

メタファーの自動生成に向けた客観的評価指標の検討

宮澤 彬^{1,2,a)} 宮尾 祐介^{1,2,b)}

概要: 本稿では、隠喩（狭義のメタファー）を生成するタスクについて、生成された表現のよさを客観的に評価するための指標の提案を行う。提案する指標はメタファー性・新規性・理解可能性の3項目と、それらをまとめた総合評価である。はじめにこれらの必要性や妥当性について論じる。次に提案する指標について適切に評価することが可能であることをクラウドソーシングを用いた実験により示す。また得られた評価から各指標の有効性やそれぞれの関係性について分析を行う。最後に提案する評価指標により「よいメタファー表現」とされた表現が、実際によいといえるかどうか論じる。

1. はじめに

自然言語処理においてメタファーを扱うタスクは、主に検出・理解・生成の3つがあるが、生成は他の2つに比べると盛んに研究されているとは言い難い。しかしメタファーの生成、あるいはメタファーへの言い換えは、詩や小説の執筆支援や自動生成において重要な役割を果たす。さらに、メタファーは抽象的なことについて具体的な言葉で語ることを可能にするため、専門的な文書を平易に書き換えるといった実用的な応用も考えられる。

また既存のメタファー生成に関する研究は、生成の対象を「子犬のような少年」といった「SのようなT」の形式の直喩表現に限定している場合が多い。しかし、メタファーには直喩以外にも、隠喩（狭義のメタファー）やメトニミーといった分類があり、日常生活でもよく使われるこれらを無視し続けていくことはできない。今後直喩以外のメタファーを生成するタスクに取り組むためには生成されるべき「よいメタファー」というものを定める必要がある。

本稿ではメタファーのよさを測るため、メタファー性・新規性・理解可能性という3つの指標と、それら総合をした評価指標を提案する。次にこれらが実際に評価できることをクラウドソーシングを使った実験によって確認する。評価の対象は「情熱を注ぐ」や「喜びが満ちる」といった、計1,360個の短い表現である。これらは、主に感情に関する名詞（「情熱」や「喜び」など40個）を集めたりストと、水に関する動詞を含む句（「Xを注ぐ」や「Xが湧く」など

34個）を集めたりストから1つずつ選び組み合わせたものである。その後、実験により集められたデータに基づき各指標間の関係性を示す。最後にこれらの指標による評価が「よいメタファー」を発見する上で有効であることを示す。

2. 関連研究

自然言語処理においてメタファー生成を扱う研究はまだ多くない。それらの多くは「SのようなT」の形式の直喩を生成するものである([1], [2])。評価方法を見ると、まず[1]は人に「比喩としての評価」を5段階でつけてもらい評価している。[2]もまた人手による評価であるが、単一の指標ではなく、適切さ・視覚化の容易さ・おもしろさ・新規性の4項目で評価を行っている。

[2]のような評価が行われる背景には、[3]などの認知心理学的な研究の成果がある。[3]は「TはSのようだ」という形式の直喩表現を対象として、表現の評価に影響を与える要因を、アンケート調査から分析している。これによると、直喩表現の適切さの評価は、表現のおもしろさと理解のしやすさに影響を受ける。そして、おもしろさは、SとTそれぞれが属するカテゴリ間の差異の大きさに、理解のしやすさは、SとTをそれぞれ美醜などの感覚的な評価から計算した類似度に影響を受ける。

[4]はこのような評価を生成の過程に取り込んだシステムを提案している。これは“a sad song and a song narrates”のような入力から“the song like S”の形式の表現を出力する過程で、“song”と候補Sの意味の違いが大きく、かつ出力の句の意味が入力の意味に近いものを選択するようにしている。

¹ 総合研究大学院大学
SOKENDAI (The Graduate University for Advanced Studies)

² 国立情報学研究所
National Institute of Informatics, Chiyoda, Tokyo 101-8430, Japan

a) miyazawa-a@nii.ac.jp

b) yusuke@nii.ac.jp

3. 提案手法

本稿ではメタファー表現のよさを測る指標として、**メタファー性・理解可能性・新規性**の3つの指標を導入する。複数の指標を導入するのは、重視する性質に応じて出力すべき表現を変えられるようにするためである。例えば、詩や小説の執筆支援では独創的な、すなわち新規性の高いメタファーへの需要が高まるだろう。一方、表現を分かりやすいものに書き換える際には理解可能性が重要になる。特に重視すべき性質がない場合のため、これらを統合した単一の指標についても後で議論する。

