意見情報の抽出 / 構造化のタスク仕様に関する考察

小林 のぞみ　　乾 健太郎　　松本 裕治
奈良先端科学技術大学院大学 情報科学研究科
{nozomi-k,inui,matsu}@is.naist.jp

ネット上に分散する個人の意見を検索・分類・要約する意見マイニングを目的として，その部分問題である意見情報抽出の課題設定のあり方を論じる．〈 評価対象，属性，評価値 〉からなる従来の構造化の枠組みを拡張し，新しい仕様に基づいて意見タグ付きコーパスを試作することにより，どの範囲の情報を意見として抽出するのが現実的か，抽出した意見をどのような構造で表現すべきかを考察した．

Designing the Task of Opinion Extraction and Structurization

Nozomi Kobayashi　　Kentaro Inui　　Yuji Matsumoto
Graduate School of Information Science, Nara Institute of Science and Technology
{nozomi-k,inui,matsu}@is.naist.jp

The task of opinion extraction and structurization is a key component of opinion mining, which would allow Web users to retrieve and summarize personal opinions scattered over the network. On the bases of our recent experiences of building a opinion-tagged corpus, this paper discusses how the task of opinion extraction and structurization should be designed.

1. はじめに

ブログジャーナリズムのような新語の出現に象徴されるように，個人がネット上に発信する意見情報の重要性が広く認知されるようになった．これに伴い，個人の意見を Web 上のテキストから収集し，分析する技術に関する研究が活発になっている．こうした研究は文書分類に基づくアプローチと情報抽出に基づくアプローチに大別することができる．前者は，テキストを記事や段落 (あるいは文) の単位で肯定的意見か否定的意見かといった意見クラスに分類する課題を設定し，機械学習に基づく様々な手法の実験結果を報告してきた[2),14),18),20)]．一方後者は，意見の構成要素とそれらの間の関係をテキストから抽出する課題に取り組むもので，情報抽出あるいは関係抽出の技術を適用する試みの報告がにわかに増えつつあり[4),5),7),13),15),17)]，我々もテキスト中の意見情報を〈 評価対象，属性，評価値 〉の 3 つ組として抽出する技術の研究を進めてきた[8),9),16)]．

ネット上の不特定の文書集合から個人が発信する意見を抽出し，構造化情報として蓄積することによって，ユーザの関心に合わせて検索したり，分類したりする技術を本稿では意見マイニングと呼ぶ．意見マイニングは，例えば，車の購入を検討しているユーザが，ネット上の広大な領域から収集され整理された個人の意見を効率的に検索・閲覧し，実際のオーナーの「生きた情報」を手かがりに購入車種を比較検討することができるような技術である．これは大きく次の 3 つの部分問題に分けて考えることができる．

意見情報抽出　テキストから意見情報を抽出し，構造化情報として蓄積するタスク

類似意見認識　「トランクが広い」と「収納力がある」のように抽出した意見どうしの同義性・類似性を判定するタスク

意見価値評価　ユーザの関心に基づいて各意見情報の関連度と信頼度を評価するタスク

本稿では，こうした文脈を踏まえながら意見情報抽出に焦点を当て，その課題設定について議論する．次節で述べるように，一口に意見といっても様々な種類の情報があり，どの範囲の情報を抽出の対象とするか，抽出した意見をどのような構造で表現すべきかは必ずしも自明でない．そこで我々は，〈 評価対象，属性，評価値 〉からなる従来の構造化の枠組みを発展させ，新しい仕様に基づいて意見タグ付きコーパスを新規に試作した．以下，課題の基本的な枠組みを 2 節で述べ，コーパスの具体的な仕様を 3 節で報告する．つぎに今回のコーパス作成から得られた知見を 4 節で論じ，最後に関連研究における位置付けに言及する．なおコーパスの仕様のなど，本研究プログラムの詳細は下記 URL で公開しているので参照されたい．

http://cl.naist.jp/~inui/opinion-mining

2. 意見情報の構造化

2.1 抽出すべき意見とは？

我々が考える意見情報抽出の目的は，商品やサービス，人物，組織，政策など，特定のクラスの実体に対して言語的に表明される意見を文書集合から抽出し，要約や検索が容易になるように構造化することである．抽出の対象として検討すべき意見の言明の例を後述の意見タグ付きコーパスから拾うと次のようなものが挙げられる．

(1) 久々に行ったマリアージュですが、相変わらずのおいしさに大感動♪
(2) もっと真剣に燃費について考えてほしいものです
(3) もう二度と行かないと思います
(4) これで定額だったら、速攻機種変なんですが...
(5) プリウスは燃費が良い
(6) ＤＳ「ロストマジック」　独特の世界観が気に入っているのですが、ややグラフィックがショボイですねｗ
(7) 噂ではケーキが美味しいらしい、今度ゆっくり来てみよう
(8) 駅前にあるので便利
(9) もう値引きしないと言われた

