

宿泊レビュー分析による特徴情報の抽出に関する研究

2014年 3月

辻井 康一

宿泊レビュー分析による特徴情報の抽出に関する研究

辻井 康一

システム情報工学研究科

筑波大学

2014年 3月

宿泊レビュー分析による特徴情報の抽出に関する研究

目次

第1章 緒論	1
第2章 インターネット宿泊予約の評価指標	7
2.1 インターネット宿泊予約	8
2.1.1 インターネット宿泊予約サイト	8
2.1.2 宿泊施設の分類	11
2.1.3 宿泊予約に伴う宿泊検討者の知覚リスク	14
2.2 インターネット宿泊予約の情報収集	17
2.2.1 提供者発信情報	17
2.2.2 利用者発信情報	18
2.2.3 宿泊レビュー	18
2.2.4 レビュー情報の特性	20
2.2.5 宿泊施設の評価	21
2.3 テキストマイニング	24
2.3.1 自然言語処理による評価表現抽出の概要	24
2.3.2 形態素解析	25
2.3.3 構文解析	39
2.3.4 意味解析	42
2.3.5 評価情報抽出	45
2.3.6 辞書編纂	46
2.3.7 宿泊レビュー分析	47
2.4 結言	49
第3章 宿泊レビューを用いたエリア特徴の分析	51
3.1 宿泊レビューの利用目的	52
3.2 宿泊エリア分類	53
3.3 数値評価の分析	54
3.3.1 数値評価項目	54
3.3.2 宿泊利用者の採点	57
3.4 感想コメントの分析	59
3.4.1 テキストマイニング	59

3.4.2	形態素辞書	60
3.4.3	類義語辞書	61
3.4.4	感想コメントの肯定・否定の評価判定	61
3.4.5	感想コメントにおける属性の確定	63
3.5	テキストマイニングによるエリア分析	65
3.5.1	分析データと分析システムの概要	65
3.5.2	エリア別特徴表現	66
3.5.3	エリア別数値評価との対応	68
3.5.4	地区別分析	71
3.6	活用方法の提案	76
3.7	結言	78
第4章	宿泊レビューを用いた宿泊施設別の特徴分析	79
4.1	訪日旅行者の宿泊予約	80
4.2	訪日旅行者の宿泊施設評価	82
4.3	訪日旅行者の宿泊レビュー	84
4.4	テキストマイニング	87
4.4.1	類義語辞書	87
4.4.2	数値評価対応辞書	88
4.4.3	評価属性辞書	89
4.5	感想コメント分析	91
4.5.1	エリア全体の特徴	92
4.5.2	個別ホテルの特徴比較	94
4.6	結言	101
第5章	宿泊レビューの評価表現からのサービス改善情報の抽出方法	103
5.1	感想コメントの宿泊施設の評価情報	104
5.2	サービス改善のための評価表現	105
5.2.1	評価極性が遷移する表現	105
5.2.2	評価の組み合わせの抽出方法	106
5.2.3	辞書の登録	107
5.3	評価極性の遷移する表現の抽出	109
5.4	評価の遷移する数値評価項目の抽出	111
5.5	宿泊利用目的の違いによる特徴分析	118
5.6	結言	121

第 6 章 結論.....	123
謝辞.....	127
参考文献.....	129
関連業績リスト.....	139

第 1 章

緒論

インターネット宿泊予約サイトの登場により、旅行における宿泊予約の方法は大きく変化した。宿泊検討者は 24 時間いつでもホテルや旅館などの宿泊施設の予約手段を手に入れることができるようになり、旅行に出かけるための手続きが簡便になった。インターネット宿泊予約サイトが登場するまでの宿泊施設の予約方法は宿泊施設に連絡し、空室の状況を確認して予約を申し込むか、駅前やショッピングセンターにある旅行会社に出向き、カウンター窓口で相談し、希望の宿泊施設を予約するのが主流であった[Adachi 2011]。

宿泊検討者は、インターネット宿泊予約サイトを利用し、希望の日程・予算・人数、旅行したいエリアを選べば、該当する複数の宿泊施設を一括で検索し、自分の希望に合った宿泊施設を簡単に比較検討し、予約することができるようになった。インターネット宿泊予約サイトの検索機能と 24 時間いつでもネット接続環境があれば予約可能なサービスにより、インターネットで宿泊予約を行う利用者は急速に増え、その利用者はこの 10 年で 10 倍を超える成長を示し、目覚ましい成長を遂げている[JATA 2013]。

宿泊施設は、自前の販売に加え、旅行検討者の集客力と販売力を持った旅行会社へ客室を提供し、代理販売を委託していたが、インターネットの登場により、インターネット宿泊予約サイトという新たな流通チャネルへの客室提供という、新しい選択肢が現れた [Kano 2008]。インターネットの普及は旅行流通に影響を与え、宿泊サービスの提供者側である宿泊施設においてもインターネット宿泊予約サイトへの客室提供を増やすなどの活用が広がっている[JATA 2004]。インターネット宿泊予約サイトは、旅行会社への客室提供と比較して、宿泊施設の空室状況の繁閑に合わせて、リアルタイムに掲載客室在庫の提供と

販売価格の管理ができるため、客室販売を自ら細かくコントロールすることができる。宿泊施設へのアンケートによると、宿泊予約の販売経路の構成比は旅行会社からインターネット宿泊予約サイトへと移行が進んでおり、特に小規模な宿泊施設でその傾向が顕著となっている[Ryoken 2010]。またインターネット上での宿泊予約の集客においては、宿泊予約サイトへの登録が最も実施効果が高いとの結果があり、自社予約サイトを構築するよりも高い評価を示している。これは宿泊施設が単体で宣伝告知をする力は限られているが、宿泊施設が多く集まり、宿泊検討者の比較検討できるインターネット宿泊予約サイトへの参加することで、施設規模の大小に関係なくそれぞれの宿泊施設の特徴を表現できるようになったからだと考えられる。

そして昨今、海外から日本に渡航する訪日旅行者の増加が顕著であるが、特にインターネットは日本への旅行情報の取得において不可欠なものとなっており、訪日旅行者の宿泊予約手段として、高い比率でインターネット宿泊予約サイトの利用が選択されており、外国人の日本訪問を大きく発展させる役割を果たしている[JNTO 2013]。一部では自社のホームページに外国語表記の実施を行うよりも高い集客効果が出ているとの結果もある[ICHSA 2012]。

インターネット宿泊予約サイトは、旅行計画する際に活用するメディアとして利用されており[goo 2013]、自らの意思で自由に宿泊施設を選択する際には便利である。その一方で、宿泊施設に直接連絡を取る場合や旅行会社の店舗での申し込みと異なり、宿泊施設やそのサービス内容についての問合せをしたり、宿泊検討者は自分の希望条件に合った宿泊施設がどれかについて旅行に詳しい専門家からアドバイスを受けたりすることが困難で、宿泊施設選択には不安が残る。そのため、宿泊検討者は自分にとって最も良い条件の宿泊施設を選ぶためには宿泊予約サイトが提供する様々な情報を活用することが必要になる。

宿泊施設を選択する際に利用される情報は大きく分けて、宿泊施設が提供する情報と、宿泊利用者が提供する宿泊レビューに分けられる。宿泊施設が提供する情報は、宿泊利用者にどのような部屋が提供し、どのような特典がついているか宿泊条件を示す商品情報、宿泊施設自体の規模や概観、設備、ブランドイメージなどの情報がテキストと画像で提供される。これらは、いわば宿泊施設がサービス提供者の立場で発信する公式情報である。自社ホームページやイ

インターネット宿泊予約サイトに記載される宿泊施設情報がこれらに当てはまる。一方で、宿泊利用者側が発信する客観的な情報として宿泊レビューがある。宿泊レビューは、インターネット宿泊予約サイトを利用して該当宿泊施設を予約し、宿泊後に書くことのできる情報で、実際の宿泊体験に基づく宿泊利用者の率直な感想の記録である。宿泊レビューは投稿後、宿泊予約サイトの当該宿泊施設の宿泊レビューを集めたページに掲載され、今後利用する可能性のある宿泊検討者に参考情報として提供されている。昨今のインターネット宿泊検討者は、宿泊施設の提供する広告や公式情報だけを一方的に鵜呑みにせず、これらの実際の宿泊サービス提供を受けた宿泊利用者が残す宿泊レビューを閲覧し、その評価を宿泊施設を選択を行う際の参考情報として重要視している[Nielsen 2012]。

またインターネット宿泊予約サイトの成長に伴い、宿泊レビューは増加し、サービス提供者である宿泊施設においても活用されるようになった。1つには宿泊利用者の満足度、利用時の評価をアンケート調査するために利用され、サービスの向上や改善点の発見に役立てられる。また宿泊利用者の声がダイレクトに宿泊検討者に届くことから、利用者によるその体験を発信してもらうこと自体が宿泊検討者への宣伝にも利用される。サービスを改善する示唆を得ること、そしてよい体験をすることができた宿泊利用者からの宿泊レビューは最高の宣伝材料ともなりえるからである[Sernovitz 2009]。

宿泊検討者、宿泊施設双方が参考にする宿泊レビューは、実際に利用した生の声であり、宿泊利用者がどのような評価を行ったかを知るために有益な情報である。宿泊施設の担当者はそこに書かれた内容を閲覧し、その感想コメントに書かれた苦情や要望を、サービス改善のための材料として活用している。

このように宿泊検討者の予約判断材料として、宿泊施設のサービス改善材料として活用される宿泊レビューの情報の掲載方法にも課題が存在する。その情報提供方法では個々の宿泊利用者の感想コメントを並べ、数値評価を集計して掲載しているだけである。そのため、利用者の多い宿泊施設では宿泊レビューが多く集まり、個々の感想コメントを閲覧するだけでも非常に労力がかかる。様々な内容が記載された感想コメントの中から、宿泊検討者はそこに埋もれた評価情報を発見し、宿泊施設を選択するための判断材料を見つけ出すのは困難

を要する。また宿泊施設のサービス担当者においても、膨大な情報の中から、宿泊レビュー投稿者の意図する内容をすべて理解することは難しく、サービス改善の材料を見つけ出すことは時間がかかりすぎ、見落としが発生する可能性もある。

また、ビジネスで利用する都市中心部の宿泊施設とレジャーで利用する郊外の温泉地の宿泊施設では、宿泊目的が異なると考えられる。そのため宿泊目的の異なるエリア、宿泊施設においては宿泊利用者が持つ感想は異なり、評価する項目は違うはずである。同様にインターネット宿泊予約サイトで宿泊施設を探す宿泊検討者が知りたい情報も異なるはずである。しかし現在のインターネット宿泊予約サイトでは、どちらの場合も同じテンプレートを使った画一的な情報提示になっており、宿泊目的の違いに合わせた宿泊レビューの情報提示にはなっていない。

以上のような背景より、膨大に集まる宿泊利用者の宿泊レビューから、宿泊施設を評価する内容を効率的に抽出し、宿泊利用目的に合わせた情報を提示するための手法を提案することは、インターネット宿泊予約の利便性の向上に重要な取り組みであると考えられる。宿泊検討者と宿泊施設双方にとって、この宿泊レビューの構造を理解し、宿泊レビューに書かれた感想コメントから有益な特徴情報を抽出し、利用者にとって使いやすい提示方法を検討する事は、旅行者の利便向上と、宿泊施設のサービス改善に直結した課題である。

そこで本研究ではインターネット宿泊予約の利便性を向上させることを目的とし、宿泊利用者から寄せられる宿泊レビューを用いて利用者のエリア、宿泊施設がどのように評価されたかの特徴を抽出する方法を検討し、情報の提示方法を提案する。具体的には、宿泊レビューの要素である数値評価と感想コメントを用いて利用者の評価情報の特徴を分析する。特に感想コメントについては、テキストマイニングの手法を用いて対象となるエリア、宿泊施設の特徴となる評価表現を抽出し、それぞれの傾向の分析を行う。また分析の結果を分かりやすく宿泊レビューの利用者に提示できる手法を検討することで宿泊レビューのさらなる活用を目指す。

本研究では、第2章においてインターネット宿泊予約の進展と宿泊予約時に

における利用者の判断材料となる宿泊レビューの効能について俯瞰する。そして本研究の分析方法であるテキストマイニングについての解説を行い、コメントからの評価表現抽出について先行研究とその課題を示す。そして従来の宿泊施設の評価方法に関する先行研究と課題を記載し、本研究の目指す方向を示す。次に第3章では、宿泊予約サイトにおける日本人宿泊利用者によって書かれた宿泊レビューを用いて、その感想コメントをビジネスとレジャーにエリアを分類し、宿泊エリアの特徴となる評価表現の抽出を行い、宿泊エリアによって異なる評価傾向の比較分析を行う。また本手法を活用した、宿泊レビュー情報に付加する新たな情報とその提示方法について提案を行う。第4章では、海外から日本に訪問する旅行者である訪日旅行者が投稿した宿泊レビューからの宿泊施設評価について分析を行う。海外インターネット宿泊予約サイトを利用し、日本に訪れた訪日旅行者によって英語で書かれた宿泊レビューの感想コメント情報を用いて、エリアの特徴分析を行う。さらに同手法を宿泊施設に適用し、個別の宿泊施設における特徴となる評価情報の抽出と分析結果を示す。第5章では、満足、不満につながる評価表現の詳細を検討する。感想コメントに出現する評価表現の中で、特に文中で肯定評価と否定評価が同時に現れ、それらが遷移する文の抽出を行う。その特徴と意味をビジネスエリアとレジャーエリアに分けて分析し、宿泊施設のサービス改善につながる情報の抽出について検討する。最終の第6章では結論を述べる。

第2章

インターネット宿泊予約の評価指標

本章では、インターネット宿泊予約の概要と宿泊予約選択時に活用する情報について概観する。そして本研究で用いる自然言語処理およびテキストマイニングについて解説する。そして宿泊レビューからの評価情報抽出についての先行研究を概観し、本研究における課題について示す。

2.1 インターネット宿泊予約

ホテルや旅館などの宿泊施設は、観光客に安全で快適な休息と睡眠の場所を提供する。一般的に、観光ビジネスは利用客の目的を満足させるために様々なサービスを提供する事業である。インターネット宿泊予約サイトは観光ビジネスの1つにあげられるが、唯一観光サービス提供事業とは性格を異にしており、直接的には何ら観光サービスを提供せず、報酬を得て宿泊者と宿泊施設を仲介する旅行業の位置付けとなる。インターネット利用者の増加に伴い、インターネットを利用し宿泊を予約する旅行者は急激に拡大し、宿泊予約の約29%がインターネットからの申し込みとなっている[JATA 2013]。本節ではインターネット宿泊予約に関する現状と課題を示す。

2.1.1 インターネット宿泊予約サイト

本項では旅行業とインターネット宿泊予約の仕組みとその利用状況について解説する。

近代旅行業の始まりは、1840年代にイギリスのトーマスクックが企画した、禁酒運動集会への鉄道利用団体旅行が始まりとされ[Kobayashi 2010]、日本における旅行業の始まりは1905年に滋賀県の草津駅構内で弁当販売を営んでいた南新助が鉄道を使った参拝団体の斡旋を始めたことが始まりとされている[NTA 2006]。旅行業はこれらの団体旅行客の募集により発展し、1960年代の東京オリンピック、1970年代の大阪万博の開催、高度経済成長の過程において、団体旅行に加えて個人旅行の増加を迎えた。宿泊サービスを提供する宿泊施設は、個人旅行客に販路を持つ旅行会社に販売手数料を支払い、旅行客の集客を依頼してきた。このような経緯があったが、インターネット利用基盤の整備と浸透により宿泊施設は自らホームページを開設し、宿泊施設は直接旅行者への客室提供を目指す動きが始まった。ただ旅行者にとって、どのエリアのどの宿

泊施設に泊まるかを選択する際には比較検討することが多く、単一宿泊施設では、多くの選択肢の提供できるものではない。そのため宿泊施設はインターネット上で多くの宿泊施設を取りそろえるインターネット宿泊予約サイトの販売ネットワークをチャンネルミックス戦略の観点からその客室在庫を提供することになる。

インターネット宿泊予約サイトの始まりは 1996 年、ビジネスホテルを予約することができる日立造船コンピュータが開設したサイト「ホテルの窓口(現、楽天トラベル)」であった。これまで電話で交渉し、予約する手段が中心であったビジネスホテルを、24 時間インターネット経由で予約ができる、インターネット宿泊予約サイトとして登場した。インターネットを介して空室を照会し予約を行い、宿泊料金の支払いは現地で直接宿泊施設に払うというシステムで、宿泊検討者は電話照会も旅行会社に来店もせずに宿泊の申し込みができるようになった。これは、宿泊予約における流通チャンネルの変化をもたらし、宿泊予約の手続きの手軽さに利用者は拡大した。旅の窓口で代表されるインターネット宿泊予約サイトの登場以降、旅行会社も追随して参入し、インターネット宿泊予約は成長している。

鈴木は、宿泊予約のインターネット化現象により、流通チャンネルの変化を促し、サプライヤーと消費者の直接売買およびインターネットを介したオンライン旅行会社、オンライン営業部門を発展させたと指摘している。またその中で低廉性商品、出発間際商品を発売することで、利用者に向けては「利便性」「時間節約」の効果、安心、廉価性、情報判断力の効果を提供し、サプライヤーとの観点では効率的販売拡大効果、確実性を提供しているとした[Suzuki 2010]。

旅行者の観点からのインターネット宿泊予約サイトの利点は 24 時間 365 日インターネット接続できる環境があれば、簡単に宿泊施設が予約できることにある。行きたい日程、人数、予算、エリアを入力すれば素早く希望に合った宿泊施設を回答してくれる。インターネットが登場する以前、宿泊施設を予約する方法は、宿泊施設への直接予約と、旅行会社を経由した予約のいずれかであった。宿泊施設に直接予約の場合、宿泊施設に連絡し、宿泊希望日の空室の状況と価格を確認した後に予約する。直接、宿泊施設の担当者と相談できるため詳細な点を聞きやすいが、宿泊施設ごとに直接相談する必要があるため、手間

のかかる。一方、旅行会社での予約の場合は、旅行会社の確保する宿泊施設の客室を、宿泊施設の案内と宿泊料金の書かれたパンフレットを参照しつつ、店舗カウンターでスタッフに相談しながら予約する方法がとられてきた。旅行の専門家のアドバイスを得ながら申し込みができる利点もあるが、旅行会社へ出向く手間や、店舗の営業時間の制約といった不便も存在する。インターネット宿泊予約サイトは立地的、時間的な制約を解消した。

一方、宿泊施設から見たインターネット宿泊予約サイトの利点は、客室を宿泊施設の需給状況に応じて自らの手で素早くコントロールでき、宿泊料金をリアルタイムに変更できることである。インターネット宿泊予約サイトは宿泊施設からシステム利用料を収受するが、利用料が旅行会社よりも比較的割安なため、魅力を感じる宿泊施設も多い。また旅行会社各社には客室数を別々に割り当てて提供する必要があるが、客室在庫管理が煩雑であったが、インターネット上では、スイッチャーと呼ばれる効率的な客室在庫管理の仕組みを使うことで一元管理ができるようになり、多くの客室を効率的に販売する仕組みがとられるようになった。[Hatakama 2004]

また宿泊施設の規模にかかわらず参加することができるため、自らをアピールする場としてインターネット宿泊予約サイトを活用することが広がっている。宿泊施設へのアンケートによると、インターネット上での宿泊予約を集客する際には、宿泊予約サイトへの登録が最も実施効果が高いとの結果も出ており、これは自社の独自予約サイトを構築するよりも高い率である。これは外国人旅行者の集客の場合においても同様であり、自社のホームページによる外国語表記の実施よりも高い結果であった。インターネットの活用は業歴が浅い企業ほど、積極的である[JFC 2013]。これらは宿泊施設が単体で宣伝告知をする力は限られているが、宿泊施設が多く集まり、比較検討できる環境を作り出すことでそれぞれの宿泊施設の特徴がより明らかになることを示していると考えられる。

また、インターネット宿泊予約サイトの大手の一角である楽天トラベルの成長の一因として、島川は、宿泊後の評価をユーザに書き込んでもらい取捨選択することなくすべてを公開したことを挙げている。これまでの旅行会社は、自分たちで評判をコントロールするというイニシアチブを手放さなかったため、

宿泊利用者の評価を宿泊予約サイトの即評価にすることができず、利用者の要求に答えられない旅行会社のマインドの限界を指摘している。楽天トラベルの支持は、市民の力を信じ、市民みんなが参加し発信者になることで、可能性は無限になることを指摘した[Shimakawa 2010]。

本項では旅行業とインターネット宿泊予約サイトの特徴を概観し、これまでの販売手法に対し、利便性と効率性の観点からも利点があり、宿泊施設および宿泊検討者から利用が拡大していることを示した。また、利用者の発信する評価をダイレクトにすべて公開するという方針が宿泊検討者の支持を集めていることを示している。

