

日本酒の味わい表現の分析のための程度表現の定量化

小澤奈泉¹ 山田剛一² 増田英孝²

¹ 東京電機大学大学院未来科学研究科 ² 東京電機大学未来科学部
ozawa@cdl.im.dendai.ac.jp {yamada,masuda}@mail.dendai.ac.jp

概要

日本語で何かを表現する際に程度表現を用いることで、より細かな度合いで表現することが可能になる。これまで日本酒の味わい表現の分析を行ってきたが、「ちょっと」「とても」などの程度表現は考慮していなかった。そこで、本研究では程度表現に着目し、それが修飾する語句の度合いをどの程度変化させるのかを定量化するために、程度表現の表す程度の範囲を問うアンケート調査を実施した。得られたデータから範囲の上限値下限値を求めることで、程度表現それぞれの表す程度の範囲を確認した。

1 はじめに

ジャンルを問わず初心者が自分にあったものを自分で選び出すことはハードルが高く困難であり、それは日本酒選びにおいても当てはまる。自分にあった日本酒を選ぶためのツールとして、推薦サービスがいくつか存在する。藍色システムズ株式会社の提供する「さけのわ」[1]や、SCENTMATIC株式会社の提供する「KAORIUM for Sake」[2]などである。また、日本酒の他の指標として、香りの高さや味の濃さを軸とした「熟酒」「醇酒」「薫酒」「爽酒」の4タイプに分類することもできる[3]。

これまで行った日本酒の味わい表現の分析では、銘柄別に口コミサイトの評価を分析することで味わい表現を抽出し、どのような銘柄でどのような表現がどの程度使われているのかをヒートマップで可視化した。これにより、「すっきり」といった単語を指定して日本酒を探すことが出来るようになり、日本酒の味わいが馴染み深い言葉で表現される。日本酒に詳しくない初心者でも味を想像することが容易になり、自分にあったものを自分で選び出す際のハードルを下げることに繋がる。

味わいを表現する際に「ちょっと」「とても」など程度を表す表現が用いられるが、我々のこれまでの研究ではそれらを考慮していなかった。そこで、本研

究では程度表現に着目し、程度表現が表す程度を定量化する。これにより、味わい表現の抽出の際に重みづけを行うことができるようになり、可視化の精度が向上する。

2 関連研究

本節では、日本酒の味わい表現に関するもの、日本語の程度表現の定量化に関する研究について述べる。

2.1 日本酒の味わい表現に関する研究

福本ら [4] は日本酒の味表現の研究を行っている。日本酒の味表現として、基本味やそのバリエーションが具体的にどのような表現が存在しているのか、どのような対象についてどのような表現が用いられているのか、味の比喻表現を中心に、どのような対象が用いられているかを分析した。日本酒の味表現データとして、酒店の日本酒の説明文を利用し、比喻を伴うような直接的な味表現と、使用米や酵母の説明と合わせた間接的な味表現に分けて解析している。また、福本ら [5] は収集した味表現をもとに、その類似性についても分析している。日本酒の説明文の構文解析結果から、基本的味表現を示す語と係り受け関係にある語のリストの類似性を判定することで、類似した味表現をもつ銘柄の日本酒を獲得した。各銘柄の類似性は、係り受け関係が得られた各銘柄データについて、味表現の類似性として各味表現の係り受け関係のある全ての語のうち、共通の語がある件数を合計することで判定した。

2.2 程度表現の定量化に関する研究

程度表現の定量化には、WordNet(英語)の語彙に対して感情極性値を付与した SentiWordNet を用いる手法、人手で程度を付与する手法、アンケート調査を行い値を決定する手法の主に3通りが用いられている。Singh ら [6] は SentiWordNet を用いて感情分類

を行っている。また、Taboada ら [7] は感情値を計算するシステムとそのための感情値辞書を作成した。辞書作成にあたり、感情値は-5から5の間の値を1名が決め、それを他の2名がレビューをする形式をとった。程度表現に関しては対象の感情値を増減させるものと考え、その割合を決定した。

アンケート調査による手法も、一対比較法や数直線に書き込む手法、感じた程度の値を答える方法など様々である。竹内ら [8] は日本語の意味計測のため、ファジィ評定法を用いて調査を行った。ファジィ評定法とは、評定調査において区間で回答させる手法であり、程度表現の表す程度の範囲を-100から100の範囲で上限と下限を回答させた。また、その中で最も適切だと感じる値も回答させ、中心値とした。結果から区間平均と度数分布を算出したところ、中心値だけで見れば同程度であっても、表現の指示範囲は語によって異なることがわかった。また、得られた度数分布は、単調に増加するもの、評価尺度が100の時点で度数が現れる単峰型、評価尺度が0と100の両端で度数が0となる単峰型、評価尺度が0の点で度数が2以上であり100の点で度数が0となる単峰型、それ以外の概ね5タイプに分類できることがわかった。