まず，(1) の「大感動♪」のような感情の表明はそれを喚起した対象 (「マリアージュ (の料理)」) に対する意見と解釈することができる．(2) や (3)，(4) のようないわゆる心的態度も意見マイニングの目的に照らせば重要な情報である．意見には，(1) の「相変らずのおいしさに」のように感情生起や心的態度の理由が述べられる場合もあり，これも意見マイニングには重要な情報である．感情生起の直接的な原因になるのは，主として「マリアージュの料理がおいしい」のような主観的評価であるが，これが (5) のように感情表現を伴わない場合も，それ単独で意見と見なすことができる．評価は，(6) のように，対象のある特定の側面 (「世界観」「グラフィック」) だけに焦点を当てる場合もあり，こうした評価の視点も構造化には重要である．(7) のように評価の主体は書き手だけに限らないので，その区別も必要であろう．評価は (8) のようにさらにその根拠を伴って出現する場合もあり，そうした根拠の記述も意見情報の重要な構成要素と考えられる．ただし，(9) のように評価 (あるいは感情生起) を伴わない言明もあり，どこまでを意見抽出の対象とすべきか検討が必要である．

2.2 意見の構成要素

感情，心的態度，主観的評価といった概念をまずは整理しておく必要がある．

2.2.1 感情と心的態度

「感情」は悲しみや怒り，喜びなど，その主体が意図的にコントロールすることのできない生理反応であり「心的態度」は感心，叱責，願望，疑いなど，主体がかなりの程度まで意図的にコントロールできる認知プロセスであると言われている．しかし，感情にも反射的な生理反応と見なせるものから高次の認知活動に伴って生起するものものまでという議論もあり (例えば文献[6] を参照)，感情と心的態度の区別はそれほど明確ではない．また，そもそも両者の区別が意見マイニングにおいても重要かどうか判断が難しいという現状もあるので，以下の議論では感情と心的態度を区別せず，まとめて「態度」と呼ぶことにする．

2.2.2 評　　価

認知心理学によると，態度 (感情) の本質は，様々な事象の「望ましさ」について主体が行う評価であると考えられている．たとえば，Lazarus[10] は，人は直面する状況に対して認知的評価を行い，その結果感情が出現すると主張している．また，Frijda[3] によると，感情は「外界や体内の様々な事象がその人にとって重要な目標の達成にどのように影響するかを評価し，適切な行動を準備する過程」と見なされる．言語学でも，Martin[11] が affect (上の議論における態度) と judgement/appreciation (評価) をあわせて attitude と呼ぶなど，態度と評価の関係が論じられている．このように，評価は態度の直接的な理由と解釈できるので，態度とともに意見の中心的構成要素として抽出すべき対象であると考えられる．実際，1 節でも述べたように，意見情報抽出の研究の多くがいわゆる評価の情報を抽出する課題に取り組んでいる．

ところで，ここで言う評価は，事象の望ましさに関する認知的な評価であるから「プリウスは燃費が良い」という認識は評価に当たるが「プリウスの燃費は 12km/l」という認識はそれ自体は評価ではない．両者の間に明確な境界を引けるかどうかは必ずしも明らかでないが，以下の議論では評価の言明をひとまず次のように定義し，客観的な属性–属性値関係の言明と区別する．

(10) 評価とは，評価の主体が評価対象のクラスに対して期待している標準値から実際の評価対象がどれくらい偏位しているか，そしてそれが「望ましい」偏位であるか「望ましくない」偏位であるかに関する評価主体の認識であり，その言明を指す．

「プリウスは燃費が良い」の例で言えば，評価対象である「プリウス」が属する適当なクラスの車の「燃費」に対して評価主体が期待する標準値がどこかにあって「プリウスの燃費」がその標準値よりも「良い (= 望ましい)」と評価主体が認識した，ということの言明と解釈できる．これに対し「燃費が 12km/l」の場合は，そういった標準値との比較に言及せず，客観的な値が述べられているだけなので，上の意味での評価には当たらない．