2.1.2 宿泊施設の分類

宿泊検討者が予約を行う宿泊サービスは「かたちあるモノ」と違い、無形性、消滅性、変動制、不可分性というサービスの4つの特徴を持つ。宿泊商品はサービス商品であり、その提供物は無形のサービスとなる。また、日々一定の客室を提供するストックができない商材であり、売れ残ってしまえばそのサービスは消滅してしまう。また定価が存在しない商品のため、同一の客室であっても繁忙期と閑散期で料金が異なる。また、宿泊という行為に対してはどの宿泊施設に泊まってもそのサービスは得られるわけで、どのような付加価値をつけて商品を提供するかを検討する必要がある。また、商品には特許や実用新案を取ることはできず差別性を発揮することが難しい。提供されるサービスと信頼性が商品の価値となる[Kotler 2003]。

本項では宿泊施設の分類について検討する。宿泊施設の特性を検討し、その特徴と提供価値について概観する。まず宿泊施設として代表的な区分である長谷らの分類に従いホテル業と旅館業の特性の違いを示す[Hase 1999]。

ホテル業

ホテル業の特性は典型的労働集約型産業であり、「宿泊」「宴会」「レストラン」という3つの商品を販売する。主に都市の中心部や主要な駅前に立地し、ビジ

ネス利用が期待されるエリアに存在する場合が多い。

商品の特性としてはそれぞれの商品を販売するにあたっては、サービスのウェイトが高い。客室販売には数量的制限があり、収容人員を上回って販売することはできない。客室商品は在庫のできない商品でありその日の客室はその日に売らないと消えてしまう。立地が重要な要素で移動することができない。基本的に24時間営業で、旅館と異なり欧米の生活様式を基本としている。

運営の特性としてはロビーに代表されるパブリックスペースの提供など公共性が高いこと。国内外の旅行者をその風俗や習慣を配慮して、平等に対処する必要があること。客の生命、財産、プライバシーを守らねばならないこと。防災、防犯、事故、衛生に対して社会的責任がある。

労務の特性として、労働集約型の産業であり、ホスピタリティの精神に基づく高度な人的サービスを求められる。語学などの専門職や、調理などの高い技術職が必要であり24時間営業のため、他産業と比べて労働調整が難しい。資本・財務の特性と役割として、開業にあたり固定資産に多額の投資が必要であることがあげられる。装置産業なのでその投資の大半は施設・設備に固定化される。

ホテルはスーパーラグジュアリーホテル、ラグジュアリークラス、アッパーミドルクラス・ミドルクラス、エコノミークラス、バジェットクラスの6つのカテゴリに分類される。帝国ホテルやリッツカールトンホテルといったスーパーラグジュアリーホテルでは、普段の生活と異なる時間を過ごしてもらうため、様々な付加価値サービスが提供され、できるだけ高く料金を支払わせることが当たり前の常識であった。一方、新しいコンセプトとして付加価値サービスを徹底的に省いて余計な料金は取らないという逆転の発想が生まれている。東横インやヴィアインといったエコノミー、バジェットクラスのホテルでは、宿泊に絞った料金体系に徹底しコストの削減と利用者の支持を獲得しようとしている。

旅館業

旅館業が一般的に宿泊業としての「旅館」の形になったのは、旅人に一夜の宿と食事を提供した、江戸時代の木賃宿や旅籠であるとされる。高度経済成長

期にクローズアップされた「観光」の基地として旅館はその社会的機能性を高めるにいたった。欧米から輸入された「ホテル」システムと対比して旅館は日本の地方の「歴史館・文化館」として、温泉地や景勝地などのレジャーエリアに立地し、地域の経済活動拠点として重要な位置を占めている。法律上の定義では旅館とホテルの区別は和室が半数以上のものを旅館と呼び、洋室が半数以上のものをホテルと呼ぶ。そのため、実際の旅館ホテルの名称と法的な区別は一致しないことも多い。旅館業の特性は利用客に対し「宿泊」「食事」「入浴」を提供することであり、旅館サイドから見れば安全で清潔な施設・設備の提供、主に日本料理となるが、高い技術を持って調理される料理の提供、おもてなしの心を持った人的サービスの提供内容が注目される。観光地における観光消費の最大の受け皿は宿泊施設であり、その大部分は旅館を通じて消費されている。この意味では旅館は観光地の地域経済にとって重要な存在で、多数の人手により成り立っている旅館は雇用の面でも大きな役割を担っている。ほとんどは古くからの宿場町の旅籠や温泉地の当時宿がその起源であり典型的な地場産業でもある。

これらのようにホテルと旅館は、在庫のできない商品を提供するという意味では同一であるが、提供するサービス内容については異なる意味合いを持つ。宿泊施設としての特徴の違いを宿泊者として利用する際の表 2-1 にまとめる。

旅行検討者は上記のような特性を持つ宿泊施設の中から自分の条件に応じた宿泊施設を選ぶことになるが、宿泊検討者のこだわる項目はホテルが立地するビジネスエリアと、旅館が立地するレジャーエリアにおいて異なる。さらに宿泊施設にも様々な数の施設があり、その特徴は千差万別である。そのため宿泊検討者にはどの施設を選ぶかで得られる満足度が異なることとなる。次項では宿泊検討者が宿泊施設を選択する際に生じる知覚リスクについて概観する。

表 2-1 宿泊施設の種類

	ホテル	旅館
客室	洋室	和室
建物デザイン	洋式，近代的空間	伝統的な和の空間
パブリック空間	外来客も利用する公共の空間	主に宿泊客だけが利用する空間．浴衣で自宅にいるようにくつろげる場所
販売方法	室料制（客室のみを販売） ただし観光地では1泊2食販売も多い	1泊2食販売
提供サービス	「宿泊」「宴会」「レストラン」	「宿泊」「食事」「入浴」
主な業態	都市立地 都市ホテル，ビジネスホテル，公共の会館 (観光地立地) リゾートホテル，小規模家族経営のペンション	観光地立地 旅館，観光ホテル，国民宿舎など様々な公共の宿，小規模家族経営の民宿 (都市立地) 小規模家族経営で室料販売する都市旅館

2.1.3 宿泊予約に伴う宿泊検討者の知覚リスク

目的地選択や旅行商品購買の意思決定に際して，観光客が感ずる不安や懸念を知覚リスク(perceived risk)という．観光客が主観的に評価する危険性でありその大きさは不確実性(uncertainty)と結果の重要性(consequences)に依存する．不確実性は選択の的確性に関わるもので選択のミスや不測の事態が発生する可能性ないし確率の高さを意味している[Hase 1999]．

旅行はサービスゆえに事前評価が困難であり，体験も天候や混雑などの目的

地の状況に左右されやすく不確実性は高い傾向にある。そして結果の重要性は選択においてミスを犯した場合に生ずる経済的損失や心理的苦痛など旅行者が被るダメージの大きさとして表される。長谷は知覚リスクには次のようなタイプがあると指摘している。

- ・機能的リスク(functional risk)…選択された観光機会において、満足いく観光体験ができるかどうかにかかわる不安・懸念
- ・身体的リスク (physical risk)…選択した観光行動における、病気、けが、その他健康上の不安・懸念.
- ・財務的リスク(financial risk)…誤った選択による時間的・金銭的費用の損失に対する不安・懸念
- ・社会的リスク(social risk)…選択に対する他社、帰属集団からの評価や承認が得られないこと、あるいは選択を誤った場合の世間体の悪さなどへの不安・懸念.
- ・心理的リスク (Psychological) …選択を誤った場合に経験する自己嫌悪や屈辱への不安, 懸念.

その対処法には3つあり、第1の方法はリスクを厭わない、またはあえてリスクを冒す行動をとるリスク耐性を持った旅行者の行動が存在する。第2の方法は選択しようとする観光行動に対して知覚するベネフィットによってリスクに対処しようとするもので、ベネフィットをより強く知覚することでリスクから目をそらそうとする動きが存在する。第3の対処法はリスクの低減を図ろうとする行動をとる旅行者が存在することを示している。知覚リスクの低減方法には、

- ・積極的情報収集,
- ・過去の経験に基づく選択,
- ・社会的評価に基づく選択 (世間の評判, 第三者による格付け, マスコミやガイドブックによる紹介記事などに基づく選択)
- ・高額な旅行商品の選択 (ただし財務的リスクを高めることもある)

- ・期待水準の低下
- ・その他（旅行傷害保険などの制度の活用）

などがある。宿泊施設に滞在を検討するにあたり、インターネット宿泊予約サイトの情報を閲覧する行為は、長谷の指摘した知覚リスクの低減に寄与していると考えられる。宿泊検討者の積極的な情報収集により、宿泊施設から発信される情報や、宿泊利用者の過去の経験に基づく情報である宿泊レビューを参照し、社会的評価に基づく選択を行うことで宿泊利用時に存在するリスクの回避を行っていると考えられる。次節ではインターネット宿泊予約をする際の情報収集にはどのようなものがあるかについて概観する。

2.2 インターネット宿泊予約の情報収集

旅行者の情報収集方法には、テレビやラジオ、新聞、雑誌などのマス媒体、インターネット、友人の体験談、旅行会社のパンフレットなどいろいろなものが存在する。その中でも旅行のきっかけとなるものは、自分自身の経験や家族や知人の体験談、テレビやラジオなどのオフラインの情報メディアが良く使われ、実際の旅行の具体的な計画を立てる時に、オンラインの情報メディアが重要とされている[goo 2013]。また、旅行の目的も様々であり、商用のビジネス旅行もあれば、帰省旅行、温泉旅行といったレジャー旅行が存在するため、それぞれの場合において宿泊検討者が知りたい情報は異なる。本節では宿泊予約時に利用する情報を、サービス提供者からもたらされる提供者発信情報と利用者発信情報に分類し特徴を検討する。

2.2.1 提供者発信情報

宿泊施設はインターネット宿泊予約サイトに対して、住所や電話番号、施設の特徴アピールなどのテキスト情報と、施設の外観画像や部屋画像といった画像情報、宿泊プランの内容や空室、料金といった商品情報を提供し、利用者の検討材料として提供している。これらは施設自らが情報公開する内容、すなわち、提供者側の発信する情報である。宿泊施設側として最大限よい印象を持ってもらい、宿泊施設選択における決断をしてもらう必要があるため、意図的に利用者にとって不利益が発生する情報は示さず、良い情報を掲載する可能性がある。また、宿泊利用者にとって利益になる点および、不利益になる点について宿泊施設に関するすべての情報を宿泊検討者に向けて盛り込むことは掲載スペースの制約からも現実的に不可能であり、その点においても宿泊検討者には伝えきれていない情報が存在する。

島川は、インターネット上の情報量の多さに起因する情報のジャンク化を指摘している。また旅の達人と称される旅行ライターの情報発信が、提供者のた

めにいい情報を提供するようになる弊害として、書かれた記事内容と実際に訪れた時の印象の乖離が生まれることを指摘した。旅行者はこの手の作られた評判に敏感になり、その評判に左右されにくくなったとしている。そして一人の達人の感想よりも多くの一般人のリアルな評価の傾向を知ることを求めるようになったとしている[Shimakawa 2010]。

そこで宿泊レビューに代表される、利用者目線で評価し、利用者の感想が書き込まれる利用者発信情報により、提供者情報では伝えきれない情報を補う役割があると考えられる。

2.2.2 利用者発信情報

宿泊の感想や利用者の声を宿泊利用者が発信する方法は、インターネット宿泊予約サイトの宿泊レビュー、ブログ、SNS(ソーシャル・ネットワーキング・サービス)、Twitter、Facebook など様々な形で存在する。旅行の感想を共有する意義として、ひとつには友人や家族といった身近な人々との関係性を深めるために発信していると考えられる。宿泊予約サイトには、ほぼ標準的に利用者が感想を投稿する宿泊レビューを表示する機能を有している。投稿された宿泊レビューは、単にその宿泊施設を利用した感想を発信したいという意味だけでなく、利用した際に気付いたことや、良かった点、改善してほしい点などを記載することで、今後の施設を利用する可能性のある宿泊検討者に対する参考情報として発信している。実際に利用するまでは内容を確認することができないサービスである宿泊において、利用者の宿泊レビュー評価は、実際に利用した人の声であり、宿泊施設が発信する情報と併せ、宿泊検討者にとって重要な参考情報、判断材料となる。

2.2.3 宿泊レビュー

一般的に利用者が宿泊レビューを投稿するには、インターネット宿泊予約サ

イトで宿泊予約し、その施設に宿泊する必要がある。宿泊後にはじめてレビューが書ける画面が表示される仕組みになっており実際に宿泊していないユーザーが、なりすましてレビューを投稿することは容易にはできない。そのため宿泊レビューは実際の利用者の体験を反映した情報であるといえる。宿泊レビューの例を図 2-1 に示す。

利用者プロフィール

G13さん
男性 / 30代

投稿日: 2011/08/24

感想コメント情報

夏休み中の家族旅行で利用。前回利用時に子ども達がプールおよび夕食(バイキング(特にチョコレートファウンテン)がお気に入りだったためリピート。プール・夕食に加え、今回はアイスクリーム製作体験もセットされていて子どもたちも大満足。大人の目から見ると夕食(バイキングレストラン)については品数は十分かつ味もまずまずだが、天ぷらが揚げたての割りにべちゃっとしていたこと、お刺身の品数が少ないことが不満。朝食(hal)は料理(品数・味)そのものは満足したが、配膳用のトレーが洗った後なのにご汚れていた(前の利用者の食べカスが残っていた)ことには驚いたし、スタッフの対応もイマイチだった。

数値評価

総合 4 ★★★★★ ☆ 部屋 3 | 風呂 3 | 料理(朝食) 1 | 料理(夕食) 4 | 接客・サービス 3 | 清潔感 3

【2011年08月宿泊 子連れ旅行】
【宿泊価格帯】13,001~14,000円(大人1人あたり)
【宿泊プラン】【じゃらん限定】夏休みは家族で満喫！選べる3特典付 ツイン 朝・夕

図 0-1 宿泊レビューの例

宿泊レビューは、概ね宿泊利用者のプロフィール、数値評価項目および自由記述欄である感想コメントの 3 種類の要素から構成される。宿泊レビューは、テンプレートに沿って宿泊利用者に記載され、宿泊施設ごとにまとめられる。それらを投稿日付順に並べ、固定的な表示形式で各宿泊施設の予約申し込みページに掲載される。数値評価項目は、宿泊した際に受けたサービスを宿泊予約サイトによって分類された代表的な項目ごとに 5 段階または 10 段階で評価する。感想コメントは自由に宿泊利用者の意見を記載する場所である。自由記述が可能なテキスト欄であり、そこには数値評価項目に対する宿泊利用者の具体的な感想、意見と共に数値評価では表現できない内容に対する宿泊利用者の感想、評価が含まれている。

通常、ビジネス宿泊、レジャー宿泊と宿泊目的が異なる場合、宿泊検討者が知りたい施設の評価情報は異なるはずである。しかし、現在の宿泊レビューの情報提供手法は、宿泊利用者がどのような利用目的で宿泊施設を利用したのかにかかわらず一様のテンプレートで表現している。そのため利用目的に応じて

どの数値評価項目が重要であるのかを示すことができていない。また、数値評価と感想コメントはそれぞれ独立した評価であるため、それぞれの評価における関係性は示されていない。そのため、宿泊利用者の目的によって異なる感想の違いをわかりやすく表現できているとは言い難い。

2.2.4 レビュー情報の特性

次に宿泊レビューなど利用者が個人の評価を投稿した情報、いわゆるインターネット上のクチコミにどのような特性があるのかを概観する。宿泊レビューを含むクチコミには書き手となる投稿者と、それを読む利用者が存在する。宿泊レビューの場合、書き手は宿泊施設に泊まった宿泊利用者であり、読み手は宿泊を検討中の宿泊検討者と宿泊サービスを提供する宿泊施設の従業員である。Rosen は、レビューが重要になった理由として、利用者が取り扱う情報量の過多によるノイズの存在や、情報発信者にとって不都合な情報が発信情報に含まれないことに対する利用者の懐疑的な態度の存在を指摘する。そしてインターネットのコミュニケーションの発達により、これまでつながることのなかった利用者どうしがつながる利用法が確立されたことで重要性が増したことを挙げている。レビューはエキサイティング、革新的、経験型、複雑、高価、可視性が高いといった消費者の関与度の高い商品において多くみられることを指摘している[Rosen 2002]。国領は、情報の非対称性の逆転現象を指摘した。かつては売り手のほうが、商品の品質や市場価格などについて多くの情報を持っており、買い手は、その情報をあまり持っていないという情報の非対称性が存在した。しかしそれがインターネットの発達により、消費者側も多くの情報量を持つようになり、消費者側からの情報発信をおこなうようになった。これをいかにビジネスに適用するかを検討するべきであるということ述べている[Kokuryo 1999]。また山本は、レビューによるユーザ行動として「購買にいたる」といったプラスの循環だけでなく、「購買を阻止する」というネガティブなレビューの循環も存在することを示し、レビューの評価は、ポジティブ、ネガティブ両方の極性に注目する必要があることを示した[Yamamoto 2008]。利用

者の評価は肯定的なものだけでなく否定的なものもあり、それらの両方の情報を閲覧し、宿泊検討者は自分の利用目的に合った宿泊施設を選んでいることになる。

濱岡は、クチコミサイトと企業のホームページの比較において、企業のホームページを情報源として利用する割合が高いものの、クチコミサイトを重視することを指摘している[Hamaoka 2006]。

これらの先行研究は、インターネットの活用が進むなか、利用者自らが情報を発信、収集できる環境が整ったことで、自らの選択をより間違いなくするためにサービス提供者側からの情報だけを鵜呑みにするのではなく、より多くの情報源からの情報を用いた判断を利用者は行うことになったことを示している。また、情報提供者側においてもこの状況を理解し、レビューを活用する必要性を示している。これらの研究においてはレビューの有用性について示すものであるが、実際にレビューに表記されるコメント内容について深く踏み込んでおらず、具体的な評価の内容についての説明に乏しい点が課題である。

2.2.5 宿泊施設の評価

前項においてレビューの特性について検討したが、本項では宿泊予約の場合にはどのような評価方法があり、どのような評価が行われているのかを概観する。世界では公的機関による宿泊施設の評価が行われており、フランスでは政府観光局がホテル業界全体のレベルアップや競争力の強化を目的とし、宿泊施設の格付け、評価を行っている。格付け機関の宿泊施設の評価として有名なものとしてミシュランガイドがあるが、こちらは匿名の覆面調査員による評価となっている。上地は、日本では評価機関による評価指標が存在せず、各旅行会社やガイドブック出版社が自分たちで集めた情報をもとに評価を行うため評価にばらつきがある点を指摘し、沖縄独自の格付け制度の評価指標を提案した[Kamichi 2012]。大内らは、宿泊予約を含む旅行は、実際に経験しないとわからない、費用と時間を消費する商材である点と宿泊施設と宿泊利用者間に存在する情報の非対称性を指摘している。自らの施設情報に対し情報の優位性を

持つ宿泊施設が、実態に即さない誇張された情報を発信する可能性を示唆し、日本の宿泊施設、特に旅館の客観的な第三者機関の評価を嫌う傾向を指摘した。この問題を解決する方法として、北海道の宿泊施設を対象に「施設」「サービス」「食」「交通至便」「眺望」「癒し」を大項目とする 177 項目からなる評価項目を、宿泊施設の従業員が自ら評価し、公開するという自己評価型の評価基準の提案を行った[Ouchi 2008]。これら研究は公平で客観的な評価基準を統一することで、提供者側の情報発信方法の向上に貢献する提案である。しかしあくまで施設側が行う評価であり、一般的な宿泊利用者が実際に宿泊した際に施設をどのように評価するかを情報提供する方法までは検討を行っていない。

一方、Gretzel らは宿泊利用者が発信する宿泊レビューの活用状況について調査を行った。その結果、利用者は旅程選択支援のためではなく、どの宿泊施設に泊まるかの意思決定の判断において活用を行っていることを挙げている[Gretzel 2008]。森川は、宿泊レビューについて、旅行会社が旅行後に回収するはがきアンケート調査をもとに調査を行い、顧客満足評価への影響力が強い特徴項目として、食事の内容、客室などの手入れ、館内の設備、従業員の接客態度を挙げている[Morikawa 2001]。

また、長島は、旅行におけるクチコミについて注目しサービス商品の特性を踏まえた評価をおこなっている。そしてクチコミが商品選択や満足度に与える影響が大きいことを指摘し、下記の 5 種類の機能を想定した[Nagashima 2009]。

- ・クチコミの意思決定推進機能
- ・クチコミの期待値形成機能
- ・クチコミの評価軸付加機能
- ・クチコミの実際値修正機能
- ・クチコミによる改善要望機能

これらはクチコミの参照機能による意思決定の支援機能の存在と、他人の目に触れる形でサービス提供者へサービスの利用者が改善要望を伝達するという新しい機能、および利用者の期待値と実際値に与える影響に関する要素が存在するとしている。

これらの調査研究は、宿泊施設の判断材料として宿泊レビューが重要視されていることを明らかにし、また利用者の感想コメントの中に、顧客満足評価に影響のあるいくつかの要素が存在することを示唆している。

