す

1. gurunaviのAPIより料理ごとに**最大1000件**のレビューを取得
2. 取得したレビューを「,」「.」, 記号で分割
3. 分析対象の料理名を含む文と, その前後3文の合計7文を取得

うどん	うなぎ	おでん	おはぎ	オムライス	お好み焼き	カツ丼	カフェラテ	カレー	きつねうどん
ケバブ	コーンポタージュ	コロッケ	しゃぶしゃぶ	シューマイ	すき焼き	ステーキ	ぜんざい	チーズケーキ	チャーハン
テールスープ	とんかつ	ハヤシライス	ハンバーガー	ハンバーグ	バンバンジー	ピザ	ブイヤベース	ポテトサラダ	マルゲリータ
メンチカツ	もつ鍋	ラーメン	ローストビーフ	ワンタンスープ	串カツ	唐揚げ	天丼	寿司	春巻き
杏仁豆腐	焼き鳥	牛タン	牛丼	納豆	肉じゃが	親子丼	豚キムチ	餃子	麻婆豆腐

※赤字の料理は1000件のレビューが取得できたものを指す

データ準備:分割したレビュー

「カレー」のレビュー例

1. gurunaviのAPIより料理ごとに最大1000件のレビューを取得
2. 取得したレビューを「,」「.」, 記号で分割
3. 分析対象の料理名を含む文と, その前後3文の合計7文を取得

3つ前の文	NULL
2つ前の文	NULL
1つ前の文	おいしいインドカレーが食べたくて訪問しました
料理名が含まれる文	辛党にお薦めのカシミール カレー を注文
1つ後の文	コクがありまたスパイスの風味が良かったです
2つ後の文	少々油っぽく感じましたが
3つ後の文	NULL

※赤字はシズルワードとする

料理名が含まれる文章の周辺には, その料理に関する情報が記載されていることが多い
→本研究では料理名が含まれる文章と対象周辺のレビューを分析対象とする

データ準備:分割したレビュー

「餃子」のレビュー例

1. gurunaviのAPIより料理ごとに最大1000件のレビューを取得
2. 取得したレビューを「,」「.」, 記号で分割
3. 分析対象の料理名を含む文と, その前後3文の合計7文を取得

3つ前の文	NULL
2つ前の文	NULL
1つ前の文	NULL
料理名が含まれる文	焼き餃子にはパリパリ系ともっちり系が有ると思いますが
1つ後の文	こちらは後者かな
2つ後の文	アンも多め少なめで言うと多めに具が入っています
3つ後の文	ビールとの相性も良く

餃子のレビューに「ビール」に関するレビューが存在
 餃子にとって「ビール」という単語は重要な評価表現だが係り受け解析では得られない
 →係り受け解析では得られない評価表現を単語の重要度を算出することで評価表現を獲得する

分析手法

1. 取得したレビューを対象とし，料理名を示す単語に対し
全文係り受け解析を行い，料理名を修飾している単語を取得
2. 取得したレビューを単語単位に分割し，
名詞，動詞，形容詞，形容動詞の4品詞に絞り込む
3. 2.で得られた単語に対して，TF-IDF法を用い，
単語の重要度を算出

分析手法:GINZAの出力

単語ID
文書単語
原型単語
品詞親ID

ハンバーグについての
レビュー

特製デミグラスソースとチーズと
ジューシーなハンバーグが
ベストマッチ

0 特製 特製 NOUN 1
1 デミグラスソース デミグラスソース PROP 3
2 とと ADP 1
3 チーズ チーズ NOUN 5
4 とと ADP 3
5 **ジューシー** ジューシー ADJ 7
6 だだ AUX 5
7 **ハンバーグ** ハンバーグ NOUN 10
8 がが ADP 7
9 ベスト ベスト NOUN 10
10 マッチ マッチ NOUN 10
EOS

ふわふわの煮込みハンバーグに
デミグラスソースが合っ
ておいしかったです

0 **ふわふわ** ふわふわ ADV 3
1 のの ADP 0
2 **煮込み** 煮込み NOUN 3
3 **ハンバーグ** ハンバーグ NOUN 7
4 にに ADP 3
5 デミグラスソース デミグラスソース PROP 7
6 がが ADP 5
7 合っ 合う VERB 9
8 てて SCONJ 7
9 おいしかっ 美味しい ADJ 9
10 たた AUX 9
11 です です AUX 9
EOS

分析手法:全文係り受け解析

料理名毎の係り受け解析によって

得られた料理名を修飾している単語の頻度上位5件を示す

1. データ準備3.で取得したレビューを対象とし、料理名を示す単語に対し全文係り受け解析を行い、料理名を修飾している単語を取得
2. データ準備3.で取得したレビューを単語単位に分割し、名詞、動詞、形容詞、形容動詞の4品詞に絞り込む
3. 2.で得られた単語に対して、TF-IDF法を用い、単語の重要度を算出