3.1 メタファー性

第一の指標として「どの程度メタファーらしいか」を測る**メタファー性**が必要である。前節で述べたように、既存のメタファー生成の研究は「*S*のような*T*」などの限られた形式で生成を行っていたため出力された表現がメタファーかどうかは、大きな問題になってこなかった。しかし、一般的なメタファー生成ではこれは重要な問題である。生成された表現がメタファーでなければ、それは単なる言い換え表現に過ぎないからである。

メタファーの定義は分野や研究によって様々である。認知言語学では領域間の構造的対応関係を(概念)メタファーと呼び、個々のメタファー的な表現はメタファー表現と区別して呼ぶことが多い[5]。これに従うと「怒りに燃える」や「怒りに油を注ぐ」は、領域《火》から領域《怒り》への概念メタファーに基づくメタファー表現となる。他の基準として、コーパス言語学における[6]の基準がある。この基準では、辞書に掲載されている語義のうち身体や五感と結びついたものを基礎語義(basic meaning)と定め、それと異なる語義で使われている語をメタファーとみなす。ただし、本研究では、[6]のような厳密な方法を用いることはしない。かわりに各評価者に自身の直感で提示された表現が「比喩だと感じられるかどうか」を5段階で評価してもらい、その平均を表現の**メタファー性**を測る指標とする。この方法の利点として、ガイドラインに従ってメタファーかどうかを判断する場合と比べ、作業者の負担が小さいこと、作業者の習熟が必要ないことが挙げられる。また2値ではなく連続的な指標にすることで他の指標との比較が容易になるという利点もある。

3.2 新規性

次に「表現がどの程度新しいと感じるか」を測るため**新規性**を導入する。「よい表現」というものを定義することは難しいが、他の表現と比較して使いたいと思わせるかどうかは1つの基準である。そしてメタファー表現を使いたいという動機が生じる場面で分かりやすいのは、まだ対象を

指す表現がないときである。例えば、不祥事の発覚をきっかけとして主にブログやSNS上でそれに対する非難が大量に投稿される状態を「炎上」と呼ぶが、このメタファー的な「炎上」の用法は、実際にそういった現象が起きるまでは存在しなかった用法である。既存の慣例的な用法が対象を記述するのに適切でないと感じた場合、新規性のあるメタファー的な用法がより適している可能性がある。文学研究では特に革新的なメタファー表現に注意が向けられる傾向がある[7]。よって詩や小説の執筆を支援する応用などを指向すると新規性は重要な指標であると考えられる。

3.3 理解可能性

第三の指標として「表現がどの程度分かりやすいか」を測る**理解可能性**を導入する。書き手がよいと感じる表現であっても、読み手にとって理解が難しければその表現をよいといえないからである。

理解可能性は特に隠喩(狭義のメタファー)において重要である。その理由として隠喩では何について述べているのかが分からない場合があることが挙げられる。直喩表現では「*S*のような*T*」という明示的に比喩を表す形式から、実際には*T*という主題について述べていることが容易に分かる。例えば「もやしのように痩せた少年」なら分かるがいきなり少年を指示するつもりで「もやしは…」と言っても相手には伝わりにくい。したがって主題*T*について述べていることを読み手に何らかの方法で伝える必要がある。これを回避する方法として「もやしのように痩せた少年」と述べたあとで「そのもやしは…」のように文脈を利用することが挙げられる。ただし評価において文脈を含めることと各表現を公平に評価することは両立が困難であるため、本研究ではこのような文脈を使った分かりやすさの改善を評価に含めない。

隠喩の伝わりやすさについては、レトリックの観点から[8]が「 Y^* によって臨時にどんな(X)^{*2}があらわされているのかということがあらかじめ相手に理解されなければならない。言いかえれば隠喩においては、(X)と Y との類似性が、語り手と聞き手の間に前もって共通化されていなければならない。」と述べている。一方でこれは認知言語学で言うところのメタファーの定義とやや矛盾している。[9]はメタファー表現には*S*と*T*の乖離([9]の用語では仮想フレームと現実フレーム)が必要であることを述べている。このことから理解可能性はメタファー性とある程度のトレードオフがあるという仮説を立てることができる。これは次節で説明する実験により確認する。

3.4 総合評価

メタファーの生成を行うシステムを考えたとき、どの指

*1 本稿における *S*。

*2 本稿における *T*。

標の評価値が高い表現が好ましいかを事前に理解している利用者は少ないと考えられる。そのため複数の指標を統合した単一の指標があることが望ましい。本研究ではその役割を果たす指標として、上で提案した3つの指標の和を用いる。これを**総合評価**と呼ぶことにする。他の方法として、平均や標準偏差を考慮して各指標を標準化した得点を利用する方法が考えられる。しかし我々はそのような方法は採らない。なぜなら、平均や標準偏差は評価値を集める際に使用する表現の選び方に大きく依存しており、かつクラウドソーシングではこれらの統計量に偏りを生じさせないほど十分に広範な表現を対象とすることは困難だからである。

