

# 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

Finding cultural differences on food using dish descriptions from social media:  
a comparative study of culinary perceptions and preferences between Japanese and Americans

中元 雪絵 \*1  
Yukie Nakamoto

大知 正直 \*1  
Masanao Ochi

榎 剛史 \*2  
Takeshi Sakaki

森 純一郎 \*3  
Junichiro Mori

坂田 一郎 \*1  
Ichiro Sakata

\*1 東京大学工学系研究科技術経営戦略学専攻

Department of Technology Management for Innovation, Graduate School of Engineering, The University of Tokyo

\*2 株式会社ホットリンク

Hotto Link Inc.

\*3 東京大学情報理工学系研究科

Graduate School of Information Science and Technology, The University of Tokyo

With the progress of globalization and the overseas expansion of the food industry, the importance of understanding regional differences in food culture is increasing. However, the traditional questionnaire-based study is costly in time and money. In this study, a new method for detecting cultural differences using social media is proposed with an experimental analysis on Japanese and American attitudes toward beverages. We gathered descriptive expressions (adjectives) about beverages with both languages from Twitter, and classified them into thirty meaning-based groups. We compared the co-occurrence frequencies of these groups between Japanese and Americans. The results reveal cross-cultural differences in impression and value standards on beverages. For example, coffee is generally recognized as 'bitter' and 'sour' in Japan, while it frequently appears in 'boring' contexts in the U.S. The proposed method enables a fast specific investigation on regional food culture, which can be applied to the localization strategy of the food industry.

## 1. 研究背景

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、アンケート調査がほとんどであったが、近年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のデータに注目が集まっている。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。Sajadmanesh らは、レシピデータを用い、味や材料からみた各国食習慣の距離を算出した [Sajadmanesh 2017]。これらの研究では文化の距離を世界規模で議論することに成功している一方で、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。

## 2. 目的

本研究では、ソーシャルメディアを用いて地域ごとの対象の感じ方の違いを明確化することを課題とし、言語的表現の違いを分析することでこれら差異を得ることが可能であるかを検証する。具体的には、描写表現に着目し、描写表現の共起頻度の違いを用いて文化間差異を自動的に比較する手法を提案する。提案手法を用いて Twitter の食に関する英日言語の感想データを比較することにより、文化間の感じ方の差異を検出し、提案手法の有効性を考察する。

## 3. 提案手法

### 3.1 着想と概要

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。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。



図 1: 英語と日本語のツイートの例

図 1 は Twitter から得られる麺についての英語と日本語のツイートの例である。英語では麺に対して「hot」や「spicy」が共起し、日本語では「真冬の夜中」という場面設定や「旨い」が共起する。ここから、英語ツイートの投稿者は麺類に対して辛くてスパシーなものを求めている、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、言語ごとに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差異を定量的に評価するこ

連絡先: 中元 雪絵, 東京大学工学系研究科技術経営戦略学専攻, 東京都文京区本郷 7-3-1, snopict@gmail.com

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、どのような文脈に対象が登場するのかといった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、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、表現同士の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、分析したい対象に対し、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。提案手法の概要を図2に示す。以降で詳細を説明する。

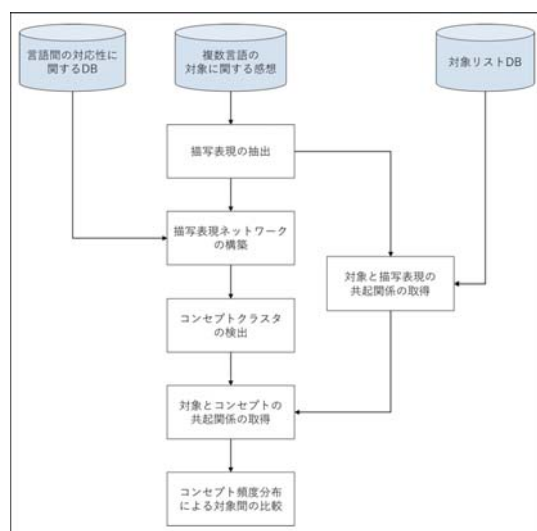


図2: 提案手法の概要

### 3.2 描写表現ネットワークの構築

入力として、分析対象のリスト、これら対象に関する複数言語における感想のデータ、言語間の対応性に関するデータベースの3つを用いる。あらかじめ分析したい対象を決定し、Twitterなどのマイクロブログやレビューサイトを用い、分析対象への感想データを取得しておく。

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。本研究では、形容詞が描写表現に相当すると考え、それぞれの感想データに対して形態素解析を行うことで、英語と日本語の形容詞群を抽出した。（描写表現の抽出）

