

比較記述テキスト中の比較表現のアノテーションと分析

飯田 諒^{1,a)} 飯田 龍^{2,b)} 徳永 健伸^{1,c)}

概要：従来の参照表現生成では、ある対象を、他の対象と区別し、曖昧性なく簡潔に指示する表現を生成することを目的に研究が進められてきた。これに対し、本研究では、複数の対象を横並びで比較し、言及する対象ごとに適切な比較対象を選び、すべての対象についてその特徴を記述するテキストの生成を目指している。このようなテキストを本研究では比較記述テキストと呼ぶ。比較記述テキストでは、対象を比較して述べる際に様々な比較の方略が採用されるが、本研究ではこの比較の種類を分類し、その分類カテゴリの情報を既に収集済みの比較記述テキスト集合へアノテーションした結果について報告する。さらに、アノテーションした内容を分析し、比較対象との関係を考慮して比較記述テキストを自動生成する際にどのような点に留意する必要があるかを調査した結果についても報告する。

1. はじめに

計算機によるテキストの生成は自然言語処理の研究の一分野として研究されてきた [10], [12]。とりわけ特定の対象を曖昧性なく簡潔な表現で他の対象と区別する参照表現の生成はテキスト生成の中心的な課題として扱われてきた [5]。参照表現生成における主な関心は、指示対象を他の対象から弁別するために、指示対象が持つ様々な属性の中からどのような属性を選び、言語化するかという点にある。このために種々の属性選択のアルゴリズムが提案されている [1], [2]。これに対して我々は複数の対象と各対象の特徴が与えられた時に、その中から聞き手が各自の規準にしたがって言及すべき対象の属性を選択し、必要に応じて対象間の差異等を比較して記述するテキストを生成することを目指している。したがって、このようなテキストでは、聞き手の選好やシステムの選好を前提とせず、聞き手が対象を比較する上で必要となる情報を、すべての対象に対して中立的に記述することになる。このようなテキストを以下、**比較記述テキスト**と呼ぶ。対象の特徴を記述するという点では、比較記述テキスト生成も既存の参照表現生成と同じであるが、その目的が他の対象からの「弁別」ではなく、他の対象との「比較」であるという点が異なっている。このため、既存の参照表現生成のアルゴリズムを適用しても、対象間の属性について差異や等価性を必要に応じて言及するような生成結果は得られないため、比較記述テキストを生成する新規の生成アルゴリズムを考案する必要がある。

比較記述テキスト生成の利用例としては、インターネット上の商品の評判サイトなどに見られる商品の比較情報などをテキスト化し、要約して情報を提示することで、ユーザが表を読み取る負荷を軽減すること等が考えられる。商品比較は表 1 に示すような表の形で提示されることが多い。このような表は各対象の各属性について属性値が与えられたデータの集合であると考えられる。一見、このような表をテキストとして言語化することは、「要約」という観点からは逆行しているように見えるかもしれない。しかし、実際に人間がこの表から情報を読み取るには、表の内容を解釈し、それぞれの商品が他の商品に比べてどのような特徴があるかを把握する必要がある。このようなデータの集合を入力として、データを解釈し、テキストを生成する研究はこれまでも行われてきた。例えば、気象情報データから天気予報士が読み上げる天気予報の原稿を生成するシステム [11] や株価の変動データをテキスト化するシステム [6] などが構築されている。本研究で対象とする生成処理は、データの集合からテキストを生成するという点では既存研究と同じであるが、既存研究は個別の場所の天気や株価について独立にテキストを生成するのに対し、本研究では対象横断的に中立的なテキストを生成する点が大きく異なる。

また、対象を説明する際に、類似した対象との相違点を説明することでわかりやすい説明ができると言われている [4], [7]。このように、他の対象と対比することで、言及したい対象を説明する方略は初期のテキスト生成でも使われてきた [8]。例えば、Direct comparison [9] という比較手法では、2つの対象を比較して、類似点と相違点を述べる戦略をとる。この際に、2つの対象をそれぞれ偏りなく記述するという点では本研究の比較記述テキストの目的と類似しているが、Direct comparison は対象が2つに限定されているため、3つ以上の対象を比較する場合の生成の戦略については言及されていない。また他の比較記述に関する

¹ 東京工業大学
Tokyo Institute of Technology

² 情報通信研究機構
National Institute of Information and Communication Technology

^{a)} iida.r.ab@m.titech.ac.jp

^{b)} ryu.iida@nict.go.jp

^{c)} take@cl.cs.titech.ac.jp

表1 実験参加者に与えられる評価表の例

属性	テレビ A	テレビ B	テレビ C	テレビ D	テレビ E	属性の説明
デザイン	★★★★★	★★	★★	★	★★★	見た目のよさ, 質感
サイズ	★★	★★	★	★★	★	省スペース性・コンパクトさ
機能性	★	★★★★★	★★★★	★★	★★	機能が充実しているか
応答性能	★★★	★	★★★★★	★★★★	★★	動画などで残像が残らないか
音質	★★	★★★	★★★	★★★★★	★	音質の良さ
画質	★★★	★	★★	★★	★★★★★	映像の鮮明さなど
操作性	★★	★★★	★★★★★	★★★★	★★	リモコンや本体の操作のしやすさ
価格	75,000	70,000	80,000	75,000	75,000	価格
満足度	★★★	★★★	★★★	★★★	★★	総合的に評価した値

研究では、注目された記述対象が他の対象とどう異なるかという比較を対象にしており [9], 比較対象すべてに対して中立的な立場でテキストを生成することについては考慮されていない。

このような背景から、我々はこれまでに表1に示すような表形式で提示された複数の商品とその属性値を対象にして人間に比較記述テキストを生成させ、その結果の分析を行った。例えば、表1については以下に示すような比較記述テキストが生成されるが、この生成されたテキストに対し、どの対象のどの属性が言及されているかをアノテーションし、その傾向を分析した [13]。

