学位申請論文　2017 年度（平成 29 年度）

食の描写表現の相違を用いた
言語圏ごとの料理の感じ方の比較手法の研究

指導教員 坂田 一郎 教授

東京大学 工学部 システム創成学科 PSI コース

03-160988　中元 雪絵

2018 年 2 月 22 日 提出

学位申請論文　2017 年度（平成 29 年度）

食の描写表現の相違を用いた
言語圏ごとの料理の感じ方の比較手法の研究

要約

グローバル化に伴い, 食産業でも海外進出が増加しているが, 進出先の地域社会へのローカライゼーション戦略を打ち出し成功する事例が多く見られるようになった. ローカライゼーション戦略の肝となるのは, 地域ごとの文化的差異を理解することである. これまで文化に関する研究は, アンケート調査や過去の文献を用いたものがほとんどであったため, 大規模な調査を行う事は難しかった. 近年, スマートフォンの普及に伴い, 生活に密着した情報発信手段であるソーシャルメディアが広く用いられるようになった. ソーシャルメディアのデータを用いて, 消費者の行動を世界規模で追うことが可能となり, トレンドの検出やマーケティングの手段として非常に期待されている.

本研究では, ソーシャルメディアを用いて文化的差異を検出する手法を提案した. ソーシャルメディアの感想データから, 分析対象と, その対象に関する描写表現を取得する. 取得した描写表現をコンセプト単位に分類し, 各コンセプトに対する共起頻度を求める. この頻度分布を比較することにより, 異なる言語圏における感じ方の違いを検出する.

実験では, 描写表現として形容詞を用い, 英語圏と日本語圏の食に関する Twitter データを対象とした分析を行った. 分析結果より, 言語圏全体の傾向として, 食に関するソーシャルメディアの使い方や, 食一般に対する考え方へのアメリカと日本の差異を提案手法により取得できることを示した. また, コーヒーなどの特定の飲料を対象とした感じ方の差異を議論し, 提案手法を用いることで, 具体的な個別の対象に関して, 言語圏ごとの感じ方の比較が可能となることを示した.

以上から, 本研究は描写表現を用いた文化間比較の有用性を示すとともに, 異なる言語圏に対し, 描写表現を統一尺度で測り, 比較に用いる有効な手法の開発に成功した.

東京大学 工学部 システム創成学科 PSI コース

03-160988　中元 雪絵

目 次

図 目 次

表 目 次

第1章　序論

地域が異なれば人々の感性が異なり, 物事の感じ方が異なる. 例え全く同じ材料で作られた一杯のコーヒーであっても, コーヒーがどの地域で消費されるかが異なれば, そのコーヒーが飲む人に与える印象は大きく変化するだろう. それぞれの地域に, 特有の飲食習慣, コーヒーへのイメージ, 好きなコーヒーの風味が存在し, これら全てが人々のコーヒーの感じ方に影響を与え, 異なる感想を抱かせる.

近年, グローバル化に伴う世界各国企業の海外進出が増加し, 地域社会へのローカライゼーション戦略を打ち出し成功する事例が多く見られるようになった. ローカライゼーション戦略の肝となるのは, 地域ごとの文化的差異を理解することである. これまで文化に関する研究は, アンケート調査や過去の文献を用いたものがほとんどであったため, 大規模な調査を行う事は難しかった. 近年, スマートフォンの普及に伴い, 生活に密着した情報発信手段であるソーシャルメディアが広く用いられるようになった. ソーシャルメディアのデータを用いて, 消費者の行動を世界規模で追うことが可能となり, トレンドの検出やマーケティングの手段として非常に期待されている. 本研究では, ソーシャルメディアにおける描写表現を用いて文化的差異を検出する手段を提案し, 特に食文化を取り上げて分析を行った.

本章では, 上記の背景について詳細に説明し, 問題意識と目的を明らかにした上で, 描写表現の相違を用いてコミュニティごとの考え方の違いを比較する手法を提案する. そして, 実験結果の概略をもとに本研究の貢献を説明する.

1.1　研究の背景

1.1.1　企業のローカライゼーション戦略と異文化理解の重要性

近年私たちの生活は多様化を続け, 人々の需要を把握するのが困難になりつつある. グローバル化に伴う市場拡大を背景に, 世界各国において貿易の自由化や規制緩和が実施され, 各国の商品やサービス, 果ては文化までもが輸出入されてきた. 消費者は様々な国に由来する商品やサービスを選択することが可能であり, 企業は国際競争力を持つことが今や生存の必須条件となりつつある.

海外進出の際に, 多くの企業はローカライゼーション戦略を用いる. その際, 事前準備として現地の地域性や文化, 法律や規則, 商品やサービスへのニーズを理解することが必須となる. 自国で売れ行きの良い製品やサービスをそのまま海外市場に持ち込んだとしても, 現地で受け入れられるとは限らず, 進出する国の国民性や習慣・特性を理解し, より消費者に合った製品やサービスの展開を行うことが海外進出成功の鍵となる.

例えば自動車メーカーのスズキ社はインドでシェア 1 位を誇るが, その背景には, 製造拠点をインドに配置し, インド人のニーズに合わせたデザインを現地で実現する, 徹底的なローカライゼーション戦略がある. また, 東洋水産のカップ麺はメキシコで大成功を喫し国民食となったが, カップ麺の食べ方は日本人とは大きく変化している. 麺は伸びてから食べ, 濃い味の上にさらに, レモンと, パレンティナと呼ばれる辛いソースで味付けする.

異なる地域における, 文化やニーズの理解には大きな労力を必要とする. ほとんどの場合, 企業は長い期間をかけて現地の言語による実地調査やアンケート調査を行う. 多大な時間とコストを払ったにも関わらず, 実際に進出した後に予期していなかった課題に直面し, 失敗に終わった事例も多く存在する. グローバル化がますます進行していくであろう今後の社会において, 異文化理解の重要性も同時に増加することが予想される.

1.1.2 ソーシャルメディアの普及と可能性

需要の多様化が進む中, 大衆の生の意見を即時に取得できるツールとして, ソーシャルメディアのデータに注目が集まっている. 本項ではその背景について説明する.

(1) ソーシャルメディアについて

ソーシャルメディアには, Twitter をはじめとしたマイクロブログの形態を取るものや, 食べログや TripAdvisor のような特定の対象に対するレビューサイトなどがある.

マイクロブログとは, 利用者が短いテキストをウェブサイトへ投稿し, リアルタイムに他の利用者と共有されるサービスである. 利用者は自身が情報を発信すると共に, 他の個人が発信した情報を閲覧することが可能であり, 多くの場合, 利用者間でコミュニケーションが発生する. 例えば Twitter では, 日々のつぶやきのメッセージを, 140 字以内という制限をつけて利用者が投稿する. 利用者が手軽かつ気軽に情報発信を行うことのできる性質から, 全世界の支持を集め, 今や 3 億人以上ものユーザー数を誇るソーシャルメディアとなっている.

レビューサイトとは, 人, 事業, 製品, サービスについて感想やコメント（レビュー）を投稿できるサービスである. 一般に利用者による投稿は匿名で行われ, レビューの平均値は対象の評価を表す一つの指標となる. 事業を提供する側にとっては, レビューの平均値を上げることが一つの広報手段となり, 同時に他の利用者にとっては, サービスや製品選択の際の参考となる.

これらソーシャルメディアはスマートフォンの普及とともに人々の生活に浸透した. 逐一自分の行動をソーシャルメディア上に投稿する行動は現代社会においては全く珍しくなく, また, 消費者がソーシャルメディアを用いて情報を得ることを前提としたマーケティングが企業には求められている.

(2)　ソーシャルメディアのデータの利活用

ソーシャルメディアから取得される情報は, 個々人がリアルタイムに投稿した膨大な生のデータであり, 有効に情報を整理して評価することができれば, その価値は計り知れない. 既存の研究では, これらソーシャルメディアにおける閲覧履歴や過去の評価内容を用いた推薦アルゴリズムの開発が盛んに行われてきた. 近年はソーシャルメディアのリアルタイム性に着目した研究が多く発表され, アメリカ大統領選など特定のニュースに対する世論調査や, トレンドを予測する手段が提案されている. 詳しくは 2 章の関連研究で説明する.

一方で, ソーシャルメディアはそのリアルタイム性の他に, 他のメディアとは異なるもう一つの特徴がある. 投稿者の主観的な考えを多く含んだ情報メディアだという点である. 消費者の動向や特徴については, 新聞やニュース記事, テレビにおいても特集されることがあるが, これらのメディアの情報は, 中立的な立場を保つ必要があるため, 主観的な考えを排除している. しかし, ソーシャルメディアは匿名で気軽に情報を発信するという特徴から, ユーザの主観的な意見を多く含んでいる. これらの豊富な主観的考えは, 対象に対する感情がポジティブかネガティブか調査する目的でその一部が用いられ, 感情を表現する単語や調査対象に関係する単語以外はノイズとして排除されることが多い. しかし, 排除されている表現もまた, その投稿者個人を説明する重要な情報である. 主観的な意見を多く含む性質を利用して, コミュニティごとの表現の偏りを定量的に評価することで, そのコミュニティの特徴を捉えることはできないだろうか.

1.1.3　ソーシャルメディアを用いた文化的特徴の抽出

ソーシャルメディアを用いて, 地域やコミュニティの文化的距離を測る研究は存在する. Silva らは, Foursquare というソーシャルメディアの飲食行動に関するデータを用いて, 世界の国と地域に対して文化境界を定める手法を提案した [1]. Sajadmanesh らは, インターネット上の世界中のレシピのデータを用いて, 味や材料などの側面からみた各国食習慣の距離を算出した [2]. これらの研究では, ソーシャルメディアのデータを用いて, 文化の距離を世界規模で議論することに成功しているが, 一方で, 各文化の間に具体的にどのような差異があるのか検出することは難しい.

1.2　本論文の目的

本研究では, ソーシャルメディアのデータを用いて, 地域やコミュニティごとの文化的差異に基づく対象の捉え方の違いを明確化することを課題とする. そして, これらの考え方の違いは, 無意識に用いている言語的表現に現れているという仮説を設定し, 検討することを目的とする.

そのため, ソーシャルメディアのデータを用いて, 異なるコミュニティの人々が同じ対象を描写する表現の違いを分析することで, 対象の捉え方の違いを比較する手法を提案する. 特に, 文化の中心であり, ソーシャルメディア上での投稿も多い食文化に着目し, 英語言語圏と

日本語言語圏の間にどのような差異があるか, 既存の研究から得られる知見と比較しながら具体的に分析し, 有効性を示すことで検証を行う.

1.3 本論文の貢献

本研究の貢献は以下の通りである.

描写表現の相違を用いて異なる言語圏の考え方の相違を比較する手法の提案

異なるコミュニティに属する人々が, 対象を描写するのに用いている, 描写表現の分布の違いを比較することで, コミュニティの間の文化的差異を検出する手法の提案を行った.

文化間距離の新たな基準としての感性の提案

人が対象をどのように感じるか, その感性の分布の相違に着目し, 文化的距離を測ることを提案し, 有効性を示した.

日米の食文化の知見とローカライゼーション戦略への応用

日本語言語圏と英語言語圏の食文化の違いについて, 新たな知見を提示し, それを元にローカライゼーション戦略への応用の可能性を示した.

1.4 本論文の構成

第 2 章では, ソーシャルメディアや, 食文化に関する関連研究と, 本研究の位置付けについて述べる．第 3 章では, 描写表現を用いて異なる言語圏について感じ方の違いを検出する提案手法について述べる．第 4 章では Twitter の食に関する英語・日本語の感想データを対象とした実験について述べる．第 5 章では実験結果について述べる．第 6 章では実験結果に対する考察を行い, 提案手法の有用性や課題について議論する. 第 7 章では本研究の結論について述べる.

第2章　関連研究

本章では, ソーシャルメディアの分析と食文化比較について関連研究を概説し, 本研究の位置づけについて述べる.

2.1　ソーシャルメディアの分析

2.1.1　Twitter の分析

Twitter を取り上げたソーシャルメディアの研究は, Twitter のサービス開始当初から現在に到るまで盛んに行われている. Twitter は Facebook や Google Plus に比べ情報伝達の速度が速いプラットフォームであることが示されており, 高いリアルタイム性を持つソーシャルメディアである [3]. 情報伝播の仕組みについての研究は広く行われている. Sadri らは危機状況下での情報拡散に特徴的なユーザの振る舞いとネットワーク構造について示した [4]. Myers らは Twitter 外部の影響を考慮した情報拡散モデルを提案し, 情報拡散の約 3 割が Twitter 外部に起因することを示した [5].

また, リアルタイム性が高いという特徴を生かし, Twitter からトレンドを検出する研究もまた多く見られる. Zhou らはソーシャルメディアの内容, 時間, 場所の情報を用いて実世界のイベントを検出するモデルを提案し, 実際に Twitter にモデルを応用することでその有用性を示した [6]. 榊らは, Twitter ユーザを「ソーシャルセンサ」として捉え, 物理センサでは観測し得なかった現象を観測する可能性を示した [7]. また, Twitter のトレンド入りキーワードを 1 時間以上前に高い正解率で予測するモデルも提案されている [8].

2.1.2　ソーシャルメディアによる感情分析と意見マイニング

テキストの感情分析は現在急速に成長している分野であり, Twitter をはじめとする, ソーシャルメディアのテキストデータは感情分析や意見マイニングの対象として近年注目を集めている [9].

感情分析の最もシンプルかつ普遍的なタスクは, 英語の文章を Positive, Negative, Neutral の三つの軸に分類するタスクである. このタスクを実現するモデルには, Pak らの提案したソーシャルメディアから自動で集積されたコーパスを用いた感情分類器 [10] や, Baccinella らの提案した SentiWordNet と呼ばれる感情分類器 [11] が存在する. 英語以外の言語を対象にしたモデルや, より複雑なタスクを実現するモデルも多く存在し, Su らは中国語文中の隠された意味を判断可能な意見マイニングのモデルを提案した [12].

感情分析や意見マイニングは, 市場調査や政治学, 社会学など, 一般に世論を調査する目的に対して幅広く応用されている [13]. Hu らは, 特定の産業やブランドに対する消費者の感情をソーシャルメディアの反応から特定した [14]. また, 複数のデータセットに対して, 伝統的な世論調査と Twitter による世論調査の間に相関傾向が見られることも知られている [15].

2.1.3 ソーシャルメディアによる文化比較

ソーシャルメディアを用いた言語間や地域間の文化比較には, Gao らの研究がある [16]. この研究では, Twitter と Sina Weibo のユーザの行動を比較し, 社会学にて提唱されるアメリカと中国の文化モデルと, ユーザ行動との間に相関が考えられることを示した. また, Hu らは, LinkedIn から集めた職業に関するデータと, Twitter から集めた言語圏, 興味, 個性に関するデータを用いて, これらの間に特定のパターンが存在することを示した [17].

食に着目した研究も多く行われている. Silva らは, Foursquare というソーシャルメディアの飲食行動に関するデータを用いて, 世界の国と地域に対して文化境界を定める手法を提案した [1]. Sajadmanesh らは, インターネット上の世界中のレシピのデータを用いて, 味や材料などの側面からみた各国食習慣の距離を算出した [2]. Thanh らは, スイスで行われた Instagram の投稿から, スイスでの食の消費パターンについて, 西欧の分類とは異なるパターンが見られることを示した [18].

