

製品の比較記述コーパスの構築

飯田 諒 飯田 龍 徳永 健伸

東京工業大学大学院情報理工学研究所

iida.r.ab@m.titech.ac.jp, {take,ryu-i}@cl.cs.titech.ac.jp

1 はじめに

近年, インターネット上の情報を様々な観点から分析する研究がおこなわれており, その結果を利用者に簡潔に提示する要求が高まっている. たとえば利用者に商品を推薦するシステムは, 利用者の嗜好に合致する商品を選び, その根拠とともに商品を推薦する (Bobadilla et al., 2013). 根拠の説明には商品の特徴, 利用者と同じ嗜好を持つ人の評価, 他の商品に対する利用者の嗜好などが利用される. このような推薦システムの出力を複数の商品の特徴を横並びで比較したテキストで伝えることができれば, 利用者に対する情報提示の選択肢の幅はさらに広がると考えられる. このように複数の対象を比較するために, 各対象の特徴をわかりやすく記述したテキストをここでは「比較記述テキスト」と呼ぶ. 我々は, 簡潔で利用者の意志決定に有用な比較記述テキストを生成することを目指している.

コンピュータ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おり (Reiter and Dale, 1997; Uszkoreit, 2010), とりわけ指示対象を他の対象と区別する参照表現の生成はその中心的な課題として研究されてきた (Krahmer and van Deemter, 2012). 与えられた状況において人間が使う参照表現を分析するために, 様々な参照表現コーパスが作成されている (Byron and Fosler-Lussier, 2006; van der Sluis et al., 2008; Viethen and Dale, 2008; Stoia et al., 2008; Gargett et al., 2010; Gatt and Belz, 2010; Spanger et al., 2012). 参照表現生成が特定のひとつの指示対象を他から区別する表現の生成を目的としているのに対し, 比較記述テキストの生成は, すべての対象を比較するのに十分な情報を読み手に伝えることを目的としている. 参照表現生成のアルゴリズムによって個々の対象を他と弁別するための特徴的な属性を選択することは可能であるが, それらの属性が複数の対象を比較する上で十分な情報となりうるかは明らかではない.

ある対象について説明する際に, それと類似した物と比較し, 相異点を説明するとわかりやすい説

明ができると言われている (Karasimos and Isard, 2004; Marge et al., 2008). このような「比較」方略は初期のテキスト生成の研究でも注目され, 使われてきた (McKeown, 1985). Milosavljevic and Dale (1996) は2種類の百科事典から動物に関する項目を収集し, それを分析することによって対象記述における比較を Direct comparison, Clarificatory comparison, Illustrative comparison の3種類に分類している. Direct comparison は2つの対象を比較してその類似点, 相異点を述べるもの, Clarificatory comparison は対象を類似した別の物と混同しないように違いを明確にするためのもの, Illustrative comparison は対象を同じクラスの他の物と区別する特徴について述べるものである. 我々の目的は複数の対象を比較するための情報を読み手に提供することなので, 記述の対象が特定の対象に偏った Clarificatory comparison や Illustrative comparison のような比較は好ましくない. 一方, Direct comparison は2つの対象を偏りなく記述するという点では我々の目的に合致するが, 従来, 比較対象は2つに限られており, 3以上の対象を比較するテキストの生成については考慮されていない. また「比較」方略を用いたテキスト生成に関する最近の研究では Clarificatory comparison や Illustrative comparison に主な関心が向けられており, そのほとんどは2つの対象の比較に限定されている (Milosavljevic, 1997; Dale et al., 1998; O'Donnell et al., 2001; Stede, 2002; Androutsopoulos et al., 2007; Isard, 2007).

以上の背景をふまえ, 本稿では3以上の対象について特定の対象に偏ることなく比較記述を生成することを目的とし, そのための基礎データとして現在構築中の比較記述テキストを収集したコーパスについて述べる. 特に, 比較記述において各記述対象のどのような特徴に言及するかという, 属性選択の問題に焦点をあて, このためのアノテーションをおこなう. 以下, 比較記述テキストの収集実験 (2節), テキストへのアノテーション結果 (3節) およびその分析 (4節) について述べる. 最後に5節でまとめと今後の課題について述べる.

