注釈事例参照を用いた複数注釈者による評判情報コーパスの作成

宮崎 林太郎[†]・森 辰則^{††}

本稿では評判情報関連タスクにおいて必要不可欠と考えられる，評判情報コーパスを人手により効率良く作成する手法について検討し，作成されたコーパスについて基礎的な分析を行う．まず，注釈付けに用いる評判情報モデルとして，項目－属性－属性値－評価の4つ組からなる2層構造モデルを提案する．次に，複数注釈者の人手によるコーパス作成について検討する．その際に，注釈者間の注釈揺れが問題となる．予備実験の結果，注釈者が他の注釈者と相談をせずに独自に注釈付けの判断を行った場合には注釈付けの一致率が十分でないことがわかった．そこで，複数の注釈者間で判断に関する情報を共有するための方法として，注釈事例参照の利用を提案し，注釈事例参照を組み込んだ注釈付け支援ツールの試作を行った．これにより，注釈付けの判断に関する情報を複数の注釈者間で緩やかに共有することができる．評価実験によれば，注釈事例の参照機能が注釈揺れ削減に効果があることがわかった．さらに，上記の手法を用いた評判情報コーパス作成について報告する．また，注釈事例参照の有効性を確認した後，1万文のレビュー文書に対して10名の注釈者が注釈付けを行い，評判情報コーパスを作成した．そして，作成したコーパスについて，評判情報の各構成要素の統計的調査を行った結果，提案した2層構造モデルを用いて評判情報を捉えることが有効であることがわかった．

キーワード：評判，コーパス，注釈付け支援

Creation of Sentiment Corpus by Multiple Annotators with an Annotation Tool that has a Function of Referring Example Annotations

RINTARO MIYAZAKI[†] and TATSUNORI MORI^{††}

In this paper, we investigated an effective way of creation of a sentiment corpus by using manual annotation. Sentiment corpora are regarded as indispensable resource in related tasks of sentiment information processing. First, we proposed a two-layered model of structure of sentiment information, each of which is a tuple of four kinds of elements, i.e., ‘item’, ‘attribute’, ‘value’, and ‘evaluation’. Second, We investigated corpus annotation by multiple annotators. In this situation, a corpus to be annotated is divided into multiple parts, and they are assigned to multiple annotators. Result of preliminary experiment shows that agreement of annotation among multiple annotators is not enough when annotators perform annotation individually. We

[†] 横浜国立大学大学院環境情報学府, Graduate School of Environment and Information Sciences, Yokohama National University

^{††} 横浜国立大学大学院環境情報研究院, Graduate School of Environment and Information Sciences, Yokohama National University

proposed utilization of example annotations in the corpus of manual annotation, and a tool for supporting manual annotation based on example annotation. The proposed tool offers annotators with a way to moderately share the criterion of annotation by referring existing annotations that have been already performed. Our experimental result shows that referring example annotations is effective to control discrepancy in annotation. Third, we reported an experiment of creating a sentiment corpus by using the tool. After confirming the effectiveness, ten annotators annotated a corpus of review text that contains of 10,000 sentences in order to create a corpus of sentiment information. According to a statistical investigation of each element in the annotated corpus, the proposed two-layered model is effective to analyze sentiment information in text more precisely.

Key Words: *sentiment, corpus, annotation with supporting tool*

1 はじめに

近年、Web を介したユーザの情報流通が盛んになっている。それに伴い、CGM (Consumer Generated Media) が広く利用されるようになってきている。

CGM のひとつである口コミサイトには個人のユーザから寄せられた大量のレビューが蓄積されている。その中には製品の仕様や数値情報等の客観的な情報に加え、組織や個人に対する評判や、製品またはサービスに関する評判等のレビューの著者による主観的な見解が多く含まれている。また、Weblog も CGM のひとつである。Weblog にはその時々に関心を持っている事柄についての記述が存在し、その中には評判情報も多数存在している。

これらの Web 上の情報源から、評判情報を抽出し、収集することができれば、ユーザはある対象に関する特徴や評価を容易に知ることができ、商品の購入を検討する際に意思決定支援が可能になる。また、製品を販売する企業にとっても商品開発や企業活動などに消費者の生の声を反映させることができ、消費者・企業の双方にとって、有益であると考えられる。そのため、この考えに沿って、文書中から筆者の主観的な記述を抽出し、解析する試みが行われている。

本研究の目的は評判情報抽出タスクに関する研究を推進するにあたって、必要不可欠と考えられる評判情報コーパスを効率的に、かつ精度良く作成すると共に、テキストに現れる評判情報をより精密に捉えることにある。

既存研究においても、機械学習手法における学習データや評価データに評判情報コーパスが利用されているが、そのほとんどが独自に作成された物であるために共有されることがなく、コーパスの質に言及しているものは少ない。また、コーパスの作成過程においても評価表現辞書を作成支援に用いるなど、あらかじめ用意された知識を用いているものが多い。

本研究においては「注釈者への指示が十分であれば注釈付けについて高い一致が見られる」

という仮説が最初に存在した。その仮説を検証するため、注釈者へ作業前の指示を行った場合の注釈揺れの分析と注釈揺れの調査を行う。4.2 節で述べるように、注釈者間の注釈付けの一致率が十分では無いと判断されたが、注釈揺れの主要な原因の一つとして省略された要素の存在があることがわかった。そのため、省略されている要素を注釈者が補完しながら注釈付けを行うことで注釈付けの一致率を向上できるという仮説を立てた。4.4 節で述べるように、この仮説を検証するために行った実験から、省略の補完という手法は、ある程度効果があるものの、十分に有用であったとはいえないという結果が得られた。

そこで、たくさんの注釈事例の中から、当該文と類似する事例を検索し提示することが、注釈揺れの削減に効果があるのではないかと仮説を立てた。この仮説に基づき、注釈事例の参照を行いながら注釈付けが可能なツールを試作した。ツールを用いて、注釈事例を参照した場合には、注釈事例を参照しない場合に比べて、高い一致率で注釈付けを行うことが出来ると期待される。

また、評判情報のモデルについて、既存研究においては製品の様態と評価を混在した状態で扱っており、評価対象—属性—評価値の3つ組等で評判情報を捉えていた。本研究では、同一の様態に対してレビュアーにより評価が異なる場合にも評判情報を正確に捉えるために、製品の様態と評価を分離して扱うことを考える。そのために、項目—属性—属性値—評価の4つの構成要素からなる評判情報モデルを提案する。

なお、本研究で作成する評判情報コーパスの利用目的は次の3つである。

- 評判情報を構成要素に分けて考え、機械学習手法にて自動抽出するための学習データを作成する
- 属性—属性値を表す様態と、その評価の出現を統計的に調査する
- 将来的には抽出した評判情報の構成要素の組において、必ずしも評価が明示されていない場合にも、評価極性の自動推定を目指す

上記の手法により10名の注釈者が作成した1万文のコーパスについて、注釈付けされた部分を統計的に分析し、提案した評判情報モデルの特徴について実例により確認する。また、提案モデルを用いることでより正確に評判を捉えられることを示す。

2 関連研究

乾ら(乾, 奥村 2006)は「評価を記述するもの」や「賛否の表明」は意見の下位分類に含まれるとしている。本研究でも、意見の一部として評判が存在すると考える。意見は主に主観的言明全般を指しているが、物事に対する肯定や否定を表明する評判はその中的一部分と考えるからである。

評判情報に関連する研究は、レビュー中の意見記述部分や評判情報記述部分を特定する問題

を扱う研究と、記述されている評判が肯定的か否定的かの極性を判断する問題を扱う研究の二つに大きく分かれる。

評判情報記述部分を特定する問題に関連する研究の一つに、文書中のある程度まとまった範囲での意見性を判定するものと、文単位での意見性判定を行うものがある。それとは別に、評判情報記述部分を特定する問題に対するアプローチの中には、評判情報の構成要素を定義し、各構成要素組を抽出しようとするものがある。

文単位での意見性判定においては Yu et al. (Yu and Hatzivassiloglou 2003) や Pang et al. (Pang and Lee 2004) が評価文書中の事実文と意見文を分け、意見文を抽出している。峠ら (峠 他 2005) においても文単位で意見性の判定を行っている。この研究では、文中に現れる単語が意見文になりやすい単語であるか否かを学習し Web 掲示板から意見文の抽出を行っている。さらに、日本語における評判情報抽出に関する先駆けの研究でもある立石ら (立石 他 2001) の研究は、あらかじめ用意された評価表現辞書を用いて、対象物と評価表現を含む一定の範囲を、意見として抽出している。

評判情報の構成要素を定義し、各構成要素組を抽出しようとする研究においては、村野ら (村野, 佐藤 2003) が評価文の文型パターンを整理し、その構成要素を“対象”“比較対象”“評価”“項目”“様態”としている。また、各構成要素毎に辞書を用意することで抽出を行っている。Kobayashi et al. (Kobayashi et al. 2007) では評判情報を“Opinion holder”, “Subject”, “Part”, “Attribute”, “Evaluation”, “Condition”, “Support” からなるものとし、対象とその属性・評価の関係を抽出している。また、構成要素組の同定に着目した研究として、飯田ら (飯田 他 2005) は評判情報における属性—属性値の組同定問題を、照応解析における照応詞—先行詞の組同定問題に類似した問題と捉え、トーナメント手法を用いて属性—属性値対を同定している。

次に、記述されている意見の極性を判断する研究について述べる。極性を判断する単位についても文書単位、文単位など様々な範囲を対象とした研究が行われている。Turney (Turney 2002) は文書中に含まれる評価表現の出現比率から評価文書全体の評価極性を求めている。Pang et al. (Pang et al. 2002) も文書単位での極性判断を行っており、これには機械学習手法を用いている。

文単位の極性判断としては Yu et al. (Yu and Hatzivassiloglou 2003) が⁵, Turney (Turney 2002) と同様の手法で評価文を肯定、否定、中立に分類している。

上記の研究に関連して、評判情報に関するコーパスの必要性に着目した研究も行われている。日本語の評判情報コーパスに関係する研究としては、小林ら (小林 他 2006) が、あらかじめ辞書引きにより評価値候補を与えた上での意見タグ付きコーパスの作成を行っている。この研究では、評価値候補に対する注釈者の判断はある程度一致したが、関係や根拠に対する判断の揺れは無視できるものではないことを報告している。また、Kaji et al. (Kaji and Kitsuregawa 2006) では箇条書き、表、定型文などを用いて HTML 文書から評価文コーパスの自動構築を行っている。

また、意見分析に関するコーパス研究においては Wiebe et al. (Wiebe et al. 2002) が MPQA コーパスを作成している。これは、新聞記事に対して、主観的な表現とその意見主が注釈付けされた大規模なコーパスである。Seki et al. (Seki et al. 2007) は日本語、中国語、英語の新聞記事に対し、文単位で意見文かどうかを注釈付けしている。また、意見文に対してその極性や意見主を注釈付けしている。

既存研究のコーパスと本研究で作成するコーパスの違いについて述べる。MPQA コーパス (Wiebe et al. 2002) においては、主観表現に対して注釈付けを行っており、この点が本研究で作成するコーパスと関連がある。評判情報を正確に捉えるには対象の項目がどのような様態かを表している表層表現に対して主観表現であることに限らず注釈付けを行う必要があるが、MPQA コーパスでは対応していない。一方で、本研究で作成したコーパスでは客観的表現にも注釈付けを行っている。次に Seki ら (Seki et al. 2007) のコーパスとの比較だが、このコーパスでは文単位で意見性の有無を注釈付けしている。評判情報の構成要素の抽出を行うためには文単位ではなく構成要素単位での注釈付けが必要であるが、これには対応していない。一方で、本研究で作成しているコーパスでは対応している。

また、Kaji ら (Kaji and Kitsuregawa 2006) のコーパスは文書源として Web データを用いている点では、本研究で作成したコーパスと同じだが、注釈付けの単位は文単位であり、評判情報の構成要素の単位で注釈付けはされていない。小林ら (小林 他 2006) のコーパスも文書源は Web データである。しかし、同論文に公表されている情報によれば注釈付けされている評価値は一語単位である。また、評価値として主観的な表現のみを注釈付けしており、客観的な値は評価値として扱っておらず、評価値に付随する根拠として扱っている。本稿では主観的な値に加え、客観的な値も属性値の一部として扱っている。小林ら (小林 他 2006) では客観的な値には注釈付けがされているわけではないため、この点が異なる。さらに、様態を表す表現も全て評価値として扱われている上、評価対象間の階層構造の注釈付けは必要性を認めながらも扱っていない。これに対して、本研究のコーパスでは様態を表す属性値と評価を分離している。さらに、評価対象間の関係についてはオントロジー情報として記述を行う。