テレビ A は他の製品に比べて圧倒的にデザイン性が優れているが、機能性は劣っている。映像や音はごく標準で、価格も平均的だ。
 テレビ B は充実した機能を持つ反面、残念な画質だ。音やリモコンの操作性に対する評価は悪くなく、価格はお手頃といえるだろう。
 テレビ C は動画が良く、リモコンが非常に使いやすい。しかし映像は粗めで本体サイズが大きい。そして他と比べて価格はやや高めだ。
 テレビ D は音質が素晴らしい。だが映像への評価は普通。見た目の悪い点がマイナスポイントとしてあげられている。
 テレビ E は映像が格段に鮮明で美しい。しかし、音に関してはいま一つだ。その他の機能評価も平均よりやや低い。テレビ本体の見た目もなかなか良いが、ユーザーの総合満足度は低めだ。
 各製品とも特徴は様々だが、テレビ A, B, C, D に対する総合的な満足度はほぼ同水準となっている。ただし、テレビ E だけがやや低い結果となった。

このような比較記述テキストを一つの表に対して複数人に記述させることで、表の中の商品と属性の組み合わせで特定される項目（以降、これを**セル**と呼ぶ）がどの程度言及されやすいかを調査し、さらに回帰分析の問題として自動的に言及される割合を特定する手法を提案した。この結果、個別のセルに対する言及率が推定可能となったため、生成すべきテキストの分量に応じて言及率が高いセルから順に生成することで、基本的には言及すべきセルを決定できるようになった。ただし、この研究を行った際にアノテーションした情報には、比較記述テキスト生成の研究で

特に扱うべき「ある対象と比較される対象」に関する情報は含まれていなかったため、どのような場合に明示的に他の対象と比較して生成すべきかという点は未解決である。そこで、本研究では、これまでに収集、アノテーションした比較記述テキストを対象に、さらに明示的に比較される対象の情報や、その比較のされ方の種類等をアノテーションし、どのような場合に明示的な比較表現が導入されるかを分析する。本稿では、まず、2節でこれまでに構築した比較記述テキストコーパスの作成の方針や収集結果についての概要を示し、3節で、明示的な比較表現に対するアノテーションの仕様とアノテーション結果について報告する。さらに、4節で明示的な比較表現を含めた比較記述テキスト生成のためにどのような特徴を捉えるべきかをアノテーションした結果をもとに調査した結果について報告し、最後に5節でまとめと今後の課題について述べる。

2. 比較記述テキストコーパス

我々の先行研究 [13] では比較記述テキストをどのように自動生成すればよいかを定量的に分析するために、複数の評価表のそれぞれに対し、複数人の記述者に比較記述テキストを記述させ、その傾向を分析した。本節では比較記述テキストの収集実験および、各対象の記述で対象内の属性に限定してどの属性を用いているかの情報をアノテーションした結果の概要について説明する。

2.1 比較記述テキストの収集実験

収集実験では、表1に示すような10種類の表（以下、この表を**評価表**と呼ぶ）を作成し、20名の記述者に各評価表について各製品の特徴を比較して述べるテキストを記述させ、合計200の比較記述テキストを収集した。この比較の対象としては、特別な背景知識を必要とせずに属性等の判断が可能だと考えられる領域として家電製品を選択した。各家電製品に対する評価項目（属性）は価格.com^{*1}のレビューに利用されている項目を用いた。比較する各製品の属性値には、表1の星で記載された5段階評価、もしくは実数値が記入されている。以降、5段階の属性値をとる属性を段階評価、実数値をとる属性を絶対評価と呼ぶ。各製品タイプの属性は価格.comで使用されているものをそのまま利用したが、属性値は異なる表の特性を調べるために、5

*1 <http://kakaku.com>

種類の要因を考慮して恣意的に決定した。5種類の要因の詳細と各評価表と要因の関係については文献 [13] を参照されたい。価格.com に掲載された家電製品のうち、テレビ、掃除機、電子レンジ、(ブルーレイ・DVD) レコーダ、炊飯器の5種類を記述対象とする製品クラスとして選択し、各製品クラスに対してそれぞれ2種類の評価表を作成し、合計10種の評価表を用意した。これらの評価表のそれぞれが与えられた際に記述者が比較記述テキストを記述する場合には、以下の指示を与えてテキストを作成させた。

- 一つの評価表について生成するテキストは400字程度とする。
- 特定の製品を薦めるのではなく、中立的な立場ですべての製品について記述する。
- 属性ごとに記述するのではなく、製品ごとにその特徴を記述する。
- 満足度と価格だけで各製品を説明しない。
- 複数の製品をまとめて記述してもよい。

表1のテレビの評価表に対してある記述者が書いた比較記述テキストが1節で示したものである。

2.2 属性選択のアノテーション

2.1節で説明した実験を通じて収集したテキストに対し、テキスト中のどの部分がどの対象のどの属性に言及しているかをアノテーションする。評価表のセルに対する言及は基本的には「デザインが良い」「値段が高い」等の評価表現をともなって記述される。そこで、このアノテーションでは、まずテキスト中の評価表現をアノテーションし、そのアノテーションされた評価表現を対象にさらに「評価対象」、「評価の属性」、「評価の極性」をアノテーションする。アノテーション例を挙げると、1節で示した比較記述テキストの「テレビAは他の製品に比べて圧倒的にデザイン性が優れているが、」という記述に対して、「優れ」を評価表現として抽出し、評価対象が「A」、評価の属性が「デザイン」、評価の極性が「肯定」であるという評価される対象の属性への言及に関する情報を付与する。アノテーションには、テキストのある区間に対してセグメント情報を付与し、そこに属性情報を設定可能なアノテーションツールである Slate [3] を使用した。