表 1: 実験参加者に与えられる評価表の例

特徴	テレビ A	テレビ B	テレビ C	テレビ D	テレビ E	特徴の説明
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

2 比較記述テキストの収集

比較記述テキストを収集するために実験をおこなった。記述の対象としては、誰もが特別な知識なしで扱える領域として家電製品を取り上げた。各家電製品に対する評価項目(属性)として価格.com¹のレビューに利用されている項目を用い、比較する各製品の各項目について表 1 に示すような 5 段階評価を付与した評価表を用意する。テレビ、洗濯機、電子レンジ、DVD レコーダ、炊飯器の 5 種類の製品クラスを選び、各クラスについて 2 つの表を用意した。また、表の作成にあたっては以下の 5 つの要因を考慮して 10 種類の表を作成した。

【製品数】 比較する製品数の違いによって比較記述に違いがでるかどうかを調べるために、製品数が 5, 6, 7 個の 3 種類の表を用意した。表の数はそれぞれ 3, 3, 4 の合計 10 である。

【評価値の偏り】 評価表の各セルの値は 1 から 5 個の星の数で表現されている。特定の属性を特に際立たせるために、星 4 つあるいは星 2 つの評価を避けて表を作成した。これにより星 5 つの属性の顕現性が高まることが期待できる。星 4 つを使わない表を 2 つ、星 4 つと星 2 つを使わない表を 2 つ用意した。

【各製品に特有な属性の有無】 各製品が他の製品にない特定の属性の評価値を持つかどうかで 2 通りのパターンを考えた。たとえば、表 1 において各テレビはそれぞれ異なる属性で唯一の星 5 つの評価を受けている。このような状況で参照表現生成のアルゴリズムを利用すると各テレビの特徴的な属性だけに言及することになる。比較記述において、人間が参照表現生成とは異なる方略で比較記述テキストを生成することを確かめるために、この要因を導入した。特徴的な属性を含む表を 4 つ、含まない表を 6 つ用意した。

【属性間の相関の有無】 互いに依存関係のある属性間の評価値に相関を持たせるかどうかで、それぞれ 5 つの表を用意した。

¹<http://kakaku.com>

【属性間の相関における反例の有無】 前述した属性間の相関に関して一部相関しない例外的な評価値を入れるかどうかの要因である。前述の相関のある評価表 5 つに対して、反例のある場合を 3 つ、ない場合を 2 つ用意した。

実験参加者にはこの 10 個の評価表と以下の指示を与えて各表について 1 つの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。
- 特定の製品を勧めるのではなく、中立的な立場ですべての製品について記述する。
- 特徴を基本として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。

現時点で 3 名の実験参加者から合計 30 の比較記述テキストを収集しており、以下ではこの 30 テキストについて分析をおこなう。

3 アノテーション

今回構築するコーパスは比較記述における人間の属性選択の傾向を明らかにすることを目的としているため、収集した 30 のテキストに対し、まず、著者のうち 2 名がテキスト中で言及されている各製品の属性を独立にアノテーションした。アノテーションは表 1 と同一形式の表のセルに「言及された」印をつけることによっておこなった。たとえば「テレビ A はデザイン性にはすぐれていますが、機能があまり充実していません。画質や音質は平均的です。」という記述に対して、テレビ A の列の属性「デザイン」、「機能・メニュー」、「画質」、「音質」のセルを言及されたことを示す印で埋める。

30 テキストに対する 2 人のアノテータの表を比較すると、いずれかのアノテータが印をつけたセルの総数は 784、そのうち両者が印を付けたものが 741 であった。したがって、一致率は 0.945 と非常に高く、安定したアノテーションとなっている。

4 コーパス分析

二人のアノテータの間で不一致が生じたセルの数は43事例あり、そこからアノテータの不注意により言及を見落したり、言及されていないのに印をつけた場合を除くと、約半数の24事例が残る。これらの事例を分析すると以下の4種類に分類できることがわかった。以下、各分類について例を用いて説明し、アノテーションの方針について述べる。

【1. 他との比較による属性への言及】

例:「炊飯器DはBと同じ価格ですが…」において炊飯器Bの価格についても言及しているか?

分類1は特定の製品の属性について記述する際に、他の製品を比較対象として使う場合である。我々の当面の関心は製品を記述する際の属性選択であり、この例では炊飯器Dの記述において炊飯器Dのどの属性を選択するかが重要となる。したがって、炊飯器Bの価格についての言及は認めない。

【2. 属性に対する暗黙的な言及】

例:「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。」においてレコーダAとCの満足度についても言及しているか?