2.2 食文化に関する研究

食文化に関する研究は盛んに行われている. 各文化圏における食の嗜好傾向や, 食事を選択する基準の解明が様々な角度からなされており, 特にアンケート調査や文献を用いた研究が広く行われてきた. Prescott らは日本とオーストラリアにおける味の認識と嗜好の違いについて議論し, 異文化間の比較における課題を示した [19]. Steptoe らは, 食を選択する動機を計測する手段として, 9 の要因と 36 の因子から成る FCQ(Food Choice Questionaire) という質問票を提案した [20]. Pearcey らはこの質問票を用いて中国人とアメリカ人大学生に対する調査を行い, 中国人学生が自然素材と食倫理を重視するのに対し, アメリカ人学生は価格と手間の少なさを重視する傾向があることを示した [21]. Freedman は, 日本の民族誌の分析から, 日本人の食選択の基準として種類の豊富さが存在することを示した [22].

2.3 本研究の位置づけ

既存研究と照らし合わせた本研究の位置づけ・有用性は以下の通りである.

- ソーシャルメディアのテキスト情報に対し, 表現の分布というより豊富な情報を用いた手法の提案
- 文化の評価方法に対し, 感性という新たな基準の提案

- 文化圏全体ではなく, 特定の対象ごとの文化間距離の比較手法の提案

第3章　提案手法

本章では, ソーシャルメディア上で用いられている描写表現を利用し, 異なるコミュニティの人々の表現分布の違いを見ることで, 文化的差異を比較する手法を提案する.

3.1　着想

ソーシャルメディアは私たちの生活に密着したツールであり, 毎日の何気ない発言であふれている. 利用者は気軽に自分の意見や感想を投稿し, 日々膨大な量の情報がインターネット上で共有される. 匿名で気軽に主観的な考えを共有できることがソーシャルメディアの特徴であり, だからこそ多種多様な人の感性や価値観がソーシャルメディアの投稿に如実に現れる.

Fig. 3.1 は, Twitter における, 英語と日本語のラーメンに関する投稿の例である. 英語のツイートでは, ラーメンは嫌いだが辛くてスパイシーな麺が食べたい, という反応が見られ, 日本語のツイートでは, 真冬の夜中にラーメンが食べたいという旨の発言が見られる.

この投稿から何か読み取れることはあるだろうか. 少なくとも, 英語ツイートの投稿者は麺類に対して辛くてスパイシーなものを求めているはずであり, 日本語ツイートの投稿者はラーメンに対して「真冬の夜中」という場面設定や, 旨味を想起することが推測できる. もしも, さらに多くの投稿を分析した時に, 英語と日本語それぞれに対し, 同じ傾向があったとしたらどうだろうか. それは英語圏と日本語圏の麺類に対する考え方や感じ方に違いがあるとは言えないだろうか.

このように, ソーシャルメディア上で人々が無意識に用いている描写表現は, 文化圏における感性や価値観を反映している可能性がある. 描写表現の使い方の違いを分析することで, 異なる文化圏において, 対象に対して一般にどのような認識があるのか, また, どのような文

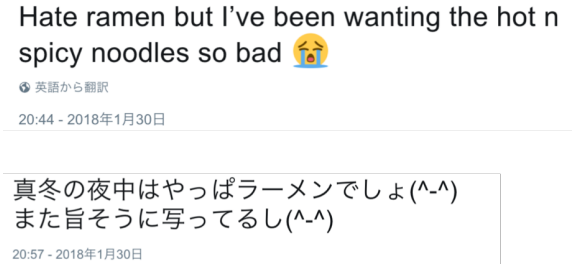


Fig. 3.1: Twitter におけるラーメンへの反応の例

脈に対象が登場するのかといった違いが見えてくる. この文化的差異を定量的に評価することで, これまでわからなかった文化圏ごとの考え方の違いを明確化することができる.

しかし, ここで問題となるのが, 異なる言語の描写表現をどのように統一尺度で測るかということである. 最も簡単なのは辞書を用いて翻訳することだが, 辞書での表記と厳密に一致しないものの近い意味を持つ表現は多く存在する. 全ての描写表現を言語間で一対一対応させることは難しく, この方法では描写表現全体を扱うことができない. また, トピックモデルを用いた文書分類法を用い, 出現するツイートを分類するという方法もあるが, この場合, 「感性の分布」によってトピックが分類され, 感じ方ひとつひとつを区別して扱うことが難しい.

そこで本研究では, 言葉の持つ意味や概念に着目し, 近い意味を持つ言葉を言語に関わらず全て同じコンセプトとして扱うことが有効であると考えた. コンセプトを定義する際は, 複数言語の言葉同士の対応関係を示すデータを用い, 意味の近さでエッジを定義したネットワークを構築し, ネットワーク上で近い位置に集まった表現群を一つのコンセプトとみなす. これにより複数の言語の表現を各コンセプトに分類し, 言語を超えた尺度で扱うことが可能となる.

それぞれの言語圏で, 分析対象に対してどのようなコンセプトを多く用いているか, その頻度を比較することで異なる言語圏における文化的差異の検出を実現できると考えた.

3.2 文化的差異の検出法

本研究では, ソーシャルメディア上で用いられている描写表現を利用し, 異なる言語圏の人々の表現分布の違いを見ることで, 対象への感じ方の文化的差異を比較する手法を提案する. 以下, 本研究の提案手法の手順を説明する. 提案手法の概要を Fig. 3.2 に示す.

入力として, 分析対象のリストと, これら対象に関する複数言語における感想のデータ, さらに言語間の対応性に関するデータベースの 3 つを用いる.

まず, 感想データから, それぞれの言語における描写表現群を抽出する. 抽出した全ての描写表現に対し, 近い意味を持つ表現の間に, 関係の近さを示す重みを付与する. この際に, 異なる言語間の言葉の対応性に関するデータベースを入力として用いることで, 異なる言語の表現同士を結ぶ.

各表現をノード, 表現同士の対応関係をエッジとみなし, ネットワークを構築する. このネットワークを描写表現ネットワークと呼ぶことにする. (**描写表現ネットワークの構築**)

構築した描写表現ネットワークをクラスタリングし, 取得されるクラスタ一つ一つをコンセプトとみなす. この工程により, 複数言語の表現をコンセプトで分類し, 同じ尺度で扱うことが可能になる. (**コンセプトクラスタの検出**)

ここで, 入力に用いた感想データと, 分析対象のリストから, 分析対象それぞれに対して, 描写表現がどのように共起したかを求めておく. (**対象と描写表現の共起関係の取得**)

対象と描写表現との共起関係を, 先で検出したコンセプトクラスタ単位で合計し, 分析対

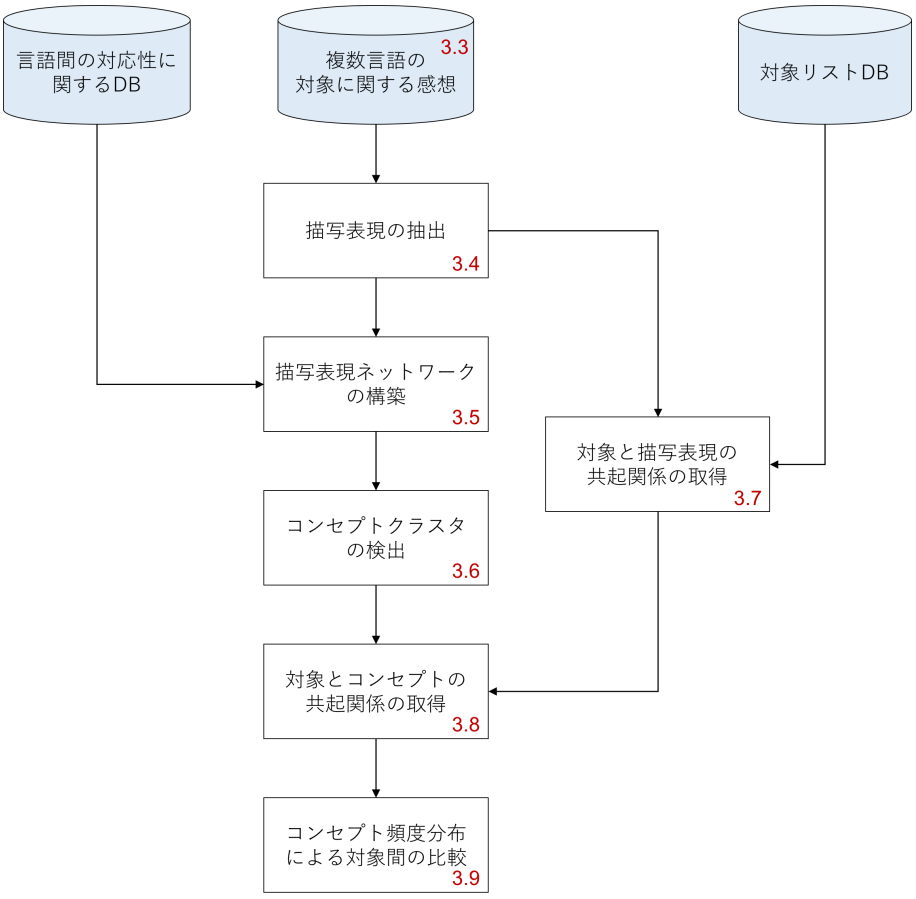


Fig. 3.2: 提案手法の概要

象リストと各コンセプトクラスタの間の頻度分布を計算する．これをコンセプト頻度分布と呼ぶことにする．（**対象とコンセプトの共起関係の取得**）

最後に, 頻度分布の類似度を計算することで, 対象リストに属する項目間の類似度を得る．また, コンセプト頻度分布の差が大きいコンセプトクラスタを比較することで感じ方の差異を検出する．（**コンセプト頻度分布による対象間の比較**）

以上が本論文で提案する手法の概要である．本研究では, 形容詞を描写表現とみなし分析を行った．以下で形容詞を取り上げながら詳細を述べていく．

Table 3.1: 形態素解析に用いたツール

	英語	日本語
形態素解析エンジン	TreeTagger	MeCab
辞書名	-	mecab-ipadic-NEologd
形容詞の品詞分類	JJ, JJR(比較級), JJS(最上級)	形容詞 (自立, 接尾, 非自立)

3.3 感想データの取得

提案手法では, 入力として, 複数言語の対象に関する感想のデータを用いる. Twitter などのマイクロブログや, レビューサイトを対象とし, 描写分布を求めたい対象について述べているデータを取得する. コミュニティにおける個人の発言を抽出することが目的であるため, Twitter における bot や RT など, ユーザの発言でないものは前処理として除外する.

本研究においては, 料理レシピを投稿するソーシャルメディアのカテゴリをクエリとし, TwitterAPI を用いて日本語と英語に対して感想データを収集した. 詳細については, 第 4 章で説明する.

3.4 描写表現の抽出

この手順では複数言語の感想データから, それぞれの言語について描写表現を抽出する. 描写表現には, 対象自体の様子を伝える言葉や, 対象に対する自分の感情表現, 対象の存在する環境の特徴や印象が例として含まれる.

本研究では, 対象に関する描写表現として, 英語と日本語の形容詞を用いることとした. 例えば, 何を「可愛い」と表現し, 何を「美しい」と表現するかは, その個人の感性に基づくものであり, 対象をどのように感じるかを伝える最も簡単な方法である. 従って, 形容詞の分布を言語圏の間で比較することで, それぞれの言語圏に属する人々が持つ感性の傾向の差異を議論できると考えた.

感想データから形容詞を抽出する手順については, まずそれぞれのデータを形態素解析により, 意味を持つ最小限の単位（単語）に分割し, 品詞が形容詞と判断された単語のみを取得する. 日本語の形態素解析には MeCab[1], 英語の形態素解析には TreeTagger[2]を用いた. また, 今回の分析対象は Twitter の短文であり, 新語の出現頻度が多いと予想されたため, 新語や固有表現を多く含む辞書である NEologd[3]を日本語の形態素解析に用いた. 用いたツールの詳細な情報について Table 3.1 に示す. また, 形態素解析により取得できる形容詞の例について Table 3.2 に示す.

なお, それぞれの描写表現をポジティブな意味で用いているか, ネガティブな意味で用いているかは区別せず, 同一の表現として扱った. その理由については 3.9.2 項で説明する.

[1]http://taku910.github.io/mecab/
[2]http://www.cis.uni-muenchen.de/ schmid/tools/TreeTagger/
[3]https://github.com/neologd/mecab-ipadic-neologd

Table 3.2: 形態素解析と取得形容詞の例

言語	原文	取得形容詞
英語	my second favorite cider is something i bought randomly in DC that i think is called Blue Ridge and it was DELICIOUS	second, favorite, delicious
日本語	なんで水切りヨーグルトってこんなに美味しいの. 400 グラム百何十円の安いやつなのにめっちゃ美味しい. 贅沢にはちみつかけてみた.	美味しい (2 回), 安い

3.5 描写表現ネットワークの構築

抽出した描写表現に対し, 近い関係のある表現同士に重みを付与して繋ぐことでネットワークを作成する. この際, 異なる言語の単語同士の関係を考慮に入れるため, 複数言語の言葉の関係を与えるデータベース入力として用いる.

本研究では, 形容詞を描写表現として分析を行う. 英語と日本語の形容詞について関係を取得するため, ConceptNet[4]を用い, 英語同士, 日本語同士, 英語と日本語の形容詞の全てのパターンについて, 近い関係を持つペアに対し, 重みの付与を行った. ConceptNet を採用したことには以下の理由がある.

- 今回の分析対象である英語と日本語双方について十分なデータ量と信頼性がある
- API を用いて関係の情報を容易に取得可能である
- 関係のある単語のペアに対し, 0 から 1 の範囲で重みが付与されている

以下に ConceptNet を用いた重みの付与と描写表現ネットワーク構築について詳細に説明する.

3.5.1 ConceptNet について

ConceptNet とは, 自然言語の単語やフレーズをノードとし, 関係性を表現するエッジで結んだネットワークである. その概略図を Fig. 3.3 に示す. 1999 年, マサチューセッツ工科大学 (MIT) メディアラボが主体となっている人工知能プロジェクトである Open Mind Common Sense (OMCS)[23] が始動したが, この OMCS データベースの情報を有向グラフによって表現したものとして, Liu らが 2004 年に初めて ConceptNet を提案した [24]. その後改良が加えられ, 2017 年に第 5 バージョンである ConceptNet5.5 が発表された [25]. この ConceptNet5.5 のノードやエッジの情報はインターネット上で無料で公開されており, API を用いてアクセスすることができる.

ConceptNet5.5 の知識は, 辞書などの専門家により与えられる知識や, OMCS などのクラウドソーシングプロジェクトによって集められた知識, ゲームを通じて知識を集める “Games with a purpose”[26] などをソースとしている.

ConceptNet5.5 は約 800 万ものノードを持ち, 83 の言語に対応している. このうち英語と日本語を含む 10 言語が中心言語として指定され, 20 万語を超える豊富な知識情報を持つ. ConceptNet5.5 の概要情報を Table 3.3 に示す. また, ConceptNet5.5 は 36 種類のエッジによりコモンセンスを管理している. 全てのエッジには関係に応じてラベリングが行われており, エッジの種類には双方向のものと単方向のものがある. ラベルの種類については Table 3.4 に示す.