	うどん	カレー	ステーキ	ハンバーグ	ラーメン	餃子
1位	カレー	チキン	ソース	ジューシー	味噌	水
2位	讃岐	インド	サラダ	煮込み	塩	焼き
3位	かけ	カツ	ボリューム	チーズ	とんこつ	ジューシー
4位	煮込み	グリーン	スープ	デミグラスソース	博多	蒸し
5位	冷たい	スープ	柔らかく	ソース	豚骨	皮

ハンバーグについての
レビュー

特製デミグラスソースとチーズと
ジューシーなハンバーグがベストマッチ

ふわふわの煮込みハンバーグに
デミグラスソースが合っておいしかったです

分析手法:全文係り受け解析

料理名毎の係り受け解析によって

得られた料理名を修飾している単語の頻度上位5件を示す

1. データ準備3.で取得したレビューを対象とし、料理名を示す単語に対し全文係り受け解析を行い、料理名を修飾している単語を取得
2. データ準備3.で取得したレビューを単語単位に分割し、名詞、動詞、形容詞、形容動詞の4品詞に絞り込む
3. 2.で得られた単語に対して、TF-IDF法を用い、単語の重要度を算出

	うどん	カレー	ステーキ	ハンバーグ	ラーメン	餃子
1位	カレー	チキン	ソース	ジューシー	味噌	水
2位	讃岐	インド	サラダ	煮込み	塩	焼き
3位	かけ	カツ	ボリューム	チーズ	とんこつ	ジューシー
4位	煮込み	グリーン	スープ	デミグラスソース	博多	蒸し
5位	冷たい	スープ	柔らかく	ソース	豚骨	皮

「デミグラスソース」は味表現ではなく、ハンバーグの属性を示す単語。
→料理の詳細を評価するための単語

分析手法:全文係り受け解析

表は料理名毎の係り受け解析によって
得られた料理名を修飾している単語の頻度上位5件

1. データ準備3.で取得したレビューを対象とし、料理名を示す単語に対し全文係り受け解析を行い、料理名を修飾している単語を取得
2. データ準備3で取得したレビューを単語単位に分割し、名詞、動詞、形容詞、形容動詞の4品詞に絞り込む
3. 2.で得られた単語に対して、TF-IDF法を用い、単語の重要度を算出

	うどん	カレー	ステーキ	ハンバーグ	ラーメン	餃子
1位	カレー	チキン	ソース	ジューシー	味噌	水
2位	讃岐	インド	サラダ	煮込み	塩	焼き
3位	かけ	カツ	ボリューム	チーズ	とんこつ	ジューシー
4位	煮込み	グリーン	スープ	デミグラスソース	博多	蒸し
5位	冷たい	スープ	柔らかく	ソース	豚骨	皮

「水」が
用いられているレビュー

皮がモチモチで肉汁あふれる
とってもジューシーな水餃子です。

「焼き」が
用いられているレビュー

香ばしい焼き餃子は、
皮は薄めですがもちっとしていて
美味しいです。

分析手法:全文係り受け解析

表は料理名毎の係り受け解析によって
得られた料理名を修飾している単語の頻度上位5件

1. データ準備3.で取得したレビューを対象とし、料理名を示す単語に対し全文係り受け解析を行い、料理名を修飾している単語を取得
2. データ準備3.で取得したレビューを単語単位に分割し、名詞、動詞、形容詞、形容動詞の4品詞に絞り込む
3. 2.で得られた単語に対して、TF-IDF法を用い、単語の重要度を算出

	うどん	カレー	ステーキ	ハンバーグ	ラーメン	餃子
1位	カレー	チキン	ソース	ジューシー	味噌	水
2位	讃岐	インド	サラダ	煮込み	塩	焼き
3位	かけ	カツ	ボリューム	チーズ	とんこつ	ジューシー
4位	煮込み	グリーン	スープ	デミグラスソース	博多	蒸し
5位	冷たい	スープ	柔らかく	ソース	豚骨	皮

調理方法である「水」「焼き」「蒸し」が上位に見られた。
調理方法は料理によって大きく変化するもの。
→調理方法は味や食感を推測するための評価表現

分析手法:品詞抽出

1. データ準備3.で取得したレビューを対象とし、料理名を示す単語に対し全文係り受け解析を行い、料理名を修飾している単語を取得
2. データ準備3で取得したレビューを単語単位に分割し、名詞、動詞、形容詞、形容動詞の4品詞に絞り込む
3. 2.で得られた単語に対して、TF-IDF法を用い、単語の重要度を算出