加えて、本研究のコーパスでは、長い表現や客観的な表現に対しても注釈付けを行っている。製品の評価に対して一語単位でのみ注釈付けを行うと、「好き」「嫌い」「良い」「悪い」などの特定の表現のみに注釈付けが行われてしまう。しかし、レビュアーがレビューを記述する際には様々な表現を用いて肯定否定を明示している。また、製品の様態についても、レビュアーは様々な様態に着目してレビューを記述している。このため、一語単位でのみ注釈を行う場合、レビュアーが着目した製品の様態を誤って注釈付けしたり、注釈付けを落としてしまうことがあると考えられる。さらに、これらの長い表現は、表現の一部の語のみに注釈付けを行っても正しい注釈にならないため、長い表現を注釈付けする必要があると考えた。長い表現に注釈付けを行う必要性がある例を図 1 に示す。

値段は8000～9000 円くらいだったので買わなければ良かった。

図 1 注釈付け対象となる長い表層表現の例

図 1 の例における下線部は一語だけに注釈付けを行うことはできないが、これらは製品の様態を表す値段や、評価であり評判情報の一部となる。

注釈付け支援に関する研究では、野口ら (野口 他 2008) がセグメント間の関係を注釈付けするためのアノテーションツール SLAT を作成している。また、洪ら (洪, 白井 2005) は対話コーパス作成の際に、事例参照を注釈付け作業の支援として用いており、有効性を報告している。他にも、翻訳の分野においては翻訳メモリと呼ばれる原文と訳文のデータベースを用いた翻訳支援が行われている。これも、ある種の事例参照と言える。本研究では、Kaji ら (Kaji and Kitsuregawa 2006) や小林ら (小林 他 2006) のコーパスのように注釈を自動的に付与せずに、注釈事例の参照を用いた評判情報コーパスの作成を行う。注釈者が注釈付けを行う際に、あらかじめ注釈付けがなされている状態で対象の文を見てもらうと、あらかじめ付与された注釈が判断の前提となってしまうことが予想される。あらかじめ付与された注釈が注釈者の判断に与える影響に関する実験については今後の課題とし本稿では行わない。本研究では注釈者が判断に迷う場合に、事例を参照しつつも独自の判断の下で注釈付けを行ってもらうというアプローチを採用した。

また、評判情報のモデル化についても、従来研究においては製品の様態と評価が混在するモデルがその多くであった。本研究では、様態に対する評価がレビュアーによって異なる場合を正確に表現するために、これを分離するモデルを提案する。

本研究では製品全体やその部分について以下の 2 種類を様態として扱う。

- 具体的な値
例：30 cm, 2 G
- 他の製品との比較から得られる値など主観的であっても、特にその表層表現のみからでは肯定的や否定的といった評価が一意に決まらないもの
例：速い, 静か, 明るい

一方、以下に示す例のように、肯定や否定といった極性が陽に記述されている部分を評価として扱う。

- 陽に記述されている極性表現
例：大満足である, いいです, ちょっと不満

3 評判情報の提案モデル

3.1 4つ組による評判情報モデル

本節では、本研究で提案する評判情報モデルについて説明する。提案モデルでは、評判情報は製品やサービスに対する個人の見解やその製品がどのようなものであるかが述べられたものであるとし、4つの構成要素から成るものとする。構成要素の各項を以下に示す。

項目 製品やサービスを構成する要素を意味する概念クラスやそのインスタンス

属性 項目の様態を表す観点

属性値 属性に対する様態の内容

評価 項目に対する極性を明示している主観的見解

提案モデルに沿った評判情報の解析例を例文1に示す。

例文1：この製品_{項目} は価格_{属性} が安い_{属性値} のが魅力です_{評価}。

提案モデルでは（属性、属性値）の組で表現される項目の様態が属する層と、評価の属する層との2層構造とすることで、従来の研究（小林 他 2005）では混同されることが多かった対象となる項目の様態と話者の評価を分離することができる。これにより、例文2、例文3のように同一の様態に対して異なる極性の評価が記されている場合にも、評判情報を正確に捉えることが出来るようになる。

例文2：このフィギュア_{項目} はフォルム_{属性} を忠実に再現している_{属性値} のがお気に入りです_{評価}。

例文3：この製品_{項目} はフォルム_{属性} を忠実に再現している_{属性値} ために魅力が薄い_{評価}。

本研究で提案するモデルにおける評価は、主観的な表現の中でもその表現のみでいかなる文脈でも変わらない極性を示す表現である。文脈に依存して異なる極性を暗示する表現は評価としない。提案モデルの構造を図2に示す。図2では製品の部分全体関係とその様態を表す層と、製品やその部分に関する評価の属する層を示している。この2層の関係は、様態を表す層に属する属性—属性値の組を評価の属する層にある評価の理由とし、様態を表す層に属する製品やその部分を対象に評価が記述することで表される。このように、製品の様態と評価を分離した2層構造にすることで、同一の属性—属性値の組からレビュアーにより異なる評価が与えられている場合には異なる評価への関連付けが可能となっている。提案モデルでは、肯定・否定は評価によってのみ決定され、属性—属性値の組についてはそれが主観性を含むものであっても、製品の様態のみを表し、評価は含まれていないとして扱う。

また、様態を表す層では、製品の項目が部分—全体関係や上位—下位関係のような階層構造を持ち、それぞれが属性、属性値を持つことを表している。これは必ずしも階層構造の葉の部分のみが属性、属性値を持つわけではなく、上位の項目の属性—属性値の組について記述され


```

[c01]<キッチン・ホーム>
+[c141]<圧力鍋>
++[c02]<この圧力鍋>
+++[c931]<付録>
+++[c1201]<蓋><ふた>

```

図 3 オントロジー情報の記述例

なお、オントロジーに関する情報は文中に付与するタグとは別に記述する方法を提案する。記述されたオントロジー情報の例を図 3 に示す。

オントロジー情報はオントロジーの木構造に現れる各ノードを先順で記述してあり、行頭にある + は深さを、その後ろの [] の中が各階層の概念番号を表している。この概念番号は item タグを注釈付けする際の class 属性の値と対応している。< > の中は表層表現であり、同一概念や指示物を表す表層表現は同一の概念番号を付与している。なお、提案モデル中ではレビューで着目している製品に対応する概念が起点になっており、それより上の側には概念の上位—下位関係が、そしてそれよりも下側には概念の全体—部分関係が記されている。図 3 の例では「この圧力鍋」が起点となっている。この方法により注釈者毎にオントロジー情報を作成してもらうことを想定している。また、各注釈者が作成したオントロジー情報を 1 つに統合することは想定していない。

また、このモデルにおいては評価の理由となるのが属性—属性値の組であると考え。先行研究のモデルの多くでは、評価と属性値とが区別無く扱われていた。本研究ではこれらを別々に扱う。また、一般的には主観的な表層表現と評価には密接な関係がある。しかし、主観的な表層表現が必ずしも極性を一意に決定するわけではない。つまり、評価ではない。提案モデルでは肯定・否定・中立の判断にかかわる表層表現を評価と呼び、主観的な表現であっても評価の理由となる様態を表現しているものを属性—属性値と呼ぶ。例を以下に示す。

例文 4: ディスプレイ項目 の画質属性 もきれい属性値 でいいです評価 ね。

なお、構成要素の各項は表層表現が文中で省略されている場合がある。特に、属性—属性値の組については片方が省略されている場合が多く存在する。例文を次に示す。

例文 5: このカメラ項目 は小さく属性値 で気に入っています評価。

例文 5 では属性値「小さい」が単独で出現している。これは本来ならば属性として現れるべき「大きさ」や「サイズ」といった表現が省略されているものと解釈する。

3.2 評判情報を注釈付けするためのタグセット

本研究では、前節で述べたモデルの項目を item、属性を attribute、属性値を value、評価を evaluation として XML のタグセットを作成した。作成したタグセットを図 4 に示す。item タグは対象となる製品やその部分を表す表層表現に付与する。同様に attribute タグ、value タグ、evaluation タグはそれぞれ、属性、属性値、評価を表す表層表現に付与する。図 4 には注釈付与のための XML タグの要素名と属性情報を示してある。なお、各タグにはそれぞれ一意に決まる識別子を id 属性として付与する。attribute タグと value タグは片方が省略されている場合を考え、組ごとに一意に決まる識別子をそれぞれの pair 属性の値として付与する。evaluation タ

```

<item>：項目
  id：この項目に対する一意な番号
  class：この項目が属する概念番号

<attribute>：属性
  id：この属性に対する一意な番号
  pair：属性—属性値の組番号
  target：属性—属性値を持つ項目の概念番号

<value>：属性値
  id：この属性値に対する一意な番号
  pair：属性—属性値の組番号
  target：属性—属性値を持つ項目の概念番号

<evaluation>：評価
  id：この評価に対する一意な番号
  target：評価の対象となる項目の概念番号
  reason：評価理由となる属性—属性値の組番号
  orientation：評価が肯定・中立・否定のどれか

例：<item id="i01" class="c01">この製品</item>は<attribute id="a01"
  pair="p01" target="c01">価格</attribute>が<value id="v01" pair="p01"
  target="c01">安い</value>ので<evaluation id="e01" reason="p01"
  target="c01" orientation="positive">好きです</evaluation>。

```

図 4 評判情報構成要素のタグセット

グにはその評価の理由となる attribute-value の pair 属性の値を reason 属性として付与する。また、評価の極性を positive, neutral, negative の 3 値で orientation 属性として付与する。

4 予備実験

本節では予備実験の内容について述べる。本研究の目的は人手による評判情報コーパスの作成である。大規模なコーパスの作成、将来におけるコーパスの拡張を考えた場合に、複数注釈者による注釈付け作業の並列化が不可欠であると考えた。

複数注釈者による注釈付け作業を行う際には、注釈揺れが問題となる。本研究では、2 章で述べたように注釈付け対象となる文書に対して、事前に機械的に注釈付けすることを想定していない。注釈者に先入観を与えずに判断をしてもらう必要があると考えたためである。さらに、あらかじめ用意された辞書等の知識を注釈付けに用いるのならば、その知識を直接抽出に用いればよいと考えたためである。本研究では、特定の表現にとらわれず、注釈者には様々な表層表現に対して注釈付けを行ってもらう。

ここで本稿では、注釈者には事前に指示を行い、それが十分ならば複数の注釈者が独立して注釈付けを行っても、注釈付けにおいて高い一致が見られるという、以下の仮説 1 を立てた。

仮説 1：注釈者への事前の指示が十分ならば、複数の注釈者が独立して注釈付けを行っても注釈付けの高い一致が見られる。

上記の仮説 1 が成立するなら、コーパスを作成するために複数の注釈者に個別に仕事を割り振り、並列作業を行うだけで十分であると言える。仮説 1 を確認するため、予備実験として複数注釈者による注釈付け作業を行い、注釈者間の注釈付けの一致率を調査した。

4.1 予備実験 1 の方法

本稿では注釈付けの対象として Amazon¹からレビュー文書を収集した。本稿の前提としては、あらゆる製品に対してレビューがあればその全てに注釈付けを行いたいと考えている。そのため、Amazon のトップページの製品分類を元に、なるべくそれらを網羅することを考えた。

コーパスに用いるテキストを Amazon から収集した時点²では Amazon のトップページにある製品分類は「本、ミュージック、DVD」「家電、エレクトロニクス」「コンピュータ、ソフトウェア」「ホーム&キッチン」「おもちゃ、ゲーム、キッズ」「スポーツ、アウトドア」「ヘルス、ビューティー」の 7 種類であった。

これらに属する製品は、製品の主要部分が有形物であるものと、無形物であるものに分類で

¹ <http://www.amazon.co.jp>

² 2007 年 2 月

きるが、これら二つに分類される製品は互いに構造が異なるため、レビューに表れる表現が異なる。有形物のレビューにおいては、物理的な特徴に関する記述が多くなると考えられる。一方、無形物に対してはその機能や物語に対する特徴が記述される場合が多い。これに関連して、Turney ら (Turney 2002) は、映画のレビュー分析において、映画は「出来事や俳優といった映画の要素」と、「様式や美術といった映画の形態」という二つの様相を持つ特徴があるため、自動車などのレビューと異なる特徴があると述べている。

Amazon の製品分類において我々が有形物と分類したものは「家電、エレクトロニクス」「ホーム&キッチン」「おもちゃ、キッズ」「スポーツ、アウトドア」「ヘルス、ビューティー」「コンピュータ」である。また、「本、ミュージック、DVD」「ソフトウェア」「ゲーム」は無形物とした。「ゲーム」を無形物としたのは、レビューの対象がボードゲームのように形を持つものではなくゲームソフトだったためである。

次に、有形物の中から、複数の機能を有しそれを実現するために自動的に動作することが多い電化製品と、自ら動作することはなく人間に道具として使われる非電化製品の2つのジャンルにさらに分類した。Amazon の製品分類の中では「家電、エレクトロニクス」に加えて「コンピュータ、ソフトウェア」の中の「コンピュータ」を電化製品として分類した。残りの「ホーム&キッチン」「おもちゃ、キッズ」「スポーツ、アウトドア」「ヘルス、ビューティー」を非電化製品とした。本稿ではこれら非電化製品の中から「ホーム&キッチン」「おもちゃ、キッズ」を注釈付けの対象とした。