2.3 属性情報アノテーション時の問題点

先行研究 [13] におけるアノテーションの際にも、比較表現を含むテキストへのアノテーションの方針に曖昧さが生じた。例えば、「レコーダA, C, Fは同一価格ですが、Fの満足度が最も高くなっています」といった記述では、商品全体に対してレコーダFの満足度が高いと述べているのではなく、レコーダA, C, Fの3商品の中でFの満足度が高いという関係を述べていると考えられる。しかし、先行研究では評価される対象自体の言及された属性に焦点を当てたため、AとCの満足度への言及に関する情報は付与しなかった。本研究ではこのような記述に対して比較表現のアノテーションを行なった。このようなアノテーションを行うことで、上記の例を元に考えると、テキスト生成の際の属性選択において、単純にFの満足度だけを選択する

のではなく、Fの満足度を記述に利用する際にAとCを比較対象として明示的に記述し、対象間の比較をより反映させた高度な内容のテキストの生成が可能となる。実際のアノテーションの指針やアノテーション結果、および結果の分析内容については3節で述べる。

2.4 属性選択に関するアノテーション結果

2.2節でアノテーションした結果を利用し、各セルの言及数について調査した。各セルに対して全記述者のうち何名が記述したかの割合(以下、**言及率**)の分布をテレビ1(表1)について調査した結果を表2に示す。表2より、言及率が1.00や0.95であり大半の記述者が言及するセルや、逆に言及率が0.00や0.05であり大半の記述者が言及しないセルがある一方で、半数程度が選択するようなセルも存在することがわかる。

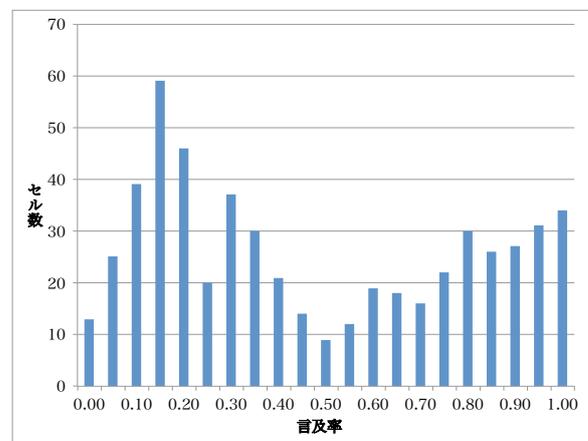


図1 言及率の分布

さらに各セルの言及率の分布を調査するために、10種の評価表の全セルに対する言及率をヒストグラムとして出力したものを図1に示す。横軸が言及率、縦軸はその言及率に該当するセルの個数を示している。この図より、言及率が0から0.15までの区間でセル数は単調増加しているものの、0.20以降で減少し、0.50から再度増加していることがわかる。この言及率の分布から、言及するか否かの判断が偏るセルが多いことがわかる。ただし、言及率が半数程度の属性も少ないわけではなく、人間の製品の特徴を述べる上での属性の選択の仕方には個人差が生じることがわかった。

表3 各属性値に対応したセルの言及率の平均

セルの属性値	1	2	3	4	5
言及率の平均	0.81	0.28	0.18	0.67	0.96

また、記述者の属性選択の傾向として、極性が否定側の属性よりも、肯定側の属性を好んで選択する傾向があることも明らかになった。表3は段階評価の各属性値における言及率の平均を示している。この表が示す通り、属性値の極性が肯定になりやすい属性値の方が言及がされやすく、人は否定寄りの評価の属性よりも肯定寄りの評価の属性を

表2 各セルと言及率の分布例 (テレビ1)

属性	テレビ A	テレビ B	テレビ C	テレビ D	テレビ E
デザイン	1.00	0.20	0.05	0.90	0.20
サイズ	0.10	0.15	0.90	0.10	0.70
機能性	0.90	1.00	0.15	0.30	0.15
応答性能	0.15	0.85	0.95	0.30	0.15
音質	0.35	0.15	0.05	0.95	0.80
画質	0.35	0.95	0.30	0.35	1.00
操作性	0.40	0.20	1.00	0.35	0.15
価格	0.30	0.95	0.95	0.40	0.20
満足度	0.45	0.35	0.45	0.30	0.70

用いて、テキストを生成する傾向にあることがわかった。ここまでで、評価される対象自体の属性選択に着目したアノテーション結果について、各セルの言及率の分布を中心に説明した。しかし、実際には属性選択の際に他対象との比較を用いたことにより、比較される対象の属性への言及が明示的に記述されていない場合がある。そのため、これまでに扱った言及率ではこの比較される対象の属性への言及を考慮できていない問題がある。そこで、本研究ではさらに正確な属性選択を実現するためには比較表現の情報をテキストに付与し、比較対象との関係を考慮した属性選択を実現することを目指す。そのために、比較記述テキストに対して比較表現のアノテーションを施し、アノテーション結果を分析することで、比較記述テキスト生成中の比較表現の利用傾向を明らかにする。

3. 比較表現のアノテーション

2節では比較記述テキストコーパス構築の第一段階として、収集したテキスト中の各セルが評価される対象として言及されるか否かをアノテーションした概要とそのアノテーション結果について述べた。ただし、1節で述べたように、比較の文脈では「Aに対してBが良い」や「A, B, Cの3つの商品の中でCが良い」のような明示的な比較をとらない評価対象について言及することで、評価対象と比較される対象についても言及することで、その対象を評価対象としてあらためて言及しないといった比較記述生成の方略(以下、比較方略)を捉えることが必要となる。そこで、本研究では、与えられた評価表内の商品、属性値の組み合わせに対して、どのように比較方略が確定し、その結果、どのように比較表現が生成されるかを捉えるために、先行研究でアノテーションした評価対象の情報に加え、比較対象の情報もアノテーションすることで、比較表現の生成がどのような場合に起こるかを分析する。そこで、まず比較表現に関するアノテーションの仕様を決定し、実際にテキストにアノテーションした。以下で、比較表現のアノテーションの仕様およびアノテーション結果について述べる。