3 味わい表現と程度表現

味わい表現と程度表現について説明する。味わい表現とは、食品の味や香りを表現する際に用いられる用語をさし、福本ら [4] の提示した味表現を修飾する、もしくはそれに係るものとする。例えば、「フルーティーな香り」というフレーズであれば「フルーティー」が味わい表現であり、「香り」は味表現にあたる。程度表現とは、「ちょっと」「たくさん」など物事の度合いを増減させる役割をもつ単語とし、織田ら [9] は、程度表現を「実現の程度量（確信）表現用語」「現実の程度量表現用語」「時間的の程度量（頻度）表現用語」「心理的時間の程度量表現用語」の4種に分類している。本研究では、程度表現は味わい表現の度合いを増減されるものと考え、味わいに限らず程度表現が度合いを変化させる表現を「形容表現」と呼ぶ。

4 程度表現の定量化

本節では、本研究にて扱う程度表現とそれらの定量化手法について述べる。

4.1 定量化の方法

程度表現の定量化をするにあたり、アンケート調査を用いる。竹内ら [8] によると、程度表現は単語により表す範囲が異なる。そこで、程度表現が付加された形容表現の表す範囲の上限値と下限値、それらを定める際に連想した形容表現の2点を質問項目とした。

範囲を質問するにあたり、評価尺度を設定した。形容表現単体での度合いを1、形容表現に「ない」を伴う語の度合いを0とし、0より大きい実数値での回答を求めた。また、形容表現の下限は「ない」を伴う語であるが、上限にあたる語は無いと考え、上限値を設けずに「表現する範囲は下限値以上の全て」となる場合には「上限なし」と回答してもらった。以下に「やや」「あまり」という程度表現の回答例を「甘い」という形容表現を用いて示す。

- やや甘い→「甘い」を1、「甘くない」を0として、「やや」は度合いを減少させ、0.3~0.7程度が「やや甘い」の表現する範囲と感じたら、下限値:0.3, 上限値:0.7
- あまり甘くない→「甘い」を1、「甘くない」を0として、「あまり~ない」は度合いを減少させ、0.2~0.6程度が「あまり甘くない」の表現する範囲と感じたら、下限値:0.2, 上限値:0.6

4.2 程度表現の選定

調査に使用する単語を選定するにあたり、過去の研究で採用されていた単語の多くがかしこまった堅い表現であった。レビューで 사용되는単語は砕けた表現が多く見られることから、レビューで用いられている単語を使用することとした。そこで、さけのわよりレビューを収集した上で、程度表現は多くが副詞なことから、その中から副詞のみを取得し、程度を表す表現のみを選択した。

織田ら [9] を参考に程度表現を「実現の程度量（確信）表現用語」「現実の程度量表現用語」「時間的の程度量（頻度）表現用語」「心理的時間の程度量表現用語」「その他」の5つに分類した。「その他」に分類された単語の中には、使用される形容表現に限りのあるものがあつた。そこで、その中でも味表現に使用できる語と、最も単語数が多く汎用性の高い「現実の程度量表現用語」を加えたもの採用した。

5 調査

本節では、行ったアンケート調査の詳細について述べる。

5.1 調査対象とする程度表現

4 節 3 章で述べた手法で単語を選定し、最も基本的な程度量を表す「現実の程度量表現用語」に分類された 25 語と、「その他」に分類された語の内、味わい表現に対して用いられる「ほんのり」「しっかり」「うっすら」を加えた 28 語とした。

5.2 アンケート調査の方法

程度表現の定量化のためのアンケート調査は 2024 年 1 月に、10 代から 40 代の男女 23 名（男性：11 名，女性 12 名）に対して実施した。

5.3 アンケート調査の結果

上限値を設けなかったが、値の付け方にあまり個人差は見られなかった。調査の結果得られたデータの下限值，上限値，中心値の中央値の一部を表 1 に示す。中心値とは，下限値と上限値の中間の値とする。下限値と上限値の中央値はすべての程度表現で算出し，上限値の半数以上が「上限なし」だった場合，上限値の中央値は「なし」としている。また，中心値は「上限なし」の場合を無限大とし，それが過半数を占める場合には中心値の中央値は「なし」としている。各単語の指示範囲を，下限値と上限値の中央値をもとに数直線に起こしたものを図 1 に示す。点線は，評価尺度が 1 の地点である。