無形物に対しては、ユーザが視聴、閲覧といった受動的な関わり方をするだけの映像・音楽と、操作等の能動的な関わり方をするソフトウェアの2つのジャンルにさらに分類した。実際の製品分類の中からは、「ミュージック、DVD」を映像・音楽として分類した。また、「コンピュータ、ソフトウェア」の中の「ソフトウェア」と「おもちゃ、ゲーム、キッズ」の中の「ゲーム」をソフトウェアとして分類した。「本」に関しては、事前の調査において、レビュー文章中に本の中の記述が引用されている場合が存在することが確認された。こうしたレビューについて、本の引用部分は評判情報の注釈付けの対象外であるので、引用部分に引用であることを示す別の注釈づけを行い、レビューの本文と分離して管理をする必要がある。しかし、本稿の目的はレビュー文に対する注釈付けであり、引用については本稿の目的の範囲外であると考え除外した。

上記の各ジャンルについて 50 文、計 200 文を実験毎に用意し、条件を変えて次に示す予備実験 1 を行った。なお item に関しては、製品名やその部分の呼称、照応表現などに限られているため、あらかじめ別の注釈者により注釈付けを行った状態で、それ以外のタグの注釈作業を行ってもらった。同時に、item 間の関係を参照できるように item タグを注釈付けした際に作成されたオントロジー情報を各注釈者に配布した。

予備実験 1 item タグが既に付与されている文書に対して、評判情報の構成要素（すなわち、attribute, value, evaluation）を注釈者に注釈付けしてもらった。また、注釈付けの対象

となるレビュー文書はあらかじめ文単位に分けて XML ファイル化し、注釈付けを行ってもらった。オントロジー情報は各注釈者が同じものを参照している。注釈付けの際に特別なツールは利用せず、テキストエディタのみを使用してもらった。注釈を行った被験者は情報工学を専攻する学生 5 名であり、2 名は本研究との直接の関係を持つ。

また、予備実験 1 の際に注釈者に行った作業指示を付録 A に示す。

4.2 予備実験 1 の結果

予備実験において、attribute, value, evaluation の 3 要素について注釈者間の注釈付けの一致率を調べた。次に示す 3 つの場合を、注釈者間の注釈付けが一致したものと判定した。

- 完全一致
- 部分文字列一致

● 一方では attribute と value を分けているが、他方では両方を一つの value としている場合
最初に、完全一致である。これは同一箇所の全く同一の文字列に対して同一のタグが付与されている場合である。次に、部分文字列一致である。これは完全一致ではないが、タグが付与されている両文字列に共通部分が存在する場合である。予備実験では、形態素区切りや文節区切りを明確にして注釈付けを行ったわけではないために、タグの範囲についてはわずかに異なってしまう場合がある。このような状況を許容して一致とみなすものである。

完全一致と部分文字列一致の例を図 5 に示す。図 5 における部分文字列一致の例では、「強さ」の attribute に対する value が「頑強」であるという点ではどちらの注釈者も注釈付けが一致している。しかし、程度を表す表現「それなりに」の部分をタグに含めているかどうかという点が異なっている。

最後に、一方では attribute と value を分けているが、他方では両方を一つの value としている場合がある。図 6 に示すように、attribute タグを付与するかどうか異なっているが value タグを付与した表層表現は部分一致している。これは、value タグの注釈付け範囲の細かさの違いから起こるものである。例を図 6 に示す。

図 6 の場合、注釈者 A の注釈付けが、我々が意図していた注釈付けである。この例について考えると、注釈者 B が省略されていると考えた attribute は「操作のしやすさ」である。また、

完全一致：注釈者 A:<evaluation>優れています</evaluation>
注釈者 B:<evaluation>優れています</evaluation>
部分文字列一致：注釈者 A:<value>それなりに頑強</value>
注釈者 B:それなりに<value>頑強</value>

図 5 完全一致と部分文字列一致の例

注釈者 A は「操作」の表層部分が「操作しやすさ」を意味する attribute であると認識している。この点を考えると、図 6 の場合は 2 人の注釈者が「しやすい」という点では同じ認識をしていると考えられる。また、図 6 の揺れは「サ変名詞+する」の組み合わせである点が揺れの理由として挙げられる。その間を区切るかどうかは注釈者によって分かれている。しかし、注釈揺れに対する注釈事例参照の効果の確認実験ではこの点については注釈者に指示をしていなかった。本稿では、このような場合は value について部分一致しているものとして扱った。

また、一致率の評価尺度には κ 値を用いた。 κ 値は主観が入る判定が偶然に拠らず一致する割合であり、0.41 から 0.60 の間ならば中程度の一致、0.80 を超えるとほぼ完璧な一致と考えられる。

$$\kappa \text{ 値} = \frac{(\text{二人の注釈者が同じ判断をしている割合}) - (\text{偶然に一致すると期待される割合})}{(1 - \text{偶然に一致すると期待される割合})} \quad (1)$$

予備実験 1 における注釈付けの一致率を調べた結果を表 1 に示す。表 1 において注釈者 2 名の組（注釈者は全部で 4 名なので全 6 組）の中で最も高い κ 値（上段）と低い κ 値（下段）を示してある。

一致度の高い注釈者の組では κ 値で 0.6 以上の値となっている部分も多数あり、ある程度の一致率となっていると考えられる。しかし、一致率の低い注釈者間では κ 値は 0.3 以下となっ

てしまっている。これは仮説 1 が成り立たなかった事を意味する。

次に、予備実験 1 で明らかになった注釈付けの揺れについて、それらが生じる場合にどのようなものがあるのかをその数とともに調査した。まず、予備実験 1 において注釈付けを行った 200 文の内、5 人の注釈者全員の注釈付けが一致した文は 42 文、4 人の注釈者のみ一致して 1 人

注釈者 B: <value>操作しにくい</value>

注釈者 A: <attribute>操作</attribute><value>しにくい</value>

図 6 value の中に attribute を含んでしまっている例

表 1 予備実験 1 の注釈付けの一致率 (κ 値)

		電化製品	非電化製品	映像・音楽	ソフトウェア
attribute	最大	0.73	0.83	0.50	0.75
	最小	0.46	0.29	0.05	0.40
value	最大	0.93	0.67	0.57	0.64
	最小	0.42	0.19	0.12	0.30
evaluation	最大	0.65	0.46	0.65	0.59
	最小	0.31	0.23	0.31	0.26

一致しなかった文は 42 文であった。以上についてはほぼ一致していると考え、残りの 116 文中に存在した注釈揺れについて調査した。

同一箇所の一文字列に対する注釈揺れに注目すると、a) 注釈者により同一箇所の同一文字列に対する注釈付けの有無が異なる注釈揺れが存在する文が 91 文、b) 同一箇所の同一文字列に対して異なるタグを付与されているために注釈揺れが存在する文が 52 文存在した。この 52 文の内、一方の注釈者が value タグを付与している箇所について、別の注釈者は evaluation タグを付与している文が 47 文存在した。

また、注釈付けを行う範囲が異なり、場合によってはタグも異なるために注釈揺れとなっている文が 17 文存在した。この 17 文の内、c) 注釈付けされた文字列の最後の文節については同じ注釈付けがなされているものが 9 文、d) 注釈者により注釈付けの粒度が異なるために揺れが複数タグにまたがるものが 8 文であった。

加えて、e) 一方の注釈者が evaluation タグを付与している箇所について、別の注釈者はこれを二つわけて attribute と value の組として捉え、attribute タグならびに value タグを付与しているために注釈揺れとなっている文が 4 文存在した。なお、同一文中に複数の揺れがあり得るので、合計数は注釈付けが一致していないとした 116 文を超えている。それぞれの注釈揺れの例を図 7 に示す。

なお、我々のモデルでは図 7 においては、注釈者 A のように注釈付けすることを意図している。

図 7 の「a) 同一箇所の同一文字列に対する注釈付けの有無が異なる揺れの例」について述べる。この例において、本来ならば「声」に attribute を注釈付けし、「機械的」に value を注釈付けするのが正しいが、注釈者 B は注釈付けを行っていない。注釈者が注釈付けを行う部分が存在すると判断するかどうか揺れている。

次に、「b) 同一箇所の同一文字列に対する注釈が異なる揺れの例」について述べる。ここでは、同一箇所の同一文字列に対して value と evaluation という異なるタグが付与されることで注釈が揺れている。この場合には本来ならば、「可愛い」という表現は必ずしも肯定的に使われるとは限らないので、value タグを注釈付けするのが正しいが、注釈者 B は evaluation タグを付与している。

「c) 末尾部分が同じ注釈付けをされている例」の場合には、注釈者 A と注釈者 B は共に「すばらしい」の部分は評価であると認識している。一方で、「その日の気分によって聞ける」の部分を value とするか、evaluation の一部とするかが異なっている。この揺れは、注釈者により注釈付けの粒度が異なるために発生すると考えられる。

「d) 複数タグにまたがる注釈揺れの例」の場合、value タグのみに注目すると部分文字列一致となっている。しかし、前の attribute タグを見ると「踊っている」をどちらのタグに含めるかが異なっている。つまり、区切りが揺れている。

a) 同一箇所の同一文字列に対する注釈付けの有無が異なる注釈揺れの例

注釈者 A: <value>機械的</value>な<attribute>声</attribute>です

注釈者 B: 機械的な声です

b) 同一箇所の同一文字列に対して異なるタグを付与されている注釈揺れの例

注釈者 A: 届いたときは<value>\verb 可愛い</value>と思わずつぶやきました.

注釈者 B: 届いたときは<evaluation>可愛い</evaluation>と思わずつぶやきました.

c) 末尾部分が同じ注釈付けをされている揺れの例

注釈者 A: <value>その日の気分によって聞ける</value>のが<evaluation>すばらしい</evaluation>

注釈者 B: <evaluation>その日の気分によって聞けるといのがすばらしい</evaluation>

d) 複数タグにまたがる注釈揺れの例

注釈者 A: <attribute>茶葉が踊っている</attribute>のが<value>目に見える</value>

注釈者 B: <attribute>茶葉</attribute>が<value>踊っているのが目に見える</value>

e) attribute-value 組と evaluation の注釈付けが異なる揺れの例

注釈者 A: <attribute>落ち着き</attribute>が<value>あります</value>

注釈者 B: <evaluation>落ち着きがあります</evaluation>

図 7 予備実験 1 の結果における注釈揺れ

「e) attribute-value 組と evaluation の注釈付けが異なる例」の場合、極性を持たない value が attribute と組になることで、注釈者が極性を持つと判断し、evaluation としてしまった事が原因となる揺れと考えられる。

上記の内 a), b) が最も多い。これらの注釈揺れの原因として、省略された要素の存在があるのではないかと考えた。省略された要素が注釈揺れの原因となっているのならば、注釈者が注釈を行う際に、省略されている要素を補完した上で注釈付けを行うことで、注釈付けを行おうとしている表現に対して、4 つ構成要素の内のどれに当たるかを判断する際の注釈揺れを削減することが期待できる。例えば、図 7 の b) の例では、attribute が省略されている。この例に対して「外観」を attribute として補うことで「可愛い」がその属性値として認識されやすくなると期待される。そこで次の仮説 2 を立てた。

仮説 2: 省略されている要素を注釈者が補完しながら注釈付けを行うことで複数注釈者間の注釈付けの一致率を向上できる。

この仮説を検証するために行った実験が予備実験 2 である。

4.3 予備実験 2 の方法

予備実験 1 と同様に Amazon から収集したレビュー文 200 文を, item タグが付与された状態で注釈者に配布した. 以下に示す条件で予備実験 2 を行った.

予備実験 2 item タグが既に付与されている文書に対して, attribute-value の組については片方の要素が省略されている場合にはその要素を補完しながら注釈付けを行ってもらった. 補完作業は注釈者が想像するだけでなく, 実際に省略されている要素の内容がどのようなものになるかを記述してもらった. 記述の方法は 3.2 節の attribute と value の注釈付け規則に, 省略であることを明示する属性を追加し, attribute が省略されていると注釈者が判断した場合には, attribute が省略されていると思われる value タグの直前に attribute タグを記述し, タグの要素として省略されていると思われる表現を注釈者に記述してもらった. 同様に value タグが省略されていると思われる場合には value が省略されていた attribute タグの直後に同様に注釈者に省略されていると思われる表層表現を記述してもらった. 用いた文書は予備実験 1 とは異なる. 予備実験 1 と同様特別なツールは利用せず, テキストエディタのみを使用してもらった. 注釈を行った被験者は本研究と直接の関係がある情報工学を専攻する学生 2 名であり, 予備実験 1 に参加している.

予備実験 2 で作成した, 省略されていた要素を補完した注釈付けの例を図 8 に示す. また, 予備実験 2 で利用した作業指示を付録 B に示す.