これら先行研究の課題は、全体的にレジャー目的での宿泊利用を主眼とした調査となっており、ビジネス目的での利用の場合の特徴については言及されていない。また、宿泊レビューの有効性を指摘しつつも、具体的なレビューに書かれた感想コメントの記載内容についての分析が乏しいことである。感想レビューを評価指標として利用するためには、実際の宿泊レビューの感想コメントにおいてどのような記述がなされ、どのような評価を受けているのかを検証する必要が残されている。次節では、宿泊レビューの感想コメントより評価表現を抽出し分析するテキストマイニングの手法について検討する。

2.3 テキストマイニング

テキストマイニングとは、テキストを自然言語処理したのち、データマイニングの技術を用い、テキストにある内容の傾向や特徴から、新たな知見を見いだす技術である。テキストマイニングの典型的な利用例としてコールセンターにおけるコンタクト履歴の分析、医療文献データベースを用いた研究成果を活用するための方法、特許文書から企業の技術戦略を分析するといった研究が行われている[Nasugawa 2006]。

また、インターネットによる双方向の利用者の交流により、いわゆるビッグデータと称される分析対象データの増加を背景としたテキストマイニングを利用した研究のニーズも高まりをみせており、巨大掲示板での投稿情報を対象とした研究、化粧品や健康食品の評価サイトにおけるレビュー、クチコミ情報からのユーザの行動を分析する研究などが存在する[Okumura 2013]。

宿泊レビューの感想コメントのテキストは、宿泊施設を利用したユーザより日々投稿され、膨大な量のデータが集まる。これらの大量に集まるテキスト情報データから宿泊施設の評価を分析する手法としてテキストマイニングを用いることが有効であり、多くの感想テキストを統計的に処理し、宿泊レビューの特徴をつかむことができる。本節では、本研究で用いる自然言語処理と分析手法であるテキストマイニングについて解説する。

2.3.1 自然言語処理による評価表現抽出の概要

本節では、自然言語処理による評価表現抽出を行うときに必要となる形態素解析、構文解析、意味解析の技術について概説する。

自然言語処理では、表 2-2 に示すように、日本語の語を分類する形態素解析技術をベースに、構文解析、意味解析などの手法が使われる。これらの手法を用いて、情報検索処理[Kawano 02]や機械翻訳処理[Yoshimura 00]などが行われる。それぞれの技術の関係は図 2-2 のようになり、各解析手法から評価表現

抽出を行うことができる。

表 2-2 自然言語処理による評価表現抽出の手法一覧

手法	説明
形態素解析	文を構成する語を切り出し，その品詞などの性質を得る。
構文解析	語の関係を解析し，文の構造を明らかにする。
意味解析	文の形式的な構造から意味表現を得る。

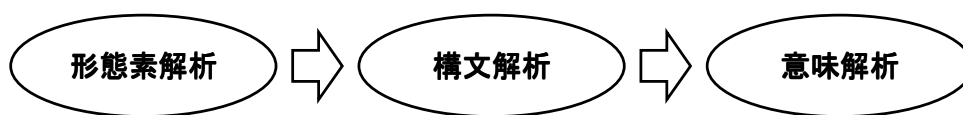


図 2-2 自然言語処理による解析法の関連

また，最近では，インターネットにおける情報検索技術の研究が特に多く行われており，自然言語処理技術の発展に大きく貢献している[Tokunaga 99]．具体的には，Web 上の豊富なテキスト情報を使って自動的に事典を構築し[Fujii 02]，ホームページの自動要約処理[Mani 01]などを行うための自然言語処理技術が研究されている。

2.3.2 形態素解析

評価表現抽出の対象となる文章は，いくつかの語の系列として表されている。

そのため、文構造の解析にあたっては、文を語の単位へ分解する必要がある。このような語へ分解するための処理が形態素解析処理である。形態素解析処理では、文が分かち書きされている英語文については接辞や屈折形を同定することであり、文が分かち書きされていない日本文については、さらに、文字系列から語を切り出す処理が加わる。現在、広く使われている形態素解析ソフトウェアに茶釜や Mecab がある[Asahara 03a].

(1) 語の種類

日本語では、語を自立語と付属語に分けている。自立語は単体で文節となるが、付属語は単体では文節とはならない。文節は、文の中で文法的な機能を持つ最小の単位のことである。

語は、名詞的であるか動詞的であるかによって分けられる。名詞的であるものを体言と呼び、動詞的であるものを用言と呼ぶ。また、同じ意味を持つ語をシソーラスと言う[Yamaguchi 03].

(2) 語の構成

語に接頭語や接尾語が付属して新しい語を形成する。この接頭語や接尾語をまとめて接辞と言う。また、複数の語が結合して複合語を構成する。

語には活用するものがある。日本語では、動詞、形容動詞、形容詞が活用する。英語では、名詞が数について、動詞が人称、数、時制について、形容詞が比較級、最上級について活用する。このような語の活用を屈折と言う。語の中で活用しない部分を語幹と言い、活用する部分を語尾と言う。

(3) 形態素

文を構成する語を分解すると、文節、単語、文字、音素のように、より細かく分解することができる。この中の単語においては、たとえば、「見られる」の

ように、「見る」と「られる」に分割される．この「られる」を抽象的な受け身の助動詞として定義すると，これは様々な文の中で「られる」や「れる」などに変化する性質を持つ．このような抽象的な存在を「形態素」と呼ぶ[Matsuoka 00]．形態素解析処理は，システム内に保存の辞書に登録されている形態素に対応して，文を形態素のレベルまで分解し，それぞれの形態素の品詞情報を特定する処理を行う．図 2.3 に形態素と品詞情報の例を示す．

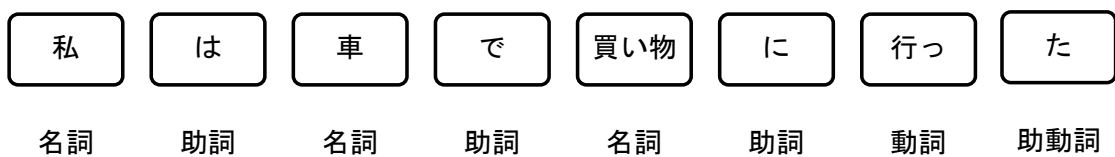


図 2.3 形態素と品詞情報の例

(4) 品詞情報

形態素解析処理では，文を形態素に分割し，品詞情報を付加して出力する．このときには，品詞の定義が必要となる．品詞については，様々な分類が行われているが，現在，情報処理の分野で最も一般的に利用されているのが，IPA(情報処理推進機構)品詞体系[Asahara 03b]である．表 2-3 に，この IPA 品詞体系を示す．自然言語処理の多くは，これらの品詞情報を利用して解析を行っている．

表 2-3 IPA 品詞体系

Type1	Type2	Type3	Type4	説明	例
名詞	一般			普通名詞	「大根」「人間」
	固有名詞	一般		一般的な固有名詞	「北穂高岳」
		人名	一般	姓と名に分けられないもの	「お市の方」
			姓	主に日本人の姓	「山田」
			名	主に日本人の名	「太郎」
		組織		組織を表す名称	「通産省」「NHK」
		地域	一般	国名以外の地名を表すもの	「アジア」「京都」
	国		国の名前	「日本」	
	代名詞	一般		代名詞, 不定詞	「それ」「ここ」
		縮約		代名詞と係助詞「は」の組合せで、 短縮した形	「ありゃ」「こりゃ」「そりゃ」
	副詞可能	時間, 量, 割合など副詞的に使う ことのできる名詞			「金曜」「一月」「少量」
	サ変接続	後ろに「する」「できる」などが 後接可能なもの			「インプット」「愛着」 「悪化」
	形容動詞語幹	「な」の前に現れるもの			「健康」「安易」「駄目」
	ナイ形容詞語幹	助動詞の「ない」の直前に現れて 形容詞的に働くもの			「申し訳」「仕方」「と んでも」
数	漢数字, 算用数字, および, 「何 (回)」「数(%)」「幾(夜)」			「0」「1」「何」	
非自立	一般	連体詞, 「の」, 活用語の基本形に 接続して使われる			「こと」「きれい」「く せ」「もの」
	副詞可能	連体詞, 「の」, 活用語の基本形に 接続して使われ, 副詞的に働く			「限り」「さなか」「う ち」

Type1	Type2	Type3	Type4	説明	例
		助動詞語幹		連体詞, 「の」, 活用語の基本形に 接続して使われ, 助動詞とされる 「ようだ」の語幹	「よう」「やう」
		形容動詞語幹		連体詞, 「の」, 活用語の基本形に 接続し, 「な」と接続可能なもの	「みたい」「ふう」
	特殊	助動詞語幹		基本形に接続し, 「そうだ(伝聞)」 の語幹部分	「そう」
	接尾	一般		名詞, あるいは, 「ガル接続」や 「連用タイ接続」に接続して複合名 詞を形成	「おき」「かた」「がい」
		人名		名詞のうち特に人名に接続して 名詞を形成する接尾語	「君」「様」「著」
		地域		名詞のうち特に地域名に接続し て名詞を形成する	「町」「市」「県」
		サ変接続		名詞に接続して「スル」に前接し 得る接尾語	「化」「視」「分け」
		助動詞語幹		他の品詞の連用形に接続し, 「そ うだ」の語幹部分	「そう」
		形容動詞語幹		他の名詞や活用語の連用形に接 続し, 「だ」に前接する	「的」「げ」「がち」
		副詞可能		他の名詞に接続する接尾語で, 副 詞的に働くもの	「後」「以後」「中」
		助数詞		数に接続して名詞を形成	「個」「cm」
	特殊		用言につく特殊な接尾辞	「(楽し) さ」	
	接続詞的			単語と単語を接続する接続詞的 な働きをするもの	「(日本) 対 (アメリ カ)」

Type1	Type2	Type3	Type4	説明	例
	動詞非自立的			「助詞-接続助詞」の「て」に接続し、動詞的なもの	「ごらん」、「ちょうだい」
接頭詞	名詞接続			名詞に前接する接頭語のうち、数に接続するもの以外	「お(水)」「某(氏)」 「ご(立派)」
	数接続			名詞に前接する接頭語のうち、数に接続するもの	「約」「およそ」「毎時」
	動詞接続			動詞の命令形あるいは「動詞 連用形」+「なる/なさる/くださる」に前接	「お(読みなさい)」「お(座り)」
	形容詞接続			形容詞に前接する接頭語	「お(寒いですねえ)」
動詞	自立	カ変			「やってくる」
		サ変・スル		「する」および「名詞 サ変接続」に接続する「する」	「する」
		サ変・ースル		和語系のサ変動詞	「接する」
		サ変・ーズル		和語系のザ変動詞	「信ずる」
		一段		上一段活用/下一段活用	「着る」
		五段・カ行イ音便		五段カ行で「て」に接続するとき にイ音便になるもの	「解く」「聞く」
		五段・ガ行		五段ガ行で「て」に接続するとき にイ音便になるもの	「継ぐ」「急ぐ」
		五段・サ行		五段サ行で「て」に接続するとき に音便化しないもの	「話す」
		五段・タ行		五段タ行で「て」に接続するとき に促音便になるもの	「持つ」
		五段・ナ行		五段ナ行で「て」に接続するとき にハツ音便になるもの	「死ぬ」

Type1	Type2	Type3	Type4	説明	例
		五段・バ行		五段バ行で「て」に接続するとき にハツ音便になるもの	「呼ぶ」
		五段・マ行		五段マ行で「て」に接続するとき にハツ音便になるもの	「進む」
		五段・ラ行		五段ラ行で「て」に接続するとき に促音便になるもの	「切る」「なる」
		五段・ラ行特殊		五段ラ行で「ます」接続および命 令形が「-い」の形	「いらっしゃる」「おっ しゃる」
		五段・ワ行ウ音便		五段ワ行で、「て」に接続すると きにウ音便になるもの	「問う」「乞う」「浴う (て)」
		五段・ワ行促音便		五段ワ行で、「て」に接続すると きに促音便になるもの	「言う」「食う」「負う」
		四段・ハ行			「云ふ」「向かふ」
		ラ変			「あり」「なり」
		上ニ・ハ行		その他, 「ダ行」	
		下ニ・ア行		その他, 「カ行」「ガ行」「サ行」「タ 行」など	
		一段・得ル		古語「得る」の活用型. 基本形と 仮定形のみ	
	非自立	カ変			「(て) くる」
		一段			「あげる」「おえる」
		五段・カ行イ音便			「(て) 頂く」
		五段・カ行促音便		五段カ行で「て」に接続するとき に促音便になるもの	「行く」「ゆく」
		五段・サ行			「いたす」「だす」
		五段・マ行			「込む」
		五段・ラ行			「終わる」「切る」
		五段・ラ行特殊			「(お読み) なさる」

Type1	Type2	Type3	Type4	説明	例
		五段・ワ行ウ音便			「給う」
		五段・ワ行促音便			「合う」
	接尾	一段		学校文法でいう助動詞	「させる」「しめる」
		五段・ラ行			「がる」
形容詞	自立	形容詞・アウオ段		語幹の最後の母音がアウオのいずれかで終わるもの	「青い」「赤い」「厚い」
		形容詞・イ段		語幹の最後の母音がイで終わるもの	「哀しい」「楽しい」「頼もしい」
		形容詞・イイ			「いい」「ええ」
		形容詞・不変化型		形容詞の中で「基本形」の用法しかもたないもの	「かつこいい」
	非自立	形容詞・アウオ段		動詞「連用タイ接続」あるいは「連用タ接続」に後接	「がたい」「づらい」「やすい」
		形容詞・イ段		動詞の「連用タイ接続」あるいは「連用タ接続」に後接	「らしい」「(て)ほしい」
		形容詞・イイ			「いい」
		形容詞・不変化型		動詞「連用タイ接続」/「連用タ接続」に後接し「基本形」の用法しかもたないもの	「かつこいい」
	接尾	形容詞・アウオ段		学校文法でいう助動詞	「(食べ)たい」
		形容詞・イ段		学校文法でいう助動詞	「(嫌味)たらしい」
副詞	一般			必ず後ろで切れるもの、連体修飾が不可能なもの	「あいかわらず」「多分」
	助詞類接続			「の」「は」「に」「な」「する」「だ」などが後接	「こんなに」「そんなに」「なにか」
連体詞				名詞を修飾する形しかもたないもの	「この」「その」「どの」
接続詞				独立に現れる接続詞	「けれども」

Type1	Type2	Type3	Type4	説明	例
助詞	格助詞	一般		格助詞	「の」「から」「を」
		引用		名詞、人物の台詞、会議の決定事項などの直後の「と」	「(~だ) と (述べた)」
		連語		格助詞と動詞との連語で、主に格助詞に相当する働き	「という」「といった」「として」
	接続助詞			「から」「からには」	
	係助詞			「こそ」「さえ」	
	副助詞			「がてら」「かも」	
	並立助詞			「と」「たり」	
	終助詞			「かい」「かしら」	
	副助詞／並立助詞／終助詞			「か」のうち、副助詞、並立助詞、終助詞いずれか	「か」
	連体化			「名詞」に接続して体言にかかる 「の」	「の」
副詞化			擬音語、擬声語、擬態語およびそれに類する名詞や副詞の直後の 「に」「と」	「に」「と」	
特殊			以上の分類に当てはまらないもの	「かな」「けむ」「にゃ」	
間投助詞			文語的な用法	「(松島) や」	
助動詞	五段・ラ行アル			動詞型の活用の助動詞「ある」	「ある」
	五段・ラ行ゴザル			動詞型の活用の助動詞「ござる」	「ござる」
	形容詞・イ段			形容詞型活用の助動詞	「らしい」
	特殊・ナイ			否定の助動詞「ない」の活用型	「ない」
Type1	Type2	Type3	Type4	説明	例
	特殊・タ			完了を表す助動詞「た」の活用型	「た」「だ」
	特殊・ダ			断定の助動詞「だ」の活用型	「だ」

特殊・デス	断定の助動詞「です」の活用型	「です」	
特殊・ドス	断定の助動詞「どす」の活用型	「どす」	
特殊・ジャ	断定の助動詞「じゃ」の活用型	「じゃ」	
特殊・マス	謙譲・丁寧を表す助動詞「ます」の活用型	「ます」	
特殊・ヌ	否定の助動詞「ぬ」の活用型	「ぬ」	
特殊・ヤ	関西方言の断定の助動詞	「(そう) や」	
不変化型	現代では活用しない助動詞,活用が想定できない口語	「う」「まい」「(いざ行か) ん (む)」	
感動詞	あいさつなど	「おはよう」	
記号	一般	以下の分類以外の記号	「○」「@」「〒」
	アルファベット	英語のアルファベット	「A」「a」
	句点	句点	「.」「。」
	読点	読点	「,」「,」
	空白	全角の空白文字	「 」
	括弧開		「(」「【」
括弧閉		「)」「>」	
フィラー	話し言葉で起こる相槌や挿入的な音声言葉	「あの」「うんと」「えと」	
その他	間投 「名詞-接尾」や「助詞-終助詞」とし難いもの	「(だ) ア」	
未知語	未知の語		

(5) 語の切り出し

日本語の文は語と語の間に空白を含まない形の連続した文字系列で表現されるので、それらを合わせたすべても時系列な語となる。文を s , 文字を c , 語を w と表すと、式 2.1 のように表すことができる。

$$s = c_1 c_2 \cdots c_n = w_1 w_2 \cdots w_m \quad (\text{式 2.1})$$

これは、文 s が n 個の文字 c_1, c_2, \dots, c_n から構成される文字系列 $c_1 c_2 \cdots c_n$ として表されており、さらに、文 s は m 個の語 w_1, w_2, \dots, w_m からなる語系列 $w_1 w_2 \cdots w_m$ として表されていることを示している。 n 個の文字系列から独立に取り出せる任意の長さの部分文字系列は、 $c_i c_{i+1} \cdots c_j (1 \leq i \leq j \leq n)$ で示すことができる。 さらに、語系列 $w_1 w_2 \cdots w_m$ は、それらの部分文字系列の中から m 個を取り出して各語 $w_i (1 \leq i \leq m)$ として対応づけ、それらを $w_1 w_2 \cdots w_m$ の順に並べたものである。

ここで、部分文字系列と語との対応づけの方法には、2 つの制約がある。第 1 の制約は、各語 w_i に対応づけられたそれぞれの部分文字系列を語系列 $w_1 w_2 \cdots w_m$ の順に再び置き換えたとき、元の文字系列 $c_1 c_2 \cdots c_n$ になっていなければならないことである。第 2 の制約は、各語 w_i は接辞や屈折などの語処理を行ったとき、辞書見出しとして登録されている語に同定できなければならないことである。これらの制約を満たす語の分割方法は、一般に複数通りある。

(6) 語処理

語処理は、日本語の接頭語や接尾語を取り除いた語本来の形式を求める処理である。この処理は、語の定義や辞書見出しの登録の方法に依存する。たとえば、接辞がついた形の語や屈折した形の語をすべて辞書見出しとして登録してある場合は、語処理を行わない。従って、前項で述べた第 2 の制約は部分文字系列が辞書見出しに登録されている語に同定できることという条件のみで表される。しかし、実際には、辞書見出しの数を少なく抑えるために、語の原形のみを辞書見出しとして登録する。そのため、語処理が必要となる。

接辞処理は、接頭語と接尾語を抽出することである。日本語の接頭語には図 2-4 のようなものがある。

非, 否, 高, 低, 最, 再, 超

図 2-4 日本語の接頭語の例

また、日本語の接尾語には、図 2-5 のようなものがある。

的, 化, 性, 間, 形, 型, 等

図 2-5 日本語の接尾語の例

日本語の接辞の数は多い。また、接辞に関連するものとして助数詞がある。

屈折処理は、活用語の原形を抽出することである。屈折には、一般に法則性があり、それを使って文の中に現れた語の屈折形を同定する。屈折が法則に従わない語の場合には、辞書に屈折形を逐一登録することによって対応する。

その他の問題に、未知語処理がある。未知語処理は、文字系列を辞書に登録されていない語として仮定することである。未知語処理が必要な理由は2つある。1つは、辞書が完備しておらず、辞書見出しに登録漏れがある場合である。もう1つは、科学技術文書などで専門語が使われている場合である。専門語には、造語や複合語が多い。従って、未知語処理に関連して、造語処理や複合語処理も必要になる。

(7) 形態素解析

形態素解析の目的は、構文解析や意味解析などの出発点を与えることにある。たとえば、構文解析は、語系列として表された文について、部分語系列を構成素とする構文構造を求める。

そのため、形態素解析では、前に述べた2つの制約に加えて、文法的制約も利用する[Masuoka 92]。たとえば、助詞「が」と助詞「に」が続くことは無いといった制約である[Matsumoto 01]。また、文字の種類に関するヒューリスティックスも使う。たとえば、漢字が続いているときは、1つの語である可能性が高いというものである。また、漢字と漢字の間にひらがながあると、2つ目の漢字の前は、語の切れ目である可能性が高いというヒューリスティックスもある。さらに、構成的ヒューリスティックスを使うこともできる。たとえば、1つ

の部分文字列が他の部分文字列の一部となっているときは、長い部分文字列を語と仮定する方がよいといったヒューリスティクスがある。また、文の中で語の出現頻度を使うこともできる。実際には、これらのヒューリスティクスは、確約された制約ではないので、一定の基準の下での確からしさに関する順序付けを行っている。

(8) 日本語の形態素解析法

日本語文の形態素解析手法には様々なものが考案されている。形態素を構成する文字列を生成する際に、文の中のどの部分文字列を優先して選択するかを検討した手法として、最長一致法、コスト最小法、最尤評価法などがある。さらに、近年では、大容量のテキストデータを比較的容易に入手可能なことから、抽出された部分文字列に対する品詞情報の付与処理などに確率的言語モデルを用いた形態素解析処理が行われるようになっている [Kin 03]。

- ・ 最長一致法

最長一致法は、文字列の最も左から始まる長さの異なるいくつかの部分文字列がそれぞれ語として切り出させるとき、最も長い部分文字列の語または文節を優先する手法である。

- ・ コスト最小法

コスト最小法は、日本語かな漢字変換システムである **Microsoft IME** などに取り入れられており、現在最も広く使われている手法である。本手法では、隣接する単語の接続可能性を示す接続表を用い、接続する可能性を可能か不可能かの 2 値ではなく、接続する可能性が高いほど小さな値を取り、接続する可能性が低いほど高い値を取る「コスト」で表現する。このコストが最小になるように文節の候補を優先する手法である。

- 最尤評価法

最尤評価法は、語に含まれる文字の数や語の出現頻度を使って評価関数を定め、複数候補の中でその評価関数の値が最も大きいものを優先する手法である。評価関数は経験的に求める。

- 確率的言語モデル法

確率的言語モデル法は、品詞タグ付けされた形態素情報を持つコーパスを用い、品詞の 2-gram モデルや隠れマルコフモデルに基づく探索処理により候補を求める手法である[Kita 99][Tsuji 99]。

(9) 英語の形態素解析法

英語では、そもそも文を分かち書きするので、語への分割を行う必要はない。従って、接辞の同定および屈折形の同定を行うことになる。英語の接頭語の例を図 2-6 に示す。

dis-, en-, fore-, im-, in-, ir-, mid-,
mis-, non-, re-, un-, vice-, well-

図 2-6 英語の接頭語の例

また、接尾語の例を図 2-7 に示す。

-ed, -en, -er, -ese, -ess, -ful, -hood,
-ible, -ic, -ical, -ion, -ish, -ist, -ity,
-ive, -ization, -ize, -less, -ly, -ment,
-ness, -or, -ous, -ry, -ship, -th, -ure,
-ward, -wards, -y

図 2-7 英語の接尾語の例

屈折処理は、たとえば、動詞については、規則変化動詞に対しては変化規則を使い、不規則変化動詞に対してはそれぞれの語の変化を示した変化表を使う [Grishman 86].