続いて、各描写表現をノードとみなし、近い意味を持つ描写表現同士に重みを付与して繋ぐことで、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。本研究では、言語間の対応性に関するデータベースとして、ConceptNet[Speer 2017]を用いて近い関係のあるペアと重みを取得した。ConceptNetは異なる言語間を繋ぐリンクの情報を持つため、描写表現ネットワークは全言語の描写表現に対し意味の近いものを結んだネットワークとなる。一例として、英語と日本語の形容詞群に対して構築した描写表現ネット

ワークを図3に示す。図はGephi<sup>\*1</sup>を用いて描画した。（描写表現ネットワークの構築）

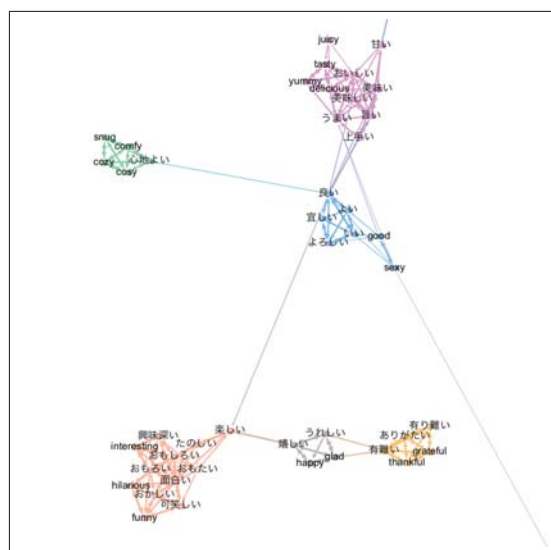


図3: 描写表現ネットワークの例

### 3.3 コンセプトクラスタの検出

構築し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。近い意味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、似た意味をもつ表現が集まる。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」と定義する。（コンセプトクラスタの検出）

### 3.4 コンセプト頻度分布による対象間の比較

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（形容詞）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）。これ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）

分析対象リストに属する二つの対象について、それぞれコンセプト頻度分布を求め、頻度分布のJS距離を用いて対象間の関係や近さを評価する。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）

## 4. 実験

### 4.1 Twitterの食データに対する提案手法の適用

飲み物に関するクエリを英語・日本語に対して設定し、Twitterからそれぞれ感想データを取得した。実験に用いたクエリを表1に示す。英語と日本語間で、クエリに用いている単語の種類や粒度が異なるが、これは飲み物群の分類の差もまた文化的差異を表していると考え、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ためである。

\*1 <https://gephi.org/>

英語は allrecipes\*2 の「Drinks Recipes」カテゴリ、日本語は 楽天レシピ\*3 の「飲みもの」カテゴリの分類を参照した。

表 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, ソードドリンク, 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

4.2 提案手法の評価

提案手法の妥当性を評価するため、Twitter から取得した飲み物の感想データを用い、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。具体的には、極性分析 (全ツイートの極性平均の比較)、形容詞の種類による分析 (松尾による形容詞の分類 [松尾 2014] に基づく、各形容詞の種類の出現割合の比較)、コンセプト頻度分布による分析 (英語全体、日本語全体のコンセプト頻度分布において、差が大きいコンセプトクラスターの検出) の三つの分析を行う。

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。

5. 結果

5.1 全体傾向の言語間比較

取得した全ての英語ツイートと日本語ツイートをを用い、描写表現を用いて英語圏と日本語圏の全体傾向の差異を正しく抽出できるか評価した。

ツイートの極性分析の結果、英語の極性平均は 0.266、日本語は 0.0137 となった。また、形容詞の種類による分析を行った結果、英語では対象の属性や性質を描写する形容詞が多く、日本語では自身の感情や評価に関する形容詞が多く用いられるという結果が得られた (表 2)。ここから、英語圏でポジティブな形容詞を用いて料理を描写する投稿が多い傾向に対し、日本語圏でネガティブな評価表現で不満を言う投稿が多いという示唆を得た。これは、日本語圏で自分に関する言及が多く、英語圏で食事の状況に関する投稿が多いという先行研究の結果と一致する [Vidal 2015, Acar 2013]。

さらに、英語全体と日本語全体のコンセプト頻度分布を比較することで、色への描写が英語圏にて多いことを検出した。これはアメリカでカラフルな食品が多い事実と合致する。

以上より、描写表現を用いて食文化に対する知見を得ることが可能であり、描写表現、中でも形容詞から文化的差異を検出する有効性を示している。

表 2: 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	<b>23.83</b>	10.39
	状況	5.39	5.79
評価表現	一般評価	19.44	<b>45.26</b>
	感情	3.02	<b>9.52</b>
その他		21.8	13.28

5.2 ネットワーク図による関係性の可視化

二つの対象について、コンセプト頻度分布の JS 距離を用いて距離の近さを定めた。この値を用いて英語のクエリ集合、日本語のクエリ集合それぞれにつ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ことを確認した。得られたネットワーク図を図 4 に示す。