4. 実験

各指標の有効性や、指標間の関係を明らかにするためクラウドソーシングで作業者を募り、いくつかのメタファー表現を評価してもらった。

4.1 評価対象

本研究では名詞と動詞をそれぞれ1つずつ組み合わせた短い表現を評価対象とする。例えば「意図を汲み取る」や「快楽に溺れる」といった表現がこれにあたる。本来は表現のよさを正確に評価するため、表現単独ではなく、その周囲の文脈を含めて評価する必要がある。しかし、本研究では評価作業を簡単にするため、また表現同士を公平な条件で比較するため、文脈は考慮しない。表現の作成には[5]が概念メタファー《感情は水》の妥当性を検討する際に用いた方法と同様に、「気持ち」など感情に関連する語を含む名詞と、水とともに使われることの多い「Xが溜まる」や「Xが湧く」といった動詞表現を組み合わせて作成した。統計的な分析がしやすいように、[5]で使われている名詞12個、動詞表現28個だけではなく、新たに名詞28個と動詞表現6個を追加した。したがって合計 $(12+28) \times (28+6) = 1360$ 表現が評価の対象となる。追加した名詞は3種類に分けられる。1つ目に「情熱」や「喜び」などの具体的な感情を表す名詞、2つ目に「音」と「声」、「光」などある種の水と似た性質をもつものを表す名詞、最後に「ネコ」や「ケーキ」など動物や固形物を表す名詞である。

メタファー表現の作成方法には、上記のような形式的な組み合わせによる方法の他に、クラウドソーシングで作業者にメタファー表現を作成してもらおう方法が考えられる。しかし、実際にそれを行ったところ、慣用的なメタファーばかりが集まってしまった結果になった。一方、上記で説明した名詞と動詞の表現を組み合わせる方法は、「不満が溜まる」や「勇気が湧く」など慣用的なものだけではなく、「希望に浸す」や「嫉妬を注ぐ」のような聞いたことがない、あるいは理解しにくい表現も生成できるため、表現間

で評価値のばらつきが大きくなることが期待できたため、形式的な組み合わせによる方法を選択した。

[5]と同じ表現を用いる別な利点として、[5]が求めた容認度（表現が認められるか、言うか言わないか）と、提案指標の比較ができることがある。理解に苦しむような表現は容認度が低いと考えられるため、我々の提案手法の1つの理解可能性と容認度にはよく似た傾向が生じることが期待される。もしこのことが確かめられたならば、クラウドソーシングにより集められた不特定に集められた作業者の評価が、[5]の大学生6名による評価と、同程度には信頼できると判断できる。

4.2 評価項目

評価項目は、メタファー性・理解可能性・新規性の3項目である。これらはすべて、1から5の5段階で評価してもらおう。なお5つの選択肢のうち、1と5のみに説明的なラベルを付与してある。以下に項目ごとの設問の例を示す。

- **メタファー性** 次の表現を比喩だと感じますか。
「感情を撒き散らす」
5. 比喩だと感じる
1. 比喩だとは感じない
- **理解可能性** 次の表現は理解しやすいですか。
「感情を撒き散らす」
5. 問題なく理解できる
1. まったく理解できない
- **新規性** 次の表現は新しい表現ですか。
「感情を撒き散らす」
5. 使ったこと、見聞きしたことがない新しい表現である
1. 広く使われている慣用的な表現である

4.3 作業者

クラウドソーシングのプラットフォームにはYahoo! JAPANクラウドソーシング^{*3}を利用した。今回、年齢や性別といった属性による作業者の制限は特に行わなかった。ただし、不正を防ぐため、指定した品詞の語を選択肢から選ぶ問題（チェック設問）を設け、これに誤った回答をした者の評価値は除外した。

設問の数は、40の名詞と34の動詞表現をそれぞれ1つずつ組み合わせてできる表現を3つの指標で評価するので計 $40 \times 34 \times 3 = 4080$ 問である。これらを9つのタスクに割り振り、1問あたり10名の回答が得られるように作業者を募集した。複数のタスクに分割したのは、作業者を集めやすくするためである。ただし同一の作業者が同一の設問に重複して回答することがないように各タスクへの応募は一