ただし，この区別は見かけほど単純ではない．たとえば次の例では「ざる蕎麦」の値段「１２００円」が

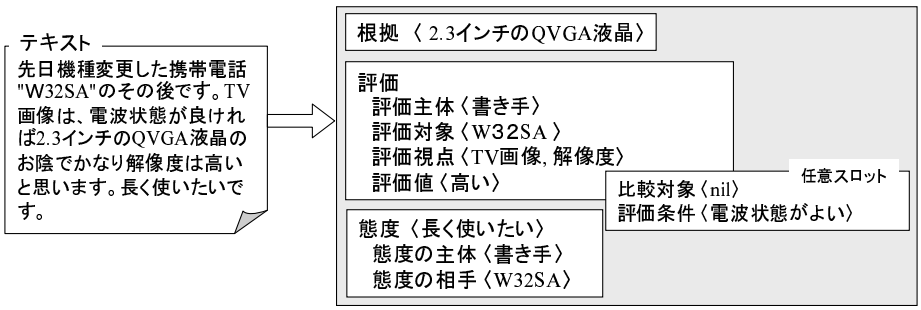

図 1　意見の構造

期待される標準値よりも高かったという主体の認識が副助詞「も」によって言語化されている．(10) の定義はこれを評価と判別する．

(11)　ざる蕎麦１枚で１２００円もした

2.2.3　態度・評価以外の意見情報

2.1 で見たように，意見抽出の対象になりうる言明には，態度や評価のような明示的なものの他に，態度や評価の原因・根拠となる事態に関する叙述がある．前掲の「駅前にあるので便利」では「駅前にある」という事態が「便利」であるという評価に対する根拠として述べられている「駅前にある」ことが評価対象の利点として認識されているという意味で，こうした情報も意見マイニングでは有用と考えられる．

ただし，次の例のように事態の記述が評価や態度を伴わずに単独で出現する場合は注意が必要である．

(12)　もう値引きしないと言われた

(13)　シートがフルフラットになる

(14)　１日で２００キロ（車で）走った

事態 (12) は店の対応に対する否定的な評価，事態 (13) は車の装備に対する肯定的な評価を含意すると解釈できなくもない．実際に文脈を見ると，そう解釈するのが自然であった．一方，事態 (14) の場合は，それが何らかの評価を含意すると解釈すべきか（すなわち，意見の一部と見なせるか）明らかでない．こうした解釈の不確かさは，たとえばコーパスに対するタグ付けの品質を落し，意見抽出モデルの性能評価を困難にする原因となるので，何らかの対策が必要である．

2.3　意見情報抽出課題

上で述べた意見の構成要素の分類を作業仮説として，次の 4 種類の情報を抽出する意見情報抽出 / 構造化問題を考える．

態度　心的態度や感情の明示的な言明，およびその態度の相手と対象

評価　主観的評価の明示的な言明，およびその評価の主体と対象

根拠　態度や評価の理由・根拠となる言明のうち，評価でないもの

理由関係　評価と態度，根拠と評価，あるいは根拠と態度の間の原因– 反応関係

図 1 に具体例を示す．図のテキストには態度，評価，根拠に相当する記述がいずれも含まれている．態度と評価の主体はいずれもテキストの書き手，対象は「Ｗ３２ＳＡ」という携帯電話の機種である．

また，図 1 に例示したように，評価についてはさらに次の下位構成素を考えることができる．

評価主体　評価を行った主体を指す表現

評価対象　製品やサービス，人物，組織，政策など，意見マイニングの対象となる特定のクラスの実体を指す表現

評価視点（任意）　評価対象のどの構成要素あるいは属性が評価の焦点になっているかを示す表現

評価値　評価対象全体あるいは評価視点に対する評価の値を示す表現

比較対象（任意）　評価対象の相対的な評価において参照される比較の対象

評価条件（任意）〈評価値〉で表現される評価が成り立つための条件

評価対象と評価視点の関係には主として全体–部分関係（製品と部品など），全体–役割関係（店と店員など），対象–属性関係（製品と性能など）がある．評価対象全体に対する評価の場合は評価視点は空であると見なす．また，評価主体や評価視点がさらにその内部に階層構造を持つ場合があることに注意する必要がある．例えば，図 1 の例で「高い（評価値）」のは「((Ｗ３２ＳＡの）ＴＶ画像の）解像度」であり，評価視点を構成する「ＴＶ画像」と「解像度」の間に対象–属性関係を認めることができる．意見の構造化表現では，これを〈ＴＶ画像，解像度〉のようなリストで表し，最上位の評価対象から評価視点の階層の末端に至る〈Ｗ３２ＳＡ，ＴＶ画像，解像度〉のようなパスを評価対象のチェーンと呼ぶ．

3.　意見タグ付きコーパス

課題仕様の問題点を明らかにするために，2.3 での議論に基づき，実際に文書集合に対してタグの付与と関係の同定を行なった．

3.1　作 業 設 定

2.3 で述べた構造の持つ情報のうち区別がはっきり

表 1　評価値表現の例

品詞	数	例
名詞	2,797	うってつけ，えり抜き，贅沢，不便
形容詞	864	大きい，高い，良い，面倒くさい，明るい
動詞	577	いかす，いらだつ，見損なう，喜ぶ
副詞	188	ごちゃごちゃ，スッキリ，ぎすぎす，ダラダラ
未知語	115	カワイイ，カンペキ，GOOD，ヒドイ
文節内，複数形態素	437	必要十分，罰金もの，平均点，捨てがたい
文節を越える表現	619	恐いもの知らず，意地が悪い，胸が躍る