2.3.3 構文解析

構文解析は、形態素に分けられた語を用いて、文の構造を明らかにするものである。一般に、どのような言語でも単語を任意に並べただけでは正しい文とはならない。単語は何らかの規則に従って文を形成する。この規則性を整理したものを文法という [Masuoka 97].

自然言語の代表的な文法に Chomsky の句構造文法がある [Nomura 88]. 句構造文法では、単語は規則に従って句と呼ばれるまとまりを作り、句は他の単語と共にさらに大きな句を作る。そして最終的に文ができあがる。句構造文法では、この規則を句構造規則と呼ぶ。規則は、句の範疇を表す非終端記号と、語に相当する終端記号を使って記述され、次のような形をしている。

$$\alpha \rightarrow \beta$$

ここで、 α と β は非終端記号または終端記号から構成される記号列である。この規則は、書き換え規則や生成規則とも呼ばれ、左辺の記号列が右辺の記号列に置き換えられることを表している。

具体的には、2.3.2 項で示した図 2-3 の文を生成するには図 2-8 のような規則が考えられる。

① S	→	PP VP
② PP	→	NP P
③ VP	→	PP VP
④ VP	→	V TENS
⑤ NP	→	NP の NP
⑥ NP	→	NP と NP
⑦ NP	→	N
⑧ N	→	私 車 買い物 先生
⑨ P	→	は が を で に
⑩ V	→	行 k 来 r
⑪ TENS	→	ru ita

S 文	PP 後置詞句
VP 動詞句	NP 名詞句
P 後置詞	V 動詞
N 名詞	TENS 時制

図 2-8 日本語の句構造規則の例

句構造文法では、左辺に開始記号を持つ句構造規則を使って右辺の記号列に置き換える。そして、新しくできた非終端記号を左辺に持つ句構造規則を使って、その記号を右辺の記号に置き換えるという操作を繰り返す。これによって文が生成される。具体的には、2.3.2 項で示した図 2-3 を構成する形態素は図 2.9 のように生成される。句構造規則の適用は非終端記号が無くなるまで続けられ、最終的に形態素だけからなる文が得られる。

	→	PP VP
S	→	NP P VP
	→	N P VP
	→	私 P VP
	→	私は VP
	→	私は PP VP
	→	私は NP P VP
	→	私は N P VP
	→	私は車 P VP
	→	私は車で VP
	→	私は車で PP VP
	→	私は車で NP P VP
	→	私は車で N P VP
	→	私は車で買い物 P VP
	→	私は車で買い物に VP
	→	私は車で買い物に V TENS
	→	私は車で買い物に行 k
	→	TENS
		私は車で買い物に行 k ita

図 2-9 形態素を生成する例

この規則は、図 2-10 のような木構造でも表現することができる。これを構文木という。このような構文木を使うことで、一次元的に並べられた単語の列では表現できない文の構造を表すことが可能となる [Matsumoto 04].

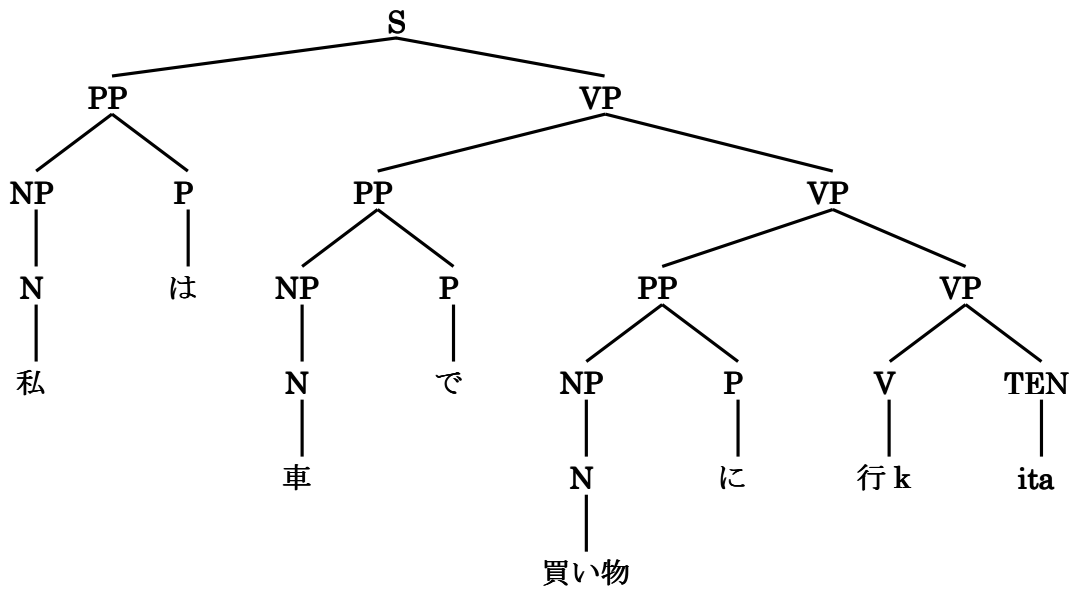


図 2-10 構文木の例

2.3.4 意味解析

文の意味は、文を構成する単位である個々の単語の意味と、それらの単語と他の単語との間にある意味的な関係によって決まる。自然言語処理における意味解析の目的は、このような文の意味を形式的に表現することである[Nagao 96]。この形式的な意味表現は、図 2-11 のような性質を備えている必要がある。

- 曖昧さを持たない。
- 解釈や推論の方法を備えている。
- 構文木のような文の形式的な構造から機械的な手順で変換できる。

図 2-11 意味解析のための形式的表現の条件

本項では、これらの条件を満たす形式的な意味表現の方法と文の形式的な構造から意味表現を求める方法について述べる。具体的には、文を構成する単語

の役割を明確化する格文法について示し、文の構造を係り受け関係で解析する係り受け文法について説明する。

(1) 格文法

構文解析により文の解析木が得られると、文を構成する語と語の間に一定の関係が成り立つ。たとえば、動詞などの述語と名詞の関係は名詞がその述語に対して果たしている役割とみることができる。この役割を格と言う。自然言語の類型論的特徴に表 2-4 に示す 3 種類があるが、膠着的な性質の強い日本語では、述語と名詞の間の文法的な関係は名詞に付加された格助詞によって表現される。

表 2-4 言語の類型論的特徴

孤立的	語が変化せず、文法的な関係は語順を用いて示される。
屈折的	文法的な関係が語形変化によって表現される。
膠着的	文法的な関係が語に付加される前置詞や後置詞によって表現される。

このように語順や語の形、付加された助詞などの表層的な手がかりから決まる構文的な役割を表層格という。これに対して、文中の述語に対して他の語が果たす意味的な役割を深層格とよぶ。表 2-5 に深層格を示す。

また、述語を中心として文中の名詞がその述語に対してもつ深層格を示した木を格構造表現という。格文法では、図 2-12 に示すように、格構造のそれぞれの格要素が表層格の主格の位置に置かれることによって多様な文を生成することができる[Nagao 99]。

表 2-5 深層格

格の名称 (記号)	意味的な役割
動作主格 (A)	動作を引き起こすもの
経験者格 (E)	心理事象を体験するもの
道具格 (I)	出来事の直接原因となるものや心理事象と関係して反応を起す刺激となるもの
対象格 (O)	移動する対象物や変化する対象物または判断や想像のような心理事象の内容
源泉格 (S)	対象物の移動における起点または状態変化や形状変化における最初の状態や形状
目標格 (G)	対象物の移動における終点または状態変化や形状変化における最終的な状態や形状
場所格 (L)	出来事が起こる場所や位置
時間格 (T)	出来事が起こる時間

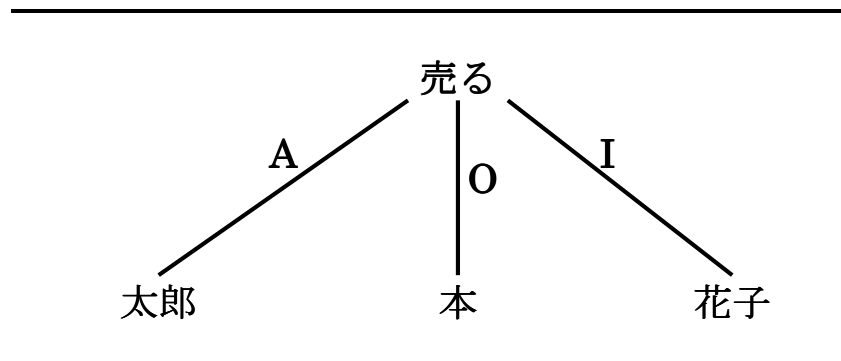


図 2-12 格構造表現の例

(2) 係り受け文法

前項の格文法では、図 2-12 で示すように、文を構成する格要素「太郎」、「花子」、「本」と述語「売る」の間に、動作主格や対象格などの格関係があると説明される。これは、知識の表現形式として提案された意味ネットワークやフレ

ームとの親和性も良く、第一階述語論理とも密接な関係がある。このように、語順に対する制約が比較的緩く、「が」、「を」、「に」などの格助詞が格関係を示している日本語には、古くから係り受け文法または依存文法という考え方がある。係り受け文法では、文の構造を、係り受けとよぶ修飾・非修飾の関係にあるとき、文頭側の文節が文末側の文節に係るといい、文末側の文節は文頭側の文節を受けるといふ [Shimooka 95]。

係り受け文法では、表 2-6 のような係り受け構造の制約を設けている。

表 2-6 係り受け構造の制約

制約	説明
係りの唯一性	文節は、文末側にある唯 1 つの文節に係る。
係り受けの非交差性	係り受け関係は基本的に交差しない。
係り受け関係の非重複性	文節は同じ係り受け関係で 2 つ以上の文節を受けない。

係り受け文法は、日本語の文を文節の連鎖としてとらえることで、日本語の構造を簡潔に記述しようとしている。最近では、EDR 電子化辞書のデータを知識元として、係り受け構造を利用した日本語意味解析システムなども研究されている [Harada 01]。

2.3.5 評価情報抽出

個人の発信する意見には、評価を記述するもの、要望・要求・提案の表明、不安・懸念・不満・満足などの感情を表すもの、認識・印象を述べるもの、賛否の表明するものが含まれる。そして題材となるテキストデータには、意見の収集、集約目的のテキストである社会調査などによる自由回答アンケートやレビューおよび潜在的に意見を含むテキストにチャットや Weblog が存在する。

評価分析とはその意見の中から評価を推定することを目指した分析で、前項で解説した係り受け文を用いて評価情報の抽出を行う。評価情報の評価極性には「良い」といった肯定的な極性を持つ評価と「悪い」といった否定的な極性を持つ評価、どちらでもない中立の評価が存在する [Inui 2006]。

小林らは、評判分析に用いる肯定・否定表現の評価は、〈対象〉、評価対象の具体的な項目である〈属性〉、および 〈評価〉の3要素から構成されるとした。そして評価を行うには対象、属性、評価の辞書を準備し、〈対象、属性、評価〉の表現を抽出し、組にまとめ上げる必要があることを指摘した。また、この3要素は対象領域に依存しており、たとえば電気製品の属性にはなり得ても、料理の属性ではないといったことが発生する。そのため辞書にノイズが発生しないように注意が必要である [Kobayashi 2003]。小林らの手法は、収集した評価表現における評価極性は決定しないため、それらを決定する必要がある。属性とは評価対象となる具体的な項目であり、宿泊施設を対象として分析を行う本研究においては、特徴を判断する重要な要素である。この構成を用いて感想テキスト情報の係り受けに対し、属性となる係り元の単語と、評価となる係り先の単語による組合せに対し、対象への評価の肯定および否定を分類した辞書を用意することで対応する必要がある。

2.3.6 辞書編纂

個人の発信する意見には同一内容の表現においてその表現方法にばらつきが存在する。そのため、出現数の多い項目についてグルーピングを行い、同一内容を表す単語の揺らぎをまとめる類義語辞書の作成の必要がある。その他にも商品名などの未知語の登録を行う形態素辞書、評価が肯定的か否定的かを判断する評価属性辞書の登録を行う。その後、改めてテキストマイニングを実施し、結果の検証を行い、必要に応じて辞書群を追加することで精度を向上させていく。

具体的には以下のような構造となる

(1) 類義語辞書

同様の意味をあらわす言葉を1つにまとめる辞書で、本研究においては、宿泊検討者の観点から単語の利用用途が同様と想定されるものをまとめ、類義語とした。

(2) 形態素辞書

本辞書は、未知語の登録に用いるものである。主に地名や施設名、商品名といった、分析システムに登録のない単語の登録を行う。

(3) 評価属性辞書

評価表現とは物事に対する人の評価を示す表現である。前項で示したように評判分析に用いる肯定・否定表現の評価は、「対象」、「属性」、「評価」の3要素から成り立つ。また、評価表現には「良い」「好き」「最高だ」といった用言が多いが、「人気」「絶品」といった名刺も該当する場合がある。評価表現には共通の表現と、その対象が属するカテゴリによって固有の表現が存在する [Tateishi 2001]。

なお、評価表現については一般的には「良い」、「悪い」といった評価単語の意味により肯定、否定の評価を与えることができる。しかし属性と評価との係り方により、肯定になる場合と否定になる場合の両方の意味を持つ単語が存在する。たとえば、係り受け表現を抽出し「料金が高い」は否定表現であるが、「品質が高い」は肯定表現であるというように評価が変化する単語に対応するため、係り受け解析の結果として紐づけられた表現を適宜確認しながら、手作業による辞書の作成を行った。また乾らの指摘もあった、肯定、否定のどちらでもない表現、「料金が普通」といった中立表現や不明瞭な表現については評価極性を与えない設定とした。

2.3.7 宿泊レビュー分析

本研究では宿泊レビューの感想コメントを分析対象として取り扱うが、そこ

には旅行者の感想，行動，事実，評価などの様々な内容の事柄が書き込まれている．そこから自然言語処理を用いて解析し，宿泊施設の評価が行われた文を抽出し，分析する必要がある．本研究に用いる，テキストマイニングによる分析手法について紹介する．

（１）係り受け頻度分析

本分析は，感想コメント中に現れる係り受け表現の頻度をカウントし，表やグラフに表すもので，対象とする施設やエリアにおいてどのような係り受け表現が高い頻度で用いられるかを分析する．

（２）評価分析

係り受け頻度分析の結果をもとに，対象への肯定，否定の分類を行うものである．感想レビューにおいて対象に肯定表現と否定表現が存在するため，評価属性辞書を用いて分類を行い，集計を行う．本研究では，感想レビューから導かれる肯定，否定評価の結果を数値評価項目へ反映し，数値評価項目別の感想レビュー情報における注目度の違いの発見を意図している．そのため，評判分析については，対象のどの数値評価項目に対する肯定，否定の評価が行われているかを検証するために分析を行う．

宿泊レビューを用いたテキストマイニングによる先行研究として，田邊らは施設特有の特徴項目の分類と宿泊施設の戦略検討への活用についての研究を行い，テキストマイニングを利用したレビューの記述から利用者の声を分析した．その結果は料理や部屋に対し，利用者は高く評価していることを示した [Tanabe 2008]．しかし宿泊施設に関する知見やテキストマイニングの際に必要な類義語などの辞書編纂を行わずに分析を実施しているため，結果の羅列のみとなっている．また，テキストマイニングを用いた分析結果をもとに，宿泊施設の利用検討者への情報提供方法についての提案およびビジネスエリアへの分析には言及されていない．

2.4 結言

本章では、インターネット宿泊予約サイトにおいて、宿泊施設を選択する際に活用する情報と宿泊レビューが閲覧者の行動に影響を及ぼす効果の検証を行った。そして宿泊レビューを用いた研究について検証し、宿泊施設側のサービス価値向上のための提案および、利用者に向けた評価情報の抽出の提案が行われてきたことが確認できた。

次に自然言語処理およびテキストマイニングに関して概観した。宿泊利用者はプロフィール、数値評価項目および感想コメント情報をそれぞれ書き込む。自由記述文である感想コメントから評価情報を集約する手法について解説を行った。

これまでの研究の課題を示すと以下のようなになる。

- (1) 数値評価と感想コメントを同時に評価する方法の検討
- (2) 宿泊エリアおよび宿泊施設それぞれの特徴の分析
- (3) 宿泊利用者の感想コメントにおける評価表現の特徴分析の深度化

利用者から寄せられる感想コメントと、数値評価情報の二種類の評価指標が同時に記載されるにもかかわらず、これらを合わせて分析し、数値評価項目と感想コメントの関係性を検証した事例はない。また研究対象がレジャー用途に旅行目的が絞られている傾向が強く、ビジネス用途の宿泊施設の特徴に関する検討を行った事例が存在しない。これらより宿泊レビューを用いたより詳細な宿泊施設の評価情報の抽出と提示方法に課題が存在することがわかる。

(1) の課題については本研究の根底となる研究課題として、各章においてその検討を行う。(2) の課題については第3章において宿泊エリアにおける検討を行い、第4章において宿泊施設における検討を行い、その特徴を示すこととする。(3) の課題については評価表現の中で、文の途中で評価表現が肯定から否定、または否定から肯定に遷移するものに焦点をあてその分析を行う。

第3章

宿泊レビューを用いたエリア特徴の分析

本章では、宿泊レビューの感想コメントと数値評価を用いて宿泊検討者が宿泊するエリアを選択する際に、重要な判断材料となる特徴情報の抽出手法と、その特徴の提示方法について提案する。

3.1 宿泊レビューの利用目的

宿泊検討者はインターネット宿泊予約サイトに、様々な旅行の目的をもって訪れる。それはビジネス出張での業務目的の利用であったり、家族や友人と観光目的での利用であったりし、利用目的により宿泊検討者が宿泊施設を選択する際に知りたい情報は異なる。また、宿泊レビューを閲覧し、判断材料として注目する評価項目は異なるはずである。そして宿泊後に、宿泊施設を利用した宿泊利用者の記載する宿泊レビューにおいても宿泊施設の存在するエリアに応じた感想と評価の特徴が表れるはずである。

エリア間の利用者の誘致競争が激しい昨今、宿泊施設の立場からも、大都市に立地する宿泊施設と、観光地の宿泊施設で対象とする顧客は異なっている。利用者の評価からエリアの特徴を分析し、宿泊利用者の評価を知ることで満足度を高めることができる事柄に対し、サービスの改善を行っていくことは重要な話題である。