また，程度表現それぞれについて下限値と上限値にどの程度ばらつきがあるのかを調べた。「やや」という程度表現の例を図 2 に示す。線上の点はそれぞれの中央値を示す。

次に度数分布を求める。まず，評価尺度を 0.1 刻みで分割し，評定値を縮約する。それぞれの単語について，被験者それぞれの回答した範囲をの要素が 1，それ以外の要素は 0 をとる行ベクトルを用意し，被験者全員分の行ベクトルを加算することで度数分布が得られる。被験者は 23 名のため，度数分布の最大値は 23 となる。

度数分布の出力結果の一部をである「やや」「とても」を図 3,4 に示す。

表 1 程度表現が表す程度の範囲

程度表現	下限値	上限値	中心値
まったく	0.0	0.1	0.05
うっすら	0.0	0.4	0.275
全然	0.1	0.4	0.15
ほんのり	0.2	0.55	0.375
少し	0.2	0.5	0.4
あんまり	0.25	0.5	0.4
ちらほら	0.25	0.55	0.425
ちょっと	0.3	0.5	0.4
いささか	0.3	0.65	0.5
そこそこ	0.5	0.8	0.65
やや	0.5	0.8	0.65
まだまだ	0.5	0.95	0.75
まあまあ	0.5	0.9	0.65
わりと	0.7	1.0	0.9
じゅうぶん	0.9	1.3	1.05
ちょうど	0.925	1.0	1.0
ぎっしり	1.05	2.5	1.95
しっかり	1.2	3.0	2.25
大いに	1.2	2.0	1.5
さらに	1.35	2.75	2.125
だいぶ	1.5	2.0	1.75
かなり	1.5	10	6.5
とても	2	なし	なし
最も	2.0	なし	なし
たいへん	2.0	なし	なし
一番	2.0	なし	なし
たくさん	2.0	なし	なし
めっちゃ	2.0	なし	なし

5.4 考察

「やや」や「ほんのり」のように範囲の狭いものと、「とても」のようにある値以上全てを範囲とするものに分かれた。単語によりとる範囲の幅が異なる，という点で竹内ら [8] と同様の結果が得られたと言える。

本研究の結果を利用する際，程度表現が表す程度の範囲ではなく，程度の代表値が使いやすいこともある。程度が 1 より小さい場合には上限値と下限値の平均，程度が 1 より大きい場合には下限値を代表として扱うことが考えられる。程度が 1 より大きい場合に下限値を使うことで，程度の範囲に上限のない程度表現についても統一的に扱うことができる。