4.4 予備実験 2 の結果

予備実験 2 を行った結果を表 2 に示す. 予備実験 2 は注釈者が 2 名だったため, 2 名の間の κ 値のみを記してある. なお, 一致の評価については予備実験 1 と同じ方法を用いている.

結果を見ると, 多くの場合で κ 値が 0.4 を超えている. このことから, 中程度の一致はして

```
<item id="i11" class="c2">このソフト</item> は <attribute id="a101" pair="p101"
target="c2" abbr="1"> 重さ <attribute><value id="v101" pair= "p10
1" target="c2">軽</value>て<evaluation id="e101" target="c2" readon="p101">気に入っ
ている</evaluation>.
```

図 8 省略されていた要素を補完して注釈付けを行った例 (下線部が省略を補完した部分)

表 2 予備実験 2 の注釈付けの一致率 (κ 値)

	電化製品	非電化製品	映像・音楽	ソフトウェア
attribute	0.72	0.44	0.38	0.51
value	0.66	0.48	0.24	0.56
evaluation	0.77	0.53	0.41	0.49

いと言える。しかし、予備実験1の最低値と比較すれば一致率は良いが、最高値と比べると一致率が安定しているとはいえず、十分であるとは言えない。この結果から、省略の補完という手法は、ある程度効果があるものの、十分に有用であったとはいえない。このため仮説2自身は成立すると思われるが、その適用範囲が限定的であると考えられる。

予備実験1, 2においては、各注釈者は注釈付けを行う際に、他の注釈者と相談等による情報共有をせず、それぞれが独自に注釈付けを行っている。しかし、各注釈者が独自に注釈付けを行った場合に、省略された要素の補完を各注釈者が行っても、一致率は十分とは言えなかった。さらなる注釈揺れの削減のためには、注釈者が完全に独立して注釈付けを行うのではなく、注釈付けを行う個別の場合について、注釈付けの判断に関する情報を共有した状態で注釈付けを行う必要があると考えられる。

そこで以下の仮説を立てる。

仮説3：注釈者が完全に独立して注釈付けを行うのではなく、注釈付けの判断に関する情報共有をすることで注釈付けの一致率を向上させることができる。

仮説3を実験するための具体的な方法として、次章で注釈事例の参照について述べる。

5 注釈事例を用いた評判情報の注釈付け

本章では注釈者が独立に注釈付けを行いつつも、判断に関する情報共有を行える手法を提案する。

4章の予備実験において一致率が十分ではなかったため、仮説3を立てた。この仮説3における注釈付けの判断に関する共有を、どのように行うかを検討した。まず、注釈者間の話し合いを定期的に設けて、注釈揺れについて議論を行う手法を検討した。しかし、主に時間的な側面からコーパス作成作業のコストが膨大になってしまう点が問題となった。そこで、本稿では過去の注釈事例を参照しながら注釈付けを行う手法を提案する。そして、仮説3を検証するために、具体的な方法として次の仮説3'の検証を行うこととする。

仮説3'：複数注釈者間で注釈事例を共有し、これを参照しながら注釈付けを行うことで注釈揺れを減らすことが出来る。

話し合い等により注釈者間の判断を共有しない場合には、複数の注釈者間で共有する判断材料としては注釈者に対する指示が基本となる。あらゆる状況を網羅して想定した上で個別事例に対する指示を用意することが理想であるが、それは現実的ではない。そこで、個別事例に対する指示を用意するかわりに、過去の注釈事例の中から注釈付け対象となる文に「似た」事例を探し出し、文単位で提示する手法を提案する。さらに事例の中には、注釈付け作業の発注者

が注釈付けを行った事例も含めることが出来るので、発注者の要求を事例として含めることも出来る。このため、個別事例に対する指示を用意するのと同様の効果が得られるのではないかと考えた。

図7に示される注釈揺れについて、注釈事例の参照を用いることによる注釈揺れの削減効果について考えてみる。注釈付け対象となる表現については、表層表現が一致した事例を提示することが出来れば、これは当然注釈揺れの削減に繋がると考えられる。例えば図7のb)における「可愛い」やe)における「あります」などは、レビュー文に出現しやすそうな表現であるが、こういった表現に対する注釈事例を提示できれば、注釈揺れが減らせる。

また、注釈付け対象となる表層表現が必ずしも完全に一致しなくても、類似した表現を事例で提示出来れば注釈者の判断の支援になると考えられる。図7のa)における「機械的」やd)における「目に見える」などは類似した表現が存在する可能性が高いと思われる。

さらに、注釈付けを行う対象として注目している表現の周囲の表現について見てみると、文単位で事例を提示することで、注釈付けの判断対象となる表現の周囲の表現も注釈者は参照可能となる。この周辺の表現が事例により参照可能になることで、どこまで注釈付けの範囲とするかの揺れの削減に効果があるのではないかと考える。例えば図7のc)における注釈揺れにおいては、両方の注釈者が「すばらしい」の部分は evaluation だと判断しているが、その前の部分のどこまでが注釈付けの範囲となるかを判断する際の支援が可能になるとと思われる。

また、図7の例にはないが、注釈付けされたコーパスの中にある、注釈付けを含まない文も事例として登録しておき、参照できるようにしておくことにより、注釈者が注釈付けを行わないという判断を支援できると考える。すなわち、注釈者が注釈付けを行おうとしている文がそのような事例に類似しているときには、それを提示することにより注釈付けが行われていないこと注釈者が確認できるので、注釈者はあえて注釈付けを行わないという判断を下すことができる。

上記のいずれの場合においても、注釈付けを行おうとしている文に対して、複数の事例を提示する事ができる。これにより、図7のb)のような場合、レビュアーが記述した主観的な表現にどのような注釈付けがされているか示す過去の注釈事例を複数提示することで、value と evaluation のどちらを注釈付けすればよいかの判断を支援する事が出来るのではないかと考えた。

これは、複数注釈者間で判断を緩やかに共有することを目的とするものである。注釈事例の参照により、注釈者間の話し合いに時間をかけずに、注釈揺れを減らすことが期待される。ここでも、4章冒頭で説明したように、前もって機械的な注釈付けを行うことはせず、注釈者が判断に悩む時に、事例を参照してもらうというアプローチを用いた。一方で、注釈事例の参照の利用は、注釈者が、すでに付与されている一つの注釈可能性をみるだけでなく、複数の注釈可能性が存在しうることがわかるという効果も考えられる。この点は、事例となる文の選択が適切に働けば、注釈付けの際に有効だと思われる。

5.1 注釈事例の提示機能を有する注釈付けツール

本稿で我々が提案する手法は、評判情報コーパス作成に従事する注釈者へ注釈事例の提示を行うことにより、注釈揺れを削減しながら複数注釈者による注釈付けができるようにする方法である。既存のエディタや注釈付けツール(野口 他 2008)では事例の提示が行えないため、新たに注釈付けツールを試作した。図9に作成した注釈付けツールの画面例を示す。

上段部には注釈付け対象となる文とその前後の文を表示している。注釈者は注釈付け対象となる文を順に見ていく。文中の範囲を指定し、中段部にあるタグ付与のためのボタンで注釈付けを行う。この際、下段部には編集中的文に類似した注釈付け済みの文を例示している。どのようなタグを注釈付けすると良さそうかを下段部の事例を参照して判断することが出来る。さらに、XMLタグの属性情報の入力を行うことにより、組となる要素の指定などを行う。

試作した注釈付けツールの特徴は、注釈付けを行おうとしている文に似た過去の事例を提示することにより、注釈者が注釈付けを行う際に参考に出来る点である。これを実現するためには、注釈付け対象文に類似した注釈事例を事例集の中から検索する必要がある。事例の提示においては、注釈付け対象となっている文において、注釈付けが必要となりそうな部分と類似し

図9 注釈付けツールの画面

た表現をなるべく多く含む事例を提示することが望ましい。そのために、注釈付けが必要となる部分と同一の表層表現に対して過去に注釈付けされた例の提示を行う。本稿では類似度計算に文字 bigram の一致の度合いとして定義される次式を用いた。

$$\text{類似度} = \frac{\text{両文での文字 bigram の一致数}}{\text{対象文の長さ} + \text{事例文の長さ}} \quad (2)$$

上記の式は単純なものだが、簡単な事例提示を行うだけで注釈付けの一致率が向上するのならば、事例提示の有効性が示せると考えられる。

また、本稿で用いた提示事例の類似度計算においては、事例中の各文と注釈対象文との一致を文字 bigram で計算しており、既に付与されたタグの有無は考慮しない。タグを付与すべきではない文を明示的に扱ってはいないが、事例中の類似度が高い文においてタグが付与されていないものが提示されれば、間接的に「現在の対象文に対してはタグを付与すべきではない」ということが例示により示せる。

今後さらなる注釈付けの一致率向上を考えた際には、表層表現の一致だけではなく、意味的に類似した注釈事例の提示を行う等の、異なる文間の類似度を用いることが必要になると思われる。また、本実験で使用した事例集合では確認できなかったが、事例中に非常に似かよった文が存在する場合には、簡単な類似度計算のみでは提示される注釈事例が類似した文で埋め尽くされてしまう可能性がある。今後コーパスサイズを拡大する際には、MMR (Carbonell and Goldstein 1998) のような手法を用いて、提示事例中の類似文を除去する必要がある。

なお、現在の設定では、新たな注釈付け対象文が読み込まれる度に、事例集として指定されたファイルに含まれる文の中から注釈付け対象文との類似度が高い、上位 5 文を検索し表示する。また、注釈事例の提示に併せて、注釈付け対象文に注釈付けがされていた文字列が存在する場合には、強調表示するようにした。これは注釈付けの見落としを少なくするためである。現在の版のツールでは、提示事例中で注釈付けがなされている文字列と、注釈対象文の文字列との間で文字 bigram が 2 つ以上連続して一致した部分については、注釈対象文中の文字を太字で強調表示している。

上記に加え、注釈者の作業を軽減するため、タグの付与作業の支援と、各タグの id 属性値の自動入力が行えるように実装した。

6 注釈揺れに対する注釈事例参照の効果の確認実験

本節では 5 章で述べた注釈事例の参照と注釈付けツールの有効性について検証するための実験について述べる。注釈付けツールを用いない場合と用いた場合、注釈事例の参照を行った場合と行わない場合の注釈付けの一致率の変化と、予備実験 1 で発見された注釈揺れが、注釈事例を参照することでどのように減少するかを調査することで仮説 3' を確認する。

6.1 注釈揺れに対する注釈事例参照の効果の確認実験の方法

予備実験 1 と同様に Amazon から収集したレビュー文書、4 ジャンル 200 文について注釈者が注釈付けを行った。予備実験 1 と同様に製品についての部分—全体関係を記述したオントロジー情報があらかじめ作成してあり、item タグがあらかじめ付与されている文書を使用した。本節で行う注釈揺れに対する注釈事例参照の効果の確認実験では、次の三つの条件で注釈者に注釈付けを行ってもらった。

条件 1 ツール無し：注釈付けツールを使用しない。注釈事例の提示も行わない。

条件 2 事例無し：注釈付けツールを使用する。しかし、注釈事例の提示は行わない。

条件 3 事例有り：注釈付けツールを使用する。注釈事例の提示を行う。

3つの条件下での注釈付けの比較を行うため、6人の注釈者を2名ずつ3つのグループに分けて各文書について条件を変えて3回注釈付けを行ってもらった。注釈者のグループ分けと注釈付けの順番を表3に示す。また、注釈揺れに対する注釈事例参照の効果の確認実験で注釈事例として使用した文書は、予備実験の過程で注釈付けされた文書の中の一人分（200文）である。

表3のように、注釈者1から注釈者4までには「事例有り」を最後に行ってもらい、注釈者5・6には「事例有り」を最初に行ってもらった。また、注釈揺れに対する注釈事例参照の効果の確認実験で注釈を行った被験者は情報工学を専攻する学生6名である。2名（注釈者3と注釈者4）は本研究との直接の関係を持ち、すべての予備実験に参加している。他の4名は本研究と直接の関係を持たず、その内の2名（注釈者1と注釈者2）は予備実験に参加していない。残りの2名（注釈者5と6）は予備実験に参加している。

本章の実験、すなわち「注釈揺れに対する注釈事例参照の効果の確認実験」で用いた作業指示は予備実験1の指示に加えて5.1節で紹介した注釈付けツールの使用説明書と、上述の実験の手順である。ここで、予備実験2に際して、追記された作業指示（補完に関するもの）は破棄していることに注意されたい。これは、予備実験2の結果、仮説2は成立するが、効果が少なかったためである。

6.2 注釈事例の効果に関する確認の実験結果

本節では、評判情報の注釈付けにおいて事例参照を用いた場合と用いなかった場合を比較した時の、複数注釈者間における一致率の変化、図7における注釈揺れの変化について述べる。

表 3 注釈付けツールを用いた実験の手順

	注釈者 1・2	注釈者 3・4	注釈者 5・6
1 回目	ツール無し	事例無し	事例有り
2 回目	事例無し	ツール無し	事例無し
3 回目	事例有り	事例有り	ツール無し