3.1 比較表現のアノテーション仕様

これまでに作成した比較記述テキストコーパスに対して、比較表現の情報を追加アノテーションする。アノテーションには、次の3つのタグ情報を用いる。

比較対象

記述者は、言及すべき評価対象に加えて、必要に応じて比較対象を適切な比較表現とともに記述する。そこで、タグ「比較対象」を利用して、このような比較対象として選択された製品名を評価対象と関連付ける。例えば、「テレビAはテレビBに比べてデザインが良いですが」という記述では、主体的に特徴を記述されているテレビAが評価対象であり、その比較の対象として明示的に記述されているテレビBが比較対象となる。

比較の種類

比較表現は「Aに対してBは～」のような対比的な記述と「(A, B, Cの中で)Aが最も～」のような評価対象を含むグループの中で評価対象がどのようなものであるかという記述に大別できる。本研究では前者を「対比」、後者を「包含」と呼び、比較対象がどちらの比較表現で言及されるかもアノテーションする。例えば、「AはBより属性Xで優れている」という記述では、評価対象Aが比較対象Bと対比的に比較されているため「比較の種類」は「対比」となる。一方、「Aは属性Xで最も良い評価を得ている」という記述では、評価対象Aを含めた全対象のグループを考え、その中で属性Xにおいて、Aがどのような関係になるかを述べているため、「比較の種類」は「包含」となる。

評価対象と比較対象の評価の差

評価対象と比較対象が比較される際に、ある属性に関して評価対象の属性値が等価であるのか、相違であるのかを区別することが可能であり、これらの違いによって生成の実現方法が異なる可能性がある。そこで、比較表現のアノテーション時には、この属性値の等価性を「相違」もしくは「等価」というラベルでアノテーションする。例えば、「AはBより属性Xで優れていますが」では、Aの属性Xによる評価はBの属性Xによる評価より優れており、属性値間に違いがあるため「相違」をアノテーションする。一方、「AはBと属性Xで同じ評価ですが」の場合は、AとBの属性Xによる評価に差はないため、「等価」をアノテーションする。

以上の3種類のタグ情報を利用して比較表現の情報をテキストにアノテーションする。以下に比較表現を用いたテキストを示し、実際のアノテーション例を説明する。

テレビ E は、画質が非常にいいですが、音質が悪く、他の点も平均を下回っています。同じ価格の A、D に比べ、満足度は低くなっています。

上の記述において、二文目の「同じ」と「低く」の評価表現が示す属性への言及は比較表現が用いられていると捉えることができる。始めに 2 節で述べた属性選択のアノテーションにより、「同じ」には「評価対象」が「E」, 「評価の属性」が「価格」, 「評価の極性」は「中立」を情報として付与し、「低く」に対して「評価対象」が「E」, 「評価の属性」が「満足度」, 「評価の極性」が「否定」を言及の情報として付与する。次に比較表現のアノテーションを行う。一つ目の「同じ」では比較対象として「A」と「D」を選択する。これは「EはAとDと同じ価格」という意味に捉えることができることによる。さらにAとDは、Eの特徴を記述する上で比較対象として明示的に記述されており、EとA、Dの関係は対比的であると捉え、「比較の種類」として「対比」を選択する。そしてEとA、Dの価格における評価に大小関係の差異はないと考えられるため、「評価対象と比較対象の評価の差」として「等価」を選択する。二つ目の「低く」も同様に考え、「比較対象」は「A」と「D」, 「比較の種類」は「対比」, 「評価対象と比較対象の評価の差」は「相違」を選択する。次に以下のテキストを考える。

テレビ A は、最も人気のあるデザインだが、機能不足が目立ち、音質と使い勝手が多少悪く感じる。

上の記述では「人気」の評価表現が示す属性への言及は比較表現が用いられていると捉えることができる。先の例同様、始めに 2 節で述べた属性選択のアノテーションにより、「評価対象」が「A」, 「評価の属性」が「デザイン」, 「評価の極性」が「肯定」を情報として付与する。この記述ではテレビ A の特徴を説明する上で、他製品との比較を踏まえた上で最上級の表現を用いて A のデザインについて言及していると言える。「最も」のような最上級（最下級）の表現は、比較対象であるその他の製品だけでなく評価対象自体も含めたグループの中で評価対象を評価し、記述内容が決定される。よって「人気」に対する比較表現のアノテーションは、「比較対象」として「その他」, 「比較の種類」として「包含」を選択する。また「評価対象と比較対象の評価の差」であるが、評価対象の A と比較対象であるその他の製品では、A のみ優れているとしており、評価に差が生じている。よって「評価対象と比較対象の評価の差」は「相違」を選択する。

3.2 比較表現のアノテーション結果の分析

比較表現のアノテーションをしたコーパスを対象に、比較表現の使われ方の分布を調査した。ここでは比較記述テキストに記述された「A に対して B が画質が優れた」のような評価表現を単位として、比較表現の種類を計上する場合と、評価表のセルを単位として比較対象が使われたセルを計上する二通りの調査を行った。前者の場合、つまり、評価表現を対象として見ると、収集したコーパス中で評価表現は全体で 4,031 箇所に出現しており、そのうち、約 21% の評価表現が比較表現をとまうことがわかった。次