している情報のみに限定して考える．

評価と態度は，2.3 で述べたように概念的な区別はあるものの，その間の境界は明確ではない．例として以下の文を考える．

(15)　室内は静かで落ちつく / 落ちついた室内

(16)　昼の西欧亭ランチはオススメ

(15) はいずれも「落ちつく」という表現を含む例であるが，前者は「室内は静か」に対する態度，後者は「室内」に対する評価と解釈できる．また，(16) の場合『オススメ』は「昼の西欧亭ランチ」に対して何らかの評価を行った結果生じた態度と考えられるが，何らかの評価が省略されているため『室内は静か』と同じ形をしている．これらの例からわかるように，実際に個々の事例で評価と態度を区別するのはそれほど容易でない．そこで今回は評価と態度を区別せずに『X に対する態度 Y』を「X に関する何らかの評価の結果生じた態度 Y」と考えて，X を評価対象もしくは評価視点とし，Y を評価値/態度としてタグを付与する．以降では，評価値と態度を併せて評価値と呼ぶ．ただし，態度に関しては未整理の状態であるので，明示的な感情や態度 (嬉しい，おすすめ，など) のみを対象とした．

以上から，今回の作業では，評価主体，評価対象，評価視点，評価値，比較対象，評価条件，根拠を付与するタグのセットとする．

実際の作業は，それぞれの情報にタグを付与する作業と，その表現と評価値，評価視点との関係を同定する作業から成る．ここでいう関係は，評価視点–評価値，評価視点–評価視点，評価対象–評価値，評価対象–評価視点の 4 つの関係である．3.2 で述べるように，作業は下から上へ，つまり評価値から始めて評価視点を経て評価対象へとすすめていくことになる．これは，評価値を基点として考えていて，評価値がない場合はタグ付与，関係同定を行わないためである．

以下では今回のタグ付与作業のために取り決めた点について述べる．

根拠:　2.2.3 で述べたように，評価を伴わずに単独で出現する根拠は，評価を含意すると解釈すべきかの判断が明らかでない．そこで，今回は評価を伴って出現する根拠のみを「根拠」とした．例えば，(8) の「駅前にある」は「便利」という評価を伴っているので根拠とみなすが『駅前にある』が単独で出現している場合は根拠とみなさない．

評価値:　どこまでを評価値とするかが揺れることを回避するために，タグの付与範囲を「評価をあらわす最小の範囲」に制限する．たとえば，

(17)　味は普通においしいと思います

の場合，評価値として『普通においしいと思います』『普通においしい』『おいしいと思います』，「おいしい」の 4 つが考えられ，どこまでを評価値とするかは判定者に依存すると考えられる．この問題を避けるために『思う』『感じる』などは評価値に含めず，かつ程度をあらわす表現は評価値に含めないとした．この例では『おいしい』のみを評価値とする．

評価対象:　2.3 で述べたように，評価対象は階層構造を持つ場合がある．例えば車のドメインを考えると，アコード，アテンザ，プリウス，レガシィなどが評価対象となるが，ボディの種類によりセダン，ワゴンなどに分かれ，さらにボディの種類はグレードを持つ．今回のタグ付与では評価対象間の階層関係は区別せずに，アテンザと書かれていればアテンザを評価対象とし，アテンザスポーツと書かれていれば，アテンザスポーツを評価対象とした．

評価視点:　評価視点は内部に階層構造を持っている．ここでいう階層構造は，全体–部分関係 (エンジン–ピストン)，対象–属性関係 (エンジン–性能) などを指す．以下，これらの関係を評価視点間の階層関係と呼ぶ．評価値と直接リンクを張るのは評価視点の階層の末端の評価視点として，階層関係が存在すればその階層関係を同定する．

評価主体:　評価主体がテキスト中に明記されている場合のみタグを付与し，評価主体が書き手である場合はタグを付与しない．

3.2　作業の手順

評価値の候補は辞書引きによりあらかじめ与えた上で，以下の手順で作業を行った．

(i)　その候補表現が評価として使われていれば候補表現に評価値タグを付与し，その評価視点を同定する．

(ii)　評価視点が評価対象でなければ，評価対象に到達するまで評価視点間の階層関係を同定する．この際，特定の商品もしくは店についての評価で

あるにも関わらず，評価対象がテキスト中に存在しない場合は「評価対象省略」を付与する．たとえば，

(18) そういえば 3 月にＮＹで食べた揚州炒飯はあっさりしてて美味しかった。

からは『揚州炒飯–美味しかった」という評価視点と評価値を同定できるが『揚州炒飯」の店名はテキスト中に存在していない．このような場合は「揚州炒飯」に評価対象省略タグを付与する．

(iii) 評価条件，根拠，評価主体が文章中に存在すれば，それらにタグを付与する．

以下，レストランに関する blog 記事にタグを付与した例をあげる．下線部は評価値を { }は評価対象を，⟨⟩ は評価視点を表している．$\langle\rangle_{n-m}$ の n は評価値との対応を，m は階層の深さを表す．m が小さいほど下位の評価視点になる．