これらの状況を踏まえて、本章では宿泊レビューを用いた宿泊エリアの特徴分析を行う。宿泊レビューの3つの構成要素であるユーザープロフィール、数値評価、感想コメントのうち、感想コメントに注目し、利用者の宿泊目的の異なるエリアを対象としたテキストマイニングによる評価表現の抽出と、感想コメントの評価表現分析を行い、利用者が重要視した評価対象項目を抽出する。そして、もう1つの評価指標である数値評価と感想コメントとの関連性に注目した宿泊レビューの評価方法を提示する。また抽出した評価表現を数値評価と関連付け、宿泊レビューの2つの評価要素を連結した、宿泊検討者に向けた情報表示手法を提示する。

3.2 宿泊エリア分類

本節では旅行における宿泊エリア分類の意味について検証する。

御園は宿泊施設を選択する際に、利用者は「いつ」「どこに」「だれと」「どのような目的で」宿泊するかを考慮する。その際に「どこに」は重要な観点であり、宿泊する宿泊施設だけの問題ではなく、その宿泊施設が属するエリア地域としての主体性が重要であるとした [Misono 2013]。大井はエリアの魅力度に関する調査の傾向を分析し、観光地のブランド調査に必要な項目として観光地の資源や独自性などの指標とともに宿泊施設が魅力の構成要素として含まれることを指摘している [Oi 2013]。また、日本全国各地の都道府県、市区町村では観光課や観光協会が設けられ、当該エリアの観光資源への誘客をアピールしている。観光庁では毎年宿泊旅行統計調査を実施し、各都道府県の宿泊旅行の実態把握を全国規模で行っている [JTA 2013b]。これまでの観光統計は、日本全体の動向や、県単位の独自調査に限られ、統一した観光統計の整備がされていない現状がある。そのため都道府県をまたがるエリアを比較し、分析することは難しいのが現状である。また、これまでの研究ではいずれの場合も地域内の観光地の名所や食事場所を含めた観光資源全般を対象とし特徴の抽出に注力されており、宿泊施設を中心に置いた検証は行われていない。

これらの問題に対し、インターネット宿泊予約サイトは全国の宿泊施設を網羅し、宿泊予約サービスを提供していることから、エリアを横断した特徴の分析を行うことができる。そこで本研究では宿泊施設のみにその対象を絞り込み、宿泊エリアの宿泊レビューの検証を行う。宿泊予約サイトを用いて実際に宿泊した利用者からの宿泊レビューを分析することで、これらの観光統計では比較が困難であった地域をまたがる観光地どうしの特徴の違いおよび、ビジネスでの利用と観光での利用の両面での特徴の違いを分析する。エリアの特徴情報を知ることで、宿泊利用者はどのエリアに旅行に行くかを検討する材料を得ることになり、宿泊施設の場合は、エリアの評判、求められる要素を知ることによって周辺施設との違いを創出したり、施設の良さをアピールする際に利用したりすることが出来る。

3.3 数値評価の分析

本節では、宿泊レビューの3要素のうち、数値評価について議論する。数値評価項目は、宿泊予約サイトごとに表示方法は異なるが、その項目に大きな差異はなく、立地、スタッフの対応、快適性、価格などの項目と、それらの項目を総合的に宿泊利用者が評価した項目である総合評価で構成されている。数値評価は一般的に利用される評価項目だが、利用者はどのような意図をもって評価しているのだろうか。数値評価項目と利用者の数値評価の性質について検証する。

3.3.1 数値評価項目

本項では、宿泊レビューの数値評価項目についてまとめる。数値評価とは、宿泊予約サイトが独自に決めた、宿泊検討者にとって重要と思われる項目を列挙したもので、宿泊利用者がその項目に対して採点を行う形式で運営されている。特に宿泊レビューの場合は、宿泊予約サイトを利用して予約し、実際に宿泊した後にしか評価が行えないものが多い。宿泊者の会員登録など利用者確認を行った上でのレビューの投稿となるため、誰でもが評価を書き込める掲示板サイトなどと比べスパムが少なく、より実際の宿泊利用者の声を反映していると考えられる。表 3-1 に国内の宿泊予約サイトの数値評価項目をまとめたものを、表 3-2 に、海外宿泊予約サイトの数値評価項目をまとめたものを示す。

表 3-1. 国内宿泊予約サイトの数値評価項目

評価項目	じゃらん net ¹	楽天トラベル ²	るるぶ ³	一休.com ⁴	宿ぷらざ ⁵
施設	清潔感				清潔度
	部屋	部屋 設備・アメ ニティ	部屋設備	客室・アメ ニティ	設備
				ひと休み 度 (満足 度)	
	風呂	風呂	風呂	バスルー ム・お風呂	お風呂
接客	接客・サー ビス	サービス	サービス	接客・サー ビス	サービス
立地		立地	立地		立地
価格					料金
食事	料理	食事	食事	お食事	お食事

¹ じゃらん net : <http://www.jalan.net/>

² 楽天トラベル: <http://travel.rakuten.co.jp/>

³ るるぶトラベル: <http://rurubu.travel/>

⁴ 一休.com : <http://www.ikyuu.com/>

⁵ 宿ぷらざ: <http://www.tabiplaza.net/>

表 3-2.海外宿泊予約サイトの数値評価項目

	Booking ⁶	Agoda ⁷	HotelClub ⁸	Expedia ⁹
施設	清潔さ	ホテルの状態/清潔さ	設備	清潔感
	快適さ	お部屋の快適さ	客室の快適さ	客室
接客	サービス			サービス
	スタッフ	スタッフの対応	ホテル・スタッフ	
立地	ロケーション	ロケーション	ロケーション	
価格	お得感	価値	コストパフォーマンス	価格
食事		お食事		

表 3-1 に示す国内宿泊予約サイトと、表 3-2 に示す海外宿泊予約サイトでは、いくつかの数値評価項目に差異がある。国内宿泊予約サイトが独自に決めた数値評価項目には、表記方法や順序に差異はあるが、その内容を大まかに分類すると「施設」、「接客」、「立地」、「価格」、「食事」の 5 つから構成され、施設における「風呂」や「食事」の数値評価は必ず存在する。一方、海外宿泊予約サイトでは、これらの数値評価はほぼ存在しない。

この理由として、宿泊場所と食事場所の分離が定着している海外とは異なり、日本では夕食と朝食の付いたレジャー目的の温泉旅館などが多く、国内宿泊予約サイトではこれらを考慮した評価項目となっていることが考えられる。

しかし、ビジネスホテルの宿泊検討者で食事の良し悪しを重要視する人は多数存在するであろうか。少なくとも宿泊利用者のビジネスホテルの感想コメントでは、食事に関するコメントは極めて少ない。また、同じ「立地」という数値評価項目に対する感想コメントでも、ビジネス利用においては駅からの距離

⁶ Booking : <http://www.booking.com/index.ja.html>

⁷ Agoda : <http://www.agoda.jp/>

⁸ HotelClub : <http://www.hotelclub.co.jp/>

⁹ Expedia : <http://www.expedia.co.jp/>

や近隣にコンビニがあるかなど利便性が重要視されるが、レジャー利用において景観を楽しむ地域では部屋から見える眺望などが重要視される。

このように、利用目的やエリアにより重要視する情報が異なるため、同じ数値評価項目でも宿泊利用者の評価する事項は多種多様となっている。

3.3.2 宿泊利用者の採点

宿泊利用者はどのように宿泊施設を採点しているのだろうか。表 3-3 は、宿泊予約サイト「じゃらん net」における数値評価の採点を集計した結果である。集計に用いたデータは、ビジネスエリア 5 地区（東京、大阪、名古屋、博多、札幌）、レジャーエリア 5 地区（箱根、湯布院、草津、登別、道後）の計 10 地区を選択した。そして各地区でそれぞれ書き込みの多い 10 の宿泊施設を選択し、各宿泊施設の宿泊レビュー中で 2011 年 11 月時点での最新のものからそれぞれ 100 件、合計で 10,000 件を抽出したものである。

表 3-3. 数値評価における宿泊利用者の採点

	評価	総合	部屋	風呂	料理	接客	清潔感
分布 頻度	1	109	143	105	85	170	112
	2	367	352	396	263	330	329
	3	1,121	1,725	1,922	935	1,521	1,485
	4	4,089	3,910	3,575	2,371	3,342	3,478
	5	4,314	3,870	3,891	3,700	4,637	4,596
分布 割合	1	1%	1%	1%	1%	2%	1%
	2	4%	4%	4%	4%	3%	3%
	3	11%	17%	19%	13%	15%	15%
	4	41%	39%	36%	32%	33%	35%
	5	43%	39%	39%	50%	46%	46%
(件数)		10,000	10,000	9,889	7,354	10,000	10,000

表 3-3 から明らかなように、宿泊利用者の採点の大半は 4, 5 の評価であり、基準である 3 よりも高い評価が中心である。1, 2 という低評価は全体の 10%に満たない。このように偏りのある数値評価では当該宿泊施設の特徴を明

確に表すのは難しく、また宿泊施設どうしの特徴の違いを示すことも難しい。

次に宿泊利用者が個別の数値評価項目をそれぞれ異なる数値評価を行っているかを調べる為、同データを用いて、すべての数値評価項目を採点している7,287件について、その採点傾向の集計を行った。その結果を表3-4に示す。

表3-4. 数値評価における宿泊利用者の採点傾向

	件数	割合
全項目で同一の採点	1,361	18.7%
1項目のみ異なる採点	1,913	26.3%
2項目以上で異なる採点	4,013	55.0%
合計	7,287	100.0%

数値評価項目の全項目に対しすべて5やすべて4といった同一の採点のみを行っていた宿泊利用者が18.7%、1項目のみ異なる採点を行った宿泊利用者が26.3%であった。ここから、宿泊利用者の採点の多くは、特に気になった項目以外は適当に4や5を付与していることが読み取れる。逆に言えば、1項目のみ異なる採点をした項目は、宿泊利用者にとって強く印象を持った採点項目であるといえる。

数値評価項目は、様々な宿泊予約サイトで採用され、ほぼ同様な評価項目を用いて利用者からの評価情報を集めている。しかしその数値評価は集計すると情報に偏りが生じることとなり、参考情報としての意味合いが低くなってしまふことがわかる。また、数値評価項目は宿泊利用者が特徴として評価した対象項目である。その対象項目がもつイメージによる宿泊施設の評価はできるが、それらの対象項目に含まれる具体的な属性要素に対する評価の詳細を知ることができない。具体的な評価を知るにはさらに別の記載情報を調べる必要がある。

3.4 感想コメントの分析

感想コメントは、宿泊利用者が実際に当該宿泊施設を利用して感じたことを文章記述したものである。それゆえ、数値評価と異なり入力に対する手間も大きい。それにもかかわらず記入した項目なので、感想コメントは宿泊利用者が強い印象を持った事項に対する率直な意見であると考えられる。

しかし、宿泊利用者の生の声であるがゆえ、分析には多くの問題が生じる。まず一番に解決すべき課題は、表記の揺れである。「風呂」と「バス」、「スタッフ」と「係員」など同義語の表記の揺れが多い。また、意図の解釈においても、その係り受けを見なければ判定はできない。たとえば、「大きい」についても、「騒音が大きい」は否定的な意見であり、「TV が大きい」は肯定的な意見である。さらには、携帯電話などからの感想コメントの書き込みができるサイトも多く、それゆえ口語調の書き込みも散在しており、自然言語処理技術による解析が困難である。

以降、本節ではこれらの課題を解決し、感想コメントから宿泊検討者にとって有意義な情報を抽出するために提案する手法について説明する。

3.4.1 テキストマイニング

前節に示したように、宿泊検討者の宿泊施設に対して重要視する事項は、利用目的により大きく異なるはずである。そのためこれらを一括して分析することは困難である。そこで本節では、テキストマイニングの手法を用いた分析を行う。図 3-1 に示すように利用目的を大きく「ビジネス」と「レジャー」の2つのエリアに分割して、各エリアの感想コメントを用いてそれぞれの評価の特徴を分析する。さらにエリア内での地区の違いに由来する詳細な特徴の違いの分析を試みるため、エリア内の地区別評判分析も実施することとした。

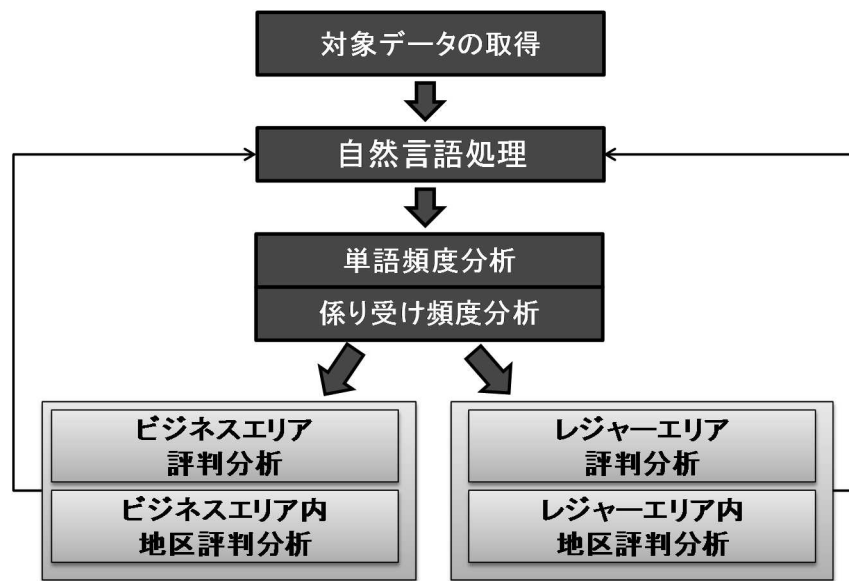


図 3-1. テキストマイニングによる分析フロー

3.4.2 形態素辞書

自然言語処理では、まず形態素解析を行い、文を形態素に分割すると共に、各形態素に品詞情報を付与する。このとき、旅行に関する専門用語は形態素辞書に登録することで対処した。形態素辞書に登録した単語は、主にブランドや商品名を中心とした造語である。これらは対象となる事物の特定につながるため、追加登録を随時行うこととした。

感想コメントには、口語調の表記が多く存在する。形態素解析において、口語調文書の処理は文語調の処理と比較して解析精度が低下する傾向にある。しかし、本研究では、当該宿泊施設やその施設が立地する地域の何に対して満足や不満を感じたかを示す特徴表現を抽出できれば、その後の処理に影響を及ぼさない。さらに、たとえば良いことを示す「グ～～」など口語調で記載された評価単語を形態素辞書に登録することで対処した。また、顔文字などの書き込みも見受けられる。これに対しては、鈴木らの研究成果を活用することで対処した[Suzuki 2008].

3.4.3 類義語辞書

感想コメントを解析するにあたり、宿泊利用者は様々な単語を使い表現するため「バス」と「風呂」などの異表記の同義語もしくは類義語が多々存在しているという課題がある。この課題に対しても、異表記の同義語もしくは類義語を類義語辞書としてグルーピングすることで対処した。表 3-5 に類義語辞書の一例を示す。

表 3-5 類義語辞書

代表語	類義語例			
宿泊施設(類)	ホテル	旅館	宿泊先	宿泊施設
部屋(類)	客室	お部屋	おへや	部屋
風呂(類)	風呂場	バス	湯船	浴室
食事(類)	料理	お料理	オカズ	ごはん
接客(類)	応対	配慮	心遣い	接客態度
掃除(類)	お手入れ	お掃除	手入れ	清掃
立地(類)	ロケーション	立地条件	立地状況	立地上

3.4.4 感想コメントの肯定・否定の評価判定

宿泊検討者が感想コメントから欲する情報とは、当該宿泊施設の情報に関する、宿泊施設が提供する情報だけでは理解できなかったこと、宿泊利用者が実際に利用して感じたこと、施設情報に掲載されていない宿泊施設周辺の情報などである。これら宿泊検討者の要望を満たすためには、宿泊利用者が当該宿泊施設やその施設が立地するエリアの何に対して満足や不満を感じたかを明らかにする必要がある。

感想コメントの特徴表現には、「良い」、「悪い」などという明瞭な肯定、否定を表す評価単語を用いた表記と、たとえば「部屋が広かった」のように事実の記載であるが、明らかに肯定と判断できる表記がある。本章では、前者を感性表現と呼び、後者を事実表現と呼ぶこととする。

感性表現の処理は、感性表現の対象物を明確にすればよい。よって、自然言語処理における構文解析により、感性表現の係り受け表現を抽出し対象物を明確にする。事実表現では、対象物と評価の係り方により、評価が異なる単語が存在する。たとえば、「料金－高い」は否定表現であり、「品質－高い」は肯定表現である。そこで、本研究では、小林らの評判分析に関する研究を、宿泊レビューに適用すると共に拡張することで対応した[Kobayashi 2006].

小林らは、評判分析に用いる肯定・否定表現の評価は、〈対象〉、評価対象の具体的な項目である〈属性〉、および 〈評価〉の3要素から構成されるとしている。本研究では、数値評価との連携を考慮し〈対象〉は数値評価の項目とし、〈属性〉を“大浴場”や“ベッド”など具体的な評価物、〈評価〉を評価物の状態とした。そして、〈属性〉と〈評価〉の係り受けに対して、肯定極性・否定極性を付与した評価属性辞書を構築した。構築した辞書の一例を表3-6に示す。

表 3-6 評価属性辞書

感性表現

〈対象〉	〈属性〉	実際の表記	〈評価〉	評価極性
接客	従業員（類）	スタッフ	良い	肯定
料理	夕食（類）	ディナー	美味しい	肯定
風呂	特徴風呂（類）	露天風呂	良い	肯定
料理	朝食（類）	朝食バイキング	残念	否定
風呂	お風呂（類）	大浴場	残念	否定
部屋	部屋（類）	部屋	残念	否定

事実表現

〈対象〉	〈属性〉	実際の表記	〈評価〉	評価極性
部屋	部屋（類）	部屋	広い	肯定
立地	最寄駅（類）	大阪駅	近い	肯定
風呂	風呂設備（類）	バスタブ	広い	肯定
部屋	部屋（類）	部屋	狭い	否定
部屋	客室寝具（類）	ベッド	堅い	否定
風呂	お風呂（類）	大浴場	遠い	否定

表中の〈属性〉に示す(類)が付いた単語は、類義語辞書に登録している代表表記であることを表している。感想コメントには表記の揺れが多いため、類義語

辞書に登録している語に対しては、その代表表記を〈属性〉として登録することで評価属性辞書への登録数を削減した。

3.4.5 感想コメントにおける属性の確定

日本語の特徴として、主語を省略する傾向がある。感想コメントにおいても同様の傾向がみられ、〈属性〉が曖昧になったり、抽出できなくなったりするという課題が生じる。また、1つの〈属性〉に対して〈評価〉が2つあるものや、同時に複数の〈属性〉に対して〈評価〉をしているものがある。

このような課題に対して、本研究では属性継承の手法を用いて対応した。属性継承の基本的な考え方は、文を前方から解析していき、一度〈属性〉を抽出すると次の〈属性〉を抽出するまで、抽出している〈属性〉を継承し続けるというものである。属性継承は、〈属性〉が属する〈対象〉ごととする。

図 3-2. 属性継承の事例

例文 No.	例文
〈対象〉	〈属性〉継承の状況
例文 1	<u>風呂</u> は狭いけど、 <u>ベッド</u> は広かった。
施設	風呂 → ベッド →
例文 2	<u>風呂</u> も <u>ベッド</u> も狭かった。
施設	風呂 → &ベッド
例文 3	<u>風呂</u> は狭いけど、 <u>予想通り</u> であった。
施設	風呂 →————→
例文 4	<u>風呂</u> は狭いけど、 <u>値段</u> が <u>安い</u> ので <u>許す</u> 。
施設	風呂 →————→
価格	値段 →————→

以下、図 3-2 に示す事例を用いて属性継承の基本的なルールを説明する。ここでは<対象>を表 3-1 に示す数値評価項目とし、例文中の一重下線は<属性>を、二重下線は<評価>を表す。

例文 1 と 2 は、施設の<属性>“風呂”，“ベッド” が出現する。例文 1 では、最初に出た“風呂”は“ベッド”が出現するまで継承し，“ベッド”は出現位置から文末まで継承する。例文 2 では，“風呂”，“ベッド”は並列句と判断されるため、この 2 つを文末まで継承し、「風呂ー狭い」と「ベッドー狭い」という 2 つの評価属性を抽出する。例文 3 では、施設の<属性>として”風呂”が抽出され、文末まで継承する。これにより「風呂ー狭い」と「風呂ー予想通り」の 2 つの属性評価を抽出する。

例文 4 では、まず“風呂”が施設の<属性>として抽出され、9文字目で“値段”が抽出される。ここで，“値段”は価格の<属性>であるので、施設<属性>の“風呂”は文末まで継承する。ここまでの「風呂ー狭い」と「値段ー安い」が抽出される。文末の“許す”に対しては“風呂”か“値段”かが曖昧になる。しかし，“許す”は否定の許容なので否定極性を示す「風呂ー狭い」への対応と判断し、「風呂ー許す」を抽出する。このような処理を実現するため，“許す”や“我慢”などの一部の<評価>に対しては、辞書にどちらの極性を優先するか
のフラグを付けることで対処した。