ネットワークにおいて、英語では juice, lemonade, tea など果物が用いられる飲み物が近い位置に存在し、日本語ではココア、甘酒、はちみつドリンク、紅茶と温かい飲み物が近い位置に存在する。全ての飲み物の関係を正しく評価できている訳ではないが、私たちの感性に合致する結果が見えていることから、感性の評価にあたり、コンセプト頻度分布の JS 距離を用いることは妥当であると考えられる。

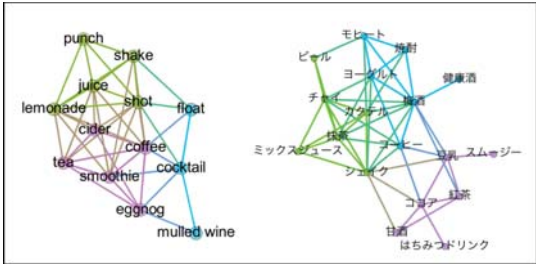


図 4: ネットワーク図 (左: 英語, 右: 日本語)

\*2 <http://allrecipes.com/>  
\*3 <https://recipe.rakuten.co.jp/>



### 5.3 同一対象への感じ方の言語間比較

クエリのうち、同一対象を扱っている4つのペア(コーヒー、紅茶、スムージー、ジュース)について差の大きいコンセプト頻度分布を比較し、感じ方の差異を検出できるか検証した。ここではコーヒーに対する分析の結果を取り上げて説明する。図5はcoffeeとコーヒーの頻度分布の差を表し、値が正のクラスは英語で共起が多く、負のクラスは日本語で共起が多いことを示す。この図において、英語全体と日本語全体における差よりもさらに大きい差を持つ(左側の棒グラフが右側の棒グラフより突出している)コンセプトクラスの内容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。

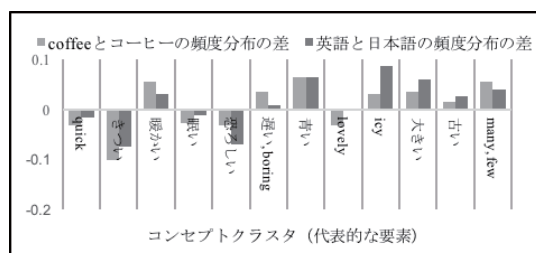


図 5: コーヒーに対するコンセプト頻度分布の差

英語ではコンセプト《暖かい, warm》や、コンセプト《遅い, boring》が日本語に比べて多く共起した。アイスコーヒーの起源は日本であり、英語圏では冷たいコーヒーを飲む習慣があまりない背景から、コーヒーに対する「熱い」という印象がアメリカの方が強い可能性がある。また、コンセプト《遅い, boring》の共起傾向から、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。

日本語では、コンセプト《lovely, 可愛い》が多く共起した。これ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、日本のようにキャラクターを描く文化は一般的ではなく、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と予測できる。

他のペアに対しても、同じように文化的な感じ方や捉え方の差異を検出することが可能であった。描写表現を分析することで得られる示唆には次のものがあげられる。

#### 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、英語でのホットサイダーがこれにあたる。

このように、提案手法により、アンケート調査や聞き込み調査の労力を使うことなく、食文化の文化間差異に対する知識を自動的に得ることが可能である。地域ごとの市場の文化や風習、嗜好を理解し、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点で、食産業が海外進出の際のローカライゼーション戦略に対する一助となると考える。

## 6. 結論

本研究では、ソーシャルメディア上の描写表現を用いて人の感性に基づく文化間比較の行うことの有用性を示した。また、複数言語の描写表現をコンセプト単位で分類した上で、共起頻度の分布を用いて比較を行うことで、異なる言語圏に関する文化的差異を明確化する手法の開発に成功した。

## 参考文献

- [Silva 2014] Silva, T. H., de Melo, P. O. V., Almeida, J. M., Musolesi, M., and Loureiro, A. A. (2014, June). 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).
- [Sajadmanesh 2017] Sajadmanesh, S., Jafarzadeh, S., Ossia, S. A., Rabiee, H. R., Haddadi, H., Mejova, Y., Musolesi, M., Cristofaro, E. D., and Stringhini, G. (2017, April). Kissing Cuisines: Exploring Worldwide Culinary Habits on the Web. In Proceedings of the 26th International Conference on World Wide Web Companion (pp. 1013-1021). International World Wide Web Conferences Steering Committee.
- [Speer 2017] Speer, R., Chin, J., and Havasi, C. (2017, February). ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. In AAAI (pp. 4444-4451).
- [松尾 2014] 松尾章子. (2014). 食べ物のおいしさを表すことばに関する研究, 京都府立大学. Ph.D. thesis.
- [Vidal 2015] Vidal, L., Ares, G., Machn, L., and Jaeger, S. R. (2015). 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.
- [Acar 2013] Acar, A., and Deguchi, A. (2013). Culture and social media usage: Analysis of Japanese Twitter users. International Journal of Electronic Commerce Studies, 4(1), 21.