分類2は、集合の中の特定の製品についてのみある属性に言及することによって、その他の製品のその属性が暗黙的に示唆される場合である。この場合もこれがレコーダFに関する記述であることを考えると分類1と同様にレコーダA, Cの満足度への言及は認めない。なお、この例の前半部分「レコーダA, C, Fは同一価格ですが」は分類1に相当することに注意されたい。

【3. 複数の属性をまとめた言及】

例:「(掃除機)Dも大型ですが、運転音が静かです。それ以外は平均的な評価になっています。」において「それ以外」がサイズ、静音性以外の属性すべてに言及しているか?

分類1, 2では、他の複数の製品が属性の記述に関わっているのに対し、分類3は複数の属性をまとめて言及している。この例のように複数の属性を「それ以外」のような表現でまとめて言及することが多い。記述している対象についての属性に言及しているという点では分類1や2とは異なるが、明示的に個別の属性に言及していないという理由で、「それ以外」のような表現で指される属性については言及を認めない。

【4. 言及している属性が曖昧な表現】

例:「(ブルーレイ・DVDレコーダの)機能も豊富ですが」において「機能」が録画機能と編集機能のどちらを指すのか?

分類4は表現の曖昧性に起因するものなので、アノテーションの方針を明確にすることによって不一致を防ぐことは難しい。本稿で対象としているテキストについては、テキストの元となる評価表を参照することによって一部、曖昧性が解消できる場合もある。

現在構築しているコーパスの目的が製品の比較記述をする上で、各製品の属性選択がどのようにおこなわれるかを分析するためであると考え、分類1から3の事例については、いずれも今回のアノテーションでは属性への言及を認めない方針をとった。つまり、我々の当面の関心は、一般的なテキスト生成の枠組という「内容の決定」に相当する処理であると言える(Reiter and Dale, 1997)。

しかしながら、我々の最終的な目的は実際に比較記述テキストを生成することであり、分類1から3の事例で使われている表現を生成する必要がある。このためには、現在のように属性に言及したかどうかという二値的なアノテーションに加え、どのような表現・方略を使って属性に言及したかという情報もアノテーションする必要がある。たとえば、分類1の例において表層的には炊飯器Bの価格が言及されているが、これは炊飯器Dの価格に言及するために比較対象として派生的に言及されたと考えられることができる。現在は炊飯器Bの価格への言及は認めない方針を取っているが、将来的にはこれを派生的な属性言及としてアノテーションし、さらに炊飯器Dの価格への言及と「対比」などの関係によって関係づけるようなアノテーションが必要となる。このような属性言及間の関係も含めたより高次のアノテーションは、テキスト生成において「内容の決定」の次に必要な「談話プランニング」の段階で必要となる情報を提供するものになると考えられる。

5 おわりに

本稿では、比較記述テキスト生成のための基礎データとして我々が構築している比較記述コーパスについて述べた。まず、テキスト収集実験について説明し、実験によって収集したテキストに2名でアノテーションをおこなった。2名のアノテーション結果を比較し、不一致の原因となる表現を分析し、4種類に分類した。さらに各分類についてアノテーションの不一致を解消するための方針を与えた。今回は製品の属性への言及があるかないかという二値素性によってアノテーションをおこなったが、実際にどのような表現・方略によって属性への言及がなされるかも合わせてアノテーションする必要性がわかった。今後は、さらに比較記述テキストを収集するとともに、現在の属性への言及の有無に加え、

言及の表現・方略に関するより高次のアノテーションもおこなう予定である。

1 節でも述べたとおり、参照表現の生成アルゴリズムを用いて各製品を他の製品と弁別するのに必要な属性について述べる記述を生成することができる。たとえば、弁別に最低限の属性のみに言及し、できるだけ簡潔な記述を生成する Full Brevity アルゴリズム (Dale, 1989) を使うと、表 1 からは、テレビ A ではデザイン、テレビ B では機能性、テレビ C では応答性能もしくは操作性、テレビ D では音質、テレビ E では画質が選ばれることになる。一方、表 1 に対して今回の実験で得られたテキストには、「テレビ A はデザイン性にはすぐれていますが、機能があまり充実していません。画質や音質は平均的です。」のような冗長な記述が見られた。他の製品と弁別するという観点からはテレビ A についてデザインについて言及すれば十分であるが、その他にも機能性、画質、音質の属性にも言及している。参照表現生成の研究でも人間が冗長な参照表現を使用する傾向にあることが報告されており (Arts et al., 2011)、参照表現で導入される冗長性と比較記述で選択される属性との比較分析も興味深い課題である。