^{*3} <http://crowdsourcing.yahoo.co.jp/>

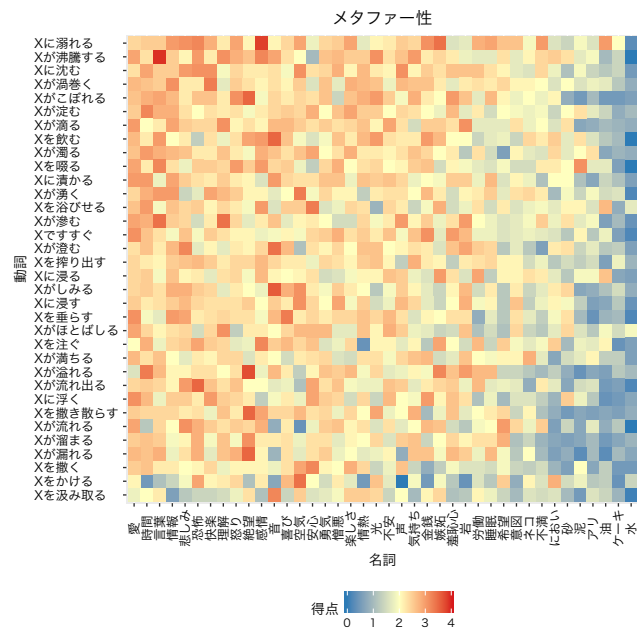


図 1: メタファー性. 名詞と動詞はそれぞれを含む表現の得点の高い順に並べてある.

度に制限した. 各作業者は 1 タスクで計 21 問 (うち 1 問はチェック設問) に回答した.

5. 評価値の分析

5.1 各指標の分析

メタファー性, 理解可能性, 新規性について, 表現ごとの評価値の平均をヒートマップ形式でそれぞれ図 1~図 3 に示す. 評価値の平均を以下では得点と呼ぶ. また, それぞれの上位と下位を表 1~表 3 に示す. 理解可能性を [5] の容認度と比較する都合上, すべての評価値は 1 から 5 ではなく 0 から 4 になるよう 1 を引いて集計している.

表 1 から, メタファー性の大きい表現には「感情に溺れる」や「感情を注ぐ」など「感情」を使った表現が多いことが分かる. これは [6] の基準でもメタファーである. 一方, 下位には「水が流れる」や「水ですすぐ」など, 水と水に関連した動詞の組み合わせが多い. また図 1 を見ると, 上位には名詞は「愛」や「時間」, 「言葉」といった抽象名詞が多く並び, 下位には「油」や「ケーキ」, 「水」といった具体的なものを指す名詞が並んでいることも分かる. これらから提案手法はメタファー性を捉えられていると考えられる.

次に理解可能性であるが, 表 2 に示すように, 上位には「水が溜まる」や「水が湧く」など「水」と水に関連する動詞の組み合わせが多い. 一方で, 下位には「感情ですすぐ」や「勇気を垂らす」といった見慣れない表現が並ぶ. このことから理解可能性が捉えられていると考えられる. さらにこれを裏付けるため理解可能性と [5] の容認度との比較を行う. 容認度は理解可能性とはやや異なるものの得点の

順位	名詞	動詞	得点
1	言葉	X が沸騰する	3.9
2	感情	X に溺れる	3.8
3	絶望	X が溢れる	3.7
4	音	X がしみる	3.6
5	恐怖	X が流れ出る	3.5
5	絶望	X がこぼれる	3.5
5	絶望	X を撒き散らす	3.5
5	絶望	X が漏れる	3.5
5	音	X を飲む	3.5
		⋮	
1351	水	X ですすぐ	0.2
1351	水	X を垂らす	0.2
1351	水	X が流れ出る	0.2
1351	水	X が湧く	0.2
1355	水	X を汲み取る	0.1
1356	声	X をかける	0.0
1356	水	X を啜る	0.0
1356	水	X が沸騰する	0.0
1356	水	X が流れる	0.0
1356	水	X を飲む	0.0

表 1: メタファー性の上位と下位.

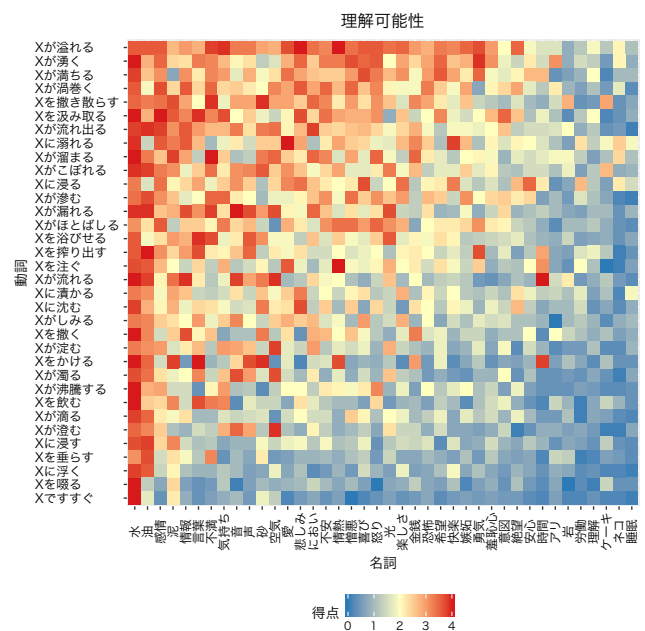


図 2: 理解可能性. 名詞と動詞はそれぞれを含む表現の得点の高い順に並べてある.