ConceptNet は単語の分散表現に対して応用することが可能であり, ConceptNet5.5 と word2vec[27] と GloVe 1.2[28] の知識を統合した ConceptNet Numberbatch 16.09 という

[4]http://conceptnet.io/

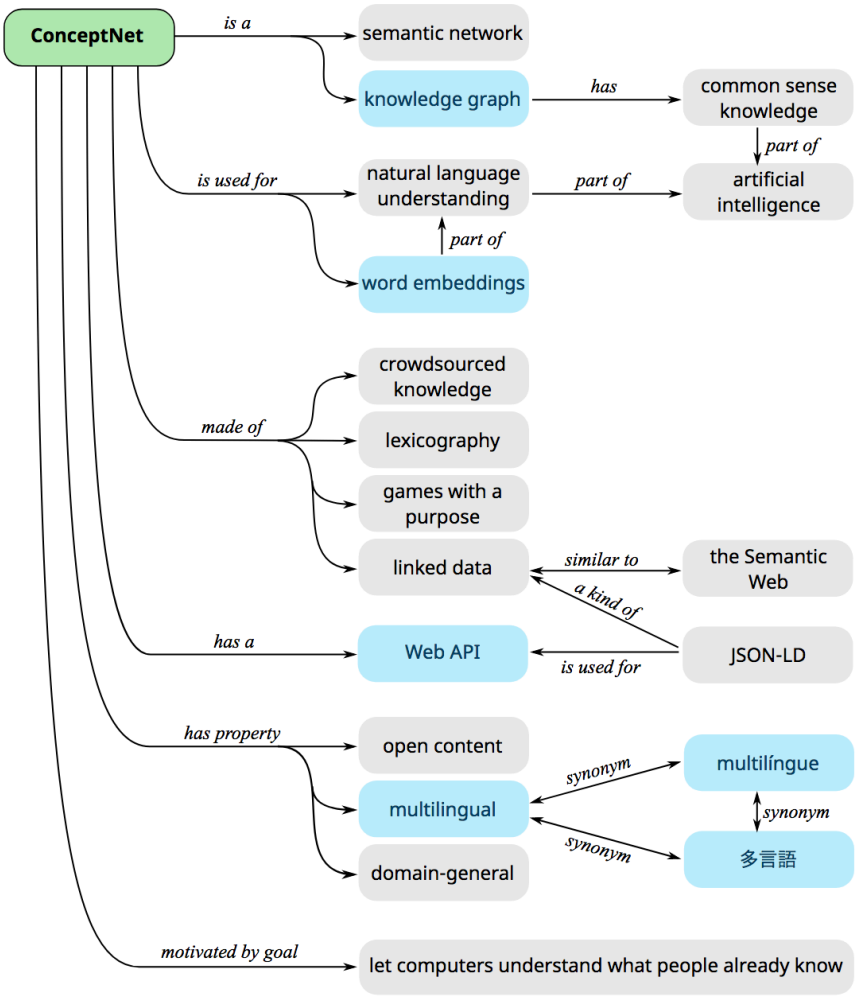


Fig. 3.3: ConceptNet の概略図

Table 3.3: ConceptNet5.5 の概要

総ノード数	約 800 万
総エッジ数	約 2100 万
対応言語数	83
中心言語数	10
中心言語のノード数	20 万以上
英語のノード数	約 150 万
その他の各言語のノード数	1 万以上

Table 3.4: ConceptNet5.5 のエッジの種類

関係	エッジの種類
Symmetric Relations	Antonym, DistinctFrom, EtymologicallyRelatedTo, LocatedNear, RelatedTo, SimilarTo, and Synonym
Asymmetric Relations	AtLocation, CapableOf, Causes, CausesDesire, CreatedBy, DefinedAs, DerivedFrom, Desires, Entails, ExternalURL, FormOf, HasA, HasContext, HasFirstSubevent, HasLastSubevent, HasPrerequisite, HasProperty, InstanceOf, IsA, MadeOf, MannerOf, MotivatedByGoal, ObstructedBy, PartOf, ReceivesAction, SenseOf, SymbolOf, and UsedFor

システムを用いると, 単語の関連度判定などのタスクにおいて, 独立したモデルを用いた場合に比べて精度の向上があることが示されている [25]. この ConceptNet Numberbatch 16.09 を用いた単語の分散表現の情報もまた, インターネット上で無料で公開されており, API によりアクセスすることができる.

3.5.2 形容詞間の重みの付与

本研究では, 取得した英語と日本語の形容詞に対し, 全てのペアを探索し, 関連のあるペア間に重みの付与を行った. 付与する重みは, ConceptNet Numberbatch 16.09 による単語の分散表現の API を用いて取得した. この API では, ひとつの言葉を入力すると, その言葉と関連のある単語の一覧と, それぞれの関連度の重みを取得することができる. 例として, 「自転車」という日本語の単語に対して取得される英語と日本語の関連単語と重みの一覧を Table 3.5 に示す.

この手順を数式を用いて説明する. n 個の描写表現 $d_1, d_2, ..., d_n$ を抽出したとする. このとき, 全ての描写表現による空間 $\mathbb{D}$ は次の式によって表される.

$$\mathbb{D} = \{d_k, \forall k \in \{1, ..., n\}\} \tag{3.1}$$

本研究では描写表現 d は形容詞であり, 空間 $\mathbb{D}$ は英語と日本語の形容詞による空間となる. このとき, 形容詞 $d_i \in \mathbb{D}$ に対し, ConceptNet API を用いて関連する形容詞 $d_j \in \mathbb{D}$ とその重み w_{ij} を取得する. これを全ての $i, j \in \{1, ..., n\}$ のペアに対して行う.

Table 3.5: 「自転車」に対し ConceptNet の API で取得できる関連単語と重み

言語	単語	重み w_{ij}
en	bicycle	1.0
ja	ちゃり	0.979
ja	チャリ	0.951
ja	ちゃりんこ	0.951
en	bike	0.95
en	pushbike	0.887
en	bikes	0.874
en	bicycles	0.862
en	bycicle	0.807
en	bicycling	0.784
en	velocipede	0.777
en	biking	0.759
ja	二輪	0.731
en	mountain bike	0.724

3.5.3 描写表現ネットワークの構築

それぞれの描写表現をノードとみなし, 近い関係のある表現同士を重み付きエッジで結ぶことでネットワークを構築する. このネットワークを描写表現ネットワークと呼ぶことにする.

このネットワークを G とすると, G は次式によって定義される.

$$\begin{cases} G = (V, E) \\ V = \{d_k \mid \forall k \in \{1, ..., n\}\} \\ E = \{e_{ij} \mid i, j \in \{1, ..., n\}\} \end{cases} \tag{3.2}$$

ただし, V はノードの集合であり, n 個の描写表現 d を要素に持つ. また, E はエッジの集合であり, 3.5.2 項において取得した近い関係を持つ描写表現のペアに対してエッジ e_{ij} を付与する. このとき, ConceptNet API により取得した重みをエッジの重み w_{ij} として定義する.

英語と日本語の形容詞を用い, 描写表現ネットワークを描画した例を Fig. 3.4 に示す. 近い意味を持つ形容詞が英語・日本語問わず近い位置に集まっていることがわかる. この図では, クラスタごとに色を分けて表示している. 図ではエッジを有向エッジとして表示しているが, ConceptNet API において, 形容詞 d_i が形容詞 d_j に対して関連性を持つとき, 逆に d_j も d_i に対して関連性を持つ. よって全てのエッジは双方向であるため, 本研究では描写表現ネットワークを無向グラフとみなして扱っている. 図のクラスタの検出方法については 3.6 節で説明する.

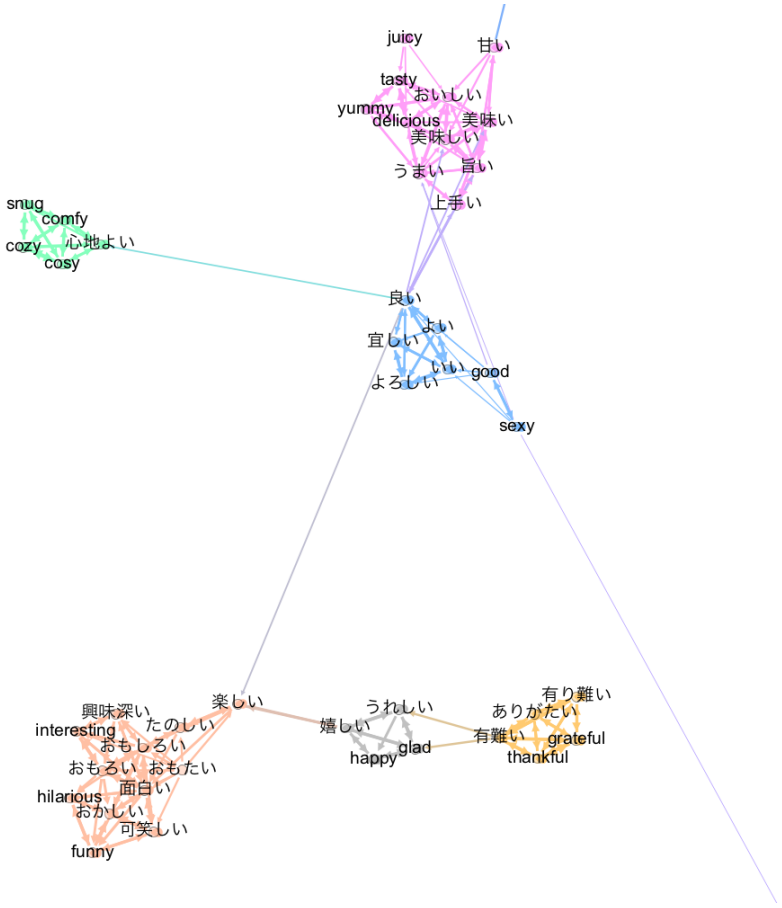


Fig. 3.4: 描写表現ネットワークの例

3.6　コンセプトクラスタの検出

3.5.3 項で求めた描写表現ネットワークに対して, クラスタリングを行い, 全ての表現を複数の部分集合に分割する. 近い関係をもつ表現の間にエッジを張っているため, それぞれの部分集合には, 近い関係をもつ表現が集まっているはずである. また, エッジには異なる言語間を繋ぐエッジが存在するため, 異なる言語の表現で意味が近いもの同士もまた, 同じ部分集合に属することとなる. クラスタリングの結果として得られる部分集合一つ一つを, 言語を超えたコンセプトを表現する「コンセプトクラスタ」と定義し, このコンセプトクラスタの分類を用いて 3.9 節以降の類似度の測定を行う. 以降では, ネットワークのクラスタリングとコンセプトの定義について詳細に説明する.

3.6.1　ネットワーククラスタリング

クラスタリングとは, 分類対象の集合を, 類似する要素を集めた部分集合に分割することであり, 分割後の部分集合をクラスタと呼ぶ. ネットワーク構造を用いたクラスタリングでは, ノードをクラスタリングの対象とし, エッジの密度の高いノードの集まりをクラスタとして検出する. 検出されたクラスタはコミュニティとも呼ばれる.

ネットワーク構造に対してクラスタリングを行うモデルとしては, Girvan と Newman によるモジュラリティ最大化 [29] が頻繁に使用される. モジュラリティとは, ネットワークの与えられた分割に対して, グループ内のノード同士が繋がるリンクの割合からリンクがランダムに配置された場合の期待値を引いた値として定義される.

ノードの数 N, エッジの数 M のネットワークに対し, ノードを $\{g_1, g_2, ..., g_C\}$ の C 個のグループに分ける. ネットワークの隣接行列を A と表し, 行列成分 A_{rs} はノード r, s 間に存在するリンクの数とする. このとき, グループ g_i に属するノードとグループ g_j に属するノードが繋がるリンク数の合計の, 全リンクに占める割合 e_{ij} は以下のように書ける.

$$e_{ij} = \sum_{r \in g_i} \sum_{s \in g_j} \frac{A_{rs}}{2M} \tag{3.3}$$

このとき, モジュラリティQ は次式によって定義される.

$$Q = \sum_i \left(e_{ii} - a_i^2 \right) \tag{3.4}$$

ここで Q はモジュラリティ, e_{ii}は グループ内のノード同士が繋がるリンクの割合, a_i^2 はリンクがランダムに配置された場合の期待値である.

モジュラリティQ は $0 \leq Q \leq 1$ を満たし, 1 に近いほどコミュニティ分割の精度が良いことを示す. 重み付きグラフについては, 式 3.3 において次数 A_{rs} を求める代わりにエッジの重みの合計を用いることでモジュラリティQ を計算する.

モジュラリティを計算するとき, 全てのエッジに対して媒介中心性のスコアを繰り返し付与する必要があり, 計算量が大きくなるのが課題である. 計算量を少なく保ったまま, モジュ

ラリティ最大化を達成するアルゴリズムが提案されており, その一つがLouvainアルゴリズム [30] である.

本研究では, Louvainアルゴリズムを用いてモジュラリティ最大化を行い, 重みを考慮した描写表現ネットワークのクラスタリングを行った.

3.6.2 クラスタのコンセプトとしての定義

描写表現ネットワークに対してクラスタリングを実行すると, 関係の近い描写表現の集合が, それぞれのクラスタとして取得できる. このクラスター一つ一つをコンセプトと定義する. それぞれの描写表現は, 必ず属するコンセプトクラスタを一つだけ持つ. この関係を式を用いて説明する.

クラスタリングにより, 描写表現が K 個のコンセプトに分割されたとき, 全ての描写表現による空間 $\mathbb{D}$ は次式によって表される.

$$\begin{cases} \forall k \in \{1, ..., K\} \\ \mathbb{D} = \{\mathbb{D}_k, \forall k \in \{1, ..., K\}\} \end{cases} \tag{3.5}$$

ただし, $\mathbb{D}_k$ は描写表現 d を要素に持つ, 空間 $\mathbb{D}$ の部分集合である. このとき, コンセプトクラスタ C_k を次式によって定義する.

$$C_k = \{d \in \mathbb{D}_k\} \tag{3.6}$$

このとき, $\mathbb{D}_k$ は,

$$\forall i, j \in \{1, ..., K\} \quad subject\ to\ i \neq j \tag{3.7}$$

において,

$$\mathbb{D}_i \cap \mathbb{D}_j = \emptyset \tag{3.8}$$

を満たす.

本研究では形容詞を描写表現として用いたため, 要素 d はそれぞれの形容詞であり, 空間 $\mathbb{D}$ は形容詞空間である. 例として, Fig. 3.4 に挙げた描写表現ネットワークを用いた場合に得られるコンセプトクラスタを Table 3.6 に示す. 各クラスタは Fig. 3.4 において異なる色で表示されている形容詞の集合であり, 意味の近い形容詞が言語に関係なく同じクラスタに所属していることが分かる.

Table 3.6: 描写表現ネットワークから得られるコンセプトクラスタの例

ID	クラスタ内の描写表現
1	juicy tasty yummy delicious 美味しい 美味い うまい 旨い おいしい 甘い 上手い
2	cozy cosy snug comfy 心地よい
3	good sexy 良い 宜しい よい いい よろしい
4	funny interesting hilarious funny 興味深い たのしい 楽しい たのしい おもしろい おもろい おもたい 面白い おかしい 可笑しい
5	glad happy 嬉しい うれしい
6	grateful thankful ありがたい 有難い 有り難い

3.7　対象と描写表現の共起関係の取得

この手順においては, 入力に用いた感想データから, 各対象ごとに, どのような表現を用いて対象を描写しているか, 各対象と各表現の間の共起関係を取得する. 対象（クエリ）q_i と表現 d_j との間の共起回数を $n(q_i, d_j)$ とした時, 全ての i, j のペアに対して $n(q_i, d_j)$ を求めるのがこの過程である.