3.5 テキストマイニングによるエリア分析

本節では、前節で提案した手法を用いて、感想コメントを分析した結果を示す。はじめにビジネスエリアおよびレジャーエリアについての特徴の違いを示す。次にビジネスエリア内、レジャーエリア内のそれぞれ地区による特徴の違いを示す。

3.5.1 分析データと分析システムの概要

分析に用いた感想コメントは、表 3-3 に示した「じゃらん net」における宿泊レビューのデータと同一であるが、その投稿あたりの平均文字数と単語数を表 3-7 に示す。

宿泊利用者の感想コメントは文字数、単語数いずれの場合もレジャーエリアでの数が大きく利用者にとって伝えたい事柄が多いことがわかる。

分析の際に利用する形態素解析は、オープンソースの形態素解析エンジン MeCab を用い、形態素辞書には IPA 辞書を登録した。本分析のために、形態素辞書には新たに 2,454 語を、類義語辞書には 158 組 1,505 語を登録した。また、評価属性辞書には類義語辞書に登録した 1 組を<属性>や<評価>における 1 語とみなして 1,825 組を登録した。

表 3-7 分析対象感想コメントの規模

エリア・地域		平均文字数	平均単語数
	東京	147.7	87.2
	大阪	146.4	86.4
	名古屋	139.8	83.0
	博多	140.5	83.6
	札幌	148.6	87.8
ビジネスエリア計		144.6	85.6
	箱根	222.6	134.4
	湯布院	226.5	136.1
	草津	217.4	131.4
	登別	191.0	114.7
	道後	195.3	118.5
レジャーエリア計		210.6	127.0
全体計		177.6	106.3

3.5.2 エリア別特徴表現

感想コメント全体をビジネスエリア，レジャーエリアに二分し，係り受け表現を集計し，それぞれのエリアで注目される特徴表現の頻度を集計した結果を表 3-8 に示す。

出張利用者の多いビジネスエリアでの集計では、「最寄駅-近い」，「部屋-広い」，「部屋-狭い」といった事実表現と，「接客-良い」，「立地-良い」，「交通アクセス-良い」，「朝食-美味しい」，と接客，駅からの立地，朝食に対する感性表現が多く出現した。一方，レジャーエリアでは「お食事-美味しい」，「お食事-良い」，「お風呂-良い」，「眺め - 良い」などの感性表現がみられ，事実表現では「部屋-

広い」が上位にみられる。レジャーエリアにおいては、「良い」という表現を用いた感性表現により、滞在中の満足度を高める要素である食事と風呂に対する肯定的な感性表現が上位を占めた。

表 3-8 エリア別の特徴表現

ビジネスエリア			レジャーエリア		
〈属性〉	〈評価〉	頻度	〈属性〉	〈評価〉	頻度
接客(類)	良い	336	お食事(類)	美味しい	462
最寄駅(類)	近い	244	接客(類)	良い	408
部屋(類)	広い	217	感じ(類)	良い	266
立地(類)	良い	217	お風呂(類)	良い	237
部屋(類)	狭い	198	気持ち(類)	良い	198
清潔(類)	良い	194	部屋(類)	広い	189
朝食(類)	美味しい	179	お食事(類)	良い	158
感じ(類)	良い	140	清潔(類)	良い	120
交通アクセス	良い	124	眺め(類)	良い	104
最寄駅(類)	便利	110	温泉(類)	良い	103

すなわち、ビジネスエリアでは、立地と部屋に関して、具体的な事実を評価する事実表現の投稿が多くみられることから、旅先で普段通りの生活を維持するため、宿泊本来の目的である睡眠をとる場所としての利便性が重要視されていることが読み取れる。

一方、レジャーエリアでは、客室や立地の良さだけではなく、食事、風呂、その施設からの眺望といった旅先の宿泊施設で受けることになる非日常的な体験から感動を得る要素への評価が重要視されていることが分かる。

以上の結果から、宿泊利用者の宿泊施設に滞在する際に印象に残る項目、また評価する項目が異なることがわかる。そしてこのことは、宿泊検討者にとっては利用目的により、必要とする数値評価項目とその内容に違いがあることを表していると考えられる。

3.5.3 エリア別数値評価との対応

本節では、ビジネスエリア、レジャーエリアそれぞれにおいて、感想コメント中に出現する〈属性〉と、数値評価項目の対応について明らかにする。前節において、感想コメントの分析により宿泊利用者は利用目的により評価をする項目が異なることがわかった。そこでまず、感想コメントから評価属性辞書を用いて抽出した肯定・否定極性を持つ特徴表現を、〈対象〉の情報を用いて宿泊レビューの数値評価項目へ割り当て、その頻度を集計した。この分析においての〈対象〉は、「じゃらんnet」の数値評価項目と対応を取るため〈部屋〉、〈風呂〉、〈料理〉、〈接客〉、〈清潔感〉の5項目に感想コメント中に書き込みが多い〈立地〉を加えた6項目とした。

ビジネスエリアの集計結果を表3-9に、レジャーエリアの集計結果を表3-10に示す。

表 3-9 ビジネスエリア数値評価平均と特徴表現分析表

数値評価	部屋	風呂	料理	接客	清潔感	
平均点	4.1	3.9	4.1	4.1	4.2	

感想評価	部屋	風呂	料理	接客	清潔感	立地
肯定表現	813	439	694	916	350	1,404
否定表現	406	91	78	47	34	40
中立・他	48	32	22	23	1	11
総計	1,267	562	794	986	385	1,455

表 3-10 レジャーエリア数値評価平均と特徴表現分析表

数値評価	部屋	風呂	料理	接客	清潔感	
平均点	4.1	4.3	4.4	4.3	4.2	

感想評価	部屋	風呂	料理	接客	清潔感	立地
肯定表現	526	1,118	1,900	1,309	274	471
否定表現	164	233	184	74	29	19
中立・他	35	65	67	33	2	5
総計	725	1,416	2,151	1,416	305	495

これらの表は、数値評価と感想評価から構成されている。数値評価は平均点で示しており、感想評価はそれぞれの数値評価項目に対する評価表現の出現頻度を示している。

まず数値評価を見てみよう。数値評価の平均点は、ビジネスエリアでは 3.9～4.2、レジャーエリアでは 4.1～4.4 と、どちらのエリアもその振れ幅は 0.3 と極めて小さい値であった。このことから、宿泊利用者が投稿する数値評価は、それぞれを平均点にしてしまうとエリアの特徴が表れないといえる。

けれども、宿泊利用者が投稿する感想コメントから抽出した評価を用いると、差異が明らかになる。たとえば、表 3-9 に示すビジネスエリアの<部屋>と<接客>の数値評価項目の平均点数は同じ 4.1 となっている。しかし、感想コメントの分析結果では、「部屋」においては、1,267 コメント中 813 コメントの肯定極性となり、部屋に関する評価表現の出現数のうち、肯定表現が出現する率を調べたところ肯定率は約 64%であった。一方「接客」においては、986 コメント中 916 コメントが肯定極性となり、肯定率は約 93%であった。これらより、利用者は特に気になった数値評価項目に関連して評価をしており、評価表現の発生頻度が異なることがわかる。また肯定評価と否定評価の頻度は数値評価項目により異なっている。また、数値評価項目に対応する特徴表現の出現数より、宿泊利用者が何に注目したかが明らかになる。また、ビジネスエリアにおいては「じゃらん net」の数値評価項目にない「立地」に関する特徴表現が多く出現することがわかった。

次に、それぞれのエリア分析の結果から得られた特徴表現分析表を考察する。

まず、表 3-9 に示すビジネスエリアの結果であるが、抽出した特徴表現の頻度は「立地」、「部屋」、「接客」に関するものの順に多かった。このことから宿泊利用者は「立地」、「部屋」、「接客」に注目し感想コメントを投稿していることがわかった。「じゃらん net」の数値評価項目には存在しない<立地>に対する肯定評価の頻度が最も高い。これは、ビジネスエリアを利用する際に、数値評価項目として存在する指標より、立地が重要な要素であることを示しており、特徴的である。

一方、表 3-10 に示したレジャーエリアで同様の集計を行った結果では、抽出した特徴表現の頻度は「料理」、「風呂」、「接客」に関する評価の順に多かった。このことから、宿泊利用者は数値評価項目のうち、これらの項目に対する注目度が高いことが分かる。特徴表現の出現数から、レジャー目的の宿泊利用者は滞在中に泊まる部屋にあまり注目しておらず、滞在中の非日常的な満足を得るための項目である「料理」、「風呂」、「接客」に注目し、評価していることは興味深い結果である。

また、ビジネスエリア、レジャーエリアいずれの場合も、「部屋」と「風呂」に関しては他の数値評価項目と比較して、否定極性を持つ特徴表現の比率が高い結果となった。この 2 つの評価項目に対して宿泊利用者は、数値評価が高いにもかかわらず、不満を持っていることを示していることは、利用者にとってこの項目は評価の分かれるものであることを示しており、興味深い結果である。

ただし、図 3-2 において示した属性継承の概念を用いて、部屋の否定極性を持つ特徴表現について検証したところ、宿泊利用者は部屋の狭さに対して否定極性の評価を行うが、料金や、清潔感、立地といった別の属性との総合的な評価で、相応の満足をしているという評価の遷移する文章が存在することを感想コメントの文脈を調べることで明らかとなった。このことについては第 5 章で詳しく考察を行う。

以上、感想コメントを分析することで、数値評価の高低とは別に、エリアにより注目される数値評価項目が異なること、数値評価項目の優先順位が存在すること、数値評価の点数だけでは読み取ることができない差異があることが明らかになった。

3.5.4 地区別分析

次にビジネスエリア，レジャーエリア内の各地区の特徴について分析を行った。それぞれのエリア内の地区において評価の特徴の違いを検証する。表 3-11 から表 3-15 は，ビジネスエリアの地区別に特徴表現分析を行ったもので，表 3-16 から表 3-21 まではレジャーエリアの地区別に特徴表現分析を行ったものである。

特にビジネスエリアの地区における特徴表現には，立地に関する評価の高さを示す当該地区の最寄り駅名が多くみられた。また，ビジネスエリアの各地区において，「部屋の狭さ」という否定極性を持つ特徴表現の頻度は，東京，大阪，名古屋では多く，博多，札幌では少ないという傾向の違いがみられた。一方，レジャーエリアの各地区は，それぞれ温泉地であるにもかかわらず，最も重要視される数値評価項目は，「風呂」ではなく「料理」に関するものであることがわかった。次に重視される数値評価項目は，湯布院，道後の各地区では「接客」，箱根，草津，登別の各地区では「風呂」に対する特徴表現の頻度が高く，宿泊利用者が重要視する数値評価項目は地区により順序が異なることが明らかとなった。レジャーエリアの地区においては最寄りの観光地名やランドマーク的名所の単語が多くみられ，宿泊利用者の感想テキストに具体的な地名が書かれることが多かった。けれども，温泉の効能や泉質，郷土料理の内容の詳細など，より地区の特徴を示す内容の感想コメントは少なかった。

表 3-11 大阪エリア数値評価平均と評価表現分析表

数値評価	部屋	風呂	料理	接客	清潔感
平均点	4.0	3.7	4.1	4.1	4.1

感想評価	部屋	風呂	料理	接客	清潔感	立地
肯定表現	156	57	122	177	62	285
否定表現	122	21	18	8	10	9
中立・他	12	4	4	9	0	1
総計	290	82	144	194	72	295

表 3-12 東京エリア数値評価平均と評価表現分析表

数値評価	部屋	風呂	料理	接客	清潔感
平均点	4.1	4.0	4.2	4.2	4.2

感想評価	部屋	風呂	料理	接客	清潔感	立地
肯定表現	125	95	147	215	59	303
否定表現	78	16	16	5	4	10
中立・他	7	3	8	2	1	0
総計	210	114	171	222	64	313

表 3-13 札幌エリア数値評価平均と評価表現分析表

数値評価	部屋	風呂	料理	接客	清潔感
平均点	4.3	3.9	4.1	4.2	4.4

感想評価	部屋	風呂	料理	接客	清潔感	立地
肯定表現	184	85	166	185	69	255
否定表現	43	19	11	8	4	7
中立・他	16	17	3	3	0	1
総計	243	121	180	196	73	263

表 3-14 名古屋エリア数値評価平均と評価表現分析表

数値評価	部屋	風呂	料理	接客	清潔感
平均点	3.9	3.7	4.0	4.0	4.1

感想評価	部屋	風呂	料理	接客	清潔感	立地
肯定表現	142	116	142	162	77	264
否定表現	103	25	20	16	9	5
中立・他	8	7	7	4	0	5
総計	253	148	169	182	86	274

表 3-15 博多エリア数値評価平均と評価表現分析表

数値評価	部屋	風呂	料理	接客	清潔感
平均点	4.3	3.9	4.1	4.2	4.3

感想評価	部屋	風呂	料理	接客	清潔感	立地
肯定表現	206	86	117	177	83	297
否定表現	60	10	13	10	7	9
中立・他	5	1	0	5	0	4
総計	271	97	130	192	90	310

表 3-16 箱根エリア数値評価平均と評価表現分析表

数値評価	部屋	風呂	料理	接客	清潔感
平均点	4.1	4.3	4.3	4.0	4.1

感想評価	部屋	風呂	料理	接客	清潔感	立地
肯定表現	150	261	374	237	75	64
否定表現	39	32	46	22	6	7
中立・他	6	10	15	10	1	1
総計	195	303	435	269	82	72

表 3-17 湯布院エリア数値評価平均と評価表現分析表

数値評価	部屋	風呂	料理	接客	清潔感
平均点	4.4	4.5	4.7	4.6	4.5

感想評価	部屋	風呂	料理	接客	清潔感	立地
肯定表現	95	198	402	307	54	91
否定表現	17	35	15	9	6	4
中立・他	6	16	9	6	0	0
総計	118	249	426	322	60	95

表 3-18 草津エリア数値評価平均と評価表現分析表

数値評価	部屋	風呂	料理	接客	清潔感
平均点	3.9	4.2	4.2	4.1	4.1

感想評価	部屋	風呂	料理	接客	清潔感	立地
肯定表現	86	231	354	267	50	117
否定表現	47	66	41	20	4	3
中立・他	8	15	16	5	0	2
総計	141	312	411	292	54	122

表 3-19 登別エリア数値評価平均と評価表現分析表

数値評価	部屋	風呂	料理	接客	清潔感
平均点	3.9	4.3	4.2	4.2	4.1

感想評価	部屋	風呂	料理	接客	清潔感	立地
肯定表現	83	280	377	237	40	38
否定表現	40	66	59	13	6	2
中立・他	8	12	18	4	0	0
総計	131	358	454	254	46	40

表 3-20 道後エリア数値評価平均と評価表現分析表

数値評価	部屋	風呂	料理	接客	清潔感
平均点	4.2	4.2	4.4	4.3	4.3

感想評価	部屋	風呂	料理	接客	清潔感	立地
肯定表現	112	148	393	261	55	161
否定表現	21	34	23	10	7	3
中立・他	7	12	9	8	1	2
総計	140	194	425	279	63	166

これらより、エリア内の各地区においてそれぞれ関心の高い数値評価項目が異なる点が見られた。また、いずれの地区においても、立地に関する表現が多いことが特徴的であった。次節ではこれらの結果を利用者に提示する場合の活用方法について検討を行う。

3.6 活用方法の提案

本節では、本研究の結果を活用した宿泊レビューの活用方法を提案する。宿泊予約サイトの多くが提供する宿泊レビューは、宿泊施設の数値評価の平均値集計を行った表と、宿泊利用者から投稿された感想コメントをリスト化し、投稿された時系列に沿って表示するだけであった。それゆえ、数値評価項目にない内容や詳しい情報を調べたい場合は、感想コメントを読まなければならなかった。

本研究で提案した手法を用いれば、感想コメントに書き込まれた特徴表現を、数値評価項目ごとに肯定極性と否定極性に分類して抽出することが可能である。この情報を宿泊レビュー中にわかり易く追加表示すれば、宿泊検討者にとって宿泊施設を選択する際の利便性が向上するであろう。そこで、本研究成果を活用した宿泊レビュー画面デザインの一例を図 3-3 に示す。

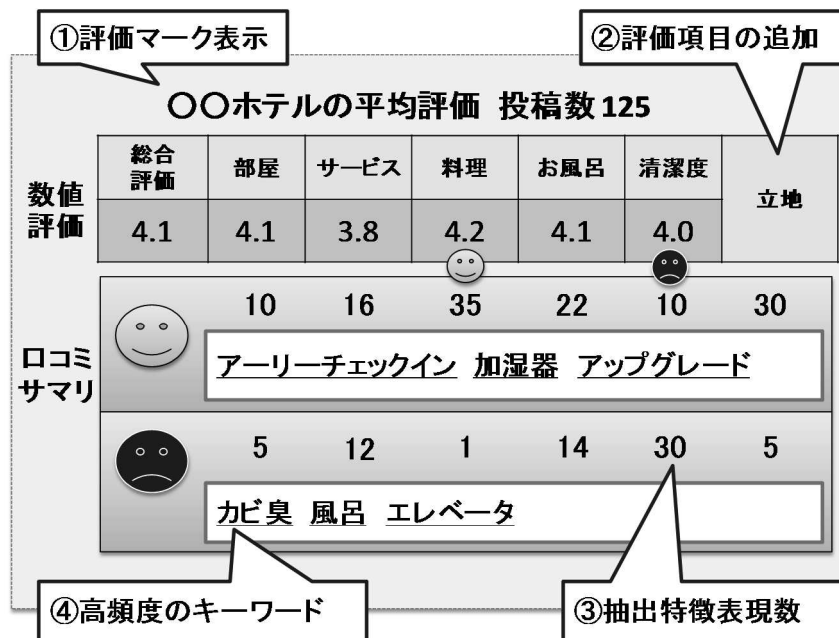


図 3-3. 宿泊レビュー画面デザインの一例

これまでの宿泊予約サイトが提供する宿泊レビューに追加する本研究からの提供情報は下記の 4 つである。

① 数値評価項目へ評価マークの表示

感想コメントから抽出した肯定極性と否定極性の書き込みの頻度、もしくは、比率により評価が良い項目と評価が悪い項目を判定し、当該項目に宿泊検討者が一目で判断できるアイコンを表示する。

② 書き込みの多い評価項目の追加表示

数値評価項目として定義されていない項目だが、感想コメント中に高頻度で出現する項目を表示する。

③ 抽出した特徴表現数の表示

感想コメントから抽出した特徴表現を、数値評価項目ごとに肯定極性と否定極性に分類して、その数をカウントして表示する。

④ 高頻度で出現した特徴表現<属性>の表示

感想コメントから抽出した特徴表現中の<属性>を、肯定極性と否定極性ごとに集計し、高頻度で出現した<属性>をキーワードとして表示する。さらには、表示キーワードにハイパーリンクを付加して、表示キーワードを含む感想コメントの原文を容易に表示できるようにする。

図 3-3 に例示した①～④の情報を新たに表示することで、数値評価だけでは宿泊検討者が把握できなかった宿泊施設の特徴が見えてくる。この例では、①の機能で評価マークが表示され直感的にこの宿泊施設が料理に評価が高く、清潔度の評価が低いことがわかるようになる。②の機能で立地を評価項目として追加することで、本来の数値評価項目にはないが、多くの感想コメントが存在し評価が高い項目であることがわかる。③の機能により、サービスの数値評価は 3.8 と最低点であるが、感想コメントでは良い評価としてアーリーチェックインやアップグレードなどがあり、運用上は良いサービスを提供していることがわかる。さらに、④の機能で部屋の数値評価は 4.1 であるが、感想コメントではカビ臭やエレベータなど施設全般に悪い評価があり、施設が古いもしくは整備が行き届いていないことがわかる。

ここに示す表示方法は一例であるが、宿泊検討者が当該宿泊施設の感想コメントの概要を容易に理解できる表示ができるようになれば、利用する宿泊施設を選択する作業が効率的になるであろう。

3.7 結言

本章では、宿泊レビューの数値評価と感想コメントを利用した評価情報の抽出と利用方法の提案について検討を行い、感想コメントから、テキストマイニングを用いて宿泊検討者にとって有益となる情報を抽出する方法を提案し、また提案方法を用いて感想コメントの分析を行った。

その分析結果を数値評価と比較すると、宿泊レビューの数値評価において同じ点数になっている項目でも、宿泊利用者の満足度に差異があることがわかった。また、数値評価にない項目でも宿泊利用者が重要視した項目があることがわかった。さらに感想コメントの内容を分析すると、同じ数値評価項目でも、ビジネスエリアとレジャーエリアの宿泊利用者では、重要視している事柄が異なることがわかった。また、同じ利用目的であっても、地区ごとに注目される数値評価の項目が異なること、駅名や観光地名などが多く出現することなどから、地域性が表れていることがわかった。