傾向は近くなるはずである. まず両者の相関係数を求めると 0.81 であった. これはきわめて高い相関があることを示す. さらに対応する表現同士の得点差の絶対値の平均は 0.64 であった. これは平均して 1 段階の差もないということである. したがってクラウドソーシングを使った我々の理解可能性の評価は, [5] の大学生による評価と同程度に

順位	名詞	動詞	得点
1	不満	Xが溜まる	4.0
1	悲しみ	Xが溢れる	4.0
1	情熱	Xを注ぐ	4.0
1	情熱	Xが溢れる	4.0
1	愛	Xに溺れる	4.0
1	感情	Xを汲み取る	4.0
1	時間	Xが流れる	4.0
1	水	Xですすぐ	4.0
1	水	Xをかける	4.0
1	水	Xを啜る	4.0
1	水	Xを撒く	4.0
1	水	Xを汲み取る	4.0
1	水	Xが沸騰する	4.0
1	水	Xが流れる	4.0
1	水	Xが湧く	4.0
1	水	Xが溜まる	4.0
1	水	Xを飲む	4.0
1	油	Xを搾り出す	4.0
1	言葉	Xをかける	4.0
1	音	Xが漏れる	4.0
		⋮	
1341	ネコ	Xを搾り出す	0.1
1341	ネコ	Xを注ぐ	0.1
1341	ネコ	Xが澄む	0.1
1341	ネコ	Xが濁る	0.1
1341	ケーキ	Xが澄む	0.1
1341	労働	Xを垂らす	0.1
1341	労働	Xが滴る	0.1
1341	労働	Xを飲む	0.1
1341	理解	Xを啜る	0.1
1341	睡眠	Xですすぐ	0.1
1341	睡眠	Xが流れ出る	0.1
1341	睡眠	Xに浮く	0.1
1341	睡眠	Xが滲む	0.1
1341	絶望	Xが澄む	0.1
1341	音	Xを啜る	0.1
1356	アリ	Xがしみる	0.0
1356	ケーキ	Xを注ぐ	0.0
1356	岩	Xを啜る	0.0
1356	怒り	Xを啜る	0.0
1356	憎悪	Xですすぐ	0.0

表 2: 理解可能性の上位と下位.

信用に足るものと結論付けられる.

最後に, 新規性を見る. 表 3 の下位を見ると, 水に関する表現だけではなく「感情を汲み取る」や「不満が溜まる」といった慣用表現が並んでいる. これはちょうど理解可能性と反対の結果になっている. このことは図 3 から分かる.

順位	名詞	動詞	得点
1	アリ	Xがしみる	4.0
1	ネコ	Xを搾り出す	4.0
1	ネコ	Xが流れ出る	4.0
1	におい	Xに浮く	4.0
1	労働	Xを啜る	4.0
1	労働	Xが滴る	4.0
1	勇氣	Xが漏れる	4.0
1	岩	Xを啜る	4.0
1	憎悪	Xですすぐ	4.0
1	時間	Xが沸騰する	4.0
1	楽しさ	Xですすぐ	4.0
1	楽しさ	Xが澄む	4.0
1	理解	Xが沸騰する	4.0
1	睡眠	Xですすぐ	4.0
1	睡眠	Xに浮く	4.0
1	言葉	Xですすぐ	4.0
1	言葉	Xが沸騰する	4.0
		⋮	
1342	不満	Xが溜まる	0.1
1342	勇氣	Xが湧く	0.1
1342	声	Xをかける	0.1
1342	愛	Xに溺れる	0.1
1342	感情	Xを汲み取る	0.1
1342	時間	Xが流れる	0.1
1342	気持ち	Xを汲み取る	0.1
1342	水	Xに浮く	0.1
1342	水	Xが滴る	0.1
1342	砂	Xを撒き散らす	0.1
1342	音	Xが漏れる	0.1
1353	情熱	Xを注ぐ	0.0
1353	時間	Xをかける	0.0
1353	水	Xをかける	0.0
1353	水	Xが沸騰する	0.0
1353	水	Xが流れる	0.0
1353	水	Xが湧く	0.0
1353	水	Xを飲む	0.0
1353	言葉	Xをかける	0.0

表 3: 新規性の上位と下位.