- 使用した形態素解析器:MeCab
- 使用した辞書:NeologD
- 使用した品詞:名詞, 動詞, 形容詞, 形容動詞の4品詞
→付属語に評価表現が含まれることが少ないと推測した為
上記4品詞のみを使用

手順3.ではメニューごとに得られた単語を**頻度順**に並べ替え整理

分析手法:重要度算出

使用したTF-IDF法の計算式

- それぞれの単語の文書内での出現頻度TF:

$$tf(t, d) = \frac{n_{t,d}}{\sum_{s \in d} n_{s,d}}$$

$n_{t,d}$: ある単語 t の文書 d 内での出現回数
 $\sum_{s \in d} n_{s,d}$: 文書 d 内のすべての単語の出現回数の和

- それぞれの単語がいくつの文書内で共通して使われているかを表すIDF:

$$idf(t) = \log \frac{N}{df(t)} + 1$$

N : 全料理数 = 50
 $df(t)$: ある単語 t が出現する文書の数

$$tfidf(t, d) = tf(t, d) \times idf(t)$$

TF-IDF値が高い単語は、その料理に頻出し、他の料理にはあまり出現しない。
 →料理に対する評価を詳細に推測しやすく **出現頻度が高い**単語

- データ準備3.で取得したレビューを対象とし、料理名を示す単語に対し全文係り受け解析を行い、料理名を修飾している単語を取得
- データ準備3.で取得したレビューを単語単位に分割し、名詞、動詞、形容詞、形容動詞の4品詞に絞り込む
- 2.で得られた単語に対して、TF-IDF法を用い、単語の重要度を算出

分析手法:重要度算出

TF-IDF表(単語の頻出順の昇順)

・ラーメン(シズルワード)

評価表現	TF-IDF値
ボリューム	0.16×10^{-2}
濃厚	0.70×10^{-2}
シンプル	0.20×10^{-2}
絶妙	0.09×10^{-2}
ジューシー	0.02×10^{-2}
新鮮	0.04×10^{-2}
あつあつ	0.00×10^{-2}
ヘルシー	0.05×10^{-2}
まろやか	0.20×10^{-2}
贅沢	0.04×10^{-2}

・ラーメン(シズルワード以外)

評価表現	TF-IDF値
スープ	1.73×10^{-2}
麺	2.07×10^{-2}
チャーシュー	1.34×10^{-2}
醤油	0.64×10^{-2}
味噌	0.48×10^{-2}

約3倍

1. データ準備3.で取得したレビューを対象とし、料理名を示す単語に対し全文係り受け解析を行い、料理名を修飾している単語を取得
2. データ準備3.で取得したレビューを単語単位に分割し、名詞、動詞、形容詞、形容動詞の4品詞に絞り込む
3. 2.で得られた単語に対して、TF-IDF法を用い、単語の重要度を算出

シズルワードとシズルワード以外の単語を比べると重要度の最も高い単語では値に約3倍もの差が見られた。
→一般的な味表現よりも、料理特有の評価表現の方がレビューを書く上で情報が多く料理を想像しやすいものと考えられる。

分析手法:重要度算出

「濃厚」が用いられている
レビュー

都内にチェーン展開する、
ラーメンブームの火付け役になった
ラーメン店の一つ。濃厚なスープが
病みつき系の味。結構お気に入り。

「麺」が用いられている
レビュー

カツオだしの効いた醤油ラーメンに、
もやしのシャキシャキが美味しい。
スープは透き通っていて
さっぱりとしたラーメン。麺は
極細で、スープがよくからみます。

1. データ準備3.で取得したレビューを対象とし、料理名を示す単語に対し全文係り受け解析を行い、料理名を修飾している単語を取得
2. データ準備3.で取得したレビューを単語単位に分割し、名詞、動詞、形容詞、形容動詞の4品詞に絞り込む
3. 2.で得られた単語に対して、TF-IDF法を用い、単語の重要度を算出

「濃厚」は料理特有の
評価表現に対しての評価となる。
(この場合は「スープ」に対して)

「麺」はラーメンそのものを
想像させる評価表現となる。

料理特有の評価表現にシズルワードを
組み合わせることで、料理に対しての
具体的なレビューを作成することが可能に

分析手法:重要度算出

TF-IDF表(単語の頻出順の昇順)

・ハンバーグ

評価表現	TF-IDF値
ソース	1.17×10^{-2}
肉汁	0.92×10^{-2}
チーズ	0.68×10^{-2}
デミグラスソース	0.86×10^{-2}
レア	0.40×10^{-2}