私たちが食べに行ったのがこのお店『{ ビストロヒロ }$_{a,b,c,d,e}$』です。
このお店、⟨ 外観 ⟩$_{a-1}$ はちょっと こじゃれた$_a$ 洋風の ⟨ 建物 ⟩$_{a-2}$ で、⟨ 店内 ⟩$_{b-2}$ はバーカウンターもあり凄く 落ち着いた$_b$ ⟨ 感じ ⟩$_{b-1}$ です。
⟨ お味 ⟩$_{e-1}$ の方ですが、まずは ⟨ 特製オムライス ⟩$_{c-2,d-2,e-2}$ から。
上に乗っかっている ⟨ 玉子 ⟩$_{c-1}$ は ふわとろ$_c$。
中の ⟨ ご飯 ⟩$_{d-1}$ も少し 硬め$_d$ でとっても 美味$_e$ ♪

上の例から評価値，評価視点，評価対象の関係を抽出すると，以下の 5 例が得られる．

a. ⟨ ビストロヒロ，⟨ 建物，外観 ⟩，こじゃれた ⟩
b. ⟨ ビストロヒロ，⟨ 店内，感じ ⟩，落ち着いた ⟩
c. ⟨ ビストロヒロ，⟨ 特製オムライス，玉子 ⟩，ふわとろ ⟩
d. ⟨ ビストロヒロ，⟨ 特製オムライス，ご飯 ⟩，硬め ⟩
e. ⟨ ビストロヒロ，⟨ 特製オムライス，お味 ⟩，美味 ⟩

3.3 評価値か否かの判定基準

基本的には，辞書引きされた評価値候補に対して，ある文脈で評価値か否かの判定を行うが，中には評価値候補ではないが，評価値である表現が存在することが予想される (たとえば『すげー」「ウザい」など) ．これらの表現に対しても上記の手順でタグを付与する．

ある表現が評価値か否かの最終的な判断は作業者の主観に任せることになるが，少なくとも『評価対象もしくはその側面に対する主観的な評価であること」が必須になる．たとえばレストランドメインで，ある表現が評価値であるのは，その表現が「店かもしくは店の側面に対する評価」として使われているときである．もし記事中に「天気が 良かっ たので，外食することにした」という記述があっても，レストランドメインでは「良い」は評価値とはしない．

3.4 評価値表現辞書

評価値候補の同定に使用する辞書について述べる．

文献[9] で提案した手法をもとに収集した表現を既存の資源を用いて拡張し，約 5,600 表現の辞書を作成した☆．ある表現を辞書に登録するかしないかは，その表現が評価値となり得る可能性があるか否かにより判定した．そのため，辞書に含まれる表現が必ず評価値になるというわけではない．

形容詞は評価を表す表現としてよく知られているが，「赤い」のように主観的な評価として使われない形容詞も存在するため，今回は形容詞に対しても人手で評価値として使われるか否かを判定した．表 1 に収集した表現の品詞とその品詞に属する表現の数，表現の例を示す．各表現は品詞付きで辞書に収められているが，この品詞は IPA の品詞体系[1] に従っている．

この辞書の他に，評価値になり得る表現のパタン (たとえば『~すぎる/やすい」) を 10 程度用意しているが，今回の作業では使用していない．

3.5 対象としたドメイン

対象ドメインとして，特徴の異なる 4 つのドメイン，レストラン，車，携帯電話，ゲームを選んだ．レストランでは建物，雰囲気，店員，料理など，さまざまな店の側面が評価の対象となり，車や携帯電話では本体と構成素，性能，デザインなどが評価の対象となる．一方ゲームはシナリオ，キャラクター，音楽など，具体的な物ではなくコンテンツが評価の対象となることが多い．特徴の違うドメインを選んだ理由は，評価値，評価対象の関係や評価対象間の階層関係の出現の傾向がドメイン間でどの程度違うかを調査するためである．

タグ付与には，livedoor Blog☆☆の blog 記事を使用した．レストランとゲームに関しては，グルメとゲームのカテゴリに属する記事を収集した．携帯電話と車に関しては，まずレビューサイトからそれぞれの評価対象のリストを収集して，評価対象を含む記事を livedoor Blog から収集した．

4. 作成したコーパスの分析と議論

表 2 に，作成した意見タグつきコーパスの規模，タグ数を示す．

この表から，レストランやゲームでは，ゲームもしくは店のある側面が評価されることがかなり多いことがわかる．この結果から，評価視点の階層関係は無視できる数ではなく，その関係を考慮していくことは重要な課題だと考えられる．また，車や携帯電話は比較対象が多く出現しているドメインであることがわかる．これは，車や携帯電話では，買い換えの際に自分が今持っているものと他の商品を比較することが多いためと考えられる．