6.2.1 注釈者間の一致率

本節では、注釈揺れに対する注釈事例参照の効果の確認実験の結果の内、同じ条件の2名の注釈者が注釈付けした文書断片の数と、注釈付けの一致率について述べる。

最初に、各注釈者が注釈付けをした文書断片の数、すなわち、付与したタグの数を調べた。結果を表4に示す。表4を見ると、注釈者1と2、注釈者3と4は、attributeタグとvalueタグについては3回目の注釈付けで最も多くの部分に注釈付けを行っている。このことから、注釈付けのために付与したタグの数については注釈者の学習による影響が懸念される。しかし、注釈者5・6が注釈のために付与されたvalueタグの数を見ると、3回目に行った「ツール無し」の場合が2回目に行った「事例無し」の場合に比べて減っている。この点から考えると、注釈者の学習による影響が全ての要因であるとは言えない。何回目に注釈付けを行ったかに関わらず、事例有りの多くの場合が注釈付けされたタグの数が最多である。このように注釈付けされた順序だけに因らないことを考慮すると、注釈事例の提示による注釈付け支援により、注釈者がより細かく注釈付けを行うことができたと推測される。

次に、注釈者間の注釈付けの一致について調べた。注釈揺れに対する注釈事例参照の効果の確認実験では4.2節と同様に次に示す3つの場合を、注釈者間の注釈付けが一致したものと判定した。

- 完全一致
- 部分文字列一致
- 一方ではattributeとvalueを分けているが、他方では両方を一つのvalueとしている場合

表4 各注釈者が付与したタグの数

	ツール無し		事例無し		事例有り	
	1回目		2回目		3回目	
	注釈者1	注釈者2	注釈者1	注釈者2	注釈者1	注釈者2
attribute	86	55	64	74	111	77
value	127	142	147	153	199	170
evaluation	114	72	135	70	97	72
	2回目		1回目		3回目	
	注釈者3	注釈者4	注釈者3	注釈者4	注釈者3	注釈者4
attribute	104	69	89	62	137	82
value	184	114	181	115	215	126
evaluation	53	79	53	81	81	88
	3回目		2回目		1回目	
	注釈者5	注釈者6	注釈者5	注釈者6	注釈者5	注釈者6
attribute	86	66	81	97	86	83
value	176	204	211	206	177	181
evaluation	66	79	54	85	66	83

上記の一致に関する判断基準に基づき、各被験者の組における注釈付けの一致率 (κ 値) をタグの種類ごとに調べた。表 5 に結果を示す。

注釈揺れに対する注釈事例参照の効果の確認実験の結果では、多くの場合で「事例有り」の場合に κ 値が高かった。注釈者 5 と 6 における value においてのみ「事例無し」の結果の方がよくなっているが、これは「事例有り」の場合から先に実験を行ったグループでのことなので、注釈者が学習をしてしまった結果とも考えられる。しかし、「事例有り」から注釈付け作業を開始した注釈者の組と、「事例無し」から注釈付け作業を開始した注釈者の組の一致率に大きな違いが無い。さらに、過去に 1 度も予備実験に参加していない注釈者の組 (注釈者 1・2) を見ると、回数を重ねた 2 回目においても 1 回目と比べて安定した一致率が得られているとはいえない。しかし、注釈事例を参照した 3 回目においては他の組と同程度の安定した一致率が得られている。これより、注釈者が注釈付け作業を経験し学習してしまった効果よりも、事例提示を行った効果の方が大きいと考えられる。

また、事例参照の有効性については、「事例有り」から注釈付け作業を開始した注釈者の組の結果に注目したい。同じデータに対して 3 回の注釈付け作業を行っていながら、注釈事例の参照を行わなくなることによって一致率が低下している。これは、注釈者の学習効果により一致率が高まるよりも、事例を参照しなくなったことにより注釈者が独自の判断で注釈付けを行ってしまい、一致率が低くなってしまうためと考えられる。このことから、注釈事例の参照は一致率を安定させるために有効であると言える。この結果、仮説 3' は正しかったといえる。

なお、本稿で用いた事例集合が十分に参考になる情報を提示しているかということを確認するためには本来ならば、個々の注釈付けにおいて、その都度表示された事例が有効であったか

表 5 注釈付けの一致率 (κ 値)

		ツール無し	事例無し	事例有り
		1 回目	2 回目	3 回目
注釈者 1・2	attribute	0.39	0.37	0.54
	value	0.51	0.59	0.62
	evaluation	0.63	0.35	0.66
		2 回目	1 回目	3 回目
注釈者 3・4	attribute	0.52	0.53	0.54
	value	0.52	0.56	0.57
	evaluation	0.57	0.67	0.69
		3 回目	2 回目	1 回目
注釈者 5・6	attribute	0.50	0.49	0.52
	value	0.56	0.67	0.64
	evaluation	0.58	0.55	0.69

という検証を行いながら、事例集合の量と質を検証することが必要である。しかし、本稿では、相対的な結果として、本稿で用いた事例集合であっても注釈揺れの削減に効果があったという提示のみとなっている。本稿で用いた事例の量が十分であるかどうかはわからないため、事例集合の量と質の検討が今後の課題として挙げられる。

6.2.2 注釈揺れの分析

本節では注釈揺れについて、細かく分析を行う。まずは、注釈付けが一致した部分がどのように変化したかを調べるために、典型的な例である、二人の注釈者が注釈付けした結果の中で、両方の注釈者が同一の文に対して完全に一致した注釈付けを行っている場合の数を調べた。ここで、「完全に一致した注釈付け」とは文中に注釈付けがなされた箇所が存在し、複数の注釈付けがあったときにはその全てが可不足なく一致している場合である。二人の注釈者が共に注釈付けを全く行わなかった文については調査外としている。表 6 に結果を示す。

表 6 においては、注釈事例を参照した場合が、注釈者間で完全に注釈付けが一致した文の数が最も多く、最も良い結果となっている。また注釈者がどのような順番で注釈付け作業を行った場合でも、「事例有り」の時に一致した文が増えたということは仮説 3' を裏付けるものである。

さらに、表 6 において、注釈者の各組について、注釈付けが完全に一致した文について分析した結果を表 7 に示す。表 7 の結果を見ると注釈付けが揺れていた部分については、事例の参照を行うことで注釈揺れが改善した。なお、表 7 中の「ツール無し」の場合に一致していなかった文の内、「事例有り」で一致した文については、同一文中にて複数種類の改善が行われたために注釈付けが一致するようになった文が存在する。このため、改善の種類の内訳について文数を合計すると、一致するようになった文数よりも多くなっている場合がある。また、注釈事例の参照を行ったために新たに注釈付けが揺れるようになった文を見ても、注釈付けの範囲が変わったため一致しなくなった文については、完全一致だったものが部分一致になるという揺れであった。部分一致を正解とする判定においては、この揺れは一致率に影響を与えない。以上の各点より、全体で見ると注釈事例の参照により注釈付けの一致率の向上が見られる。

表 6 完全に一致した注釈付けが行われた文数

	ツール無し	事例無し	事例有り
	1 回目	2 回目	3 回目
注釈者 1・2	42	45	55
	2 回目	1 回目	3 回目
注釈者 3・4	58	51	60
	3 回目	2 回目	1 回目
注釈者 5・6	54	68	70

表 7 完全に一致した注釈付けが行われた文数の変化

		注釈者 1・2	注釈者 3・4	注釈者 5・6
「ツール無し」「事例有り」いずれの場合にも注釈付けが完全に一致した文数		19 文	37 文	30 文
「ツール無し」の場合に一致していなかった文の内、「事例有り」で一致した文数		36 文	23 文	40 文
内 訳	一方の注釈者しか注釈付けを行っていない部分が改善されたために一致するようになった文数	22 文	16 文	16 文
	同一箇所の同一文字列に対して異なる注釈を付与していた部分が改善されたため一致するようになった文数	8 文	4 文	2 文
	attribute と value の組と evaluation の注釈付けが異なっていた部分が改善されたため一致するようになった文数	2 文	0 文	0 文
	注釈付けの範囲が異なっていた部分が改善されたために一致するようになった文数	5 文	5 文	24 文
「ツール無し」の場合に一致していたが「事例有り」で一致しなくなった文数		23 文	21 文	24 文
内 訳	一方の注釈者のみが新たに注釈付けを追加したことにより、一方の注釈者しか注釈付けを行っていない部分が新たにできた文数	14 文	16 文	21 文
	注釈付けの範囲が変わったため一致しなくなった文数	9 文	5 文	3 文

次に、図 7 に示した a)～e) の注釈揺れが注釈事例の参照を用いることでどのように変化したかについて述べる。まず最初に、予備実験 1 における図 7 の「a) 注釈付けを行うかどうか異なる揺れ」がどのように変化したかを調べるために、一方の注釈者のみがタグを付与した部分の数をタグごとに調べた数を表 8 に示す。

この表によると、実験の順番や、ツールの有無、注釈事例の有無の各条件と、一方の注釈者のみが注釈のために付与したタグの数の間には相関が見られない。しかし、表 4 と比較してみても、付与されたタグの数が増えた場合に一方のみが注釈付けしたタグの数が増えているというわけではないため、注釈事例の提示による悪影響は考えられない。

次に予備実験 1 で確認した図 7 の「b) value と evaluation の注釈が異なる揺れ」の変化を調べる。「b) value と evaluation の注釈が異なる揺れ」は同一箇所の同一文字列に対して異なるタグが付与された注釈揺れの一つである。表 9 に本章の実験、すなわち「注釈揺れに対する注釈事例参照の効果の確認実験」において同一箇所の同一文字列に対して異なるタグが付与された時のタグの違い方とその数を示す。

表 9 を見ると、value タグと evaluation タグの揺れが最も多かった。これは予備実験 1 の結果と同じである。注釈揺れに対する注釈事例参照の効果の確認実験によると、この問題点についても必ずしも「事例有り」の場合に揺れが減っているとは言えない。しかし、これについても

表 8 一方の注釈者のみが注釈付けしたタグの数

		ツール無し	事例無し	事例有り
		1 回目	2 回目	3 回目
注釈者 1・2	attribute	58	34	52
	value	77	44	73
	evaluation	34	49	22
		2 回目	1 回目	3 回目
注釈者 3・4	attribute	56	39	60
	value	76	66	81
	evaluation	26	19	28
		3 回目	2 回目	1 回目
注釈者 5・6	attribute	19	25	29
	value	56	48	51
	evaluation	25	23	13

表 9 同一箇所の同一文字列に異なるタグが付与された数

		ツール無し	事例無し	事例有り
		1 回目	2 回目	3 回目
注釈者 1・2	value ↔ evaluation	16	28	25
	value ↔ attribute	1	4	4
	attribute ↔ evaluation	0	3	1
		2 回目	1 回目	3 回目
注釈者 3・4	value ↔ evaluation	21	20	18
	value ↔ attribute	5	7	10
	attribute ↔ evaluation	0	1	0
		3 回目	2 回目	1 回目
注釈者 5・6	value ↔ evaluation	27	22	27
	value ↔ attribute	1	6	0
	attribute ↔ evaluation	0	5	0

表 4 と比較して、「事例有り」で注釈付けされたタグ数が増えた場合に注釈揺れが増えているわけではない。

続いて、予備実験 1 で確認した残りの注釈揺れの事例についても調査した。図 7 で述べた「c) 末尾部分が同じ注釈付けをされている注釈揺れ」, 「d) 複数タグにまたがる注釈揺れ」, 「e) attribute-value 組と evaluation の注釈付けが異なる揺れ」のそれぞれの数を表 10 に示す。

表 10 を見ると、予備実験 1 でもそうであったように、この 3 種類の注釈揺れは図 7 の a), b) の揺れに比べて少ない。しかし、注釈者 1 と注釈者 2 の結果を見ると事例の参照の効果がうかがえる。

表 10 事例を参照した場合の図 7 における c), d), e) の注釈揺れの変化

		ツール無し	事例無し	事例有り
		1 回目	2 回目	3 回目
注釈者 1・2	c) 末尾部分が同じ注釈付けをされている注釈揺れ	2	4	0
	d) 複数タグにまたがる注釈揺れ	0	0	0
	e) attribute-value 組と evaluation が異なる揺れ	10	2	1
		2 回目	1 回目	3 回目
注釈者 3・4	c) 末尾部分が同じ注釈付けをされている注釈揺れ	0	0	1
	d) 複数タグにまたがる注釈揺れ	0	1	0
	e) attribute-value 組と evaluation が異なる揺れ	2	0	0
		3 回目	2 回目	1 回目
注釈者 5・6	c) 末尾部分が同じ注釈付けをされている注釈揺れ	2	1	1
	d) 複数タグにまたがる注釈揺れ	1	2	0
	e) attribute-value 組と evaluation が異なる揺れ	2	0	0