本研究においては, 分析対象として日本語・英語の料理, 表現として形容詞を用いているため, 日本語・英語それぞれの料理に対し, Twitter のテキストデータにおいて各形容詞が共起した回数を求めた.

3.8　対象とコンセプトの共起関係の取得

3.7 節にて, 各表現と各対象との共起分布を求めたが, 複数言語の表現をコンセプト単位で扱うため, コンセプトごとにこの共起分布を合計する. さらに, それぞれの対象とコンセプトとの共起回数を, 対象と全表現との共起回数の合計で割ることで, 対象と各コンセプトクラスタとの間の頻度分布を求める. この過程の概念図を Fig. 3.5 に示す. 以下具体的に数式を用いて説明する.

対象 q_i と表現 d_j との間の共起回数を $n(q_i, d_j)$ とした時, 対象 q_i とコンセプト c_k との間の共起回数は次式によって求められる.

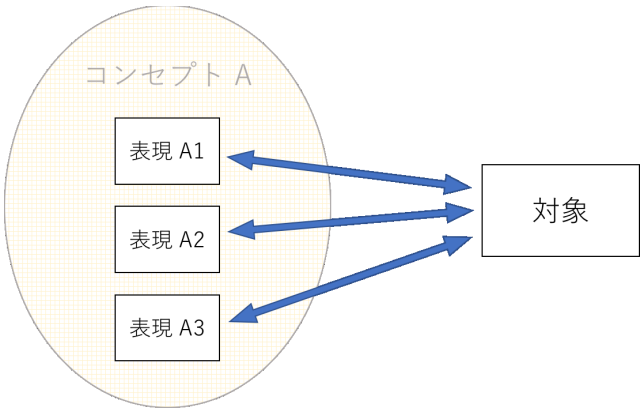
$$N(q_i, c_k) = \sum_{d_j \in C_k} n(q_i, d_j) \tag{3.9}$$

対象 q_i について, コンセプト c_k に対して発生する割合を P_{q_i,c_k} とする. このとき, P_{q_i,c_k} を次式によって定める.

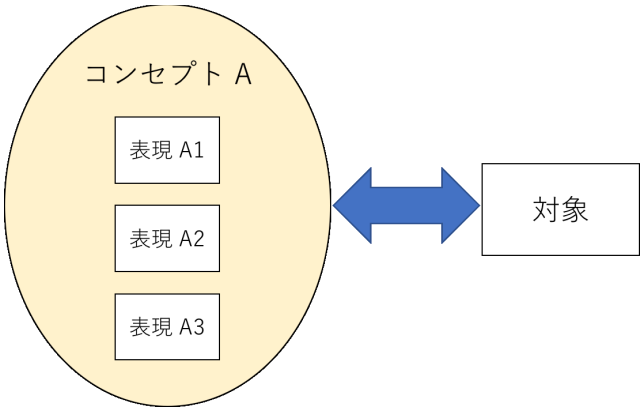
$$P_{q_i,c_k} = \frac{N(q_i, c_k)}{\sum_{k=1}^{k_{max}} N(q_i, c_k)} \tag{3.10}$$

これにより, 対象 q_i について, それぞれのコンセプト c_k の共起頻度を表す頻度分布 P_{q_i} は次式のように求まる. これを対象のコンセプト頻度分布と呼ぶことにする.

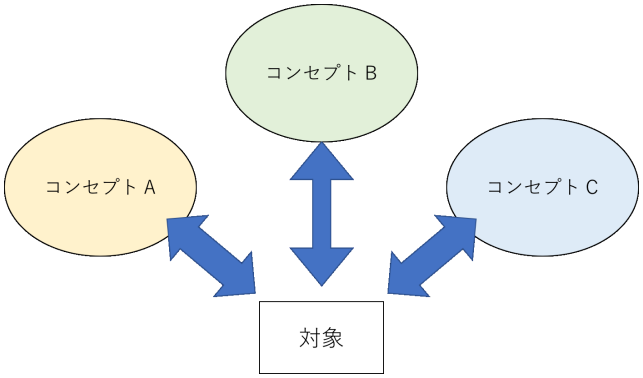
$$\begin{cases} P_{q_i} = \{P_{q_i,c_k}, \forall c_k \in \mathbb{C}\} \\ \sum_{c_k} P_{q_i,c_k} = 1 \end{cases} \tag{3.11}$$



(a) 各表現と対象の共起



(b) コンセプトと対象の共起



(c) 頻度分布の算出

Fig. 3.5: 対象とコンセプトクラスタ間の頻度分布の算出 (a) 各表現と対象の共起. (b) コンセプトと対象の共起. (c) 頻度分布の算出

3.9 コンセプト頻度分布による対象間の比較

この手順においては, 求めたコンセプト頻度分布をもとに, 対象間の類似度を測定する. 具体的な手順について以下に示す.

3.9.1 JS 距離の測定

確率分布間の近さを示す指標には, カルバック・ライブラー情報量（KL ダイバージェンス, Kullback-Leibler divergence）, JS 距離（Jensen-Shannon divergence）, コサイン類似度などがある. 本研究では, JS 距離を用いて各対象のコンセプト頻度分布を比較した.

P, Q を離散確率分布とするとき, P の Q に対するカルバック・ライブラー情報量は以下のように定義される.

$$D_{KL}(P \parallel Q) = \sum_i P(i) \log_2 \frac{P(i)}{Q(i)} \tag{3.12}$$

カルバック・ライブラー情報量は, 等しい確率分布に対して 0 となり, 分布の差異が大きくなるにつれて大きな値を取る. JS 距離は, カルバック・ライブラー情報量に対称性を持たせたものであり, 次の式で定義される.

$$D_{JS}(P \parallel Q) = \frac{1}{2} D_{KL}(P \parallel M) + \frac{1}{2} D_{KL}(Q \parallel M) \tag{3.13}$$

ただし, このとき,

$$M = \frac{1}{2}(P + Q) \tag{3.14}$$

である. JS 距離は

$$0 \leq D_{JS} \leq 1 \tag{3.15}$$

を満たし, 値が小さいほど確率分布の距離が近いことを示す. JS 距離は P と Q に関する対称性を持ち, また, 必ず有限の値を取るため, 確率分布の比較において扱いやすく, 自然言語処理の分野では文書間の距離を測る目的で多く用いられている. また, Sajadmanesh らもまた, レシピデータの材料や味の情報を用いて文化間距離を比較する際に, JS 距離を用いて文化間距離を定めている [2].

本研究では, 対象 q_a と対象 q_b の間の距離を, コンセプト頻度分布の JS 距離 $D_{JS}(P_{q_a} \parallel P_{q_b})$ により測定した. この JS 距離の値により, 描写表現を用いて対象間の関係やその近さを評価する.

3.9.2 コンセプト頻度分布の差による感じ方の比較

比較したい対象について, 異なる言語圏のコンセプト頻度分布がどのように異なるかを具体的に比較することにより, 対象への感じ方の文化的差異を取得する.

例として, 英語圏における犬への感じ方と, 日本語圏における犬への感じ方を比較したいとする. 上記の方法により, 英日の感想データから, 「dog」に対する描写表現と, 「犬」に対する描写表現を取得し, コンセプト頻度分布を得る. この時, 「dog」のコンセプト頻度分布と「犬」のコンセプト頻度分布において, 共起確率の差が大きいコンセプトクラスタを比較することによって, 感じ方の文化的差異を分析することができる. 例えば, 「格好いい」や「handsome」の属するコンセプトクラスタについて, 「犬」に比べて「dog」の方がより共起頻度が高いとすれば, 英語圏の方が犬に対してより頻繁に「handsome」という概念を用いていると分かる. ここから, 英語圏と日本語圏の人々の犬に対する印象や評価基準の違いを推測することができる.

本研究においては, それぞれの描写表現を肯定的に用いているか否定的に用いているかは区別していない. 上の例を用いると, 「handsome」の所属するクラスタの「dog」に対する共起頻度は, 「handsome」という表現を「not handsome」というように否定的に用いる場合を含んでいる. これは, 特定の対象が好きか嫌いか, という世論調査でではなく, 対象に対する評価基準や印象, ステレオタイプを比較したいためである.

犬に対し, 「not handsome」という感想を述べるということは, 犬を「handsome」かどうかで評価していることを意味し, 英語圏で「handsome dog」という考え方や感じ方が一般的だということを表している. 一方で, 温度で犬を評価する捉え方が一般的でないとすれば, 「warm」や「cold」というコンセプトは, 肯定的に用いるか否定的に用いるかに関わらず, 出現頻度が低くなると予想される.

提案手法では, 対象に対する好みの世論調査ではなく, 対象をどのようなコンセプトを用いて描写することが一般的であるのか, その評価基準や印象の違いを異なる言語圏に対して捉えるというものである. コンセプト頻度分布の差が大きいコンセプトクラスタを取り上げることで, 文化的差異を評価することが可能であると考える.

第4章　実験

本章では提案手法による各実験の設定と, 実験結果を用いた提案手法の評価方法について述べる.

4.1　対象データの決定と前処理

4.1.1　データ収集対象となるメディアの決定

本研究では, 食の描写表現の違いを用いて, 異なる言語圏について料理の感じ方の違いを分析する. 分析に際し, データの収集には Twitter を用いた. Twitter のデータを採用したのには以下の理由がある.

- プライベートな話題を日常的に発信する手段として, 非常に普及している
- テキストでの投稿が多い
- 匿名性があり, 利用者の率直な感想が表出する
- 言語を特定して API からデータを収集することが可能である

上記のような Twitter の特徴を用い, 異なる言語圏における料理への感じ方について分析を行う.

4.1.2　データ収集のクエリ設定

本実験においては, アメリカと日本を代表するレシピサイトにおけるカテゴリ分類を元に, クエリを設定した. 英語のクエリは, allrecipes[1]の「Drinks Recipes」カテゴリ内の分類を参照にし, 日本語は楽天レシピ[2]の「飲みもの」カテゴリの分類を参照にした. 具体的なクエリは Table 4.1 のようになる. 英語と日本語間で, クエリに用いている単語の種類と粒度が異なるが, ワインに詳しい地域が多くのワインの分類法を持つように, カテゴリ分類の方法は地域の文化に大きく依存し変化することが当然である. 本実験では, 飲み物に対する文化的差異を正しく評価するため, 統一でないクエリを用いた.

[1] http://allrecipes.com/
[2] https://recipe.rakuten.co.jp/

Table 4.1: 実験に用いた英語・日本語のクエリ

英語	日本語
beer cocktail, cider, cocktail, coffee, eggnog, hot chocolate, juice, lemonade, liqueur, mocktail, mulled wine, punch, sangria, shot, shake, float, slushy, smoothie, tea	コーヒー, ココア, 紅茶, 抹茶, 豆乳, ヨーグルト, はちみつドリンク, チョコレートドリンク, しょうがドリンク, お酢ドリンク, サワードリンク, ソーダドリンク, ミックスジュース, シェイク, セーキ, スムージー, グリーンスムージー, チャイ, ビール, 焼酎, 梅酒, 甘酒, カクテル, モヒート, ジントニック, 卵酒, 健康酒

Table 4.2: 収集したツイートの個別情報

収集データ	データ内容	データの例
id	ツイートの id	9**********
created at	作成日時	2017-12-25 00:22:19
id str	ツイートの id(str 型)	8**********
username	投稿者のユーザーネーム	K**********
text	本文内容	Just tried to follow a cocktail recipe and it tastes absolutely bogging drinking it anyway. Cheers
source	ツイートの url	Twitter for iPhone

4.1.3 データの収集

Twitter の API により収集したデータのフォーマットは Table 4.2 の通りである.

本実験で取得したデータの期間, データ数は表 4.3 の通りである.

Table 4.3: 実験のデータ収集期間とデータ数

	英語	日本語
期間 (年月日)	2017/11/14〜2017/12/4	2017/11/5〜2017/12/7
データ数	1,040,854	34,337
データ数 (前処理をした後)	256,264	12,327

4.1.4　データの前処理

ユーザの限定

収集したデータの source（ツイートの url）情報に, ユーザがどのアプリを用いて Twitter への投稿を行ったかという情報がある. Twitter では, bot と呼ばれる定期的に同じ内容を自動投稿するようなアカウントが存在する. 今回, 個人が自分の感想や意見を投稿したツイートに限定した分析のため, bot による投稿を排除する. また, スマートフォンを用いた投稿を行っている場面のみを分析の対象とし, クライアントを Twitter for iPhone 及び Twitter for Android に限定した.

リツイートの除外

提案手法では, 各個人が自分の感想や意見を投稿するツイートを入力としている. リツイートと呼ばれる, 他人の投稿を再共有する行動はこの対象に含まれない. これらのツイートは text（本文内容）情報の冒頭に RT の文字を伴って取得され, 区別できるため, 形態素解析の過程で RT の文字を持つツイートを除外した.

4.2　分析対象データの制限

4.2.1　一部クエリの除外

beer cocktail, liqueur, mocktail の三つのクエリについては, データが得られなかったため, 分析対象外とした. また, hot chocolate はクエリ自体に hot という形容詞が含まれてしまい, 描写表現分布を正しく評価することが難しいと判断したため, こちらも分析対象外とした.

4.2.2　形容詞群の制限

描写表現ネットワークの構築とそのクラスタリングにかかる計算時間を短縮するため, 描写表現ネットワークの構築に用いる形容詞の種類を制限した. 本実験では, 日本語は取得した全ての形容詞, 英語は全ツイートの合計で 50 回以上出現した形容詞を用いた. 英語と日本語で 50 倍という制限条件の差を設けたことには以下の理由がある.

- 英語と日本語の取得データ量の差（英語が日本語の約 25 倍）
- 用いた形態素解析ツールの違いによる品詞体系の差
- 英語と日本語の言語構造の差

Table 4.4: 形容詞群の制限

	英語	日本語
制限前の形容詞の種類数	55,830	476
制限の条件	50 回以上出現	制限なし
制限後の形容詞の種類数	367	476

形態素解析ツールについて, 英語の形態素解析に用いた TreeTagger では, 絵文字や, 一部の修飾語として用いられた名詞[3]が形容詞として取得される. 日本語の形態素解析に用いた MeCab では, これらは形容詞とは判断されない. また, 英語と日本語の言語構造の差について, 一般に文章中に出現する形容詞率が日本語は英語などの他言語に対し相対的に低いことが知られている [31].

全てのデータから取得した形容詞数と, 制限の結果として, 描写表現ネットワーク構築に用いた形容詞数について Table 4.4 に示す.

4.2.3 分析に用いるコンセプトクラスタの選択

描写表現ネットワークに対してクラスタリングを行い, 得られたコンセプトクラスタのうち, 以下の基準により, コンセプト頻度分布の計算に用いるコンセプトクラスタを制限した.

- 英語, 日本語双方の単語がクラスタに含まれている
- 分析に十分な共起回数が得られている
- 抽象的な形容詞でない

三点目の抽象的な形容詞については, 日本語での「良い」や「ない」, 英語での「good」や「all」, 「any」などの形容詞がこれに該当する. これらの単語はあらゆる場面で用いられ, 特に日本語では補助形容詞として名詞に付随する形で用いられるものもある. 共起回数が非常に多い一方で, 描写表現としての役割は強くないと判断し, 本研究においては分析から除外することとした.

以上の制限の結果, 30 のコンセプトクラスタが得られた.