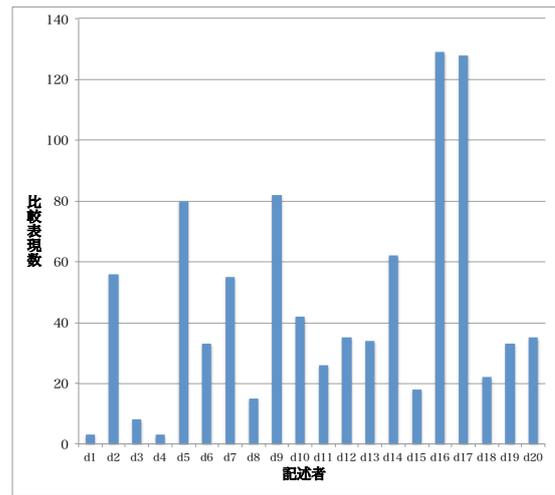


図 2 各記述者が用いた比較表現数

に、後者のセル単位での調査では、例えば、「A は B に比べて属性 X と Y で優れている」では比較される B の属性 X と Y の 2 つのセルが比較対象となるため、2 つ比較されていると計上する。本研究で採用する比較記述のアノテーションでは対比/包含、等価/相違の組み合わせで 4 通りが考えられ、それぞれがどのくらい出現したかを評価表現単位で計上したものを表 4 に、セル単位で計上したものを表 5 にまとめる。

表 4 比較表現の分布

	対比	包含
等価	66	0
相違	261	577

※評価表現単位

表 5 比較表現の分布

	対比	包含
等価	83	0
相違	303	692

※セル単位

まず、表 4 から「包含・相違（包含かつ相違の項目を指す）」の比較が最も多く存在することがわかった。「包含・相違」が最も多いのは最上級（最下級）を用いた比較表現が最も多く存在することによることが影響していると考えられる。また、「包含・等価」に該当する比較は今回構築したコーパスでは存在しなかった。「比較の種類」の観点から見ると、「対比」と「包含」はおよそ 1:2 の割合で出現していることがわかる。「評価対象と比較対象の評価の差」の観点から見ると、9 割以上が「相違」であり、比較表現を用いる際は比較対象と評価の異なる比較対象を選択する傾向が強いこともわかる。

次に記述者間で比較表現の利用傾向に差があるかを調査した。図 2 は記述者ごとに何件の比較表現を利用したかをヒストグラムで表したものである。図 2 から記述者によって差があることがわかる。各記述者の比較表現の数を平均すると 45.0 で、標準偏差は 35.4 であった。

次に評価表間で比較表現の出現回数に差があるかを調査した。図 3 は評価表ごとに何件の比較表現が出現したかをヒストグラムで表したものである。記述者の場合と比べると評価表間での大きな差はない。各評価表の比較表現の数の平均は 89.9、標準偏差は 13.5 であった。

次にセル単位の分析結果について述べる。各セルの比較表現に関する調査をすることで、各製品の特徴を説明する

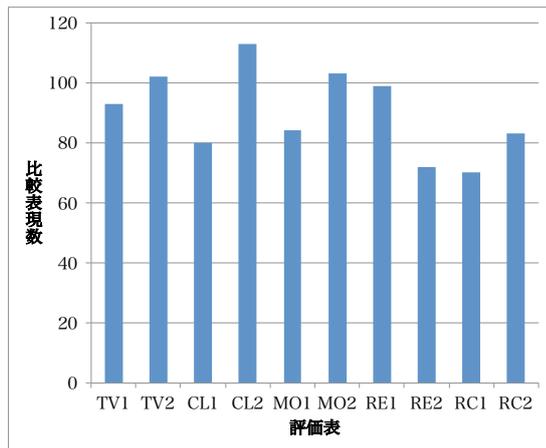


図3 各評価表で出現した比較表現数

表6 比較表現の事例数が10件以上のセル

評価表	属性	属性値	比較表現の事例数
テレビ1 (TV1)	価格	80,000	13
テレビ2 (TV2)	価格	60,000	13
レコーダ1 (RE1)	価格	35,000	12
掃除機1 (CL1)	価格	35,000	11
掃除機2 (CL2)	満足度	★★★★★	11
掃除機2 (CL2)	価格	30,000	11
掃除機2 (CL2)	取り回し	★★★★★	11
レコーダ2 (RE2)	満足度	★★★★★	11
テレビ1 (TV1)	価格	70,000	11
掃除機2 (CL2)	パワー	★★★★★	10
掃除機2 (CL2)	デザイン	★★★★★	10
掃除機2 (CL2)	サイズ	★★★★★	10
炊飯器1 (RC1)	炊き上がり	★★★★★	10
レコーダ2 (RE2)	操作性	★★★★★	10
テレビ1 (TV1)	デザイン	★★★★★	10
テレビ1 (TV1)	満足度	★★	10

際にある属性が言及されたときに、記述の戦略としてその言及上で比較表現がどの程度使われるのかを明らかにすることができる。予備調査としてどのようなセルは比較表現が用いられやすいかを分析した。表6は比較表現の事例数が10件以上のものを示す。この表からわかる傾向としては、段階評価の場合は基本的に属性値が最大の星5である場合が多い。また、属性が絶対評価である「価格」のセルが上位に集中しており、絶対評価は多少属性値に依存するものの、比較表現が利用される可能性が高いということも明らかになった。

既に述べた通り、比較表現の利用には記述者間で差がある。そこで、さらなる分析として記述者が利用した比較表現数に基づいて記述者を3つのグループに分類し、グループ間で比較表現を利用する傾向について分析した。図2に表す比較表現数に基づいて、比較表現数が80件以上のグループ、9件以下のグループ、それ以外(10件以上79件以下)のグループの3つに分類した。その結果、比較表現数が80件以上(以下「高頻度」)のグループは4名、10件以下のグループ(以下「低頻度」)は3名、それ以外(以下「中頻度」)のグループは13名となった。ここで、低頻度のグループに対しては比較表現数が極端に少なく、傾向を調査するのが困難であるため、分析対象から除外した。

表7 各グループの比較表現の利用に基づくセル数

評価表	高頻度のみ	中頻度のみ	高頻度かつ中頻度
テレビ1 (TV1)	6	1	12
テレビ2 (TV2)	14	1	22
掃除機1 (CL1)	5	3	19
掃除機2 (CL2)	1	2	14
レンジ1 (MO1)	13	2	16
レンジ2 (MO2)	6	8	20
レコーダ1 (RE1)	5	5	19
レコーダ2 (RE2)	6	9	19
炊飯器1 (RC1)	7	11	14
炊飯器2 (RC2)	9	5	18
合計	72	47	173