最後に、注釈揺れに対する注釈事例参照の効果の確認実験において問題となった注釈揺れをまとめると以下ようになる。

1) 同一箇所の同一文字列に注釈を付与するかどうかの揺れ

これについては注釈者に因るところが大きく、完全に無くすことは不可能と考える。揺れを減らすために各注釈者に、なるべく細かく注釈付けを行うように依頼することとする。なお、現在行っている事例の提示は注釈付けがなされた表層表現に注目しているが、注釈付けがなされなかった表層表現にも注目して事例の提示を行うことが今後の課題として考えられる。

2) 同一箇所の同一文字列に異なるタグが付与されている揺れ

特に value タグと evaluation タグの違いが多かった。これについては、注釈事例を増やすことで注釈揺れを減らすことが出来ると予想される。注釈事例を増やした場合の効果として、類似した表層表現を検索・提示する際のヒット率が上がる事が考えられる。また、注釈付けの判断対象となる周囲の表現についても、事例を増やすことで事例中に表示することが可能になると予想される。このような表現は注釈者間の判断を揺れなくす材料になると考えられる。

3) 注釈付けの範囲が異なる

注釈付けの範囲については、表現を注釈付けに含めるか含めないかの判断が揺れている。そのため、どの場合に含め、どの場合に含めないかを示すため、注釈付け対象となる表現の周囲の文脈を提供することで、注釈揺れの削減が期待できる。しかし、十分な量の注釈事例が無く、事例の探索についても簡単な方法を用いている現在では、注釈付

けの範囲についての規則を追加することが揺れを減らすために効果的と考える。

またさらに注釈揺れを減らす方法として、提案した評判情報モデルを実例に適応する際の判断の基準を洗練する等が考えられる。これは、注釈者のモデルに対する理解度の差による揺れを解消しようとするものである。これについては、注釈者の判断を確認するテスト等を行うなどの方法を今後検討する必要がある。

7 評判情報コーパスの作成実験

本章では試作した注釈付けツールを用いた、評判情報コーパスの作成実験について述べる。

7.1 評判情報コーパスの作成実験の方法

最初に、各注釈者へ事前説明を行った。注釈者には4章や6章の実験と同じく3.1節で述べた評判情報モデルについての解説をし、過去の実験で得られた注釈事例から「注釈揺れが起きた実例」と「判断が難しかった実例」を示しながら事前説明を行った。

6.2.2節の分析から、注釈付けの量については、注釈付けが必要かどうか判断に迷った場合には、その部分にはなるべく注釈付けを行うように依頼した。同様に、6.2.2節において分析した注釈付けされた文字列の最後の文節における注釈揺れ、複数タグにまたがる注釈揺れの2点に関しても、注釈付けの粒度をなるべく細かくすることを注釈者に注意した。なお、これらの説明を行う際にも、6章の実験で得られた注釈事例から実例を提示して説明を行った。

また、当初から評判情報の構成要素が省略される場合としては、attribute が省略され value や evaluation が現れる場合か、attribute が記述されているが value が省略され evaluation が現れる場合か、もしくは attribute と value の両方が省略され evaluation のみが現れる場合が想定されていた。しかし予備実験1の結果 attribute が記述されているにもかかわらず、value も evaluation も記述されない場合が発見された。以下に例を示す。

例文6：この製品は値段が…

例文6のような場合は、文として不完全である。また、評判情報という性質上、“どのような製品であるかという記述”，“レビュアーの評価”のどちらも記述されていない部分については注釈を付与する必要性が無いと考えられる。このため、本章の実験では、最初に value タグもしくは evaluation タグを注釈付けする部分を探しながら注釈付けを行うように、各注釈者に指示した。また、その際に value と evaluation に該当する文字列がどちらも存在せずに item や attribute のみが現れた場合には注釈付けを行わなくてもよい旨を説明した。

さらに、6.2.2節で述べた注釈揺れ3)のような注釈揺れや、6章の結果では部分文字列の一致に基づく一致として扱った事例を減らすために注釈付けの規則を追加した。これは6.2.1節の分

析結果等から追加したものである。この規則には、本稿で提案した評判情報のモデルで意図した注釈付けを行ってもらうための規則と、注釈の範囲に関する注釈揺れを減らすための規則がある。

前者について追加した規則と追加した理由を図 10 に示す。図 10 は、一方の注釈者が我々が意図しない注釈付けを行ってしまったために注釈付けの範囲が揺れている例であり、実例を示しながら注釈者への指示を指示書に追加した。

また、後者について追加した規則を図 11 に示す。図 11 の例は、評判情報の注釈付けという本来の目的においては、どちらでもよい。一方で、注釈付けの一貫性を保つためには、注釈付けの発注者が指示を与えて、一方の注釈付けに合わせる必要がある。図 11 の規則はそのためのものである。

加えてモダリティに関して追加した規則を図 12 に示す。図 12 の規則はモダリティを除いた命題部分のみに注釈付けしてもらうために追加した規則である。疑問については、その命題についてレビュアーが直接言及していないので除外した。

上記の説明を行った後に、コーパス作成に利用するレビューとは別に、次に述べる各製品ジャンルから 40 文ずつ、計 200 文のレビューを Amazon から収集し、コーパス作成に参加する注釈者に注釈付けの練習を行ってもらった。また練習の後、注釈間違いや注釈揺れについては筆者

- ・副詞的表現が直近にある場合は注釈付けの範囲に含める。

副詞的表現が直近にある場合には属性値や評価に対してその程度や範囲を表していることが確実であるため、属性値や評価のタグの範囲に含めるようにした。

例：

正：<attribute>バランス</attribute>が<evaluation>とても悪いのです</evaluation>

誤：<attribute>バランス</attribute>がとても<evaluation>悪いのです</evaluation>

- ・「～ならば」のような条件部分は注釈付けの範囲に含めない。

条件部分は、属性値や評価の範囲を表すものではなく、4 つ組となった評判情報全体全体に対して範囲を表すものであるため、注釈付けの外とした。

例：

正：妥協ができるならば<evaluation>買いだと思います</evaluation>

誤：<evaluation>妥協ができるならば買いだと思います</evaluation>

- ・ attribute と value とをなるべく細かく分解する。

attribute と value を細かく注釈付けしてもらうことで、評判情報をより精密に捉えることが可能となるため、指示した。

例：

正：<attribute>満足度</attribute>が<value>変わる</value>

誤：<value>満足度が変わる</value>

図 10 意図した注釈付けを行ってもらうために追加した規則

・注釈付けされた文字列の末尾が句読点、感嘆詞、疑問符、助詞だった場合、それらを除外するように終了タグの位置を調整する.

例:

正: <value>対応してません</value>!

誤: <value>対応してません!</value>

・詳細な説明を表現する文節は注釈付けの範囲に含めない.

例:

正: きれいな元のガラスの輝きに<value>戻ります</value>

誤: <value>きれいな元のガラスの輝きに戻ります</value>

・サ変動詞の語尾は注釈付けの範囲に含める.

例:

正: <value>可動する</value>

誤: <value>可動</value>する

・value-attribute の順番で出現し、一語になっている部分はまとめて value とする.

例:

正: <value>黄色</value>

誤: <value>黄</value><attribute>色</attribute>

図 11 注釈の範囲に関する注釈揺れを削減するために追加した規則

・「たぶん～」等、モダリティの呼応要素は注釈付けの範囲に含めない.
 ・「～よ.」「～ね.」等、終助詞等の伝達態度のモダリティは注釈付けの範囲に含めない.
 ・「～でしょう」等、推定推量を表すモダリティ部分（真偽判断のモダリティ）は注釈付けの範囲に含めない.
 ・疑問部分は表現タイプのモダリティの一部であるが、これには注釈付けを行わない.

図 12 モダリティ部分に関する規則

を中心とし注釈者間で指摘をし合い、注釈者間で話し合いを行い、判断の統一を試みた.

次に、注釈付けに用いた文書について述べる. 注釈付けに用いた文書は、Amazon から収集したレビュー 1 万文である. 使用した製品のジャンルは基本的に予備実験と同様であるが、非電化製品を Amazon の製品ジャンルを基に「ホーム・キッチン」、「ホビー・おもちゃ」の二つに分けて、電化製品、ホーム・キッチン、映像・音楽、ソフトウェア、ホビー・おもちゃの 5 ジャンルとした. それぞれのジャンルから 2,000 文ずつを収集した. さらに、1 万文の内の 1,000 文 (5 ジャンルから 200 文ずつ) は事前に第一著者が注釈付けを行い、事例用コーパスとして用いた.

注釈者は情報工学を専攻する学生 10 名である. 注釈付けは 9,000 文を 200 文程度に分割し、作業時間に余裕がある人に順次配布していく形で行った. 作業量は注釈者により 2,000 文から

200 文まで様々である。

なお、過去の実験の知見を元に、本章の実験すなわち「評判情報コーパスの作成実験」における注釈者への作業指示を付録 C に示す。

7.2 コーパス作成実験結果

7.2.1 コーパス修正作業

10 名の注釈者によるコーパス作成を行った後、注釈者の入力ミス、同一文字列に対する注釈揺れ等の明らかに分かる誤りに関しては筆者が修正作業を行った。

最初に注釈者の入力ミス、入力忘れが判明した部分を修正した。入力ミスが確認されたタグは全部で 705 個、コーパス作成作業で注釈付けされた全タグの内の 4.42%であった。入力ミスの内容としては、「タグに付与する属性情報の入力忘れ」、「入力する属性項目が入れ替わってしまっている」、「注釈付けツール使用時にマウスの操作を誤ったために注釈範囲が誤っている」等が主なものであった。以上の入力ミスは注釈付けツールのインターフェースに起因すると思われるものもあり、注釈付け作業者の使い勝手という面から注釈付けツールの改良を検討する必要性がある。

次に、同一表層表現に対し、異なるタグが注釈付けされている部分について、修正を行った。具体的には、同一表層表現に対し、異なるタグが注釈付けされている部分を確認し、正しいと考えられる注釈付けへと統一した。表 11 に注釈揺れの種類と数を示す。

さらに、各注釈付けを確認し、事前説明に沿わない部分、明らかに一部の注釈者のみが異なるタグを付与している部分については修正をし、正しいと考えられる注釈付けに揃えた。

次に、同一文字列に対する異なるタグの中で、上記の方法によっても一方のタグに統一することができなかった部分について述べる。item に関連する部分では、製品の特徴を表現し得る語句が名称の一部に用いられている場合が 12 個で最も多かった。例を以下に示す。

例文 7: 少なくとも観た事のあるシリーズの中では <evaluation> ベスト </evaluation> です。

表 11 同一文字列の同一範囲に対する注釈揺れ

注釈揺れの種類	修正前の個数	修正後の個数
item ↔ attribute	76	11
item ↔ value	16	15
item ↔ evaluation	1	1
attribute ↔ value	29	7
attribute ↔ evaluation	6	0
value ↔ evaluation	79	0
計	207	34

例文8: 注目を集めた一彼女の <item> ベスト </item> も初回盤で <value> 5040 円 </value>.

例文9: 決して寄せ集めではない <item> ベストアルバム </item>.

例えば, 「ベスト」は一般的には評価を表す表現と考えられ, その事例は例文7のようなものである. 一方で, 例文8における「ベスト」という表層表現は評価を表す「ベスト」ではなく, 「ベストアルバム」という商品を指し示すため用いられている. つまり, 例文9と同じ商品を指して入る. この場合, 例文7のように評価を示す「ベスト」には evaluation タグを, 例文8のように商品の指し示す「ベスト」には item タグを付与した状態を正しいものとした. 同様の例として「コンパクト」, 「ブラック」等が挙げられる. また, 今注目している item が別の item と関係をもって記述されるときには attribute として扱われる場合がある. その場合においては, 一方のタグに統一することができないものが6個あった. 同様の理由で item と value が同じ表層表現になっているものが4個あった. このような問題は表現が指し示す概念が多義であるために起こると考えられる. 例を図13に示す.

attribute と value について, 一方のタグに統一することが出来ずに残ったものに, サ変動詞の省略が行われているために表層表現が同じになっているものがあった. 例を図14に示す. 他に, 動詞や形容動詞が活用形により名詞と同じ表層表現になってしまったもの, 「～的」という表現等, 同音異義語, 言葉の意味自体がよく分からないために修正すべきではないとした部分がある.

最後に, さらに追加で行った修正について述べる. 我々は作成したコーパスを教師情報とした

・ item-attribute の揺れ

<item>こ の PC</item>は<item>メ モ リ</item>の<attribute>容 量</attribute>は <value>512MB</value> です.

<item>このソフト</item>は<attribute>メモリ</attribute>が<value>256MB 必要</value>.

・ item-value の揺れ

<item>紙</item>の<attribute>色</attribute>が<value>黄色い</value>よ.

<item>台座部分</item>の<attribute>素材</attribute>は<value>紙</value>だろう.

図 13 視点となる item に起因する揺れの例文

<attribute>対応</attribute>も<value>はやかった</value>.