4.1 評価値表現辞書の被覆率

作成した評価値表現辞書は，どのドメインにおいても，評価値としてタグが付与された表現の約 9 割をカバーできていた．残り 1 割は辞書になかった表現だ

☆ この辞書は 2006 年 4 月に公開する予定である
☆☆ http://blog.livedoor.com

表 2　意見タグ付きコーパス

	レストラン	車	携帯電話	ゲーム
記事数 (括弧内は文数)	1,445(25,500)	564(14,593)	494(12,326)	361(6,823)
評価値候補	14,077	10,026	7,897	3,878
評価視点-評価値	4,504	1,017	1,144	551
評価対象-評価値	622	577	584	242
評価視点-評価視点	2,054	280	304	221
評価対象-評価視点	3,253	876	881	451
評価対象省略	61	35	15	27
根拠	68	86	80	95
条件	113	86	76	41
比較対象	32	66	75	8
評価主体 (書き手以外)	111	24	22	2

表 3　作業者間の一致

	評価と判定	評価でないと判定
作業者 A	374	1365
作業者 B	337	1402
判断が一致	277	1304

表 4　評価対象チェーンの一致

	完全一致	一部一致	その他	合計
作業者 A	157	51	74	282
作業者 B	157	51	84	292

が，それらは「アツアツ」¦ウザイ」などのカタカナ語や「食べやすい/にくい」などの表現であった．後者は 3.4 で述べたパタンで対応し，前者は辞書に追加するつもりである．

4.2　作業者間の一致率

どの程度揺れなくタグを付与できるかを調べるために，2 名が同じデータに上述のタグを付与し，その一致率を評価した．

評価データはコーパス作成に使用したデータとは別の blog 記事 116 件で，ドメインはレストランとした．3.2 と同様に評価値候補はあらかじめ与えた．データ中の評価値候補の数は 1,756 表現であった．

4.2.1　評価値の判定の一致率

表 3 は，評価値候補に対して二名の作業者のタグがどの程度一致したかを示している．評価値の一致率を計算したところ，κ 値は 0.7 となった．それゆえ，評価値に関してはある程度判断が一致したといえるだろう．

一致しなかった評価値のうち，代表的な例をいくつか示す．下線はその表現が評価値であることを示す．

(19)　このから ~い麻婆豆腐と 小さい ご飯のセット。

(20)　ハンバーグは ふんわり 分厚く焼かれており...

(20) は「ふんわり」を評価値としたか，しなかったかの判断が揺れた例である．この揺れは，評価値表現が動詞を修飾している場合に，評価値タグを付与するか否かをきちんと指示していなかったことが原因と考えられる．このような場合は「ふんわり」を評価値¦ハンバーグ」を評価視点とするように統一したい．

4.2.2　評価視点間の同定の一致率

二人の判定者が評価値と判定した 277 の評価値に対して，評価対象のチェーンがどの程度一致したかを調べた．評価対象のチェーンは，評価対象から末端の評価視点までのパスである．

表 4 は，評価対象のチェーンがどの程度一致したかを示している．表中のチェーンの数が一致した評価値の数より多いのは，

(21)　〈スッポンスープ〉も〈小龍包〉も 絶品 でしたのように，一つの評価値 (この例では「絶品」) が複数の評価視点を持つ場合がある (「スッポンスープ」，「小龍包」) ためである．

表中の「一部一致」は，チェーンの一部が抜けている場合を指している．以下の例は一部一致として数えた例である．

(22)　渋谷のラーメン屋に入った。面倒臭いので日替わりラーメン定食をたのんだ。その日はレバニラ炒めで...

この例に対して，作業者 A は〈渋谷のラーメン屋，日替わりラーメン定食，レバニラ〉をチェーンとし，作業者 B は〈渋谷のラーメン屋，レバニラ〉をチェーンとしていた．この問題は，評価視点間の階層関係に関する仕様が不十分であることに起因している．この例の場合は，作業者 A のチェーン，B のチェーンのいずれでも良いと考えられるが，たとえば¦アコードのエンジンはいい音だと思う」の場合は〈アコード，エンジン，音〉とする必要があり，〈アコード，音〉は誤りになる．どのような場合に一つの解釈しかできないのかについては今後調査，検討していく必要がある．

表中の「その他」に分類している例をさらに細かくみると，全体に対する評価か一部に対する評価かの判定が揺れた例が 36 件 (その他の約 2 割)，評価視点や評価対象 (店名) のチャンクがずれた例が 77 件 (その他の約 6 割) あった．

(23)　まあつけ麺がかなりおいしかったので 満足しております。

(24) 串モノ以外は奴、枝豆、キムチ、合鴨スモーク、ちぢみなど。串焼きに欠かせないキャベツは、ザク切り二切れ。マヨネーズ、味噌付き。いずれも３００円以内と リーズナブル。

(23)，(24) は全体に対する評価か一部に対する評価かがゆれた例である．(23) は店に満足したのか，つけ麺に満足したかで揺れていて，(24) はリーズナブルが店全体に対してか，列挙された料理に対してかで揺れていた．