[3]‘strawberry cake’ という表現における ‘strawberry’ など

4.3 提案手法の評価

本研究では, 感想データから取得した描写表現を用いて, 分析対象に対する言語圏ごとの感じ方の差異を比較する手法を提案した. この提案手法の妥当性を評価するため, 以下の三つの点から評価を行った.

全体傾向の言語間比較
: 描写表現から文化的差異を取得することが実際に可能であるか, また, 描写表現として形容詞を使うことが妥当であるかを検証するため, 取得した全ての英語ツイートと日本語ツイートを用い, 英語言語圏の傾向と日本語言語圏の傾向の差異を正しく抽出できているか評価する.

ネットワーク図による関係性の可視化
: 本研究では, 分析対象のコンセプト頻度分布間の JS 距離を求めることで, 対象間の距離とし, 対象間の関係を捉えることを提案した. この手法の妥当性を確認するため, JS 距離で対象間の距離を定めたネットワーク図を作成し, 実際に対象同士の関係を捉えることが可能であるかを検証する.

同一対象への感じ方の言語間比較
: 本研究では, 対象のコンセプト頻度分布の差を具体的に分析することで, 異なる言語圏における感じ方の差異を得ることを提案した. この手法の妥当性を検証するため, 分析に用いたクエリのうち, 日本語と英語で同一対象を扱っているものについてコンセプト頻度分布の具体的な比較を行い, 言語圏における感じ方の差異が抽出できていることを確認する.

4.3.1 全体傾向の言語間比較

この手順においては, 感想データの描写表現の差から実際に文化的差異を取得することが可能であるのか, また, 描写表現として形容詞を使うことが妥当であるかの二点を検証するため, 取得した全ての英語ツイートと日本語ツイートから, 英語言語圏の傾向と日本語言語圏の全体傾向を抽出できているかを評価する. 具体的には, 以下の三つの分析を行い, 得られた知見を既存の研究と比較することにより評価する.

- 極性分析
- 形容詞の種類による分析
- コンセプト頻度分布による分析

特に, Twitter において英語を使用するユーザのほとんどはアメリカに由来する[4]ため, アメリカと日本のソーシャルメディアの用法の差異や, 料理全体に対する文化的差異を中心に議論を行った.

[4] http://growthhackjapan.com/2013-11-21-half-of-twitter-mau-now-live-in-5-countries/

(1) 極性分析

取得した全ての形容詞[5]に対し, 単語感情極性対応表 [32] を用いて極性分析を行った. 単語感情極性対応表は語彙ネットワークを用いて自動的に計算された感情極性値を各単語について振り分けた辞書であり, 辞書の単語群の出典は異なるものの, 同じモデルを用いて英語と日本語の双方に対して極性値の付与を行っている. よって, 英語と日本語の極性値を比較する目的に対して適していると考えた.

極性値の計算に関しては, ひとつのツイートに出現する形容詞のうち, 極性辞書に存在するものについて, 極性値を合計した値をそのツイートの極性値とした. 英語と日本語それぞれ, 全てのツイートに対して, 極性値の平均値を計算し, 比較に用いた.

(2) 形容詞の種類による分析

出現頻度の高い形容詞に対し, 形容詞の種類による分類を行い, 分類ごとの共起頻度を元に形容詞を用いた文化的差異の抽出に関する評価を行った.

現代日本語の形容詞は, 一般に, 客観的な性質や状態を述べる属性形容詞と, 主観的な感覚や感情を述べる感性形容詞に大きく二分される. さらに具体的な分類法も多く提案されており, 例えば細川は形容詞の意味による分類として, 二種類の感覚形容詞と, 感情形容詞, 属性形容詞, 評価性形容詞の分類を提案した [33]. また, 形容詞に限らず, 食に対する表現を分類する研究も行われており, 瀬戸による味ことばの分類 [34] や, Fig. 4.1 に示す松尾によるおいしさ表現の分類 [35] がある.

本研究においては, 食を描写する形容詞について包括的かつ具体的な議論を行うため, 松尾の分類を基盤とし, 一部をさらに細分化した分類を用いた. この分類と, 各分類に属する代表的な形容詞を Table 4.5 に示す. 松尾の分類では, 「性質表現」の項目として食の状況や食品の情報を全て並列に扱っていたが, この項目を, 食品自体の形容か食事環境の形容かによって「性質」と「状況」に二分した. さらに, 英語ではその品詞体系から, 日本語よりも多くの表現が形容詞として取得される. これを区別して扱うために, 性質と状況をそれぞれ以下のように細分化し, 英語のみで出現する表現を特定することとした.

- 性質表現の細分化

  一般 英語と日本語の双方で出現する一般的な属性形容詞（大小, 強弱, 高低など）

  ジャンル 食べ物の分類を表す英語形容詞（healthy, alcoholic など）

  状態 食べ物の状態に関する英語形容詞（icy, frozen, raw など）

  製法 食べ物の由来や製作者を示す英語形容詞（homemade, local など）

  国 国名を表す英語形容詞（Hawaiian, German など）

  食材 食材を表す英語形容詞（mint, peach など）

- 状況表現の細分化

[5] ここでは, 4.2.2 項における形容詞の制限は行っていない

環境 店の環境や天気など食事環境に関する日本語・英語形容詞

自分 自分の状況に関する英語形容詞（drunk, hungry など）

また, 種類の判定は, 手動で行ったため, 共起回数の高いもののみを対象とした. この条件と種類数を Table 4.6 に示す.

(3) コンセプト頻度分布による分析

英語の全てのツイートと, 日本語の全てのツイートを対象とし, コンセプト頻度分布の差を分析する. 差が大きいコンセプトクラスタの内容を見ることにより, 英語言語圏と日本語言語圏の飲食全体に対する文化的差異を抽出できるかを検証した.

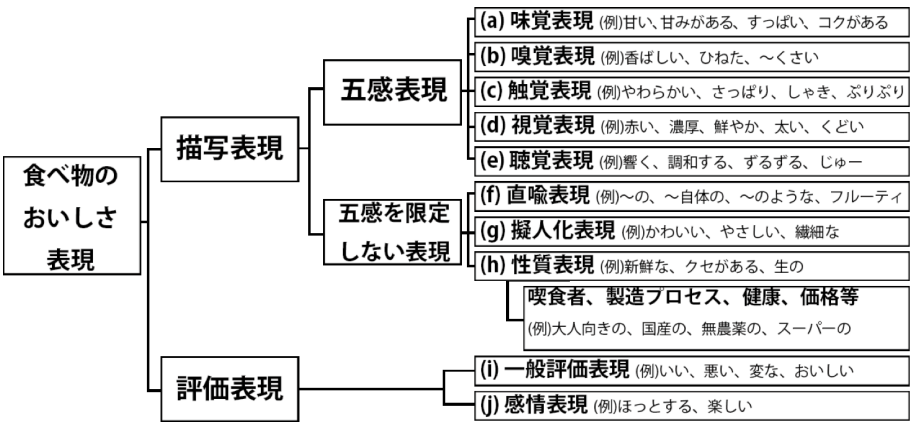


Fig. 4.1: 松尾による言葉の基本義に基づくおいしさ表現の分類 [35]

Table 4.5: 形容詞の種類による分類と対応する英語・日本語の例

			英語	日本語
五感表現	味覚		sour, sweet, salted	甘い, 旨い, 苦い, 酸っぱい, 濃い
	嗅覚		-	臭い
	触覚		cold, warm, hot, hard, soft, smooth	硬い, 熱い, 温かい, 冷たい, 痛い
	視覚		black, pink, green, red, blue	青い, 赤い
	聴覚		-	-
五感を限定しない表現	直喩		like	っぽい, ぽい,
	擬人化		lovely, beautiful, ugly, stupid	大人っぽい, 可愛い, 優しい
	性質	一般	large, wide, mini, high, long, fresh, old, cheap, free	新しい, 珍しい, 強い, 多い, 大きい, 重い, 深い
		ジャンル	healthy, organic, tropical, alcoholic	-
		状態	dry, frozen, tidal, fuzzy, raw	-
		製法	homemade, original, local	-
		国	Hawaiian, German	-
		食材	mint, peach, ginger	-
	状況	環境	snowy, dark, outside, early, cozy	明るい, 狭い, 暖かい, 忙しい, 近い
		自分	sleepy, sick, hungry, drunk	-
評価表現	一般評価		fine, disgusting, perfect, yummy	良い, 美味しい, やばい, おかしい
	感情		happy, mad, angry, sad	辛い, 楽しい, 嬉しい, さみしい, 怖い
その他			second, other, dead	くい, うい, がたい, 無い

Table 4.6: 形容詞の種類による分類における形容詞の制限

	英語	日本語
出現回数の制限	5 回以上	100 回以上
分類に用いた形容詞の種類数	94	191

Table 4.7: ネットワーク図におけるエッジ付与の域値とエッジ数

	英語	日本語
JS 距離の域値	0.2 未満	0.1 未満
エッジ数	43	48

4.3.2　ネットワーク図による関係性の可視化

描写表現の類似性により, 日本語の対象（クエリ）リスト, 英語の対象リストそれぞれについて, JS 距離を用いてネットワーク図を描画し, 対象間の関係を正しく評価できているかどうか検証した.

ネットワーク図の描画は次の方法により行った. まずクエリ集合に対し, 各クエリをノードとみなし, ある域値以下の JS 距離をもつクエリ同士をエッジで結んだ. 用いた域値とエッジ数について Table 4.7 に示す. ノード q_i と q_j を結ぶエッジ e_{ij} の重み w_{ij} には次の値を用いた.

$$w_{ij} = 1 - D_{JS}(P_{q_i} \parallel P_{q_j}) \tag{4.1}$$

ここで, P_{q_i} と P_{q_j} はそれぞれクエリ q_i と q_j のコンセプト頻度分布である. 構築したネットワークに対し, モジュラリティ最大化を用いたクラスタリングを行い, ネットワーク図と各クラスタの内容を用いて, 対象間の関係に対する分析を行った. モジュラリティ最大化によるクラスタリングについては, 3.6.1 項で説明したアルゴリズムと同じものを用いている.

4.3.3　同一対象への感じ方の言語間分析

クエリのうち, 日本語と英語で同一対象を選択しているものがある. このような対象について, 英語と日本語のコンセプト頻度分布の距離と, 頻度分布において差が大きいコンセプトクラスタの内容を分析した. これにより, 日本とアメリカの文化的差異を調査し, 既存の文献と比較することにより, 手法の正当性や有効性の評価を行った. 同一対象とみなした 4 つのクエリのペアを次に示す.

- coffee とコーヒー
- tea と紅茶
- smoothie とスムージー
- juice とミックスジュース

コンセプト頻度分布に対し, 差が大きいコンセプトクラスタを比較する際は, 4.3.1(3) 項における英語と日本語の全体傾向の差異を参考とし, 全体傾向における頻度分布の差よりもさらに大きい差を持つコンセプトクラスタを取り上げて議論を行った. これは, あるコンセプトクラスタで, ソーシャルメディアの用法による違いなどの全体傾向を排除してなお差があるということは, その対象への認識自体に差がある可能性が高いと考えたためである.

第5章　結果

5.1　コンセプトクラスタ

取得したコンセプトクラスタのうち, 分析に用いたコンセプトクラスタの一覧と, それぞれのクラスタに属する代表的な形容詞を Table 5.1 に示す. 以降のグラフではこの表における番号を用いてコンセプトクラスタを指定する.

Table 5.1: 取得したコンセプトクラスタの一覧

番号	クラスタに属する形容詞
1	fast, snap, swift, quick, 素早い, はやい, 早い,
2	salted, salty, しょっぱい, 塩辛い
3	difficult, strong, serious, bitter, sour, 硬い, 強い, 重い, 苦い, 塩っぱい, 酸っぱい
4	weak, soft, gentle, cheap, simple, easy, light, 柔らかい, 軽い, 脆い, 安い, 優しい
5	little, poor, mini, slight, wide, 幅広い, 細かい, 薄い, 狭い, 貧しい
6	nervous, jittery, excited, warm, hot, 熱い, 手厚い, 暑い, 懐っこい, 暖かい
7	interesting, hilarious, curious, exciting, 面白い, 微笑ましい, 興味深い, 楽しい
8	asleep, sleepy, ねむい, 眠い, 眠たい
9	fantastic, dangerous, awful, scary, やばい, 危ない, 素晴らしい, 凄い, 恐れ多い
10	bad, evil, tired, ugly, awkward, 悪い, 不味い, 面倒臭い, かっこいい, 鬱陶しい
11	blunt, boring, sharp, late, obtuse, くだらない, 遅い, つまらない, 緩い, 鈍い
12	high, expensive, long, tall, short, 短い, 細長い, 低い, 尊い, 浅い, 貴い, 長い
13	white, young, blue, yellow, black, dark, green, 黒い, 白い, 幼い, 赤い, 暗い, 青い, 大人っぽい, 黄色い
14	near, close, 近い
15	pretty, very, adorable, cute, beautiful, dear, lovely, nice, handsome, gorgeous, 恋しい, 可愛い, 懐かしい, 美しい, 愛しい
16	spicy, bloody, 臭い, 生臭い, えぐい
17	refreshing, fresh, cool, icy, frosted, chilly, new, 寒い, 新しい, 涼しい, 冷たい,
18	thick, giant, massive, fat, great, deep, huge, 太い, 濃い, 油っこい, 分厚い, 深い
19	like, ぽい, っぽい
20	old, ふるい, 古い
21	proper, decent, 程よい, 程良い
22	clean, 清々しい
23	brown, 茶色い
24	mad, crazy, insane, 狂おしい
25	rare, めずらしい, 珍しい
26	much, few, many, multiple, several, 数少ない, 多い
27	snug, relaxing, comfy, soothing, friendly, cozy, 心地よい, 気持ちいい, 快い
28	busy, 忙しい
29	sorry, 申し訳ない
30	flat, 平べったい

5.2 全体傾向の言語間分析

5.2.1 極性分析

極性分析の結果は Table 5.2 のようになった. 英語の極性平均は 0.266, 日本語の極性平均は 0.0137 という値を得た. 英語の方が日本語に比べ 10 倍以上極性値が高い傾向にあるとわかる.

Table 5.2: 全ツイートの極性平均

言語	平均極性値
英語	0.266
日本語	0.0137

5.2.2 形容詞の種類による分析

形容詞の種類ごとの出現割合の結果は, Table 5.3, これをグラフにしたものは Fig. 5.1 のようになる. ここから, 五感表現については, 日本語は味覚の言及が多く, 英語は触覚や視覚の言及が相対的に多い. また, 五感を限定しない描写表現については, 英語は日本語に比べて性質についての言及がはるかに多いことがわかる. このうちの一部は, 英語の形容詞の幅が広いことによるものだが, 日本語と共通する「性質一般」のみを比較しても, 英語の方が性質について言及する割合が高いことがわかる. 最後に, 評価表現については, 一般評価においては日本語が英語の 6 倍以上, 感情については, 日本語が英語の 3 倍以上と, どちらも日本のユーザの方が非常に高い頻度で用いていることがわかる.