各グループで少なくとも1人が比較表現を用いたセルが各評価表でどの程度あるかを調査し、各グループに対してこれらのセルの分布について分析した。

表7に分析結果を示す。「高(中)頻度のみ」の列は、高(中)頻度のグループでのみ少なくとも1人が比較表現を用いたセル数を表し、「高頻度かつ中頻度」の列は、高頻度のグループ、中頻度グループの両方で少なくとも1人が比較表現を用いたセル数を表す。この表から、どの評価表でも共通して比較表現を用いるセル(「高頻度かつ中頻度」列)が一番多いことがわかる。しかし、高頻度のグループのみが比較表現を用いるセルや、逆に中頻度のグループのみが比較表現を用いるセルも存在する。評価表によって高頻度のグループのみの方が多い場合や、中頻度のグループのみの方が多い場合があり、必ずしも高頻度のグループが中頻度のグループが比較表現を用いたセルに対して比較表現を利用するわけではない。一般的に考えると高頻度のグループが中頻度のグループを包含するように思えるが、実際には中頻度のグループのみが比較表現を利用するセルも47件存在することがわかった。

そこで、この47件についてどのような傾向があるか調査した。47件のうち39件は中頻度のグループの中で1人が、7件は2人が、1件は3人が比較表現を利用した事例であり、多くは1人のみが比較を用いた事例であった。また1人のみが比較表現を用いた39件を対象に、比較の種類の観点から調査したところ、「対比・相違」が25件、「対比・等価」が8件、「包含・相違」が6件であった。この分布は表5に示した全体の比較表現の分布と比較すると、傾向が大きく異なっている。これらの調査結果から、高頻度のグループの記述者が比較表現を利用しないセルで、中頻度のグループの記述者が行う比較では、「対比」の比較が割合として最も多く、全体的な傾向と異なることが明らかになった。また、今回の調査結果では中頻度のグループのみが比較表現を利用するセルにおいて、大半が1人のみの比較の利用に留まっており、上記の調査結果を考慮すると、「対比」の比較を用いるかどうかの方略は、個人差が大きく影響する可能性も大いに考えられる。

4. 比較表現の利用傾向と比較方略の実現に向けた課題

3節では、比較表現のアノテーション方法およびアノテーション結果について述べた。本研究では比較表現のアノ

表 8 各セルの比較表現が利用される割合

属性	テレビ A	テレビ B	テレビ C	テレビ D	テレビ E
デザイン	(0.3, 0.0, 0.2, 0.0)	-	-	(0.0, 0.0, 0.06, 0.0)	-
サイズ	-	-	(0.0, 0.0, 0.06, 0.0)	-	(0.0, 0.0, 0.14, 0.0)
機能性	(0.16, 0.0, 0.11, 0.0)	(0.1, 0.0, 0.15, 0.0)	-	-	-
応答性能	-	(0.0, 0.0, 0.06, 0.0)	(0.11, 0.0, 0.32, 0.0)	-	-
音質	-	-	-	(0.05, 0.0, 0.37, 0.0)	(0.0, 0.0, 0.13, 0.0)
画質	-	(0.05, 0.0, 0.05, 0.0)	-	-	(0.1, 0.0, 0.35, 0.0)
操作性	-	-	(0.05, 0.0, 0.35, 0.0)	-	-
価格	-	(0.16, 0.0, 0.42, 0.0)	(0.16, 0.0, 0.53, 0.0)	-	-
満足度	(0.0, 0.0, 0.0, 0.0)	-	(0.0, 0.0, 0.0, 0.0)	-	(0.57, 0.0, 0.14, 0.0)

各割合は (対比・相違, 対比・等価, 包含・相違, 包含・等価) の順

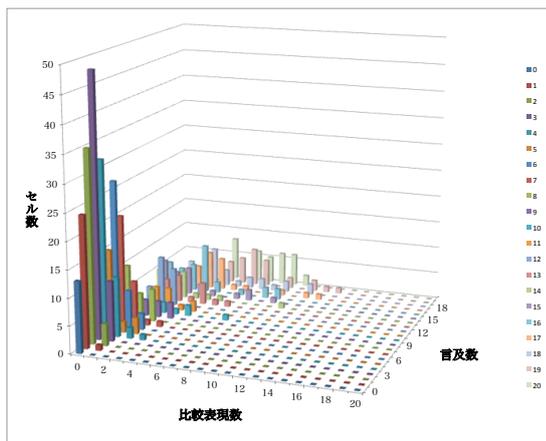


図 4 言及数と比較表現数から見たセルの分布

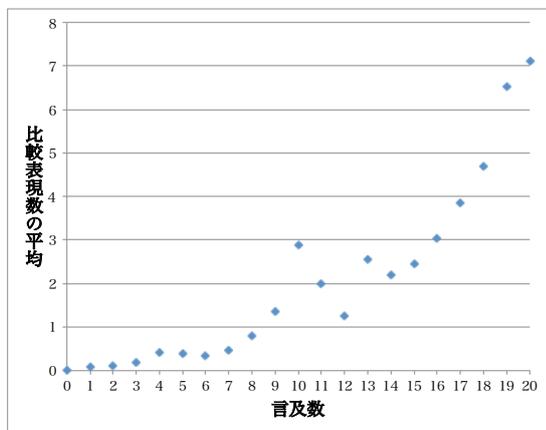


図 5 各言及数の比較表現数の平均

テーションを施したコーパスを元に、属性選択に加えて比較方略を実現したテキスト生成を目指す。先行研究 [13] では、評価対象として言及される場合の各セルの言及率を予測するという問題を扱ったが、それだけでは、あるセルを言及した上で比較表現を利用するかどうかの問題までは扱えていない。しかし、実際の比較記述テキスト生成では、あるセルを言及するだけでなく、そのセルが言及された際に他のセルも比較という文脈で言及するかどうかを扱う必要がある。このため、評価表のあるセルが言及されたときに他のセルがどのように言及されるのかを調べる必要がある。これについては様々な観点から調査が可能である