<attribute>再生</attribute>に<value>対応</value>な点も

図 14 サ変動詞の省略が行われているために表層表現が同じになっている例

学習型の抽出手法を用いた評判情報の構成要素の抽出を検討している。文字を単位としたチャンク同定手法を用いて、評判情報の構成要素の抽出予備実験を行ったところ、item の抽出精度が著しく低い結果となった。評判情報コーパスの作成実験におけるコーパス作成では、「value も evaluation も無い文には注釈付けを行わなくてもよい」と注釈者に説明したため、ある場所では item タグが付与されているが、別の個所に現れる同一の表層表現には付与されていない場合があった。このために学習がうまく進まなかったと考えられたため、item に関しては全文を第一著者が見直し、追加で注釈付けを行った。修正前と修正後の item タグの数と抽出精度を表 12 に示す。抽出には固有表現抽出に用いられるチャンク同定手法を用いた。表に示す抽出精度は、ジャンルごとに item の自動抽出を行った際の F 値の平均である。

7.2.2 基礎統計

本節では、コーパス作成実験の結果得られたコーパスの全体的な傾向を把握するための基礎的な分析を行う。

最終結果として得られた、1 万文のコーパスに注釈付けされた表現の数を表 13 に示す。

表 13 をみると、ジャンルごとに注釈のために付与されたタグの数に多少の違いがある。製品のジャンルごとの特徴を見てみると、映像・音楽における attribute タグを付与した回数が

表 12 item タグの修正前後の比較

	item タグの数	抽出精度
修正前	2,152	20.69
修正後	6,128	44.88

表 13 注釈付けされた表現の数

	注釈付けされた表現の数	item	attribute	value	evaluation
電化製品	延べ数	1,374	1,077	1,746	513
	異なり数	871	598	1,407	331
ホーム・キッチン	延べ数	1,077	875	1,838	363
	異なり数	648	476	1,523	242
映像・音楽	延べ数	1,187	614	1,026	367
	異なり数	694	378	871	253
ソフトウェア	延べ数	1,243	821	1,585	302
	異なり数	710	560	1,316	209
ホビー・おもちゃ	延べ数	1,247	785	1,687	436
	異なり数	797	453	1,441	285
全データ	延べ数	6,128	4,172	7,882	1,981
	異なり数	3,522	2,069	5,821	971

他ジャンルに比べてわずかに少ない。一方で、value タグについてはソフトウェアや、ホーム・キッチンよりも多くなっている。映像・音楽については製品の性質上、製品の部分の性能について細かく説明するよりも、その製品についてどう思うかという感想が述べられやすいという特徴が注釈付けの量にも現れていると考える。

次に、評判情報の構成要素の組について、ある要素が省略されている場合について考察する。提案モデルでは要素の省略を認めている。各要素が省略されていた場合の組数を表 14 に示す。なお、提案モデルでは要素の 1 対多関係も認めており、要素の組の数を延べ数で調べた結果は 9,287 組であった。表中の ϕ が省略されていた要素である。実験では注釈者に、まずは value と evaluation のどちらかを探してから注釈付けを行うように依頼した。そのため、本来ならば、item-attribute- ϕ - ϕ という組は存在しないはずである。しかし、注釈者が作業中に覚書として注釈付けをした部分の消し忘れ、コーパス全体の修正作業をした際に誤って注釈付けされていた value を削除した際の残りである。これらについては、item との組は正しいので付与した情報を削除せずに残してある。その他に、製品の様態のみを記述し、評価について言及していない部分が多数存在した。これにより、コーパス全体として value タグが非常に多くなる一方で、evaluation は少ない結果となった。これはレビューを書く際にはっきりと評価を書くことを避ける傾向があるためと考えられる。

次に、出現回数の多い表層表現に着目する。attribute, value, evaluation の各タグに頻出した表層表現の上位 10 位を表 15 に示す。また、各製品ジャンルごとの各タグに頻出した表層表現の上位 5 位を表 16 に示す。

各ジャンルごとに頻出している表層表現を見ると、その製品ジャンルに特徴のある表現が上位にきている。しかし、表 16 で示した全コーパス中で出現回数の多い表現の中には、複数の製品ジャンルにおいて出現頻度の高い表現が多く存在する。特に、attribute に注目して見ると、「音」や「値段」のように、製品のジャンルに関わらずレビュアーがよく言及してる表現が存在していると考えられる。

表 14 4 つ組における各要素の省略

要素の組	個数
item- ϕ - ϕ -evaluation	634
item- ϕ -value- ϕ	3,763
item-attribute- ϕ - ϕ	42
item-attribute-value- ϕ	3,064
item-attribute- ϕ -evaluation	564
item- ϕ -value-evaluation	599
item-attribute-value-evaluation	621

表 15 attribute, value, evaluation の頻出表層表現上位 10 個

attribute		value		evaluation	
表層表現	出現回数	表層表現	出現回数	表層表現	出現回数
「音」	123	「ある」	57	「良い」	71
「デザイン」	103	「あります」	52	「いい」	62
「値段」	77	「あり」	39	「いいです」	34
「機能」	61	「高い」	36	「お勧めです」	29
「価格」	58	「安い」	33	「素晴らしい」	26
「音質」	47	「ない」	31	「良いです」	22
「内容」	42	「シンプル」	30	「良く」	22
「色」	41	「便利です」「簡単」	29	「気に入っています」	19
「大きさ」	37	「ありません」	22	「良かった」	17
「サイズ」	31	「なく」「十分」	22	「お勧め」「よい」 「オススメです」 「驚きました」	16

evaluation については、頻出表層表現を見ても、肯定的な意見がコーパス中に多く存在しているように見える。そこで、evaluation タグの orientation 属性について、値として現れる positive, negative, neutral の個数を調べてみた。orientation 属性には、対応する評判が肯定なのか否定なのかを表す値が与えられている。結果を表 17 に示す。

表 17 を見ると、positive が圧倒的に多くなっている。この理由の 1 つとしては、製品レビューという情報源そのものの特徴が挙げられる。すなわち、レビューを書く動機が、「この製品を他の人に紹介したい」や「この製品を他の人に勧めたい」といった、肯定的なものであることが多いからではないかと考えられる。

item についても頻出する表層表現を調査した。7.2.1 節で述べた修正作業により注釈付けされた部分が増え、総数としては value タグに次いで多くなっている。しかし、一方でレビュアーが製品名を連呼することで頻度が上がってしまうような例もあり、統計的な傾向が記述形式に拠るところが大きい。

最後に、同じ value に対して異なる極性の evaluation が記述されている部分について調べた。3.1 節で述べた提案モデルでは、製品の同一の様態に対して評価が異なる場合を考慮し、value と evaluation を分離している。同じ表層表現の value に対して極性の異なる evaluation が組になっている部分がコーパス中に存在する事を確認する。value と evaluation を分離した提案モデルを用いたことにより、より正確に評判を捉えられた部分であり、提案モデルの有効性が示されている。この場合の value と evaluation の組を表 18 に示す。

評判情報コーパスの作成実験で作成したコーパス中にも、同じ value に対して極性の異なる evaluation が組となっている部分が存在した。上記の「高い」などは attribute が「値段」と「コ

表 16 各製品ジャンルにおける attribute, value, evaluation の頻出表層表現上位 5 個

	attribute		value		evaluation	
	表層表現	出現回数	表層表現	出現回数	表層表現	出現回数
電化製品	「音」	59	「ない」「便利です」「ある」	12	「良い」	21
	「音質」	42	「安い」	11	「いい」	18
	「デザイン」	28	「ありません」「できる」 「簡単」「十分」	10	「良く」	11
	「値段」	24	「あります」	9	「よい」「満足です」	8
	「機能」	21	「きれい」「高い」「使える」 「便利」「問題なく」	7	「いいです」「よかった」	7
ホーム・キッチン	「デザイン」	59	「シンプル」	17	「いい」	13
	「音」	33	「あります」	12	「いいです」	12
	「値段」	24	「高い」	9	「良い」	9
	「機能」「吸引力」 「色」	17	「安い」「簡単」「便利です」	8	「お勧めです」	7
	「価格」	15	「コンパクト」「軽い」	7	「おすすめです」「大満足です」	6
映像・音楽	「内容」	19	「あります」	13	「素晴らしい」	14
	「声」	12	「あり」「楽しめる」	9	「いい」	13
	「演技」「歌詞」	10	「高い」「凄い」	8	「お勧めです」「良かった」	7
	「ストーリー」	9	「ある」	7	「残念」「魅力的」「良い」	5
	「雰囲気」	8	「ポップ」「面白い」	6	「好きです」「残念」	4
ソフトウェア	「機能」	21	「ある」	16	「良い」	10
	「動作」	16	「あり」	12	「素晴らしい」	17
	「操作」「内容」	12	「あります」「高い」	8	「お勧めです」	6
	「使い勝手」	11	「あって」「ない」「重い」 「多い」	7	「おすすめします」「お勧め」	5
	「値段」	9	「安い」「簡単」「十分」 「速い」「便利です」 「出来る」「豊富」	6	「良いです」「いいです」「良かった」 「おすすめです」「悪い」「いい」 「驚きました」「気に入りました」 「良かったです」「オススメです」	4
おもちゃ・ホビー	「音」	21	「ある」	16	「良い」	26
	「色」	19	「あります」	12	「いい」	14
	「大きさ」	16	「あり」「小さい」	8	「いいです」	9
	「出来」	15	「可愛い」「楽しい」「単純」	7	「残念です」	6
	「値段」	13	「安心です」「楽しめます」 「大きい」	6	「気に入っています」「大満足」 「素晴らしい」「大喜びです」 「良かったです」	4

表 17 evaluation の極性

極性	出現回数
positive	1,894
negative	351
neutral	172

表 18 同一の様態に対する異なる極性の evaluation

value	evaluation	極性
1 万円を切る	結構ショックです	negative
	正解だった	positive
ありません	不安を感じました	negative
	おすすめです	positive
わからない	失格です	negative
	良い	positive
わかりやすく	不満	negative
	特にお薦めです	positive
違って	残念	negative
	気に入っています	positive
簡単	不満	negative
	好感が持てます	positive
軽く	不満	negative
	大満足です	positive
高い	少し残念だった	negative
	おすすめです	positive
小さい	残念です	negative
	大喜びです	positive
少ない	好みではなかった	negative
	お薦めです	positive
多く	オススメしません	negative
	よかった	positive
忠実に再現されており	魅力の薄い	negative
	とてもよい	positive
低い	残念だ	negative
	良い点	positive
無い	不満が残る点です	negative
	良い点	positive

ストパフォーマンス」の場合であり、attribute が異なるために極性が異なっている場合もある。これは組となる attribute を考慮することで極性の違いを捉えることが出来る。一方で、同種の製品の同一 value に対して異なる極性の evaluation が付与されている場合もある。例を図 15 に示す。このような場合には特に、製品の様態と評価を分離する提案モデルにより評判情報をより正確に捉えることができると考えられる。

着ぐるみを<value>忠実に再現している</value>がために, <item>この玩具</item>も<evaluation>魅力の薄い</evaluation>品物となってしまったように思います.

<value>忠実に再現されており</value><evaluation>とてもよい</evaluation>.

図 15 同種の製品の同一 value に対して異なる極性が付与されている例文

8 おわりに

本稿では, 製品の様態と評価を分離した評判情報のモデルを提案し, 同様の様態に対する評価がレビュアーにより異なる場合にも対応可能とした. 注釈付けについては, 注釈事例を用いた複数注釈者の人手による評判情報コーパス作成手法を提案した. また, 提案手法を用いてコーパス作成を行った.

コーパス作成における複数注釈者間の注釈揺れについては, 注釈者に指示を与えるだけでは一致率が十分ではないということがわかった. 省略されている要素の補完を行いながら注釈付けを行うことで一致率の向上は見られたが, 十分とは言えなかった. しかし, 注釈事例の参照という注釈者間の緩やかな知識共有を行うことにより注釈付けの一致率が向上することが分かった.

作成したコーパスについて特徴を統計的な面から調査し, 作成されたコーパス中にも, 提案モデルが想定していた様態と評価の分離が必要な記述が存在したことが確認され, 提案モデルにより評判情報をより正確に捉えることができた.

今後は, 評判情報の抽出システム等でコーパスを利用することを考えている. コーパスを教師データとする機械学習手法による評判情報の抽出において, 1 万文というコーパスサイズが十分なものであるかの確認を行う必要がある. さらに, 本コーパスの公開についても検討している. これについては, Web 上の文書を元データとする際の 2 次配布方法について, さらなる検討が必要と考えられる.

謝 辞

本研究の一部は, (独) 独立行政法人情報通信研究機構の委託研究「電気通信サービスにおける情報信憑性検証技術に関する研究開発」プロジェクトの成果である.

また, 本研究の一部は横浜国立大学環境情報研究院共同研究プロジェクトの援助により行った.

参考文献

Carbonell, J. and Goldstein, J. (1998). “The Use of MMR, Diversity-Based Reranking for Re-ordering Documents and Producing Summaries.” In *Proceedings of the 21st Annual Inter-*

national ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 335–336.