Table 5.3: 形容詞の種類ごとの出現割合（%）

		英語	日本語
五感表現	味覚	1.25	6.16
	嗅覚	0	0.21
	触覚	17.75	2.95
	視覚	4.95	0.37
	聴覚	0	0
五感を限定しない表現	直喩	0.34	1.21
	擬人化	2.23	4.87
	性質	**23.83**	10.39
	性質 1　一般	13.75	10.39
	性質 2　ジャンル	1.57	0
	性質 3　状態	2.86	0
	性質 4　製法	2.15	0
	性質 5　国	0.42	0
	性質 6　食材	3.08	0
	状況	5.39	5.79
	状況 1　環境	3.14	5.79
	状況 2　自分	2.25	0
評価表現	一般評価	19.44	**45.26**
	感情	3.02	**9.52**
その他		21.8	13.28

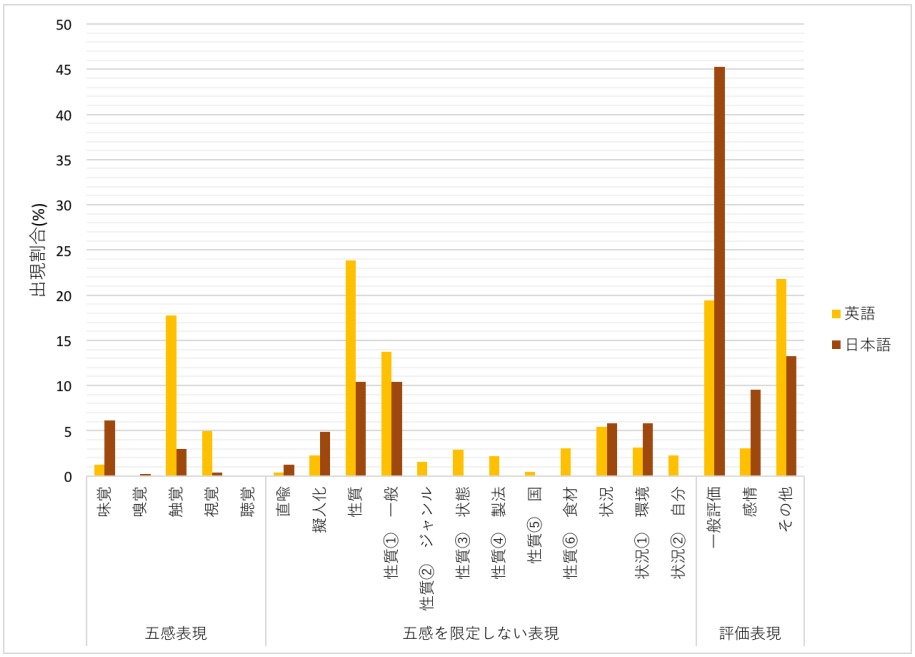


Fig. 5.1: 形容詞の種類ごとの出現割合

5.2.3　コンセプト頻度分布による分析

日本語の全てのツイートと, 英語の全てのツイートを対象とし, それぞれコンセプト頻度分布を求め, その差を調べた. 結果を Fig. 5.2 に示す. このグラフは (英語での共起頻度) – (日本語での共起頻度) を表しており, 値が正のコンセプトクラスタは英語で多く共起し, 値が負のコンセプトクラスタは日本語で多く共起したことを示す.

グラフより, 次のコンセプトクラスタでは, 英語の共起頻度が日本語に比べて高くなっている.

- **クラスタ 13**　white, young, blue, 黒い, 白い, 幼い
- **クラスタ 17**　refreshing, cool, icy, 寒い, 新しい, 涼しい
- **クラスタ 18**　thick, giant, massive, 太い, 濃い, 油っこい
- **クラスタ 26**　much, few, many, 数少ない, 多い

一方で, 次のコンセプトクラスタでは日本語の共起頻度が英語に比べて高い.

- **クラスタ 3**　difficult, strong, serious, 硬い, 強い, 苦い
- **クラスタ 4**　weak, soft, gentle, 柔らかい, 軽い, 安い
- **クラスタ 7**　interesting, hilarious, curious, 面白い, 興味深い, 楽しい
- **クラスタ 9**　fantastic, dangerous, awful, やばい, 素晴らしい, 凄い

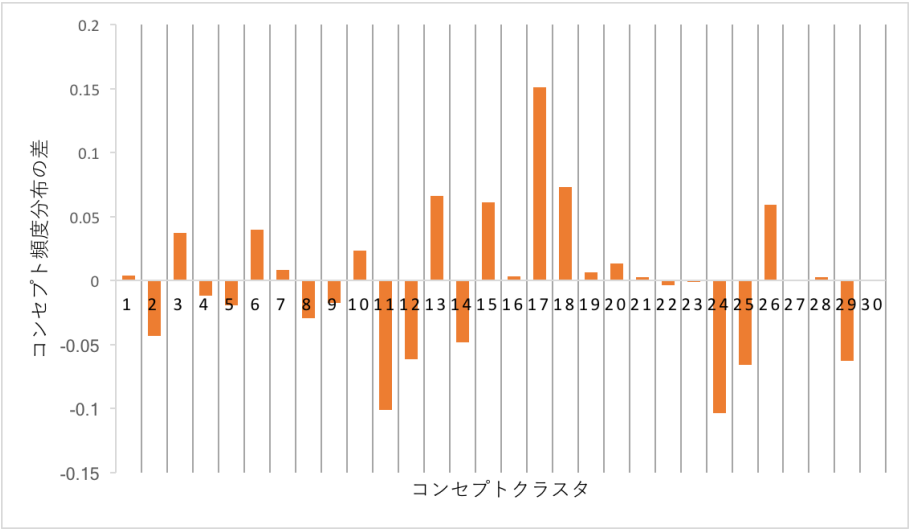


Fig. 5.2: 英語と日本語のコンセプト頻度分布の差

5.3　ネットワーク図による関係性の可視化

クラスタリングした英語のネットワーク図は Fig. 5.3 のようになる. 日本語のネットワーク図は Fig. 5.4 のようになる. 二つ以上のクエリが属するクラスタは, 英語においては次の 3 つが確認された.

- punch, juice, lemonade, shake, shot
- cider, tea, smoothie, coffee, eggnog
- float, cocktail, mulled wine

一方で, 日本語においては次の 3 つが確認された.

- ビール, チャイ, ミックスジュース, シェイク, 抹茶, カクテル
- モヒート, 焼酎, 梅酒, ヨーグルト, コーヒー, 健康酒
- 甘酒, ココア, はちみつドリンク, 紅茶, 豆乳, スムージー

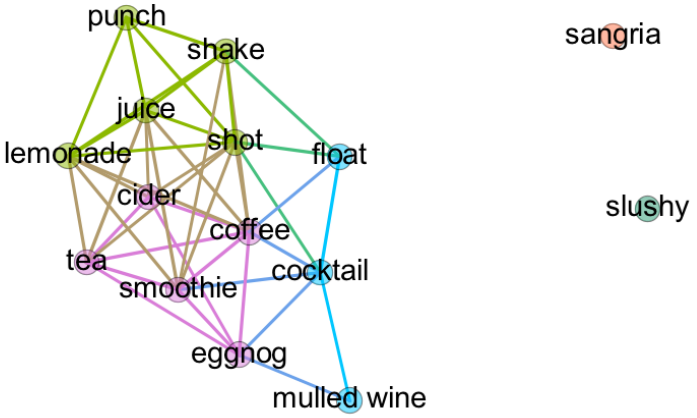


Fig. 5.3: 英語ネットワーク図

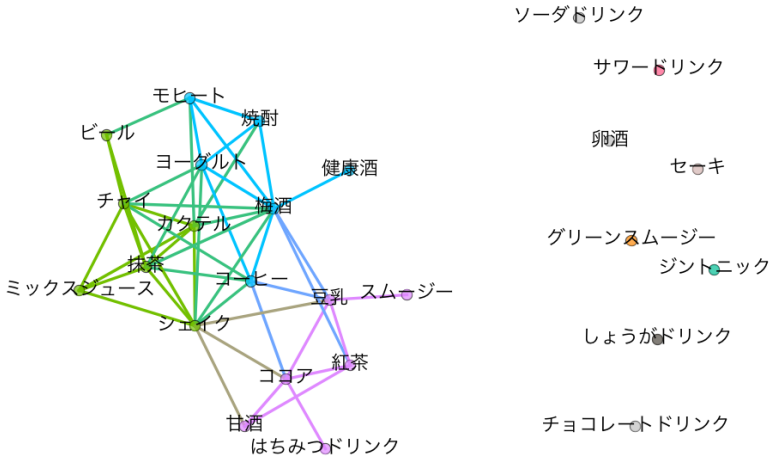


Fig. 5.4: 日本語ネットワーク図

5.4　同一対象への感じ方の言語間分析

次の 4 つのペアを同一対象とみなし, コンセプト頻度分布の具体的な差を分析した.

- coffee とコーヒー
- tea と紅茶
- smoothie とスムージー
- juice とミックスジュース

これらの JS 距離は Table 5.4 のようになった. JS 距離は値が小さいほど分布の距離が近いことを示すため, 表の順に距離が近いという結果が得られた.

Table 5.4: 同一対象のペアの JS 距離

ペア	JS 距離
coffee とコーヒー	0.206
smoothie とスムージー	0.329
tea と紅茶	0.334
juice とミックスジュース	0.415

5.4.1 coffee / コーヒー

coffee とコーヒーの頻度分布の差を Fig. 5.5 に示す．グラフより, 次のコンセプトクラスタでは, 英語の共起頻度が日本語に比べて高くなっている．

- **クラスタ 6**　warm, excited, nervous, 熱い, 手厚い, 暖かい
- **クラスタ 11**　blunt, boring, sharp, くだらない, 遅い, つまらない
- **クラスタ 26**　much, few, many, 数少ない, 多い

一方で, 次のコンセプトクラスタでは日本語の共起頻度が英語に比べて高い．

- **クラスタ 3**　difficult, strong, serious, 硬い, 強い, 苦い
- **クラスタ 15**　pretty, very, adorable, 恋しい, 可愛い, 懐かしい

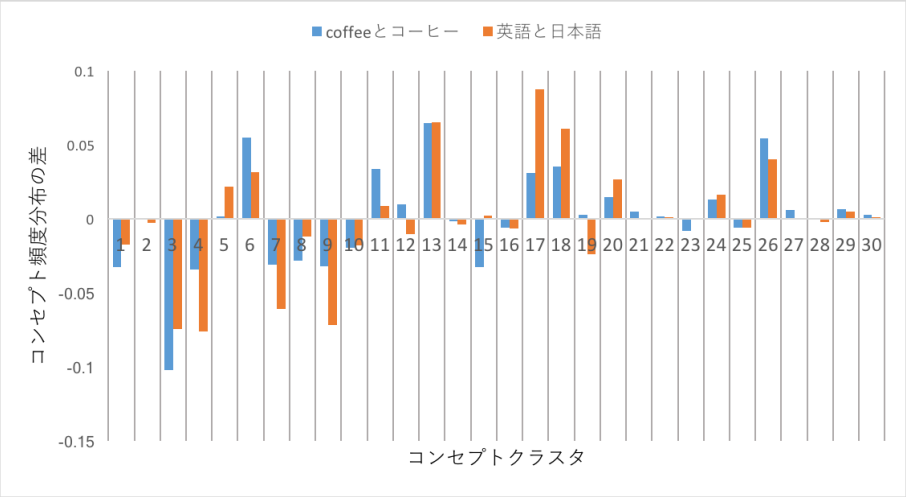


Fig. 5.5: coffee とコーヒーの頻度分布の差

5.4.2 tea / 紅茶

tea と紅茶の頻度分布の差を, Fig. 5.6 に示す. グラフより, 次のコンセプトクラスタでは, 英語の共起頻度が日本語に比べて高くなっている.

- **クラスタ 13** white, young, blue, 黒い, 白い, 幼い
- **クラスタ 15** pretty, very, adorable, 恋しい, 可愛い, 懐かしい

一方で, 次のコンセプトクラスタでは日本語の共起頻度が英語に比べて高い.

- **クラスタ 9** fantastic, dangerous, awful, やばい, 素晴らしい, 凄い

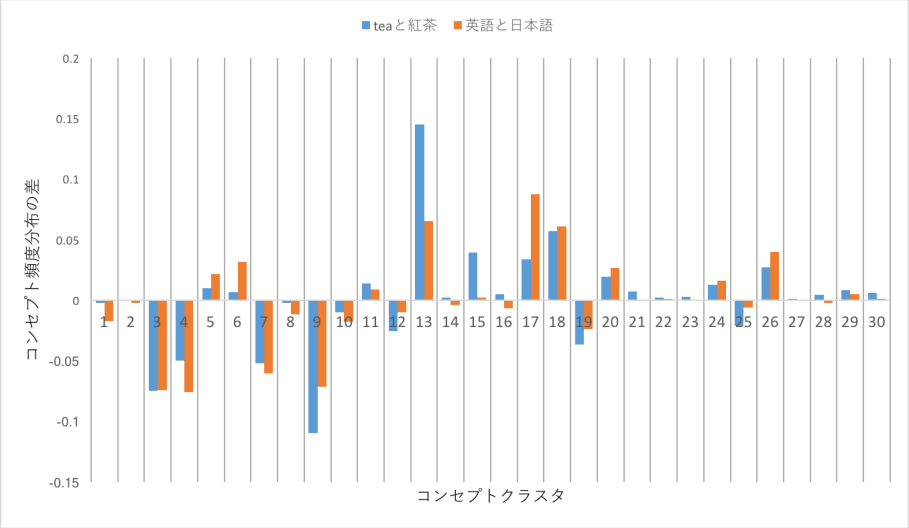


Fig. 5.6: tea と紅茶の頻度分布の差

5.4.3 smoothie / スムージー

smoothie とスムージーの頻度分布の差を, Fig. 5.7 に示す. グラフより, 次のコンセプトクラスタでは, 英語の共起頻度が日本語に比べて高くなっている.

- **クラスタ 5** little, poor, mini, 幅広い, 細かい, 薄い
- **クラスタ 13** white, young, blue, 黒い, 白い, 幼い
- **クラスタ 15** pretty, very, adorable, 恋しい, 可愛い, 懐かしい
- **クラスタ 18** thick, giant, massive, 太い, 濃い, 油っこい

一方で, 次のコンセプトクラスタでは日本語の共起頻度が英語に比べて高い.

- **クラスタ 8** asleep, sleepy, 眠い
- **クラスタ 9** fantastic, dangerous, awful, やばい, 素晴らしい, 凄い
- **クラスタ 10** bad, evil, tired, 悪い, 不味い, 面倒臭い
- **クラスタ 19** like, ぽい, っぽい

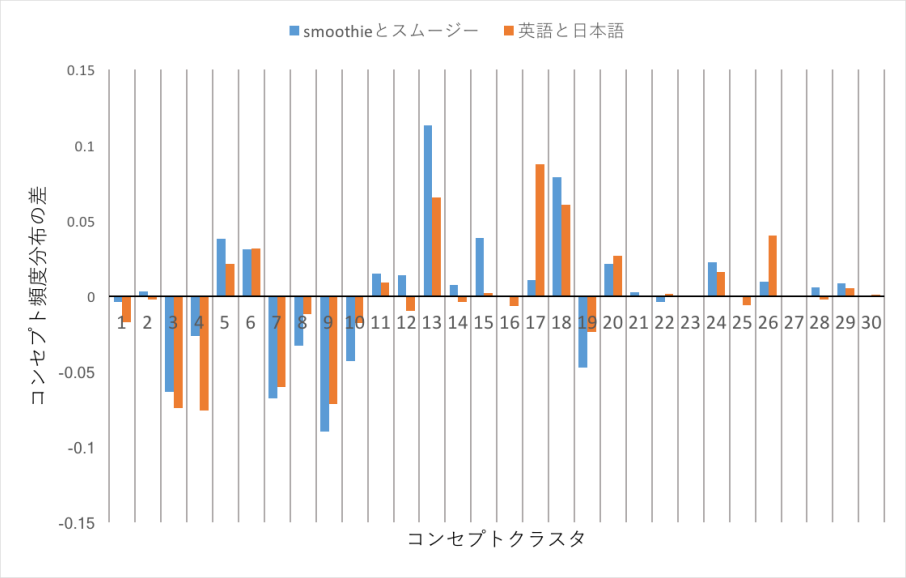


Fig. 5.7: smoothie とスムージーの頻度分布の差

5.4.4 juice / ミックスジュース

juice とミックスジュースの頻度分布の差を, Fig. 5.8 に示す. グラフより, 次のコンセプトクラスタでは, 英語の共起頻度が日本語に比べて高くなっている.