また、これらの研究成果の活用方法として、数値評価の項目へ評価マークの表示、書き込みの多い評価項目の追加表示、抽出した特徴表現数の表示、高頻度で出現した特徴表現〈属性〉の表示と、宿泊検討者にとって有益となる情報を宿泊レビューに掲載する手法を提案した。

第4章

宿泊レビューを用いた宿泊施設別の特 徴分析

本章では，訪日旅行者によって英語で書かれた宿泊レビューの感想コメントを用いて，訪日旅行者が宿泊エリアおよび宿泊施設に対しどのような項目に注目しているのかを抽出する方法を提案する．

4.1 訪日旅行者の宿泊予約

日本は国家キャンペーンとして2003年よりVISITJAPANを提唱し、海外から旅行目的で日本を訪問する訪日旅行者数を拡大するための広報活動を行っている[JTA 2013d]。これは日本国民一人あたりの国内旅行宿泊数の減少が続く中において、訪日旅行者の獲得は、観光産業の振興という観点からも重要な取り組み事項として認識されつつある。観光庁の観光立国推進基本計画によると、海外からの旅行者を日本に集めることは、観光立国を目指す我が国の観光振興にとって非常に重要な取り組みであると位置づけがなされ、計画の中では東日本大震災からの復興、国民経済の発展、国際相互理解の増進、国民生活の安定向上を大項目として掲げ、展開している[JTA 2012]。

訪日旅行者数は、諸外国の紛争や伝染病の蔓延など国際情勢や、東日本大震災の影響による大幅な減少もあったが、2012年以降は回復基調にあり、2013年12月には訪日外国人数が初めて1000万人を超えたとの発表がなされた[JTA 2013e]。一方で、日本の外国人旅行者入国者数を順位で見ると、世界で33位、アジアで8位となっており、諸外国と比較して日本の外国人訪問者数は少ない状況である[JTA 2013d]。さらに訪日旅行者の訪問を促すための誘致対策が必要とされている。

外国人に日本訪問を検討させるには、訪日旅行に対する積極的な情報発信および広報活動を行うと同時に、実際に日本を訪れた外国人がどのような感想を持ったのかを知り、それに合わせた観光サービスの改善や魅力の高いプロモーションを行うことが重要だと考えられる。日本政府観光局(JNTO)の調査によると、個人訪日旅行者の宿泊予約は66.4%の旅行者がインターネット宿泊予約サイトを参考にし、合わせてクチコミを22.6%の旅行者が参考にしている。また、実際の宿泊予約についてもインターネット宿泊予約サイトを経由したものが58%となっている[JNTO 2013]。インターネットは旅行の参考情報として、宿泊施設の予約手段として非常に利用の広がりを見せており、その際に宿泊検討者は、宿泊レビューの評価を宿泊施設選択の情報として重要視して予約を行っていることがわかる。

訪日旅行者が滞在中に利用した宿泊施設に対し、どのような要望や印象を持ったかを知るには、この急速に利用頻度の高まっているインターネット宿泊予約サイト宿泊レビューの感想コメントを利用することができる。宿泊予約サイトに訪日旅行者から投稿される宿泊レビューは、海外から訪れた旅行者が実際に宿泊施設に滞在し、その際に感じた生の意見が感想コメントとして掲載される。

この宿泊レビューを用いれば、訪日旅行者が日本で宿泊施設に泊まる際、何に注目したかを検証することができる。これは利用者にとっても重要な参考情報であるだけでなく、訪日旅行者の受け入れ側である宿泊施設にとっても外国人がどのような内容に注目しているか、今後の提供サービス時の強みとなる項目や、改善すべき項目を見つけるための重要な参考情報となるはずである。

また VISIT JAPAN キャンペーンにより訪日旅行者の来訪の増加が期待される中、一方では外国人の受け入れ時に不安を感じる宿泊施設も存在する。国際交流サービス協会の調査によると、その原因は言葉の問題であり、コミュニケーション上の問題から訪日旅行者の受け入れに消極的な宿泊施設が多い [ICHSA 2012]。その様な宿泊施設であっても、訪日旅行者の生の意見を聞き、宿泊施設を利用するにあたってどのような情報を求めているのかを知ること、訪日旅行者の受け入れのための障壁を取り外し、受け入れに向けた意識の変革を促すことが期待できると考えられる。

このような背景より本章では、訪日旅行者によって英語で書かれた宿泊レビューの感想コメントを用いて、訪日旅行者がどのような項目に注目しているのかの抽出する方法を提案する。そして提案手法を用いて宿泊施設の特徴を分析する。

4.2 訪日旅行者の宿泊施設評価

旅行者が、旅行を検討する際にインターネットの役割は非常に大きい。旅行の行先を検討し、旅行の行程を決め、旅先の情報を集めるなどあらゆる面で活用されている。その中でも実際に旅をした旅行者が発信するクチコミ情報は、生の情報であり、ガイドブックなどと比べてリアルタイムな現地情報が手に入るので人気である。特にその中でも宿泊施設を選択し、決定する際に宿泊レビューは重要な役割を果たしている[Zhang 2009]。そして、宿泊レビューの重要性は旅行を検討している旅行者のみならず、それを活用することがホテルや旅行会社にとっても重要な情報となっている[Gretzel 2008]。その際、多くの訪日旅行者から投稿される膨大な数の宿泊レビューを網羅的に閲覧することは困難である。そのため、その中から特徴となる項目を抽出するための研究がおこなわれており、田辺らは、宿泊レビューのテキストマイニングにより旅行者の注目情報を抽出する方法を提案しており[Tanabe 2008]、辻井らは日本語で書かれた宿泊レビューを用いて、利用者の感想コメントのテキストマイニングを行い、利用者が宿泊するエリアごとに注目される要素が違うことを明らかにした。また、感想コメントの特徴表現を宿泊予約サイトが持つ宿泊レビューの数値評価項目に分類し、その表現を肯定および否定の属性を加えて分類を行った。これにより、利用者は感想コメントにおいて数値評価の高低とは関係なく、注目する評価項目が存在することを示した[Tsujii 2012]。これらの研究では、宿泊エリアごとの評価については検証が行われているが個別の宿泊施設での評価の違いや用いられる特徴表現の違いが示されていない。そのため、さらに具体的に宿泊施設ごとの特徴の違いを明確にすることが必要である。また、海外での宿泊レビューの評価情報を研究した事例では Kasper らは、トリップアドバイザー¹⁰のユーザレビューを利用し、宿泊施設の評価を肯定、否定、中立に分類し、それらを集約するシステムの構築の提案を行った[Kasper 2011]。また Wu らは、宿泊レビューを用いて、評価の可視化を試みている。意見の輪とタ

¹⁰ <http://www.tripadvisor.com>

クラウドを用いて宿泊利用者の意見を表示させることで、利用者にわかりやすく情報を提示する試みを行った[Wu 2010].

これらの研究では、利用者は宿泊レビューを宿泊予約時の検討材料としており、情報の重要性を指摘し、その膨大な数にのぼる宿泊レビューの活用方法に対しての提案を行っている。しかし言葉の違いの存在する訪日旅行者が日本に訪れた際に感じる注目点の違いや、感想コメントと数値評価を相互に用いた情報提示方法についての利用可能性について言及されていない。

本研究では、これらの先行研究を参考とし、訪日旅行者によって英語で書かれた日本国内の宿泊施設への宿泊レビューを用いて、旅行者がどのような項目に対して注目しているのかを抽出する方法を提案する。具体的には、訪日旅行者から寄せられる宿泊レビューの感想コメントを用いてテキストマイニングによる宿泊施設に対する評価表現を抽出する。そして訪日旅行者が持つ感想の特徴を検証するとともに、宿泊レビューのもう一方の評価指標である数値評価項目を結合させ、訪日旅行者が投稿する宿泊レビューにおける特徴的な評価項目を検証する。

4.3 訪日旅行者の宿泊レビュー

インターネット上には旅行に関する情報があふれている。その中で、宿泊レビューは宿泊施設の特徴を知る重要な方法である。宿泊予約サイトに提供される宿泊レビューは、実際に宿泊した利用者による感想を書かせる方法で提供されている。図 4-1 に示すような投稿テンプレートとなっており、投稿者のプロフィール、感想コメント情報、数値評価項目評価の 3 要素から構成される。これは、ほとんどの宿泊予約サイトにおいても同様である。宿泊施設ごとにまとめた宿泊レビューは、該当の宿泊施設への予約する際の参考評価指標として、訪日旅行者に提供されている。宿泊レビューの要素の 1 つである数値評価項目は、宿泊予約サイトが宿泊施設に対する主な評価項目を決め、宿泊利用者がそれらの項目に対して数値での採点を行う形式である。

The screenshot displays a section titled "Recent customer comments" with a filter by traveler type and a summary of reviews. The filter shows "All reviews" selected with 1594 total reviews. The summary shows a "Fantastic" rating of 8.3 from 1594 reviews. A table lists various evaluation items with their respective scores: Value for Money (8.2), Location (9.1), Staff Performance (8.2), Hotel Condition/Cleanliness (8.6), Room Comfort/Standard (8.4), and Food/Dining (7.2). Below this, three individual reviews are shown, each with a title, traveler information, a short comment, and a rating.

Filter by Traveler Type	Count
All reviews	1594
Solo travelers	250
Couples	831
Business travelers	75
Families with young children	100
Families with older children	132
Groups	205

Item	Score
Value for Money	8.2
Location	9.1
Staff Performance	8.2
Hotel Condition/Cleanliness	8.6
Room Comfort/Standard	8.4
Food/Dining	7.2

Traveler Name	Traveler Type	Review Title	Rating
Anh B.	Families with young children	Enjoyable stay in this beautiful city	7.7
Kritakorn K.	Couples	Good hotel, Not expensive, Nice staff	8.7
Kevin P.	Families with young children	Great location for a family with small children	7

図 4-1 海外宿泊予約サイトの宿泊レビュー画面

次に、数値評価項目にはどのような要素があるのかを調べることにする。表 4-1 に海外の主要な宿泊予約サイトの数値評価項目をまとめたものを示す。

宿泊予約サイトが提供する数値評価項目は、表記方法や順序に差異はあるが、その内容を大まかに分類すると「施設」、「部屋」、「接客」、「立地」、「価格」、および一部のサイトで「食事」の 6 つで構成されている。また、3.3.1 項の表 3-1 で示したように、日本の宿泊予約サイトにおいてはこれらの評価項目に加えて「風呂」が追加される。これは日本人の旅行者の風呂に対する注目度の高さを反映しているからといえる。

表 4-1 海外の主要な宿泊予約サイトの数値評価項目

	Booking	Agoda	HotelClub	Expedia
施設	清潔さ	ホテルの状態/清潔さ Hotel Condition/Cleanliness:	設備	清潔感
部屋	快適さ	お部屋の快適さ Room Comfort/Standard:	客室の快適さ	客室
接客	サービス			サービス
	スタッフ	スタッフの対応 Staff Performance:	ホテル・スタッフ	
立地	ロケーション	ロケーション Location:	ロケーション	
価格	お得感	価値 Value for Money:	コストパフォーマンス	価格
食事		お食事 Food/Dining:		

次に宿泊レビューの感想コメントは、宿泊利用者が宿泊施設への滞在時に体験した内容を感想として記入できる自由記述欄である。感想コメントは宿泊利用者にとって数値評価の記入より手間のかかる作業であり、そこには数値評価では表現できない宿泊施設の特徴や、実際に利用した生の感想が書かれている。人気ホテルでは、多くの宿泊レビューが投稿され、利用者が体験した様々な感想が書かれている。しかし、これらの感想コメントを通読し、その施設の何が

良くて、何が悪いかを理解するのは非常に困難である。また、数値評価項目は3章でも示したようにすべての項目に評価を行うため、どの数値評価項目が宿泊利用者にとって重要であったのかを知ることが難しい。

そこで本研究では、感想コメントから宿泊施設が持つ特徴をテキストマイニングの手法を用いて抽出し、そして数値評価項目別に分類することを試みる。そして感想コメントの特徴と数値評価項目を融合させることで、どの項目が重要であったかを分析し、利用者にとって分かりやすい提示方法を提案することを目的としている。次節では本研究で提案する感想コメントより評価表現を抽出する手段であるテキストマイニングの手順を提示する。

4.4 テキストマイニング

本節では感想コメントのなかに出現する宿泊施設を評価する特徴表現を、テキストマイニングにより分析するために必要な手順について説明する。

本研究の流れを図 4-4 に示す。まず対象となる宿泊レビューから感想コメントデータの取得を行い、感想コメントを形態素解析により単語別の頻度を集計し、投稿者が注目する特徴となる単語の傾向を把握する。次に、宿泊施設がどのような評価を受けているのかを知るため、利用者が注目する特徴語と、その記載内容を係り受け解析により特徴表現として分析する。

そしてそれらを京都エリア全体で分析を行い、さらにエリア内の宿泊施設を個別に見た評価の分析を行う。

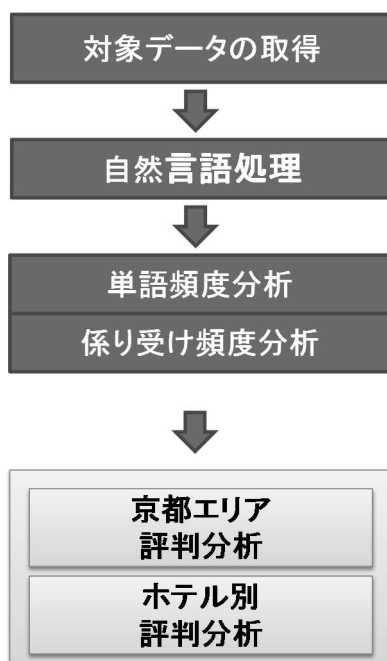


図 4-4 テキストマイニングによる分析フロー

4.4.1 類義語辞書

宿泊レビューにおける文章は、投稿者それぞれに同一内容の事柄を示す場合であってもその表現方法にばらつきが存在する。そこで同一評価対象物と想定される同様の属性を持つ項目ごとに、出現数の多い単語のグルーピングを行う。これにより同一内容を表す単語の揺らぎをまとめる類義語辞書を構成した。類義語辞書は、テキストマイニングを実施し分析結果の検証を行いながら、必要に応じて修正することで分類の精度を向上させた。類義語辞書の一例を表 4-4 に示す。

表 4-4 類義語辞書の一例

Synonym	impression			
bath	bathroom	bathtub	bath/spa	shower
staff	concierge	porter	hotel clerk	bell boy
bed	bedding	mattress		
internet	wifi	wi-fi	lan	

4.4.2 数値評価対応辞書

宿泊レビューにおける利用者の施設への評価は、感想コメントと数値評価から成り立っている。そこで感想コメントの中にあられる評価内容を数値評価に対応させることを検討する。これにより感想コメントから注目される数値評価項目を見つける。感想コメントに現れる単語およびその類義語で構成される特徴語が持つ属性がどの数値評価項目に対応するかを判断するために、数値評価対応辞書を作成した。図 4-5 にその概念を示す。今回の分析においては、4.3 節で示したホテル予約サイトの数値評価項目として頻繁に使われる「部屋」「ホテル」「料理」「接客」「価格」「立地」の 6 種類の評価項目を用いて、感想コメントの特徴語がいずれに分類されるかを判断した。これらの処理により、代表的

な特徴を表す数値評価項目と、その具体的な内容としての特徴語としての単語および類義語といった階層的な概念を持たせた。また分類の際、これらの数値評価項目に当てはまらない単語・類義語については、辞書登録を行わなかった。

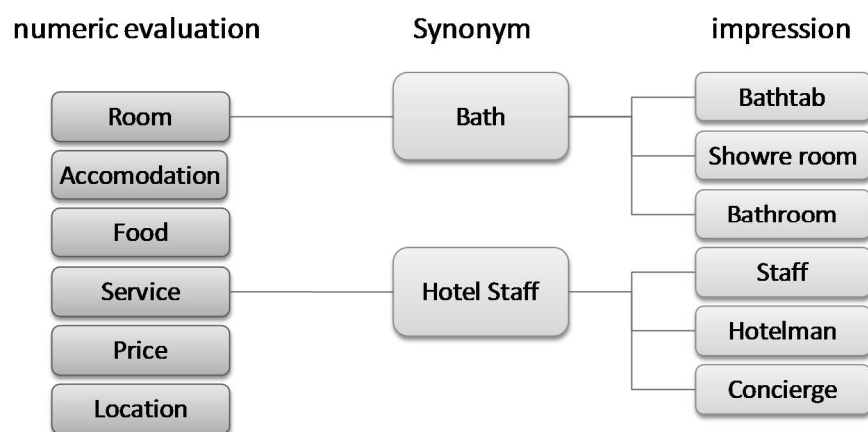


図 4-5 数値評価対応辞書

4.4.3 評価属性辞書

次に、感想コメントから評価表現を抽出するため、係り受け表現に注目する。特徴語と評価語との係り受け表現を、肯定および否定に評価極性をラベリングする評価属性辞書を構築した。その一例を表 4-5 に示す。

この評価属性辞書と数値評価対応辞書を用いることで、利用者の感想コメントを自然言語処理の技術を用いて数値評価項目と対応させ、肯定・否定いずれの評価を行っているか分類できるようにした。

その過程において評価対象項目が不明確なもの、対象項目に対する中立表現、要望表現など評価極性を持たないものは除外することにした。

本節では、英語でのテキストマイニングを用いた評価表現の抽出方法について解説を行い、その際に利用する辞書の内容について解説した。次節では京都に宿泊した訪日旅行者によって英語で書かれた感想コメントを用いて実験を行った結果およびその内容を検証する。

表 4-5 評価属性辞書

positive	negative
location-great	room-small
location-good	size-small
staff-helpful	hotel-old
room-clean	hotel-small
staff-friendly	hotel-bad
hotel-nice	bath-small
location-convenient	price-expensive
hotel-great	value-bad
hotel-good	hotel-expensive
service-good	room-overpriced

4.5 感想コメント分析

本節では、テキストマイニングを用いた、宿泊レビューの感想コメントの特徴表現を分析する。感想コメントは、宿泊利用者が実際に当該宿泊施設を利用して感じたことを記述したものである。数値評価と異なり、感想コメントの入力は労力が大きい。それにもかかわらず記入した項目なので、感想コメントは宿泊利用者が強い印象を持った事項に対する、率直な意見であると考えられる。

本研究では、訪日観光客に人気の高い日本の伝統都市である京都エリアを対象とした。宿泊予約サイトに英語で書かれた京都で人気の高い 10 軒の宿泊施設を選び、それぞれの施設の 100 件のデータを集め合計 1,000 件の宿泊レビューを利用した。本研究では、感想コメントの分析器として enju を利用した [Tsujii 2011]。利用した宿泊レビューの構成を表 4-6 に示す。

表 4-6 対象宿泊レビュー数

Hotels	10
comments	100
Total comments	1,000
Total Words	75,172
Avarage Words	75

最初に、テキストマイニングを行い、前節で示した数値評価対応辞書による分類方法を用いて、海外宿泊予約サイトで代表的な数値評価項目である「部屋」「ホテル」「料理」「接客サービス」「価格」「立地」との関連付けを行う。次に評価属性辞書を用いてそれぞれの項目に対する肯定評価か否定評価かを判断する。

この手法を用いて京都エリア全体の評価表現の分析を行い、エリアで注目される数値評価項目を分析する。次に京都エリアのホテルそれぞれに同様の手法を用いて、どの数値評価項目に対する評価表現が多いのかを検証していく。

4.5.1 エリア全体の特徴

本項では京都エリア全体の感想レビューをまとめ、エリアとしての分析を行い、その特徴を検証する。本研究の対象となる京都エリアの全宿泊レビューデータをテキストマイニングし、出現した評価表現の頻度を集計した。その結果を表 4-7 に示す。

表 4.7 京都エリア全体における評価表現の頻度

evaluation expression	frequency
location-great	210
location-good	176
room-small	92
staff-helpful	92
room-clean	91
staff-friendly	73
hotel-nice	68
location-convenient	47
hotel-great	41
hotel-good	39
service-good	38
room-good	33
room-nice	32
room-comfortable	32
internet-free	31
room-big	31
hotel-clean	29
room-spacious	29
location-nice	28
Total	3,410

この結果、全体として、評価極性が肯定的な評価表現が上位を占めており、否定的な評価表現は多くないことがわかる。“Location-great”や“Location-good”といった立地に関する評価が多く見られ、訪日旅行者は立地条件に対する感想を書く傾向が強いことがわかった。次に、部屋とスタッフの

サービスへの評価も多く見られた。だが一方では、部屋が狭い“room-small”といった否定的な評価表現が上位にみられ、部屋は利用者にとって評価の分かれる項目であることを示している。

次に、どのような数値評価項目への評価が多く集まるか、また肯定表現と否定表現がどのような割合で存在するのかを調べるため、京都エリア全体の評価表現を分類したものを表 4-8 に示す。

表 4-8 数値評価項目別評価

	Location	Hotel	Value	Food	Room	Service	Total
Positive	897	413	38	109	529	305	2291
Negative	27	33	3	11	132	8	214
Total	924	446	41	120	661	313	2505