が、まずは、各セルが言及された場合にどの程度比較表現をともなって生成されたのかを調査する。まず、各セルの言及数と各言及に対して比較表現をともなう個数を求め、言及された上でどの程度比較表現が用いられるかの割合を調査する。ただし、そもそもセルに対して言及される個数が少ない場合、そのセルに対して高い割合で比較表現が使われても、そのセルが比較をともなって生成されることに関する信憑性が少なくなると考えられる。そこで、言及数とその言及の中で比較表現の個数が使われる個数の関係を図 4 に示す 3 次元グラフにプロットし関係性を調査する。図 4 では、横軸が比較表現数を、縦軸が言及数を表し、高さがその比較表現数および言及数に当てはまるセルの数を表す。図 4 の比較表現の数については、3 節に示した比較の種類の情報 (対比と包含, 等価と相違) は捨象して頻度を求めた。図 4 から言及数が増加すると比較表現の事例が増加する傾向は確かに確認できるが、言及数が多い場合でも比較表現の事例が少ないセルは数多く存在することがわかる。しかし、言及数が多い場合、例えば、言及数が 19 や 20 といったほとんどの記述者が言及するようなセルでは、約半数の記述者が比較表現を利用するセルも多く存在する。上記の結果から、各セルが言及された際に、比較が利用されるかどうかはセルによって異なるため、対象とするセルだけではなく、評価表内の他の情報がどのように与えられた場合に比較表現が多く使われやすいかといった調査が必要であることがわかった。ここで、言及された上で比較が用いられる割合を全セルに対して共通して算出して予測するのは、言及数が少ないときに比較表現が用いられていたときなどの場合に比較表現が用いられる割合が極端に高くなってしまいう問題が存在する。例えば、2 人が言及して 1 人が比較表現を用いた場合と 20 人が言及して 10 人が比較表現を用いた場合では、同じ割合 0.50 でも信頼性に差異が生じると考えられる。そこでセルに対する言及数を閾値として、ある閾値以上の回数で言及されたセルに対してのみ比較表現が利用される割合を予測する問題を考える。閾値を決めるにあたって、各言及数に該当するセルの集合での比較表現の事例数が平均していくつ見られるかを分析した。図 5 に分析結果を示す。図 5 から事例数が平均して 1 を超えるのは、言及数 9 以上のセルであることがわかる。そこで、あるセルが言及された際に比較表現が利用される割合を予測する問題を解く上で対象とするセルを 9

以上の言及が観測されたセルに限定する。ここで言及数が9以上に限定した場合は、合計で990件（セル単位）の比較表現が観測され、全体の比較表現の数が1,078であることを踏まえると、9割以上を網羅している。表8は言及数が9以上のセルで比較される割合を各セルの値として導出した分布図である。セルの値が「-」であるものは言及数が9未満であることを示す。本研究ではこれらの値を評価表から導出可能な素性から予測する問題を解く。また、既存研究で属性選択の予測問題でモデルを開発する際に利用した回帰分析を、今回の比較表現の予測モデルでも利用する。比較表現のための回帰モデルを開発する上で次の4つの素性を導入した。

属性値の唯一性 (F1)

各セルの属性値が属性内の観点で唯一の属性値かどうかを表し、0と1の二値で表現される値である。例えば、表1のデザインにおけるAとDとEはそれぞれ★, ★★, ★★★★★で唯一の属性値であるため1を取り、BとCは★★で唯一の属性値ではないため0を取る。

属性内で最大もしくは最小の属性値 (F2)

各セルの属性値が属性内の観点で最大もしくは最小の属性値かどうかを表し、0と1の二値で表現される値である。例えば、表1のデザインにおけるAとDは1を取り、B, C, Eは0を取る。

対象の属性値と属性内の属性値の平均からの差 (F3)

属性内で属性値の平均を計算し、各セルの属性値と差を取った値である。例えば、表1の(A, デザイン)のセルの場合、デザインの属性値の平均を計算し、(A, デザイン)の属性値との差を取った値が該当する。

対象の属性値と製品内の属性値の平均からの差 (F4)

製品内で属性値の平均を計算し、各セルの属性値と差を取った値である。例えば、表1の(A, デザイン)のセルの場合、Aの属性値の平均を計算し、(A, デザイン)の属性値との差を取った値が該当する。

各素性と各分類の比較表現が用いられる割合にどの程度関係性があるのか分析した。関係性があるかどうかの判断材料として相関係数を用いた。上記の素性を考える上では、比較表現を利用するのは、他製品と比較したときに目立つセルの場合であるという考えに基づき、F1, F2, F3を設定した。また製品内で目立つ場合はとりわけ他の製品と比較する際に利用する傾向にあるという考えに基づきF4を設定した。またF3とF4を計算する上で、段階評価と実質評価を同等に扱うために、正規化を評価表に対して適用した。正規化についての詳細は文献[13]を参照されたい。分析結果を表9に示す。包含・等価の比較表現に対しては、事例が検出されなかったため分析対象から外してある。

表9 各分類の比較表現の割合と各素性の相関

素性	対比・相違	対比・等価	包含・相違
F1	0.23	-0.36	0.48
F2	-0.04	0.00	-0.10
F3	0.18	-0.13	0.59
F4	-0.025	-0.06	0.23