- **クラスタ 5** little, poor, mini, 幅広い, 細かい, 薄い
- **クラスタ 17** refreshing, cool, icy, 寒い, 新しい, 涼しい
- **クラスタ 26** much, few, many, 数少ない, 多い

一方で, 次のコンセプトクラスタでは日本語の共起頻度が英語に比べて高い.

- **クラスタ 7** interesting, hilarious, curious, 面白い, 興味深い, 楽しい
- **クラスタ 9** fantastic, dangerous, awful, やばい, 素晴らしい, 凄い
- **クラスタ 15** pretty, very, adorable, 恋しい, 可愛い, 懐かしい

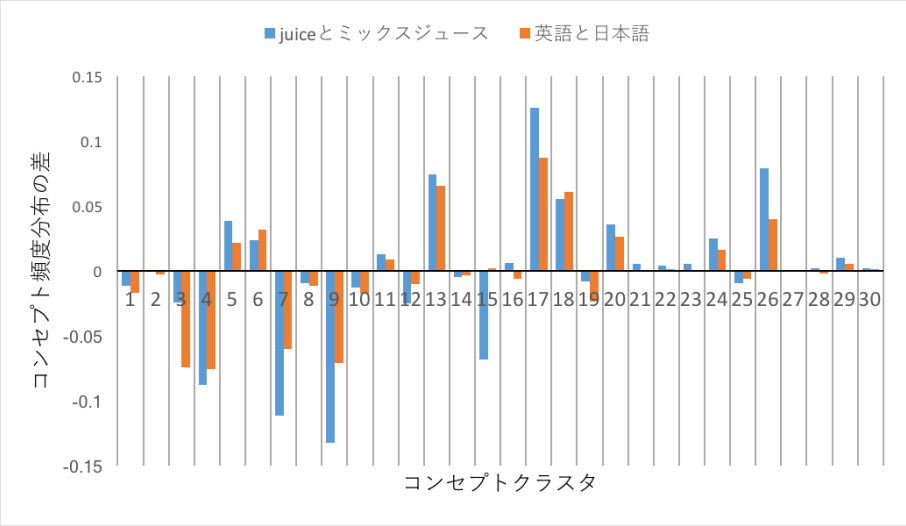


Fig. 5.8: juice とミックスジュースの頻度分布の差

第6章　考察

この章では, 実験結果の考察について述べる. 以下の三つの分析によって得られた知見と提案手法の評価について順に説明する.

- 全体傾向の言語間分析
- ネットワーク図による関係性の可視化
- 同一対象への感じ方の言語間分析

6.1　全体傾向の言語間分析

極性分析の結果から, 日本語に比較して英語の方がポジティブな形容詞を 10 倍以上多く用いていることがわかる. また, 形容詞の種類による分析から, 性質表現に関して, 英語が日本語よりも使用頻度が高いことがわかる. 英語のみに出現する性質表現を除き, 性質一般の表現のみを比較してもなお, 英語の方が性質表現の共起頻度が高いため, 英語圏では, 食の対象を描写する投稿が多いという傾向があると判断できる. 一方, 評価表現に対しては, 「良い」などの一般評価, 「楽しい」などの感情表現の双方において, 日本語が英語の 3 倍以上の差が確認された. 日本語圏では, 食事に対する自分の評価や感情についての記述が多い傾向があると判断できる. 以上の結果から, 食に関するソーシャルメディアの使い方に対し, **英語では, ポジティブな形容詞を用いて料理を描写し良さを伝えるような用法が多い一方で, 日本語では, 食事に対してネガティブな評価表現を用いて不満を言うような用法が多い**と推測した.

Vidal らは, 英語のツイートを対象に, breakfast, lunch, dinner, snack の 4 つの状況における Twitter の用法について分析し, 具体的にツイートの内容を分析し, 自分の食事の状況に関するツイートが多い一方で, 感情に対する言及は少なく, 代わりに顔文字が多く用いられていることを示した [36]. 一方で, Acar らは, 日本人大学生は自分に関する言及が多く, アメリカ人大学生は質問が多いことを示した [37]. 今回得られた結果は, これら既存研究から得られるアメリカ人と日本人のソーシャルメディアに関する用法の差と一致する. Acar らが, ツイートの内容を手動で確認して得た知見に対し, 形容詞の種類の分類を用いることで近い知見を得たことから, 描写表現, 特に形容詞の持つ情報量がわかる.

コンセプト頻度分布による分析の結果, 英語ではコンセプト 13（white, young, blue）, コンセプト 17（refreshing, cool, icy）, コンセプト 18（thick, giant, massive）, コンセプト 26（much, few, many）において特に日本語より共起頻度が高いという結果が得られた. このうち, コンセプト 26（much, few, many）は, 英語の品詞体系の特徴として, 比較級にこれ

らの言葉を用いることが多いためだと推測される.

コンセプト 13（white, young, blue）の色に関する描写が英語圏で多く用いられるという結果は, **アメリカにてカラフルな食品が多いという事実と合致する**. 古くから, 食べ物の変色を防ぐ目的や, 人の目を引く目的で商用に食用色素で着色することがアメリカでは浸透しており, アメリカでは, カラフルな食品に対して食欲をそそると感じる傾向がある. 例えば, アメリカのブランドであり, スポーツドリンクの中で世界 1 位のシェアを誇るゲータレード（GATORADE）という飲み物は, 種類ごとにカラフルな色とりどりの着色がされている. ゲータレードは, 世界 70 カ国以上で愛飲されているにも関わらず, 日本では現在生産されていない. 日本では着色料を想起するカラフルな色に対し, 不健康だ, 体に害がありそうだ, という印象が強く, 食品に対する派手な着色は好まれない.

一方で, 日本語では, コンセプト 3（硬い, 強い, 苦い）, コンセプト 4（柔らかい, 軽い, 安い）, コンセプト 7（面白い, 興味深い, 楽しい）, コンセプト 9（やばい, 素晴らしい, 凄い）において特に英語よりも高い共起頻度が見られた.

極性分析と形容詞の種類による分析にて, 日本語圏ではネガティブな言及や, 自分の感情についての記述が多いという結果があったが, 同じ傾向がコンセプト 3 やコンセプト 7, コンセプト 9 の共起頻度の高さからわかる. コンセプト 4（軽い, 安い, 優しい, 柔らかい, か弱い）が多く共起していることは, 一概には言えないが, 安い飲み物, 軽い飲み物など, 日常的な飲み物への日本人の関心度の高さを表している可能性がある. 日本では飲料の自動販売機の選択肢がアメリカに比べ多く, 日常的に触れる飲み物の種類が多い. コンビニエンスストアで見かける飲み物の種類もまた, アメリカに比較して多い. コンセプト 4 の共起の多さには, こういった背景が影響している可能性がある.

以上のように, コンセプト頻度分布に差異のあるコンセプトクラスタを比較することで, アメリカと日本の食文化の違いについて知見を得ることができることが確認できた. 違いを裏付ける根拠とはならない場合もあるが, 少なくとも, そのコンセプトにおいて, 違いが存在する可能性に気づくことができ, 描写表現から文化的差異を検出する有効性を示している.

6.2 ネットワーク図による関係性の可視化

英語のネットワーク図においては, juice と lemonade, tea が近い距離にあり, 果物が多く用いられる飲み物としての共通点が現れている. 日本語のネットワーク図においては, ココアや甘酒, 紅茶, はちみつドリンクと, 暖かい飲み物が近い距離にあり, 同じクラスタに属していることが分かる. また, モヒートと焼酎, 梅酒とカクテルなど, 酒類が比較的近い距離に集まっている. 豆乳とスムージーの距離が近いことも, 健康食としての認知や, 牛乳を使っている点での味覚の近さから, 直感とも合致する結果である.

これら二つのネットワーク図を比較すると, 例えばスムージーの位置付けの違いが確認できる. 英語のネットワーク図において smoothie は tea や cider, coffee 等の飲み物との間にエッジを持ち, ネットワークの中心に位置する. 一方で, 日本語のネットワーク図においては, スムージーは豆乳とのみエッジを持っており, ネットワークの端に位置する.

他のノードとエッジを持たないノードも存在するため, ネットワークで全ての飲み物の関係性が可視化できているとは言えない. しかし, エッジを持つノード同士や, ネットワーク上で近い位置にあるノード同士には確かに一定の関係性が見受けられる.

今回の分析に用いているのは, 人の分析対象に対する感性の部分である. ネットワーク図上に, 実際に私たちの感性に合致する関係性が見えていることから, コンセプト頻度分布のJS 距離を対象間の距離として用い, 対象同士の関係性の把握に用いる提案手法に関して, 一定の有効性が確認できる.

6.3 同一対象への感じ方の言語間分析

英語と日本語で同一対象をクエリとしている 4 つのペアについて, コンセプト頻度分布の差を分析した. 一つ一つの結果について考察した後, 4 つの結果を包括して, 提案手法の有効性やマーケティング戦略への応用について評価する.

6.3.1 coffee / コーヒー

英語では, コンセプト 6（warm, excited, nervous）が多く共起している. アイスコーヒーの起源は日本であり, 冷たいコーヒーを飲む習慣があまりないという背景から, コーヒーへの「熱い」や「暖かい」といった印象がアメリカの方が強い可能性がある. しかし一方で,「冷たい」という概念が所属するコンセプト 17（refreshing, cool, 寒い, 冷たい）もまた, 英語が日本語に比べて共起頻度が高い傾向にある. これは「refreshing」という概念が同じコンセプト 17 として分類されていることが原因である可能性があり, このコンセプトクラスタをより精度良く分割した上で更に議論する必要がある.

また, 英語でコンセプト 11（blunt, boring, sharp）の共起頻度が日本語より高い傾向にあることから, アメリカ人にとってコーヒーが「boring」な文脈に存在することが多く,「exciting」なコーヒーに対する欲求がある可能性が示唆される. 実際に,「boring」でないコーヒーを作る方法を紹介する記事は多く存在する[1].

一方, コンセプト 3（硬い, 強い, 苦い）やコンセプト 15（恋しい, 可愛い, 懐かしい）が日本語では多く共起している. コンセプト 3（硬い, 強い, 苦い）から, 日本におけるコーヒーへの「苦い」という一般認識が予想される. また, コンセプト 15（恋しい, 可愛い, 懐かしい）が多いことは, 日本でラテアートが独自の変化を遂げたためだと思われる. アメリカにもラテアートは存在するが, 日本のように動物やキャラクターを描く風潮はなく, コーヒーやラテアートに対して「可愛い」という一般認識はないものと予測できる.

[1]Storing Coffee to Avoid Boring Coffee.
https://pausecoffee.co.za/2017/07/06/storing-coffee-avoid-boring-coffee/

6.3.2 tea / 紅茶

英語では, コンセプト 13（white, young, blue）やコンセプト 15（pretty, very, adorable）が日本語に比べて多く共起している. アメリカでは, フルーツティーが好んで飲まれ, 種類も色も多種多様な紅茶が専門店にて販売されている. 茶葉の色合いや, 紅茶を注いだ時の水色も重視され, 目で紅茶を楽しむ文化が存在する. コンセプト 15（pretty, very, adorable）が多く共起していることも踏まえると, アメリカでは女性が紅茶をプレゼントとして贈る習慣が強く, カラフルで可愛いパッケージの装飾や, 紅茶関連の小物類が多く製造され販売されている傾向が見えてくる.

6.3.3 smoothie / スムージー

英語では, コンセプト 5（little, poor, mini）, コンセプト 13（white, young, blue）やコンセプト 15（pretty, very, adorable）, コンセプト 18（thick, giant, massive）の共起が日本語に比較して多く見られた. コンセプト 13（white, young, blue）やコンセプト 15（pretty, very, adorable）に関しては tea と紅茶の違いに通じる傾向であり, アメリカでのスムージーが多種多様な果物を用いたカラフルな認識であることと関係すると考えられる. コンセプト 18（thick, giant, massive）の共起頻度が高いことから, アメリカ人がスムージーの濃さをスムージーの評価基準に用いていることが予想できる. 実際, 濃いスムージーを作るためのレシピがインターネット上で紹介されている[2].

一方で, 日本では, コンセプト 8（眠い）や, コンセプト 9（やばい, 素晴らしい, 凄い）, コンセプト 10（悪い, 不味い, 面倒臭い）, コンセプト 19（ぽい, っぽい）が多く共起する傾向がある. コンセプト 10（悪い, 不味い, 面倒臭い）の共起から, スムージーに対して不味いという一般認識が日本において存在し, また, コンセプト 19（ぽい, っぽい）を多く用いている点から, スムージーの評価を他の飲食と比較して行う傾向にあると推測される. スムージー自体はアメリカほど日本で浸透しておらず, 他の食品の代替品として認識されている可能性がある.

6.3.4 juice / ミックスジュース

英語では, コンセプト 5（little, poor, mini）やコンセプト 17（refreshing, cool, icy）, コンセプト 26（much, few, many）が多く共起している. コンセプト 5（little, poor, mini）とコンセプト 26（much, few, many）の共起から, アメリカ人はジュースの量を気にする傾向があると推測される. また, コンセプト 17（refreshing, cool, icy）の共起から, ジュースに対して一般に冷たいものだという認識が強い可能性がある.

一方で, 日本語では, コンセプト 7（面白い, 興味深い, 楽しい）やコンセプト 9（やばい, 素晴らしい, 凄い）, コンセプト 15（恋しい, 可愛い, 懐かしい）の共起が多い. 今回, 日本

[2]How to Make the Thickest, Frostiest Smoothie Possible This Summer
http://www.onegreenplanet.org/vegan-food/how-to-make-the-thickest-frostiest-smoothie-possible/

語のクエリに「ジュース」ではなく「ミックスジュース」を使っているため, ジュースへの概念の違いと断定することは難しい面も残るが, 日本ではジュースに対してポジティブな感情を想起させる場面に登場する飲み物との印象があると予想することができる.

6.3.5 まとめ

4 つのペアに対する比較を行った結果, コンセプトに対する共起頻度の違いを分析することで, 歴史背景や, 飲料の出現する文脈に関わる, 文化的な違いに繋がる知見を得た. ここから, 対象に対して用いられる描写表現の頻度を分析することで, 今まで明確化していなかった対象に対する感じ方や捉え方の差異を取得でき, これら描写表現の情報を活用することに非常に意義があると考える. また, 異なる言語圏における感じ方の差異を捉えるにあたり, 本論文の提案手法は有効であると考える.

提案手法を用いることで, 描写表現から示唆として得られる差異には次のものがある.

一般認識
: 一般にその対象が文化圏においてどのような印象で認識されているか. 日本語でのスムージーに対する不味いという印象や, コーヒーに対する苦いという印象, 英語でのジュースに対する冷たいという印象がこれにあたる.