チャンキングの問題では，括弧書きされた部分を含めるか含めないかの揺れが多かった．例をあげると，「チャプチェ(春雨炒め)」に対して，作業者 B は「チャプチェ」だけを評価視点とし，作業者 A は括弧まで含めて評価視点としていた．チャンクのずれに関しては，いずれかに統一する必要はあるものの，一致したと見なしてよいと考えられる．その数を完全一致に含めると，作業者の出力したチェーンのうち約 7 割は一致するため，一致率は必ずしも低くはない．

4.2.3 根拠同定の一致率

つぎに，評価値の判定が一致した 277 件のうち，その根拠の同定がどの程度一致したかについても調査した．同定した根拠の数は作業者 A は 7 例，作業者 B は 16 例といずれも少なく，さらに判断が一致したのはわずかに 4 例であった．一致しなかった例のほとんどは，作業者のいずれかが評価に対して根拠を選んでいなかったことが原因であった．以下の (25) は根拠が一致した例で，(26) は作業者の一方しか根拠としなかった例である（〈〉内は同定された根拠を表す）．

(25) 店員さんの冷たい態度と同じく、〈トマトも冷えていた〉ので 減点。

(26) 〈その場ですり下ろした〉わさびも最高に美味しい。

(26) は「その場ですり下ろした」ことが「わさびが美味しい」ことに影響するか否かの解釈が揺れた例だと考えられる．その店で出されたわさびそのものが美味しかったと考えると「その場ですり下ろした」ことは理由ではない．しかし「その場ですり下ろした」からこそ「美味しかった」と考えると，美味しかったことの根拠となる．どちらの解釈をとるかは判定者に依存するため難しい問題である．

つぎに，一例だけ二人が異なる根拠を選んでいたケースがあったのでその記事をあげる．

(27) 添え付けの季節の野菜なんてカボチャフライだけでしたよ。茄子とかピーマンとかつけなくなったのはなぜなんでしょうか。ちょっと がっかり でした。

評価値「がっかり」の根拠として，作業者 A は「茄子とかピーマンとかつけなくなった」ことをあげ，作業者 B は「添え付けの季節の野菜なんてカボチャフライだけ」の部分をあげていた．この二つは，内容的には同じことを述べているためどちらも「がっかり」の根拠であるといえる．しかしながら，この二つが同じ内容であることを機械が判定するのは難しいと考えられる．

4.3 得られた知見

タグ付与作業を通して明らかになった点を以下にまとめる．

- 与えられた評価値候補が評価値か否かの判断はある程度一致することがわかった．また，少なくとも今対象としているテキストでは，これまでに収集した評価値表現がドメイン横断的に利用でき，かつほとんどの評価値を被覆できていることがわかった．
  一方で，明示的でない心的態度については今回はタグを付与していないため，感情以外の態度がどのような形で出現するかの調査は今後の課題である．
- 評価視点間の階層関係の数は無視できる数ではなく，評価視点間の階層関係を考慮することは重要である．
  しかしながら，現在の評価視点の階層関係に関する定義は十分ではなく，どのような場合を階層関係としてとらえるべきかは議論が必要である．今後はオントロジに関する最近の研究成果[12] などを参考に調査，検討する．
- 評価条件や比較対象に関してもその出現数は無視できる数ではない．その二つに関する作業者間の一致を調査することは今後の課題である．
- 評価の理由として出現している場合であっても，根拠を同定することは難しい問題である．今回の結果を見るかぎりでは，単独で根拠を抽出する問題はさらに困難であることが予測される．

5. 関連研究

意見をどのように構造化するかについては，最近さまざまな議論がなされてきている．

たとえば，Hu ら[5] は評価視点と評価視点に対する肯定/否定の値を抽出するタスクに取り組んでいて，Popescu ら[15] は，評価視点と評価値，その肯定/否定の値を同定する問題を考えている．一方，村野ら[13] は，主観的な評価を含む文を人手で分析し「対象，比較対象，項目，様態，評価」の 5 つの構成要素から成る 8 つの文型を，評価を含む文のパタンとして定義している．構成要素のうち，対象，比較対象，項目はそれぞれ我々のいう評価対象，比較対象，評価視点と対応する．彼らは「評価」と別に「安い」「弱い」など評価視点の持つ性質を表す表現として「様態」という要素を定義している.