1. データ準備3.で取得したレビューを対象とし、料理名を示す単語に対し全文係り受け解析を行い、料理名を修飾している単語を取得
2. データ準備3.で取得したレビューを単語単位に分割し、名詞、動詞、形容詞、形容動詞の4品詞に絞り込む
3. 2.で得られた単語に対して、TF-IDF法を用い、単語の重要度を算出

・ステーキ

評価表現	TF-IDF値
ソース	0.65×10^{-2}
サラダ	0.54×10^{-2}
リーズナブル	0.36×10^{-2}
鉄板	0.43×10^{-2}
焼き	0.38×10^{-2}

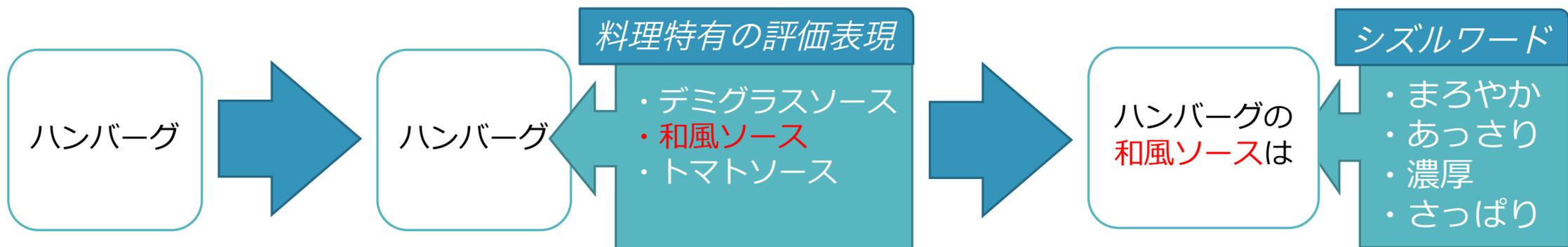
「肉料理」という同一ジャンルで分析結果を比較。
ステーキとハンバーグの周辺単語で「ソース」に関わる単語の重要度が高く見られた。
→料理のジャンルごとに**重要となる単語が共通**する可能性が示唆された。

まとめと今後の展望

全文係り受け解析と単語における重要度計算により、
料理を評価する語彙の獲得をした

<今後の展望>

- ・ 多言語レビューでの分析も検討
 - ・ 外国人が日本食に対する評価は味覚の違いによって変わることがある。
→実際に味噌カツのソースの表現に「バーベキューソースのようだ」が存在
- ・ レビュー記述支援システムの構築
 - ・ 語彙力の低いレビュアーでも質の高いレビューが可能に



分析手法:全文係り受け解析

全文係り受け解析によって

形態素解析された単語と料理名を修飾している単語が明らかになる。

- ・全文係り受け解析

↳ □自然言語処理ライブラリ「GiNZA」を利用

- ・GiNZAの特徴

- 1)高度な自然言語処理をワンステップで導入完了
- 2)高速・高精度な解析処理と依存構造解析レベルの国際化に対応
- 3)国立国語研究所との共同研究成果の学習モデルを提供

1. データ準備3.で取得したレビューを対象とし、料理名を示す単語に対し全文係り受け解析を行い、料理名を修飾している単語を取得
2. データ準備3で取得したレビューを単語単位に分割し、名詞、動詞、形容詞、形容動詞の4品詞に絞り込む
3. 2.で得られた単語に対して、TF-IDF法を用い、単語の重要度を算出

GiNZAを用いることで精度の高い係り受け解析が可能

まとめ

- ・ **料理を評価する語彙の獲得**をした

全文係り受け解析

- ・ 料理そのものの属性を示す単語が上位に見られ、
それらは**料理の詳細を示すための単語**として捉えられた
- ・ 調理方法として捉えられる単語が上位に見られ、
それらは**”味や食感を推測するための単語”**と推測される

単語における重要度計算

- ・ シズルワードよりも**料理特有の評価表現**の方が料理ごとに用いられる単語の重要度として高く見られた
- ・ 料理特有の評価表現にシズルワードを組み合わせることで
料理が想像しやすいレビュー作成することができる可能性が示唆された
- ・ 料理のジャンルごとに**重要となる単語が共通**している可能性が示唆された

今後の展望

- 分析対象の料理数の増加
 - 分析した料理を「肉料理」「魚料理」「サラダ」などにジャンル分けすることにより、**ジャンルごとの評価表現の重要度を算出**
- 多言語レビューでの分析も検討
 - 外国人が日本食に対する評価は**味覚の違い**によって変わることがある。
→実際に**味噌カツのソースの表現**に「バーベキューソースのようだ」が存在
- レビュー記述支援システムの構築
 - **語彙力の低いレビュアー**でも質の高いレビューが可能に