評価基準
: それぞれの文化圏が対象にどのような評価基準や選択基準を用いているか. 英語でのスムージーに対する濃さへの言及や, ジュースに対する量への言及, 紅茶に対する色への言及がこれにあたる.

文脈
: 対象がどのような場面で用いられているか. 英語でのコーヒーに対する「boring」という言及や, 日本語でのミックスジュースに対する「楽しい」等の言及がこれにあたる.

独自のジャンル
: その文化圏のみで存在する対象のジャンル. 日本語でのラテアートや, 結果では取り上げなかったが, 英語でのホットサイダーがこれにあたる.

上記のような言語圏ごとの文化的差異を取得することで, 企業が海外展開を行う際のローカライゼーション戦略に応用することが可能である.

例えば, 東洋水産のカップ麺がメキシコで大成功した例をあげて考える. メキシコで国民食となった東洋水産のマルちゃん製麺は, チリソースをかけたりライムを絞ったりして味をアレンジするメキシコの食習慣を反映し, 薄味となっている. さらに, 現地消費者の要望に応え, 「レモン & ハバネロ」というメキシコ人の好む酸味と辛味を強調した製品を投入している. その成功の裏には, メキシコ麺類市場の徹底調査, メキシコ人の嗜好や評価基準に対する分析と探求, さらに商品開発の試行錯誤における多大な努力がある.

2017 年の日本貿易振興機構（ジェトロ）によるメキシコでの日本食品消費動向調査[3]によれば, メキシコ人は一般に保守的な嗜好を持つと言われている. レストランでは自分の好きな料理がある程度決まっており, 続けて頼む傾向がある. 時々新しいメニューを試しに注文し, 気に入れば続けて同じものを注文する. 料理の色彩では派手ではっきりしたものが好まれ, 味覚では, 酸っぱくて辛い味を好む. また, 濃い目の味が好きで, ぼんやりとした味は好まない.

本論文の提案手法を用いることで, 上記の調査の結果得られるような, 色彩に対する反応や, 味覚の嗜好を把握することは難しくない. よって現地調査に必要な時間と労力を節約することが可能である. 加えて, 麺類に対する一般認識や文脈を把握することで, 既存の不満やステレオタイプから新たなビジネスチャンスを生む可能性や, 麺類に対する評価基準を把握することで, 製品の流通にあたり将来生じるであろう課題を事前に知り, 未然に防ぐ可能性もまた存在する.

ローカライゼーション戦略は, 現地市場の文化や風習, 嗜好を徹底的に調査し, 市場に対して自社の製品をどのように溶け込ませると受け入れられるかを探索する過程である. 本論文の提案手法では, 個別の製品が市場で受け入れられるかを検証することは難しい面もあるが, 少なくとも現地市場の傾向や現地住民の感性を自動的に明確化し, 地域に特化した製品開発を行う際の多くの気づきを与えることができる.

[3] https://www.jetro.go.jp/ext_images/_Reports/02/2017/

第7章　結論

7.1　本研究の結論

グローバル化に伴い, 企業のローカライゼーション戦略が多く見られる中, 地域の文化的差異に基づく対象の捉え方の違いを明確化することは重要な課題である. 本研究の目的は, 人々が無意識に用いている言語的表現から, 文化ごとの考え方の違いを明確化することであった. この目的に対する本研究の結論について述べる.

本研究では, ソーシャルメディアから得られる描写表現に対し, 複数言語の描写表現をコンセプトで分類した上で, コンセプトに対する頻度分布を求め, この頻度分布の差を用いて対象の文化間差異を評価する手法を提案した. 提案手法の妥当性を検証するため, 英語圏と日本語圏の食に対する Twitter データを例にあげ, 次の三つの分析を行った.

- 全体傾向の言語間分析
- ネットワーク図による関係性の可視化
- 同一対象への感じ方の言語間分析

全体傾向の言語間分析では, 形容詞を描写表現として用いることで, ソーシャルメディアの用法の言語間差異や, 食全体に対する, 文化ごとの傾向を抽出できることを示した. これにより, ソーシャルメディアから得られるテキスト情報は, コミュニティの傾向性に関する豊富な情報を持っており, 描写表現の分布を分析することが非常に有意義であると示した.

ネットワーク図による関係性の可視化においては, 描写表現の出現頻度の分布を用いて, 分布の近さによる飲み物群のマッピングを行った. この結果は, 全ての飲み物の関係を代表できる訳ではないものの, 私たちの感性と矛盾しない飲み物間の距離を表現できており, 人の感性を用いて文化を評価することの可能性を示した.

同一対象への感じ方の言語間分析では, 具体的な飲み物の種類に対し, 提案手法を用いて英語圏と日本語圏に存在する感じ方や認識の違いを分析した. 実際にコンセプト頻度分布に差があるコンセプトの内容を見ることで, 食文化に関して隠れた知識を明確化した. これにより, 提案手法により, 文化全体の距離だけでなく, 特定の対象に対し, コミュニティの間に存在する差異を明確化できることを示した.

以上から, 本研究は描写表現を用いて文化間比較を行うことの有用性を示すとともに, 複数言語に対して比較を行う手法の開発に成功した.

7.2 課題と今後の展望

本研究では, Twitter の複数言語のデータを用いた分析を行い, 描写表現の分布を比較することで文化間の感じ方の違いを抽出できることを示した. 使用するデータや分析の手順において, より一層の研究の方向性があるため, 今後の展望について記述する.

7.2.1 描写表現の拡張

本研究においては, 形容詞のみを描写表現とみなし, 分析を行ったが, 一部の名詞や動詞もまた, 対象を描写する重要な情報となることがある. これら表現を加えた分析を行うことで, さらに詳細かつ幅広い特徴を捉えることが可能となるだろう.

7.2.2 係り受け解析を用いた描写表現の抽出

本研究においては, 文章中の全ての形容詞を分析の対象に用いており, 一部, 対象とは直接関係の薄い形容詞もまた分析に影響を与えている. 係り受け解析を用いて, 文章中で, 分析対象にかかる単語や, 分析対象が主語となる際の述語のみを抽出し, 描写表現として用いることで, 各言語圏の人々の感性をより高い精度で得ることができるだろう.

7.2.3 コンセプトの細分化

本研究においては, ConceptNet API によって得られる重み情報を用いて, 関連のある単語ペアに対する重み付けを行った. しかし, 得られたクラスタからもわかるように, 一部の反意語や, 複数の意味をもつ単語の存在から, コンセプトの分類の精度が低いクラスタが存在する. 描写表現ネットワークをクラスタリングしてコンセプトを抽出するという提案手法に対し, ConceptNet に加え単語の分散表現やトピック分類などの手法を合わせて用いることで, 単語間の関係をより正確に把握できる可能性がある. これによりさらに精度の良いコンセプト分類が実現できると推測され, より具体的な文化間距離を比較することが可能となると考えられる.

7.2.4 分析結果の検証

考察にて, コンセプト頻度分布で差が大きいコンセプトに対し, その背景にある考え方の文化的差異を推測した. インターネット上の記事などを参考にしたものもあるが, 中には検証が難しいものもあった. これらに対し, アンケート調査などを用いて厳密な検証を行うことで提案手法の正当性を確認する必要がある.

7.2.5　実用面での課題

提案手法から得られる文化的差異を実際にローカライゼーション戦略に用いるためには, 過去のローカライゼーション戦略における成功例や失敗例の原因と, 本論文の提案手法で得られる知見とを照らし合わせ, より詳細かつ厳密な検討と議論を行う必要がある.

謝辞

本研究を進めるにあたり, 多くの方々にご助力を頂きました. ここに感謝の意を述べさせて頂きます.

指導教員の坂田一郎教授には, 大変ご多忙なお立場にも関わらず, 貴重なお時間を割いてご指導頂きました. 私の興味を生かした形での研究の課題設定や, 研究の質を高めるための方法等, 研究の最も重要な部分に対して親身なアドバイスを頂きました. 先生の幅広い知識や知見は本研究の大きな糧となりました. 心より深く感謝致します.

共同研究室の森純一郎准教授には, 特に手法に関して貴重なアドバイスを頂きました. 人工知能を専門とされる立場からの的確なご指摘を頂き, 本論文の完成度が高まりました. 感謝致します.

株式会社ホットリンク兼本学研究員の榊剛史氏, 本学研究員の大知正直氏, 浅谷公威氏には, 研究アイデアのブラッシュアップや基礎知識の取得, 論文校正等で大変お世話になりました. 特に大知さんには, ご自身の研究でご多忙であるにも関わらず, 週次のメンターミーティングにて研究に行き詰まった際のご相談や, 今後の方向性に対する的確なアドバイスを頂き, 大変感謝しております.

坂田森研究室のメンバーの諸先輩方には, 研究会でのアドバイスから日頃の研究室生活に至るまで, 様々な面で助力を頂きました. また, 同輩である鈴木くんと山本くんとは, 互いの研究について議論したり, 研究の息抜きをしたりと, 互いに切磋琢磨しながら楽しく研究を行うことができました. 本当にありがとう.

最後になりますが, 半年間の研究生活を支えてくれた家族と友人, また関係者の方々に心より深く感謝致します. 今後の研究室の更なる発展と, お世話になった皆様の益々のご活躍をお祈りいたします.

参考文献

[1] Thiago H Silva, Pedro OS Vaz de Melo, Jussara M Almeida, Mirco Musolesi, and Antonio AF Loureiro. You are what you eat (and drink): Identifying cultural boundaries by analyzing food and drink habits in foursquare. In *Proceedings of the 8th AAAI International Conference on Weblogs and Social Media (ICWSM' 14)*, 2014.

[2] Sina Sajadmanesh, Sina Jafarzadeh, Seyed Ali Ossia, Hamid R Rabiee, Hamed Haddadi, Yelena Mejova, Mirco Musolesi, Emiliano De Cristofaro, and Gianluca Stringhini. Kissing cuisines: Exploring worldwide culinary habits on the web. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 1013–1021. International World Wide Web Conferences Steering Committee, 2017.

[3] Miles Osborne and Mark Dredze. Facebook, twitter and google plus for breaking news: Is there a winner? In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, 2014.

[4] Arif Mohaimin Sadri, Samiul Hasan, Satish V Ukkusuri, and Manuel Cebrian. Understanding information spreading in social media during hurricane sandy: User activity and network properties. *arXiv preprint arXiv:1706.03019*, 2017.

[5] Seth A Myers, Chenguang Zhu, and Jure Leskovec. Information diffusion and external influence in networks. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 33–41. ACM, 2012.

[6] Xiangmin Zhou and Lei Chen. Event detection over twitter social media streams. *The VLDB journal*, 23(3):381–400, 2014.

[7] 榊剛史 and 松尾豊. ソーシャルセンサとしての twitter: ソーシャルセンサは物理センサを凌駕するか?(<特集> twitter とソーシャルメディア). 人工知能学会誌, 27(1):67–74, 2012.

[8] Stanislav Nikolov and Devavrat Shah. A nonparametric method for early detection of trending topics. In *Proceedings of the Interdisciplinary Workshop on Information and Decision in Social Networks (WIDS 2012). MIT*, 2012.

[9] Mika Viking Mäntylä, Daniel Graziotin, and Miikka Kuutila. The evolution of sentiment analysis-a review of research topics, venues, and top cited papers. *arXiv preprint arXiv:1612.01556*, 2016.

[10] Alexander Pak and Patrick Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In *LREc*, volume 10, 2010.

[11] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 2200–2204, 2010.

[12] Qi Su, Xinying Xu, Honglei Guo, Zhili Guo, Xian Wu, Xiaoxun Zhang, Bin Swen, and Zhong Su. Hidden sentiment association in chinese web opinion mining. In *Proceedings of the 17th international conference on World Wide Web*, pages 959–968. ACM, 2008.

[13] Preslav Nakov. Semantic sentiment analysis of twitter data. *arXiv preprint arXiv:1710.01492*, 2017.

[14] Guoning Hu, Preeti Bhargava, Saul Fuhrmann, Sarah Ellinger, and Nemanja Spasojevic. Analyzing users' sentiment towards popular consumer industries and brands on twitter. *arXiv preprint arXiv:1709.07434*, 2017.

[15] Brendan O'Connor, Ramnath Balasubramanyan, Bryan R Routledge, and Noah A Smith. From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM*, 11(122-129):1–2, 2010.

[16] Qi Gao, Fabian Abel, Geert-Jan Houben, and Yong Yu. A comparative study of users' microblogging behavior on sina weibo and twitter. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 88–101. Springer, 2012.

[17] Tianran Hu, Haoyuan Xiao, Jiebo Luo, and Thuy-vy Thi Nguyen. What the language you tweet says about your occupation. In *ICWSM*, pages 181–190, 2016.

[18] Trung Phan Thanh and Daniel Gatica-Perez. # healthy# fondue# dinner: Analysis and inference of food and drink consumption patterns on instagram. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*, number EPFL-CONF-233587, 2017.

[19] John Prescott. Comparisons of taste perceptions and preferences of japanese and australian consumers: overview and implications for cross-cultural sensory research. *Food Quality and Preference*, 9(6):393–402, 1998.

[20] Andrew Steptoe, Tessa M Pollard, and Jane Wardle. Development of a measure of the motives underlying the selection of food: the food choice questionnaire. *Appetite*, 25(3):267–284, 1995.

[21] Sharon M Pearcey and Ginny Q Zhan. A comparative study of american and chinese college students ' motives for food choice. *Appetite*, 123:325–333, 2018.

[22] Irith Freedman. Cultural specificity in food choice–the case of ethnography in japan. *Appetite*, 96:138–146, 2016.

[23] Push Singh. The public acquisition of commonsense knowledge. In *Proceedings of AAAI Spring Symposium: Acquiring (and Using) Linguistic (and World) Knowledge for Information Access*, 2002.

[24] Hugo Liu and Push Singh. Conceptnet—a practical commonsense reasoning tool-kit. *BT technology journal*, 22(4):211–226, 2004.

[25] Robert Speer, Joshua Chin, and Catherine Havasi. Conceptnet 5.5: An open multilingual graph of general knowledge. In *AAAI*, pages 4444–4451, 2017.

[26] Luis Von Ahn, Mihir Kedia, and Manuel Blum. Verbosity: a game for collecting common-sense facts. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 75–78. ACM, 2006.

[27] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[28] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[29] Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.

[30] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[31] 玉村文郎. 語彙論から見た形容詞. **その他**, 50:1, 1975.

[32] 高村大也, 乾孝司, and 奥村学. スピンモデルによる単語の感情極性抽出. **情報処理学会論文誌**, 47(2):627–637, 2006.

[33] 細川英雄. 形容詞の主観性について: 対象内容による形容詞の分類とその位置づけ. **早稲田日本語研究**, 1:78–65, 1993.

[34] 瀬戸賢一. **ことばは味を超える: 美味しい表現の探求**. Kaimeisha, 2003.

[35] 松尾章子. **食べ物のおいしさを表すことばに関する研究**. PhD thesis, 京都府立大学, 2014.

[36] Leticia Vidal, Gastón Ares, Leandro Machín, and Sara R Jaeger. Using twitter data for food-related consumer research: A case study on “what people say when tweeting about different eating situations”. *Food Quality and Preference*, 45:58–69, 2015.

[37] Adam Acar and Ayaka Deguchi. Culture and social media usage: analysis of japanese twitter users. *International Journal of Electronic Commerce Studies*, 4(1):21, 2013.