また，新聞記事を対象としたタグ付与が Wiebe らによって行われている．その仕様については文献[19] で詳しく述べられている．Wiebe らは private state に着目し，陽に private state が述べられている場合とそ

うでない場合についてそれぞれ議論している．private state とは意見，信念，考え，気持ち，目標，感情など，心的，情緒的な状態を包含する一般的な用語である．

言語学の分野では，主観性に関する言語理論が提案されている．Martin[11] は attitude（態度）を affect, judgement, appreciation から成ると定義していて，さらに人に対する評価を juegement，物に対する評価を appreciation として区別している．

このように，構造化に関するさまざまな知見が得られてきているため，今後はそれらの情報を比較し，得られている知見を整理する必要がある．

6. おわりに

本稿では，意見をどのように構造化すべきかについて議論し，一つの案として〈根拠，評価，態度〉に分類してこの問題にアプローチする枠組みについて述べた．また，この枠組みに基づいて実際に意見タグ付きコーパスを作成し，課題仕様の問題点を調べた．評価値の識別については作業者間である程度判断が一致する（$\kappa = 0.7$）という結果が得られたが，評価視点の定義の不備や根拠の認識の難しさといった問題も明らかになった．今回調査では複数のドメインを取り上げたが，それでも対象とした文書集合は比較的均質なものであった可能性は否めない．たとえば新聞記事や論説文，自由回答アンケートのような文書に出現する意見は 2 節で述べた構造化の枠組みに馴染まない可能性もある．意見の構造化表現とその実現方法について一般性と現実性のバランスを探る努力を今後も続ける必要がある．

現在，2.3 で述べた課題設定に基づいて，評価を抽出するためのモデルを文献[8] のモデルをベースにして開発中である．その結果は別の機会に報告したい．

参考文献

1) 浅原正幸, 松本裕治: ipadic version 2.7.0 ユーザーズマニュアル (2003).
2) Dave, K., Lawrence, S. and Pennock, D. M.: Mining the peanut gallery: opinion extraction and semantic classification of product reviews, *Proceedings of the 12th International World Wide Web Conference* (*WWW*), pp. 519–528 (2003).
3) Frijda, N. H.: *The Emotions*, Cambridge University Press (1986).
4) Gamon, M., Aue, A., Corston-Oliver, S. and Ringger, E. K.: Pulse: Mining Customer Opinions from Free Text, *Proceedings of the 2005 Conference on Intelligent Data Analysis* (*IDA*), pp. 121–132 (2005).
5) Hu, M. and Liu, B.: Mining and summarizing customer reviews, *Proc. of the Tenth International Conference on Knowledge Discovery and Data Mining*, pp. 168–177 (2004).
6) 乾健太郎, 徳久雅人, 徳久良子, 岡田直之: 感情の生起とその反応, 日本ファジィ学会誌, Vol. 12, No. 6, pp. 741–751 (2000).
7) Kanayama, H. and Nasukawa, T.: Deeper Sentiment Analysis Using Machine Translation Technology, *Proc. of the 20th International Conference on Computational Linguistics*(*COLING*), pp. 494–500 (2004).
8) Kobayashi, N., Iida, R., Inui, K. and Matsumoto, Y.: Opinion extraction using a learning-based anaphora resolution technique, *The Second International Joint Conference on Natural Language Processing* (*IJCNLP*), *Companion Volume to the Proceeding of Conference including Posters/Demos and Tutorial Abstracts*, pp. 175–180 (2005).
9) 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: 意見抽出のための評価表現の収集, 自然言語処理, Vol. 12, No. 2, pp. 203–222 (2005).
10) Lazrus, R. S.: *Pshchological stress and the coping prosess*, New York : McGraw-Hill (1966).
11) Martin, J. R.: Introduction, Special Issue: Negotiating Heteroglossia: Social Perspectives on Evaluation, *Text*, Vol. 23, No. 2, pp. 171–181 (2003).
12) 溝口理一郎, 池田満, 來村徳信: オントロジー工学基礎論–意味リンク，クラス，関係，ロールのオントロジー的意味論–, 人工知能学会, Vol. 14, No. 6, pp. 87–100 (1999).
13) 村野誠治, 佐藤理史: 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 9 回年次大会 発表論文集, pp. 67–70 (2003).
14) Pang, B. and Lee, L.: A Sentiment Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts, *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics* (*ACL*) , pp. 271–278 (2004).
15) Popescu, A.-M. and Etzioni, O.: Extracting product features and opinions from reviews, *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), pp. 339–346 (2005).
16) 立石健二, 福島俊一, 小林のぞみ, 高橋哲朗, 藤田篤, 乾健太郎, 松本裕治: Web 文書集合からの意見情報抽出と着眼点に基づく要約生成, 情報処理学会研究報告 NL-163, pp. 1–8 (2004).
17) 立石健二, 石黒義英, 福島俊一: インターネットからの評判情報検索, 情報処理学会研究報告 NL144-11, pp. 75–82 (2001).
18) Turney, P. D.: Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (*ACL*), pp. 417–424 (2002).
19) Wiebe, J.: Instructions for annotating opinions in newspaper articles, Technical report, Department of comuputer science, University of Pittsburgh, TR-02-101 (2002).
20) Yu, H. and Hatzivassiloglou, V.: Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences, *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), pp. 129–136 (2003).