5.2 指標間の関係の分析

次に 3 つの指標間の関係を分析する. 相関係数は表 4 の通りである. 第 3.3 節で理解可能性とメタファー性にトレードオフの関係があるという仮説を立てたが, これらの相関係数は -0.19 であり, 強い相関は見られなかった. むしろ新規性とメタファー性の間に強い相関が見られる. これは慣用性が高いとメタファーとは感じにくいことを表していると推測できる. 例えば「情熱を注ぐ」は [6] の基準でメタファーであるし, [5] の基準でも《感情は水》に基づくメタファーとみなせるが, メタファー性は低い. 最後に, 理解可能性と新規性の相関係数を見ると -0.92 であり, 強

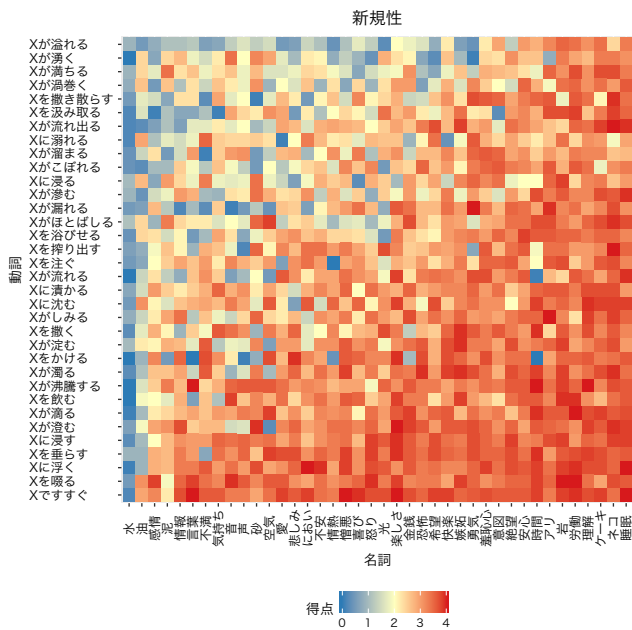


図 3: 新規性、理解可能性との対比のため、名詞と動詞は図 2 と同じ順に並べてある。

	メタファー性	理解可能性	新規性
メタファー性	1.0	-0.19	0.28
理解可能性	-0.19	1.0	-0.92
新規性	0.28	-0.92	1.0

表 4: 各指標間の相関係数。

い負の相関がある。これは理解しにくい表現は使われにくく、結果として新規性を高く感じやすいという事情を反映しているものという推測できる。ただ一方で「気持ちが沸騰する」や「悲しみが流れ出る」のようにどちらも高い表現も存在するため、理解可能性を新規性で完全に置き換えられるわけではない。

5.3 総合評価の検証

総合評価の上位と下位は表 5 の通りである。上位を見ると慣用句として定着はしていないものの、意味が理解できる表現が並ぶ。例えば「空気に沈む」の意味は、「空気」が「その場の人たちを支配する志向のあり方（雰囲気）」([I0])という語義で、また「沈む」が「気持ちのはれない状態におちこむ。」([I1])という語義で用いられており、全体として「その場の雰囲気によって落ち込む」という意味であると理解できる。また「気持ちが沸騰する」や「感情が沸騰する」は怒りや興奮を表す表現と推察できる。これは日本語に「怒りが燃え上がる」や「怒りが燻ぶる」といった怒りを火で表す表現が多く存在することや^{*4}、「燃える」に

^{*4} [5] はこれらを《怒りは火》という(概念)メタファーに基づいていると捉える。

順位	名詞	動詞	得点
1	空気	X に沈む	8.9
2	気持ち	X が沸騰する	8.8
3	恐怖	X が流れ出る	8.7
4	感情	X が沸騰する	8.6
4	羞恥心	X がこぼれる	8.6
6	怒り	X がこぼれる	8.5
6	感情	X に溺れる	8.5
6	絶望	X が溢れる	8.5
9	悲しみ	X が流れ出る	8.4
9	音	X が澄む	8.4
		⋮	
1349	時間	X をかける	4.3
1349	水	X を汲み取る	4.3
1349	泥	X が漏れる	4.3
1352	水	X を垂らす	4.2
1352	水	X が流れ出る	4.2
1352	水	X が湧く	4.2
1352	泥	X が溢れる	4.2
1356	ケーキ	X を注ぐ	4.0
1356	水	X が沸騰する	4.0
1356	水	X が流れる	4.0
1356	水	X を飲む	4.0
1360	声	X をかける	3.8