参考文献

- Ion Androutsopoulos, Jon Oberlander, and Vangelis Karkaletsis. 2007. Source authoring for multilingual generation of personalised object descriptions. *Natural Language Engineering*, 13(3):191–233, 9.
- Anja Arts, Alfons Maes, Leo Noordman, and Carel Jansen. 2011. Overspecification facilitates object identification. *Journal of Pragmatics*, 43(1):361–374.
- Jesus Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. 2013. Recommender systems survey. *Knowledge-Based Systems*, 46:109–132.
- Donna K. Byron and Eric Fosler-Lussier. 2006. The OSU Quake 2004 corpus of two-party situated problem-solving dialogs. In *Proceedings of the 15th Language Resources and Evaluation Conference (LREC 2006)*.
- Robert Dale, Stephen J. Green, Maria Milosavljevic, Cécile Paris, Cornelia Verspoor, and Sandra Williams. 1998. The realities of generating natural language from database. In *Proceedings of the 11th Australian Joint Conference on Artificial Intelligence*.
- Robert Dale. 1989. Cooking up referring expressions. In *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, pages 68–75.
- Andrew Gargett, Konstantina Garoufi, Alexander Koller, and Kristina Striegnitz. 2010. The give-2 corpus of giving instructions in virtual environments. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC 2010)*, pages 2401–2406.
- Albert Gatt and Anja Belz. 2010. Introducing shared tasks to NLG: the TUNA shared task evaluation challenges. In Emiel Kraahmer and Mariët Theune, editors, *Empirical methods in natural language generation*, volume LNAI 5790, pages 264–293. Springer-Verlag.
- Amy Isard. 2007. Choosing the best comparison under the circumstances. In *Proceedings of the International Workshop on Personalization Enhanced Access to Cultural Heritage*.
- Athanasios Karasimos and Amy Isard. 2004. Multilingual evaluation of a natural language generation system. In *Proceedings of 4th International Conference on Language Resources and Evaluation (LREC 2004)*, pages 829–832.
- Emiel Kraahmer and Kees van Deemter. 2012. Computational generation of referring expressions: A survey. *Computational Linguistics*, 38(1):173–218.
- Matthew Marge, Amy Isard, and Johanna Moore. 2008. Creation of a new domain and evaluation of comparison generation in a natural language generation system. In *Proceedings of the 5th International Natural Language Generation Conference (INLG 2008)C*, pages 169–172.
- Kathleen McKeown. 1985. *Text Generation*. Cambridge University Press.
- Maria Milosavljevic and Robert Dale. 1996. Strategies for comparison in encyclopedia descriptions. In *Proceedings of 8th International Natural Language Generation Workshop (INLG 1996)*, pages 161–170.
- Maria Milosavljevic. 1997. Content selection in comparison generation. In *Proceedings of the 6th European Workshop on Natural Language Generation (ENLG 1997)*, pages 72–81.
- Mick O'Donnell, Chris Mellish, Jon Oberlander, and Alistair Knott. 2001. Ilex: an architecture for a dynamic hypertext generation system. *Natural Language Engineering*, 7(3):225–250.
- Ehud Reiter and Robert Dale. 1997. Building applied natural language generation systems. *Natural Language Engineering*, 3(1):57–87.
- Philipp Spanger, Masaaki Yasuhara, Ryu Iida, Takenobu Tokunaga, Asuka Terai, and Naoko Kuriyama. 2012. REX-J: Japanese referring expression corpus of situated dialogs. *Language Resources and Evaluation*, 46(3):461–491.
- Manfred Stede. 2002. Polibox: Generating descriptions, comparisons, and recommendations from a database. In *Proceedings of The 19th International Conference on Computational Linguistics (Coling 2002)*, volume Project note, pages 1–5.
- Laura Stoia, Darla Magdalene Shockley, Donna K. Byron, and Eric Fosler-Lussier. 2008. SCARE: A situated corpus with annotated referring expressions. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, pages 28–30.
- Hans Uszkoreit. 2010. Language generation. In *Survey of the State of the Art in Human Language Technology*, chapter 4. Cambridge University Press.
- Ielka van der Sluis, Paul Piwek, Albert Gatt, and Adrian Bangerter. 2008. Towards a balanced corpus of multimodal referring expressions in dialogue. In *Proceedings of the Symposium on Multimodal Output Generation (MOG 2008)*.
- Jette Viethen and Robert Dale. 2008. The use of spatial relations in referring expression generation. In *Proceedings of 5th International Natural Language Generation Conference*, pages 59–67.