表9から最も相関があるのは、素性F3と包含・相違の比較表現の割合であることがわかる。また、F1と包含・相違の比較表現の割合も相関があるが、全体を通して素性F2はどの分類の比較表現が利用される割合とも相関があまりない。これらの結果が示すように、現段階の素性は包含・相違の比較表現に対してのみ有効なものが多く、対比の比較表現の割合にはほとんど有効でないことがわかる。ただ「対比・相違」と「対比・等価」の比較表現は全体のおよそ36%を占めており、「包含」だけを特定できても、全体の64%しか適切に比較表現を生成できないことになる。この問題に対処するためには、「対比」の比較に対して適切な素性を導入することが必要となるため、今後の課題として各分類ごとに適した素性の開発が挙げられる。さらに「対比」の比較に関しては比較される対象の選択が重要である。特に、評価対象と比較する対象としてなぜある対象を選択するのかといった要因には、セル間の情報に限定して導入した素性だけではなく、対象間のよりグローバルな情報から導入した素性が大きく関係してくると考えられる。現段階では比較表現を利用するかどうかに限った分析であり、今後はこの比較対象をどのように選択するかについても詳細に分析していきたい。

5. おわりに

本稿では、比較方略を実現した比較記述テキストの生成を目的として、収集した比較記述テキスト中の比較表現のアノテーションの仕様およびアノテーション結果について説明した。先行研究で構築した比較記述テキストでは、評価対象に対して比較される対象に関する情報をアノテーションしていなかったため、本研究で新たに比較表現のアノテーションを実施した。アノテーションでは、比較される対象、比較のされ方等の3つのタグを設定し、記述者が利用する比較表現について4つのカテゴリーに分類した。

アノテーション結果を複数の観点から報告し、明示的な比較表現を含めた比較記述テキストの生成に向け、比較表現の利用傾向について分析した結果についても報告した。比較表現のアノテーション結果としては、記述者の間で比較表現の利用に差異が生じることや、あるセルを言及したときに約21%の割合で比較表現が利用されることも明らかとなった。

比較記述テキスト生成で比較方略を実現するための第一段階となる分析では、今回設定した比較の4つの分類に対して比較表現が用いられる割合を各セルごとに分析し、各分類でその割合を予測する問題について調査した。比較表現が利用される割合を予測する上で、言及数を閾値として、ある一定数以上の言及数を得たセルを対象とすることで、基本的な属性選択を考慮し、なおかつ信頼度の増す比較表現が用いられる割合を予測した。この問題に取り組む上で、評価表の情報から導出可能な4つの素性を考え、各分類の比較表現が用いられる割合との相関を調査した。その結果、「包含・相違」の比較表現に対して、「対象の属性値と属性内の評価値の平均との差」および「属性内で対象のセルの属性値が唯一であるか」の2つが高い相関がある

ことがわかった。しかし、「対比・相違」と「対比・等価」の比較表現に対して有効な素性が不足していることも判明し、今後の課題として対比の比較に対して有効な新しい素性を導出することが望まれる。

本研究では、各セルに対して言及された上で比較される割合といったアノテーションされた比較表現の全体の傾向は分析したが、個別の事象については調査できていないため、今後調べる必要がある。また、現段階の比較記述テキストコーパス構築に用いた評価表の属性はほとんどが段階評価であり、絶対評価に関する傾向を詳しく分析できていない。そこで、絶対評価の属性値を多く含む場合についても新規に比較記述テキストを記述させ、その結果に対してこれまでと同様のアノテーションを行うことが考えられる。これについては、既に新しい比較記述テキストの収集実験を実施しており、今後はこの実験で収集したテキストをもとに絶対評価に関する生成がどのように行われるかを調査する予定である。このような相対評価や絶対評価の生成の傾向、個別事象の詳しい分析を通じて、頑健な比較記述テキストの生成モデルを構築する予定である。

参考文献

- [1] Dale, R.: Cooking up referring expressions, *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, pp. 68–75 (1989).
- [2] Dale, R. and Reiter, E.: Computational interpretation of the Gricean maxims in the generation of referring expressions, *Cognitive Science*, Vol. 19, No. 2, pp. 233–263 (1995).
- [3] Kaplan, D., Iida, R., Nishina, K. and Tokunaga, T.: Slate – A tool for creating and maintaining annotated corpora, *Journal for Language Technology and Computational Linguistics*, Vol. 26, No. 2, pp. 89–101 (2012).
- [4] Karasimos, A. and Isard, A.: Multi-lingual evaluation of a natural language generation system, *Proceedings of 4th International Conference on Language Resources and Evaluation (LREC 2004)*, pp. 829–832 (2004).
- [5] Kraahmer, E. and van Deemter, K.: Computational generation of referring expressions: A survey, *Computational Linguistics*, Vol. 38, No. 1, pp. 173–218 (2012).
- [6] Kukich, K.: Design of a knowledge-based report generator, *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics (ACL 1983)*, pp. 145–150 (1983).
- [7] Marge, M., Isard, A. and Moore, J.: Creation of a new domain and evaluation of comparison generation in a natural language generation system, *Proceedings of the 5th International Natural Language Generation Conference (INLG 2008)C*, pp. 169–172 (2008).
- [8] McKeown, K.: *Text Generation*, Cambridge University Press (1985).
- [9] Milosavljevic, M. and Dale, R.: Strategies for comparison in encyclopedia descriptions, *Proceedings of 8th International Natural Language Generation Workshop (INLG 1996)*, pp. 161–170 (1996).
- [10] Reiter, E. and Dale, R.: Building applied natural language generation systems, *Natural Language Engineering*, Vol. 3, No. 1, pp. 57–87 (online), DOI: doi:10.1017/S1351324997001502 (1997).
- [11] Sripada, S. G., Reiter, E., Hunter, J. and Yu, J.: Exploiting a parallel TEXT-DATA corpus, *Proceedings of the Corpus Linguistics 2003 conference*, pp. 734–743 (2003).
- [12] Uszkoreit, H.: Language generation, *Survey of the State of the*

Art in Human Language Technology, Cambridge University Press, chapter 4 (2010).

- [13] 飯田 諒, 飯田 龍, 徳永健伸: 比較記述テキスト生成のための属性選択, 研究報告自然言語処理 (NL), Vol. 2014-NL-218, No. 8, pp. 1–7 (2014).