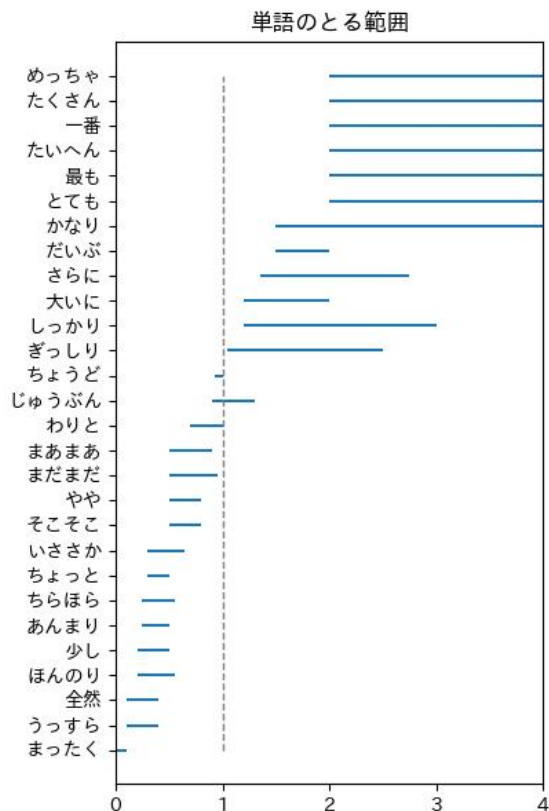


図1 各単語の指示範囲

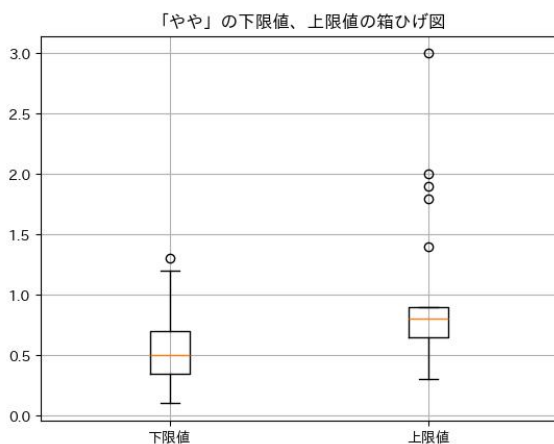


図2 「やや」の回答の箱ひげ図

本研究では程度の文脈依存性について検討をしていない。「ちょっと」は婉曲的な表現として使われることがある [10]。本研究で実施したアンケートでは文脈が与えられていないため語本来の程度が回答されていると考えられるが、実際の場面では、かなりつまらなくても「ちょっとつまらない」と言うことがある。本研究で用いた程度の指標では、程度表現を伴わない場合を1、否定を伴う場合を0とした。「面白い」を1、「面白くない」を0としたが、「面白くない」と「つまらない」の関係については扱わな

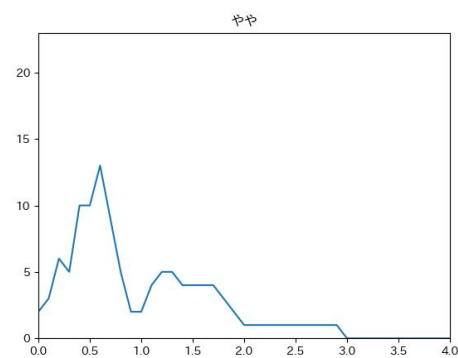


図3 「やや」の度数分布

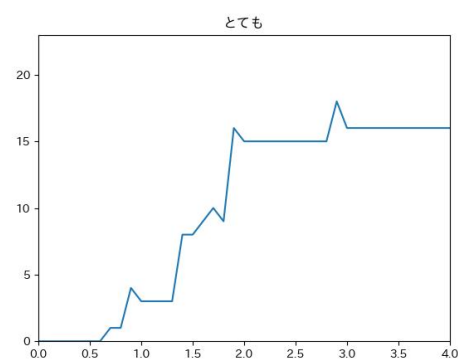


図4 「とても」の度数分布

かった。この関係は語によって異なることが知られている [10]。

6 おわりに

本研究では、日本酒のレビューより程度表現を抽出し、アンケート調査を用いて程度表現が形容表現の度合いをどの程度増減させるのか、その定量化を試みた。調査の結果得られたデータより、指示範囲の下限値、上限値、中心値それぞれの中央値を算出し、数直線に反映させた。これにより程度表現の表す程度の範囲を得られた。今後、それぞれの代表値を用いて重みづけをするなどして活用していきたい。

参考文献

- [1] 藍色システムズ株式会社. さけのわ. 2023年12月21日参照.
- [2] SCENTMATIC 株式会社. 言葉は、味わいを変える。kaorium for sake. 2023年12月21日参照.
- [3] 日本酒サービス研究会・酒匠研究会連合会 (SSI). 日本酒の香味特性格別分類 (4タイプ). 2023年12月21日参照.
- [4] 福本淳一. 日本酒の味表現の分析. 人工知能学会全国大会論文集 第30回 (2016), 3M4OS20b3. 一般社団

- 法人人工知能学会, 2016.
- [5] 福本淳一. 日本酒の味表現の類似性の分析. 人工知能学会全国大会論文集 第 32 回 (2018), 3B2-OS-22b-02. 一般社団法人人工知能学会, 2018.
 - [6] V.K. Singh, R. Piryani, A. Uddin, and P. Waila. Sentiment analysis of movie reviews and blog posts, 2013.
 - [7] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. Lexicon-based methods for sentiment analysis. **Computational linguistics**, Vol. 37, No. 2, pp. 267–307, 2011.
 - [8] 竹内晴彦. ファジィ評定法による程度表現用語の意味計測. 計量国語学, Vol. 17, No. 8, pp. 365–376, 1991.
 - [9] 織田揮準. 日本語の程度量表現用語に関する研究. 教育心理学研究, Vol. 18, No. 3, pp. 166–176, 1970.
 - [10] 川端元子. 程度副詞を分類する視点の考察. 愛知工業大学研究報告, Vol. 47, pp. 115–124, 2012.