表 5: 総合評価 (メタファー性、理解可能性、新規性の和) の上位と下位。

「ある感情に動かされて、気持ちが高ぶる。」([I1])という語義があることから類推できる。一方で表 5 の下位には水に関する非メタファー表現が並ぶ傾向がある。

総合評価の有効性を確認するためには、上位の表現が実際に「よいメタファー」になっているかどうかを検証する必要がある。ここでは「よいメタファー」とは「よい表現」かつメタファーである表現と定める。また「よい表現」であることは、使いたいと感じる表現であることと定める。

表現のよさに関する検証は以下のように行った。まず総合評価のランキングに基づき、すべての表現をを上位 10%のグループと下位 90%のグループに分ける。次に各グループからランダムに表現を 1 つずつ抽出し 10 個のペアを作る。そして作業者 1 名に、より使いたいと感じる表現を各ペアから 1 つ選んでもらう。このとき表現がメタファーかどうかは考慮しない。最後に上位グループのものが、より使いたい表現に選ばれることを確認する。作業者は言語学の知識をもつ大学院生 1 名が行った。併せて、ランキングのどの程度上位までがよい表現であるかを調べるため、上位グループの境界を上位 20%, 30%, 40%, 50%に変更した場合の検証も同様に行った。結果を表 6 に示す。

上位 10%と下位 90%から抽出した 10 個のペアのうち 8 個で上位の表現がより使いたいと判断された。上位の表現

	上位	下位	選好
	不満を飲む (23)	油を汲み取る (1087)	上位
	怒りがこぼれる (6)	岩に溺れる (1117)	上位
	羞恥心が溜まる (44)	羞恥心を注ぐ (856)	上位
	情報が濁る (106)	空気を撒き散らす (212)	上位
10	悲しみがしみる (32)	90 理解が流れる (721)	上位
%	楽しさが渦巻く (81)	% 不満に漬かる (1241)	下位
	言葉が滲む (14)	恐怖が流れる (307)	下位
	感情を注ぐ (44)	意図に漬かる (654)	上位
	不安が流れ出る (44)	情熱を汲み取る (165)	上位
	情報に溺れる (23)	油が溜まる (1241)	上位
	感情に溺れる (6)	喜びが濁る (721)	下位
	声に沈む (106)	砂を浴びせる (371)	上位
	においをかける (134)	金銭を撒く (1202)	下位
	恐怖を注ぐ (165)	ネコを撒き散らす (787)	上位
20	悲しみが渦巻く (165)	80 金銭が溢れる (572)	上位
%	怒りが沸騰する (44)	% においを吸る (721)	上位
	愛が溜まる (81)	嫉妬がしみる (1009)	下位
	恐怖に浸す (212)	喜びが流れる (962)	上位
	岩が滴る (61)	言葉が淀む (257)	下位
	恐怖が沸騰する (11)	水を垂らす (1352)	上位
	勇気が溜まる (44)	楽しさが満ちる (721)	下位
	恐怖を注ぐ (165)	ネコを飲む (572)	上位
	音が濁る (257)	喜びを汲み取る (856)	上位
	感情が淀む (212)	気持ちに沈む (572)	上位
30	情熱が淀む (212)	70 砂が溢れる (1276)	上位
%	恐怖がこぼれる (165)	% 言葉に浸す (787)	上位
	恐怖を搾り出す (212)	憎悪に浮く (572)	下位
	情熱が流れ出る (307)	不満を注ぐ (914)	上位
	楽しさを垂らす (371)	愛に浮く (423)	下位
	音を飲む (106)	絶望に浸る (787)	下位
	絶望を吸る (307)	油に漬かる (1009)	上位
	安心がこぼれる (371)	言葉を搾り出す (572)	下位
	岩が滲む (307)	においが湧く (1117)	下位
	快楽に浸す (212)	声を注ぐ (1009)	上位
40	快楽が溢れる (81)	60 嫉妬が澄む (1046)	上位
%	砂が滴る (307)	% 砂が湧く (787)	上位
	情報がこぼれる (371)	気持ちを浴びせる (962)	上位
	岩を飲む (499)	金銭が沸騰する (787)	下位
	羞恥心が滴る (423)	睡眠が渦巻く (962)	上位
	理解がほとばしる (61)	喜びに沈む (962)	下位
	喜びが淀む (165)	意図に沈む (914)	上位
	情熱が満ちる (572)	音を吸る (787)	上位
	憎悪が流れる (307)	砂を吸る (914)	上位
	恐怖がこぼれる (165)	岩が淀む (1009)	上位
50	声が満ちる (371)	50 泥がしみる (1046)	上位
%	油に沈む (307)	% においに溺れる (1009)	下位
	労働が溢れる (212)	ケーキが溜まる (1276)	上位
	嫉妬が渦巻く (572)	怒りを撒く (962)	下位
	労働に漬かる (499)	嫉妬ですすぐ (1087)	下位
	声を撒く (499)	意図に浮く (1179)	上位