飯田龍, 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一 (2005). 意見抽出を目的とした機械学習による属性—評価値同定. 自然言語処理研究会報告 2005-NL-165-4, 情報処理学会.
乾孝司, 奥村学 (2006). テキストを対象とした評価情報の分析に関する研究動向. 自然言語処理, **13** (3), pp. 201–241.

Kaji, N. and Kitsuregawa, M. (2006). “Automatic Construction of Polarity-tagged Corpus from HTML Documents.” In *Proceedings of the 21st International Conference on Computational Linguistics (COLING/ACL2006)*, pp. 452–459.

洪陽杓, 白井清昭 (2005). 対話行為タグ付きコーパスの作成支援. 言語処理学会第 11 回年次大会発表論文集, pp. 815–818. 言語処理学会.

Kobayashi, N., Inui, K., and Matsumoto, Y. (2007). “Opinion mining from web documents: extraction and structurization.” 人工知能学会論文誌, pp. 227–238. 人工知能学会.

小林のぞみ, 乾健太郎, 松本裕二 (2006). 意見情報の抽出／構造化のタスク使用に関する考察. 自然言語処理研究会報告 2006-NL-171-18, 情報処理学会.

小林のぞみ, 飯田龍, 乾健太郎, 松本裕治 (2005). 照応解析手法を利用した属性-評価値対および意見性情報の抽出. 言語処理学会第 11 回年次大会発表論文集, pp. 436–439. 言語処理学会.

村野誠治, 佐藤理史 (2003). 文型パターンを用いた主観的評価文の自動抽出. 言語処理学会第 11 回年次大会発表論文集, pp. 67–70. 言語処理学会.

野口正樹, 三好健太, 徳永健伸, 飯田龍, 小町守, 乾健太郎 (2008). 汎用アノテーションツール SLAT. 言語処理学会第 14 回年次大会発表論文集, pp. 269–272. 言語処理学会.

Pang, B. and Lee, L. (2004). “A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts.” In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL)*, pp. 271–278.

Pang, B., Lee, L., and Vaithyanathan, S. (2002). “Thumbs up? Sentiment Classification using Machine Learning Techniques.” In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 79–86.

Seki, Y., Evans, D. K., Ku, L.-W., Chan, H.-H., Kando, N., and Lin, C.-Y. (2007). “Overview of Opinion Analysis Pilot Task at NTCIR-6.” In *Proceedings of 6th NTCIR Workshop*, pp. 265–278. NII.

立石健二, 石黒義英, 福島俊一 (2001). インターネットからの評判情報検索. 自然言語処理研究会報告 2001-NL-144-11, 情報処理学会.

峠泰成, 大橋一輝, 山本和英 (2005). ドメイン特徴語の自動取得による Web 掲示板からの意見文抽出. 言語処理学会第 11 回年次大会発表論文集, pp. 672–675. 言語処理学会.

- Turney, P. D. (2002). “Thumbs Up or Thumbs Down? Semantic Orientation Applied to Un-supervised Classification of Reviews.” In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 417–424.
- Wiebe, J., Breck, E., Buckley, C., Cardie, C., Davis, P., Fraser, B., Litman, D., Pierce, D., Riloff, E., and Wilson, T. (2002). “Final Report—MPQA: Multi-Perspective Question Answering.” In *NRRC Summer Workshop on Multi-Perspective Question Answering Final Report*. NRRC.
- Yu, H. and Hatzivassiloglou, V. (2003). “Towards answering opinion questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences.” In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 129–136.

付録

A 予備実験 1 における注釈者への作業指示

作業指示中にて四角で囲われている部分は、実際に注釈者への作業指示へ記述した文ではなく、本稿の他の部分に記述されている文言を挿入した部分である。

目的

Web 上にある、製品レビューに対して、評判情報のモデル化に必要な要素に注釈（タグ）を付与し、コーパスの試作を行う。

今回の作業は、注釈者間におけるタグの付与の違いを調べるのが目的である。

結果

- ・タグ付けが行われたテキストデータ
- ・1 ファイルごとの作業時間
- ・気になる点などがあったら記録しておく

注意点

- ・タグの付与に迷いが生じたときに、他の人に相談をしない
- ・ファイルの中を一回見たら、途中でやめなくてタグ付けを行う
- ・配布したテキストエディタを用いて注釈付けの作業を行う

評判情報のモデルの説明

3 章と同一の評判情報モデルの解説

次ページに続く

評判情報コーパスの作成手順

対象にはあらかじめ item タグが付与してあります。また、対象の階層構造を示したデータも一緒に用意してあります。

* 注意点

item タグが付与されるべきであろう場所に item タグが付与されていないために、作業に支障をきたすような場合。

A: 階層構造はすでに存在する場合→ item タグを追加してしまってもかまいません。

B: 階層構造にも存在しない場合→その度、質問をしてください。

評判情報の構成要素がテキスト中に存在する場合には、下記の注釈付けを行ってください。

1～3 の数字は注釈付けの順番を意味するものではありません。

1: 属性にタグを付与

`<attribute id="a01" pair="p01" target="c01">属性</attribute>`

id: a01, a02…のように id をつける。

pair: p01, p02…のように属性—属性値組を示すための番号をつける。組になる属性値にも同じ番号をつける。

target: 属性を有する `<item>` の class を指定。複数指定の場合はスペースで区切って列挙。

2: 属性値にタグを付与

`<value id="v01" pair="p01" target="c01">属性値</value>`

id: v01, v02…のように id をつける。

pair: `<attribute>` の pair を参照。

target: `<attribute>` の target を参照。

3: 評価にタグを付与

`<evaluation id="e01" target="c01" reason="p01" orientation="positive">評価</evaluation>`

id: e01, e02…のように id をつける。

target: `<evaluation>` が評価している `<item>` の class を指定する。

reason: `<evaluation>` の理由となる `<attribute>`—`<value>` の pair の値を指定する。複数指定する場合にはスペースで区切って列挙。

B 予備実験 2 における注釈者への作業指示

予備実験 2 で利用した作業指示は付録 A に以下の各項目を作業指示の末尾に追加したものである。

attribute-value の組については片方の要素が省略されている場合にはその要素を補完する
省略されている要素の内容がどのようなものになるかを以下の方法で記述する

attribute が省略されている場合には attribute が省略されていると思われる value タグの直前に attribute タグを記述する

value タグが省略されている場合には value が省略されていると思われる attribute タグの直後に value タグを記述する

省略されていた要素を補完して注釈付けを行った例（下線部が省略を補完した部分）

```
<item id="i11" class="c2">このソフト</item>は<attribute id="a101" pair="p101" target="c2" abbr="1">重さ </attribute><value id="v101" pair="p101" target="c2">軽く </value>で <evaluation id="e101" target="c2" readon="p101">気に入っている </evaluation>.
```

C 評判情報コーパスの作成実験における注釈者への作業指示

評判情報コーパスの作成実験における注釈者の作業指示は付録 A に以下の変更と追加を行ったものである。作業指示中にて四角で囲われている部分は、実際に注釈者への作業指示に記述した文ではなく、本稿の他の部分に記述されている文言を挿入した部分である。

- ・予備実験 1 の作業指示から変更した点

目的

今回の作業は、今後利用する本コーパスの作成である。

注釈揺れの分析を行うための作業ではない

次ページに続く

注意点

タグの付与に迷いが生じたときに、注釈者間のみで相談して判断を決めずに、第一著者へ訪ねること。

注釈付けツールを用いて注釈付けを行うこと。

注釈付けの粒度をなるべく細かくすること。とくに attribute と value の粒度に注意すること。

図 6 をここに挿入

注釈付け対象データの配布について

200 文の注釈付けが終わった注釈者へ次の 200 文を配布する。

上述の注意点について補足説明をする。注釈揺れの分析を目的とした実験ではないので、作業者から発注者へ質問があった場合は、発注者が質問とその回答を全注釈者へ連絡している。

・予備実験 1 の作業指示から追加した点

注釈付けの際の規則の追加

図 10, 図 11 と図 12 をここに挿入

注釈付けの順番について

最初に value タグ、もしくは evaluation タグを注釈付けする部分を探しながら注釈付けを行う。

value, evaluation に該当する文字列がどちらも存在せずに、item や attribute のみが現れた場合には注釈付けを行わなくてもよい。

注釈揺れの事例の提示

注釈付けが必要な場所には忘れずに注釈付けを行ってください

例のような場合には注釈付けを行ってください

例：

○：<value>機械的</value>な<attribute>声</attribute>です

×：機械的な声です

次ページへ続く

評価はその表現のみで肯定否定がいかなる場合でも決定できる表現です

例のような value だけでは肯定否定は決定できません

例：

○：届いたときは<value>可愛い</value>と思わずつぶやきました

×届いたときは<evaluation>可愛い</evaluation>と思わずつぶやきました

○：<item>画面</item>の<attribute>色</attribute>は<value>きれい</value>

×：<item>画面</item>の<attribute>色</attribute>は<evaluation>きれい</evaluation>

○：<attribute>色</attribute>が<value>はっきり出て</value>いて<value>見やすい</value>

×：<attribute>色</attribute>が<value>はっきり出て</value>いて<evaluation>見やすい</evaluation>

○：<value>使いやすい</value>です。

×：<evaluation>使いやすい</evaluation>です。

例のような evaluation はその表現のみで肯定否定が決定できるので評価として扱います

例：

○：<item>シール</item>の<attribute>品質</attribute>は<evaluation>かなり悪い</evaluation>

×：<item>シール</item>の<attribute>品質</attribute>は<value>かなり悪い</value>

○：<attribute>内容</attribute>は<evaluation>良い</evaluation>だけに<evaluation>残念</evaluation>

×：<attribute>内容</attribute>は<value>良い</value>だけに<evaluation>残念</evaluation>

次ページへ続く

様態を表す値が属性値で観点しか述べていない部分は属性です

例の場合は attribute として注釈付けしてください

例：

○：購入の決め手としては<attribute>大きさ</attribute>と<attribute>デザイン</attribute>が挙げられます。

×：購入の決め手としては<value>大きさ</value>と<attribute>デザイン</attribute>が挙げられます。

○：その<attribute>シンプルさ</attribute>に<evaluation>感心してしまいます</evaluation>。

×：その<value>シンプルさ</value>に<evaluation>感心してしまいます</evaluation>。

注釈付けはなるべく細かく行ってください

例：

○：<value>その日の気分によって聞ける</value>のが<evaluation>すばらしい</evaluation>

×：<evaluation>その日の気分によって聞けるのがすばらしい</evaluation>

○：<value>熱を持ってしまい</value>とてもではないですが<value>手では持てなくなる</value>

×：<value>熱を持ってしまい、とてもではないですが手では持てなくなる</value>

属性値は様態の内容なので、観点は属性に含めてください。

例：

○：<attribute>茶葉が踊っている</attribute>のが<value>目に見える</value>

×：<attribute>茶葉</attribute>が<value>踊っているのが目に見える</value>

○：<attribute>操作が慣れるまで</attribute><value>ほんの少しかかった</value>

×：<attribute>操作</attribute><value>慣れるまでほんの少しかかった</value>

次ページへ続く

attribute-value 組はそのものがどのようなモノであるかを表すだけであり、その組に対する肯定・否定は人によって異なります

例：

○：<attribute>落ち着き</attribute>が<value>あります</value>

×：<evaluation>落ち着きがあります</evaluation>

○：<item>このディスプレイ</item>は<attribute>解像度</attribute>が<value>高い</value>

×：<item>このディスプレイ</item>は<evaluation>解像度が高い</evaluation>

特に属性は省略されている場合があります。省略を想定しながら注釈付けしてください

例の場合は「サイズ」という属性が省略していると考えられます

例：

○：<item>画面</item><value>大きい</value>

×：<attribute>画面</attribute><value>大きい</value>

上記の作業指示に加えて、別紙にて注釈付けツールの使用説明書を配布。

略歴

宮崎 林太郎：2002 年神奈川大学理学部情報科学科卒業。2004 年同大学大学院理学研究科情報科学専攻博士課程前期修了。現在、横浜国立大学大学院環境情報学府情報メディア環境学専攻博士課程後期在学中。自然言語処理に関する研究に従事。

森 辰則：1986 年横浜国立大学工学部情報工学科卒業。1991 年同大学大学院工学研究科博士課程後期修了。工学博士。同年、同大学工学部助手着任。同講師、同助教授を経て、現在、同大学大学院環境情報研究院教授。この間、1998 年 2 月より 11 月まで Stanford 大学 CSLI 客員研究員。自然言語処理、情報検索、情報抽出などの研究に従事。言語処理学会、情報処理学会、人工知能学会、ACM 各会員。

(2009 年 10 月 3 日 受付)

(2009 年 12 月 15 日 再受付)

(2010 年 6 月 29 日 再々受付)

(2010 年 7 月 27 日 採録)