表 6: 総合評価とより使いたいと感じる表現（「選好」と表示）の対応、括弧内の数字は総合評価の順位である。作業にはどちらが上位でどちらが下位は分からないように提示した。

非メタファー表現	動詞の意味の理解が難しい表現
においをかける	?岩が滴る
油に沈む	?岩が滲む
岩を飲む	?砂が滴る

表 7: 表 6 の上位の表現のうちメタファーと判断されなかった表現。

がより使いたい、すなわちよりよいと判断されているため、総合評価は表現のよさを評価できていると言える。上位の範囲を 20%, 30%, 40%, 50% に変更した場合は、それぞれ 6 個、6 個、6 個、7 個が上位から選ばれており 10% の場合と大きな差はなかった。下位が選ばれた事例 17 件を分析すると、まず 4 件は下位の方に感情に関する名詞が含まれている事例が含まれていた。また 3 件は上位に「岩」を含む分かりにくい表現（例えば「岩が滴る」など）があった。これらの事例で上位を高く評価するためには名詞の具体性・抽象の評価を含める方法が考えられる。他の下位が選ばれた事例については明らかな傾向はないように見受けられるが、今後より多くの事例を分析して詳しく調べていく必要がある。

最後に表 6 中の上位の表現がメタファーであるかどうかを著者 1 名が判定した。判定の基準は、[6] のガイドラインに基づいて動詞を MRW (metaphor-related word) と判断できるならばメタファーとみなすというものである。判定作業を行った結果、表 7 に示したいくつかの表現を除きすべてメタファーであると判断された。これらの非メタファー的な表現ではいずれも具体的な名詞が使われている。したがってこれらも、下位のほうが好まれた事例と同様に、名詞の具体性・抽象性の評価を総合評価に含めることで改善されると考えられる。

表 6 で抽出された事例を全体的に見ると、全 50 事例のうち、上位の表現がより使いたいと判断された事例が 33 事例であった。また上位の表現のうちメタファーであると判断されたものは 44 事例であった。このことから総合評価は「よいメタファー」を見つけるために有効であると結論付けられる。

6. おわりに

本稿では直喩に限定しないメタファーの生成を行う上で必要となる評価指標の提案を行った。次にそれらの指標に基づいた実際の評価を、名詞 1 つと動詞 1 つから成る表現を対象として、クラウドソーシングを利用して行った。さらにその結果から各指標間の関係性を明らかにした。最後に総合評価が高い表現が、実際によりメタファー表現であることを検証した。今後は生成システムを構築し詩の生成や分かりやすい文章の生成などの応用を行う予定である。

参考文献

- [1] 北田純弥, 萩原将文: 電子辞書を用いた比喻による文章作成支援システム, 情報処理学会論文誌, Vol. 42, No. 5, pp. 1232–1241 (2001).
- [2] Abe, K., Sakamoto, K. and Nakagawa, M.: A computational model of metaphor generation process, *Proc. of the 28th Annual Meeting of the Cognitive Science Society*, pp. 937–942 (2006).
- [3] Kusumi, T.: Effects of categorical dissimilarity and affective similarity between constituent words on metaphor appreciation, *Journal of Psycholinguistic Research*, Vol. 16, No. 6, pp. 577–595 (1987).
- [4] Terai, A., Abe, K. and Nakagawa, M.: Examining an evaluation mechanism of metaphor generation with experiments and computational model simulation, *International Conference on Artificial Neural Networks*, Springer, pp. 403–410 (2012).
- [5] 鍋島弘治朗: 日本語のメタファー, くろしお出版 (2011).
- [6] Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A., Krennmayr, T. and Pasma, T.: *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*, John Benjamins Publishing (2010).
- [7] 中村 明: 比喻表現辞典, 角川書店 (1995).
- [8] 佐藤信夫: レトリック感覚, 講談社 (1992).
- [9] 鍋島弘治朗: メタファーと身体性, ひつじ書房 (2016).
- [10] 山田忠男, 柴田 武, 酒井憲二, 倉持保男, 山田明雄, 上野善道, 井島正博, 笹原宏之 (編): 新明解国語辞典第七版, 三省堂 (2011).
- [11] 小学館国語辞典編集部 (編): 精選版 日本国語大辞典, 小学館 (2006).