この表は京都エリアの感想コメントの評価表現を、数値評価対応辞書を用いて「部屋」「ホテル」「料理」「接客サービス」「価格」「立地」の数値評価項目にそれぞれ分類し、その評価を、評価属性辞書を用いて肯定評価と否定評価に分類したものである。なお表中の数字は出現した評価表現を頻度で示したものである。

この表からは、利用者は数値評価項目の中でも「立地」に対する評価表現が最も多く、続いて「部屋」、「ホテル」、「接客サービス」の順に評価表現が多く集まることがわかる。

結果からは特に、「立地」に対する評価表現が多いことがわかった。地理に不案内な異国の地での滞在時に、最も利用者が重要視する項目は、立地条件であることを示している。一方で、訪日旅行者の感想コメントには「料金」や「料理」に対する評価は、あまり記載されていないことがわかる。訪日旅行者はこれらの数値評価項目に強い印象を受けることもなく、注意を払っていないといえるだろう。これらのことから、訪日旅行者が感想コメントを書く際に、印象に残り、注目していた数値評価項目の順位であるといえ、数値評価項目は一律に評価されているのではないことがわかる。

次項では、今回対象とした京都エリアの宿泊施設を個別に検証した場合の特徴の違いと、それぞれの具体的な評価の対象となった宿泊利用者に注目される

数値評価項目について検証を行う。

4.5.2 個別ホテルの特徴比較

本節では、京都エリアの感想コメントの全データを使い、10軒の宿泊施設それぞれの評価表現を数値評価項目ごとに分類し、評価の特徴傾向を分析する。

表 4-8 で示した結果に基づき、評価表現の出現数が多かった、「立地」、「部屋」、「ホテル」、「サービス」の数値評価項目に対する評価表現を見てみよう。結果を表 4-9 に示す。

まず「立地」の数値評価項目について見てみると、エリア全体での分析の際と同様に、どのホテルにおいても高い頻度で「立地」に対する評価表現が書かれている。その中でも宿泊施設 F および G は立地条件に対する評価表現が多く現れることがわかった。これらの施設を利用した宿泊者は、特に該当施設の立地条件に対して高く評価されているといえる。

評価表現としては”location-good”, ”location-great”といった表現が多く見られた。他に立地条件に関する評価表現としては、最寄駅からの距離に関する表現が多くみられた。また、繁華街や有名観光地への距離や周辺のショッピングモールへの利便性についての評価も書かれており、宿泊利用者にとって立地条件にも数々の意味があることがわかる。

次にエリア全体での分析において 2 番目に評価表現の出現頻度が高かった「部屋」に対する評価内容の検証を行う。「部屋」に関する評価表現は、肯定的な評価では、具体的な表現として”room-nice”, “room-clean”といった部屋の状態に対する表現が多くみられた。一方で、室内に設置されたベッドやテレビといった部屋内設備や備品への評価表現はあまり見られなかった。

ただし一方で、この「部屋」の数値評価項目は、否定表現が多く見られる。特に宿泊施設 F では多くの否定表現が書き込まれており、内容は”room-small”といった表現であった。実際にこの宿泊施設の部屋はコンパクトに設計されており、利用者が宿泊した部屋の広さに不満がある感想を表現している。このような評価表現を読むことで、宿泊検討者は部屋の広さを確認することを意識する

ようになるだろう。

続いて、「ホテル」の数値評価項目についての評価を見てみると、ホテル設備およびホテル自身への評価に対する評価表現が多くみられた。実際の表現としては”hotel-nice”, “hotel-great”といったホテル自体の印象についての表現がほとんどであった。一部、ホテルの公共スペースであるロビーやエレベータ, ランドリールームなどのホテル付帯設備に関する評価が一部見られたが、それらを対象とした評価表現は少数であった。

最後に、「サービス」の数値評価項目について見てみると、ほぼホテルスタッフの対応に対する評価が上位にあらわれた。具体的な表現としては”Staff - helpful”, “Staff - friendly”といった、ホテルスタッフの質の高さや親近感に対して、利用者は評価を行っていることがわかった。さらに、詳細に調べていくと、ホテルスタッフは英語がほとんど使えないが、よいサービスを提供してくれたという感想が存在していた。この点については日本の宿泊施設はスタッフの外国語対応力を改善する必要がある、言葉の壁を超えることでさらに高い評価を得られることを示している。

表 4-9 ホテルごとの数値評価項目別評価

	Location	Hotel	Value	Food	Room	Service	Total
A	59	49	5	9	54	37	213
Positive	58	43	5	6	48	37	197
Negative	1	6		3	6		16
B	81	39	4	10	80	26	240
Positive	77	34	4	9	74	26	224
Negative	4	5		1	6		16
C	98	47	6	11	62	34	258
Positive	97	44	6	10	50	33	240
Negative	1	3		1	12	1	18
D	103	26	1	16	43	35	224
Positive	99	22	1	15	40	34	211
Negative	4	4		1	3	1	13
E	94	41	4	16	74	21	250
Positive	92	39	4	15	59	20	229
Negative	2	2		1	15	1	21
F	121	32	4	4	100	37	298
Positive	117	28	3	3	63	35	249
Negative	4	4	1	1	37	2	49
G	111	43	9	12	64	30	269
Positive	105	40	9	12	47	29	242
Negative	6	3			17	1	27
H	72	59	3	19	57	36	246
Positive	72	58	3	17	48	35	233
Negative		1		2	9	1	13
I	103	76	2	17	73	28	299
Positive	100	74	2	16	59	27	278
Negative	3	2		1	14	1	21
J	82	34	3	6	54	29	208
Positive	80	31	1	6	41	29	188
Negative	2	3	2		13		20
Total	924	446	41	120	661	313	2505
Positive	897	413	38	109	529	305	2291
Negative	27	33	3	11	132	8	214

次に、京都全体の数値評価項目に対する肯定・否定表現の平均を計算し、それぞれのホテルの評価表現との差を比較していく。その結果を表 4-10 に示す。

各ホテルの数値評価項目別の肯定表現の頻度と平均との差は、その数値評価項目への関心の高さを示す。また各ホテルの数値評価項目別の否定表現の頻度と平均との差は、その数値評価項目への不満の強さを示す。たとえば、これらの宿泊施設の中で、高い評価を得ているのは、宿泊施設 I である。この宿泊施設はホテルと立地の数値評価項目で高い評価を得ており、否定表現の頻度が低く、数値評価項目全体として高い評価を得ている。これは立地条件だけで比較すると宿泊施設 A の 2 倍の肯定評価を得ており、総合的にみて肯定評価の頻度が少なかった宿泊施設 J と比較した場合、約 1.5 倍の肯定評価を得ていることになる。一方、宿泊施設 F および G は部屋の評価表現数における否定表現の出現比率が他の施設と比べて高かった。また、宿泊施設 F は肯定および否定評価の両方の評価を受けている。このホテルは、立地で肯定評価を得ており、一方で、部屋の数値評価項目では、否定評価を多く得ている。この宿泊施設 F は利用者を選ぶ宿泊施設であり、立地を重視し部屋にこだわらない利用者にとっては便利な宿泊施設であると考えられる。この評価を見た旅行者は、感想コメントの原文を読んだり、宿泊施設が提供する情報を参照したりすることで、自分にとって満足できる宿泊施設かどうかの適切な判断を行うことができるだろう。

次に、宿泊施設の運営者に対してエリア全体の評価傾向と自施設との比較を行うための指標を提供する方法を検討する。表 4-10 はエリア全体の評価表現の平均値を計算し、各宿泊施設の評価と評価表現の平均値の差をそれぞれ示したものである。宿泊施設の肯定・否定評価を相対的に参照させることで、エリア内での宿泊施設の特徴と、宿泊施設の運営者にとって長所とすべき項目と、改善すべき項目を明らかにすることが出来ると想定している。宿泊施設 I は、肯定評価において平均を上回っており、否定表現においては平均を下回っていることから、このエリアにおいては評判の高い宿泊施設と考えられる。一方、宿泊施設 F は肯定評価において平均を上回っているが、否定評価においても平均を上回っているため、改善を行うことでさらにエリア内での地位を築ける可能性がある。宿泊施設 J は肯定評価において平均を下回っており、否定評価においても平均を下回ることから、目立たない宿泊施設になってしまっている。

表 4-10 ホテルごとの数値評価項目別の相対評価

	Location	Hotel	Value	Food	Room	Service	Total
A							
Positive	-31.7	1.7	1.2	-4.9	-4.9	-30.5	-32.1
Negative	-1.7	2.7	-0.3	1.9	-7.2	-0.7	-5.4
B							
Positive	-12.7	-7.3	0.2	-1.9	21.1	-4.5	-5.1
Negative	1.3	1.7	-0.3	-0.1	-7.2	-0.7	-5.4
C							
Positive	7.3	2.7	2.2	-0.9	-2.9	2.5	10.9
Negative	-1.7	-0.3	-0.3	-0.1	-1.2	-0.6	-3.4
D							
Positive	9.3	-19.3	-2.8	4.1	-12.9	3.5	-18.1
Negative	1.3	0.7	-0.3	-0.1	-10.2	-0.7	-8.4
E							
Positive	2.3	-2.3	0.2	4.1	6.1	-10.5	-0.1
Negative	-0.7	-1.3	-0.3	-0.1	1.8	-0.6	-0.4
F							
Positive	27.3	-13.3	-0.8	-7.9	10.1	4.5	19.9
Negative	1.3	0.7	0.7	-0.1	23.8	1.2	27.6
G							
Positive	15.3	-1.3	5.2	1.1	-5.9	-1.5	12.9
Negative	3.3	-0.3	-0.3	-1.1	3.8	0.2	5.6
H							
Positive	-17.7	16.7	-0.8	6.1	-4.9	4.5	3.9
Negative	-2.7	-2.3	-0.3	0.9	-4.2	0.2	-8.4
I							
Positive	10.3	32.7	-1.8	5.1	6.1	-3.5	48.9
Negative	0.3	-1.3	-0.3	-0.1	0.8	0.2	-0.4
J							
Positive	-9.7	-10.3	-2.8	-4.9	-11.9	-1.5	-41.1
Negative	-0.7	-0.3	1.7	-1.1	-0.2	-0.8	-1.4

本節では、我々は感想コメントを用いて、テキストマイニングによる宿泊施設分析を実施した。訪日旅行者の感想コメントから特徴表現を集計し、感想コメントの特徴表現と数値評価項目との結合を行い、どのような評価が書かれているかの分析を行った。

そして、われわれは、独立した数値評価と感想コメントの2つの評価項目を相互に補完できる方法を提案した。宿泊利用者は、数値評価項目のすべてに注意深く評価しておらず、特定の数値評価項目に注目し、感想コメントを書いていることが分かった。そしてテキストマイニングにより特徴表現を抽出し、それらを肯定評価と否定評価に分類することで、数値評価項目ごとの重要性を判断できることを示した。

表 4-9 に示す手法を用いて、宿泊検討者に対し、各宿泊施設の特徴を提示することで、どの数値評価項目が利用者に注目されているかを知らせることができる。我々は、利用者の旅行目的や興味に合わせた選択肢を選ぶための指標を提供することができるだろう。また、利用者や宿泊施設が、数値的な情報を把握したあと、より具体的な評価内容が知りたければ、感想コメントを参照すれば良い。数値評価項目が示す特徴表現を利用者の感想コメントの中から詳しく知ることができる。

表 4-10 に示す手法によりエリア全体の数値評価平均と宿泊施設の数値評価項目に対する肯定評価、否定評価の出現頻度のギャップを比較できるようにすることで、エリア内での相対的な評価を行うことができる。宿泊施設の運営者にこれらの情報を提示することにより、強みとなる特徴を用いたプロモーションや、改善すべき項目を見つけることが出来ると考えられる。

また、日本特有の宿泊施設の形式として旅館がある。旅館は、基本として夕食と大浴場や露天風呂といった公衆浴場を楽しむために利用される場合が多い。訪日旅行者から得られた旅館への感想コメントの分析の結果、食事・風呂に対する評価が多いことがわかった。本来の意図すべき提供内容に対して満足を得られていることから、日本の文化的要素が、訪日旅行者にも受け入れられていることがわかる。

これらの方法を用いることで、訪日旅行者の注目する内容を理解することで、利用者および宿泊施設の運営者、これまで以上に有益な宿泊レビュー活用が促

進され、有益な情報提供が行うことが出来ると考えている。

4.6 結言

本章では、訪日旅行者によって英語で書かれた宿泊レビューから感想コメントのテキストマイニング分析を行い、注目する項目とその評価表現の抽出を行った。また、感想コメントの評価表現を数値評価項目に対応させることで、どの数値評価項目に対して利用者の関心が高いかを検証した。

今回、京都エリアでの分析を行い、数値評価項目には「立地」、「ホテル」、「部屋」の順に宿泊検討者の評価が高いことを示した。またエリア内のホテルごとにも評価される数値評価項目を分析し、それぞれの施設の特徴に違いがあることを示した。エリアの特徴だけでなく宿泊施設の特徴を抽出する際にも、また、日本語で書かれた感想コメントだけでなく、英語で書かれた感想コメントにおいても本手法が有効であることを示した。

第5章

宿泊レビューの評価表現からのサービス改善情報の抽出方法

本章では、感想コメントの中で複数の評価表現を伴う文からの情報抽出についての手法を検討する。まず感想コメントの評価表現の遷移について述べ、次に評価表現の遷移を抽出する手法を示す。そして本手法からのサービス改善情報の抽出についての結果を述べる。

5.1 感想コメントの宿泊施設の評価情報

前章までの研究では、宿泊レビューの感想コメントから、肯定評価と否定評価の特徴表現を抽出する方法を検討し、宿泊レビュー利用者に向けた提示方法についての提案をおこなった。それは感想コメントを形態素解析および係り受け解析し、評価表現の抽出を行い、それらの表現を肯定と否定の極性に分類し、数値評価項目と対応させる事により、エリアと宿泊施設の特徴を表現するという手法であった。この手法を用いた分析の結果、感想コメントから得られる特徴を表す評価表現は、その宿泊エリアや施設における宿泊利用者からの関心の強い数値評価項目に集中することを示した。

感想コメントの文中に出現する評価表現は、宿泊利用者が体験したさまざまな対象物を評価しており、なかには宿泊サービスを改善につながる内容が含まれている。感想コメントの文中に現れる評価表現は、肯定評価、否定評価に分類されるが、文中に肯定評価と否定評価が同時に書かれることがある。「部屋は狭かったけど、料金が安かったです」といったように、否定的な評価表現が書かれたのちに肯定的な評価表現も書かれている場合、それらの表現は、部分的な感想と全体としての結論とが異なる状態となっている。この場合、解釈として料金が安かったことは、部屋が狭かったことを補い全体としては満足していると考えられる。このような表現を、評価極性が遷移する表現と呼ぶ。本章では、宿泊エリア単位での評価極性の遷移する表現に特徴がみられるかを検証する。宿泊レビューは、サービス改善するために必要な項目を提供する有益な情報であり、評価が文中で遷移する表現から肯定と否定評価部分を抽出し、宿泊施設のサービス改善につながる特徴項目を発見し、その情報から宿泊施設が認識しておくべき長所となる項目と、改善すべき項目を明らかにすることができる可能性を示す。

5.2 サービス改善のための評価表現

本節では、宿泊利用者が投稿する宿泊レビューの感想コメントのテキスト情報を利用し、評価極性が遷移する表現の内容とその抽出方法を検討する。肯定評価から否定評価にまたは、否定評価から肯定評価に評価表現が遷移する文章を感想コメントから抽出し、サービス改善の対象となる情報を抽出する。

5.2.1 評価極性が遷移する表現

宿泊利用者が記載する感想コメントには様々な内容が記載され、なかには肯定、否定の評価表現を伴う文が出現する。それらの評価表現は多くの場合、文中に1回、示されるが、下記のように文中の前後で評価表現の極性が異なる場合がある。

①部屋が狭かったけど、料金が安かったので満足です。

②料理はおいしかった、けれども、量が多すぎる。

①の場合は、前半部分に否定評価が書かれ、後半部分に肯定評価が書かれている。部屋が狭かったことは不満であるが、料金が安かったことへの満足のほうが強いため、全体としては満足であった内容と考えられる。

一方、②の場合は、前半が肯定評価であり、後半に否定評価が書かれている。料理はおいしかったにもかかわらず、量が多すぎたために不満が生じてしまったことを示している。

結論が最後に記載される傾向が強い日本語において、宿泊利用者は文中の後半の評価表現部分を高く評価したと想定される。後半の評価表現の重要度と比較して、前半の評価表現の重要度は高くはない。しかし、前半の評価表現は、後半の結論を導くための特徴的な評価項目である可能性がある。

また、次のように、評価表現の対象物と評価語との組合せが同一の評価対象である場合と、評価対象の組合せが異なる場合がある。

③部屋は狭いが、部屋の雰囲気は良かった。

④風呂は広くない、しかし部屋は広くて満足でした

③の場合、前半と後半の評価表現は、どちらも「部屋」を対象とした評価表現であり部屋の属性である広さが狭かったことへの不満と、部屋の属性である雰囲気が良かったことへの評価表現を前後に並べて比較し、評価している。この場合、宿泊利用者は「部屋」という評価対象を重視しており、その「部屋」の属性としての部屋の広さと部屋の雰囲気を比較し、評価したものと考えられる。

④の場合、評価の対象が前半部分と後半部分で異なっている。前半の評価は「風呂」の広さに対する否定評価であり、後半の評価は「部屋」の広さに対する肯定評価が記載されている。この場合は、宿泊利用者が「風呂」と「部屋」という評価対象を横断的に評価していると考えられる。

この様に、評価極性が遷移する表現には、その評価する対象が同一の評価対象であるものと、評価対象が異なるものが存在する。また、評価極性が肯定表現から否定表現へ遷移するものと否定表現から肯定表現に遷移するものとに分類することができる。そして、評価対象に紐付く属性を加味した場合、評価の組合せは多岐にわたる。

これらの評価表現の抽出には、テキストマイニングの手法を用いることが有効である。しかし膨大な数の投稿が寄せられる宿泊レビューの感想コメントから評価表現を抽出する際には辞書構築により類義語の統合を行い、評価表現の対象と属性の取りまとめを行うことでこれらの情報の分析がしやすい環境を作ることが必要になる。

5.2.2 評価の組合せの抽出方法

感想コメントから評価表現の遷移する表現を抽出するため、テキストマイニングを用いて評価対象と評価の組合せを抽出する。まず、形態素解析を行い、単語を分類する。分類の際、宿泊利用者による表現方法の違いは、類義語辞書を使いまとめる。次に係り受け解析を行い、係り受け表現の中から評価属性と評価の組合せを抽出する。その際、評価属性となる名詞と、評価を示す形容詞や形容動詞や動詞の係り受けを抽出する。そして係り受けの組合せに評価表現辞書を適用し、評価表現の対象と肯定、否定極性を判別する。

本研究では、肯定評価と否定評価が同時に記載される文を注目するため、宿泊利用者が投稿する感想コメントから肯定評価から否定評価に遷移する文と、否定評価から肯定評価に遷移する文を抽出する。これらの手法を用いて宿泊利用者の体験した出来事が書かれた感想コメントより評価表現を肯定表現、否定表現に分類する。前章までとは異なり、単純に評価の肯定評価と否定評価の頻度を用いるのではなく、文中に評価極性の遷移する表現のみを抽出し、その評価の頻度を用いる。これにより肯定評価と、否定評価の関係性と重要度が分かると想定する。その評価表現を集計し、その特徴を示すことで、宿泊施設に対し、サービスの改善項目の検討材料を提供することが出来ると想定している。

5.2.3 辞書の登録

感想コメントは、宿泊利用者によって自由に記述される文章であり、その文章の表現の揺れは大きい。そのため、類義語辞書を構築し、表現のゆらぎを抑える処理をする。そして形態素解析と係り受け解析ののち、評価表現を抽出し、肯定と否定極性を判定するために評価表現辞書を構築する。

これらの辞書は 3.4 節で提示した類義語辞書と評価表現辞書を用いることで対応し、評価の判定には評価の<対象>、評価対象の具体的な項目である<属性>および<評価>の 3 要素で構成されるものとする。<対象>は、数値評価項目とし

て多くに採用される「食事」、「風呂」、「サービス」、「清潔度」、「施設」、「部屋」、「立地」、「料金」の8種類に分類する。<属性>は“大浴場”や“ベッド”など具体的な評価物と対応させ、<評価>は実際の肯定評価語、否定評価語を分類した。これらの辞書を基に、感想コメントからの評価表現を抽出する。実際の評価の遷移する文例を図5-1に示す。

否定表現から肯定表現への遷移	
	<対象は同一>
	<u>お部屋は狭い</u> のですが、 <u>ベッドが快適</u> でした。 部屋：否定 部屋：肯定
	<u>温泉は大きくありません</u> が、 <u>お湯がとてもよかった</u> です。 風呂：否定 風呂：肯定
	<対象は別>
	<u>部屋は狭い</u> ですが、 <u>料金が安い</u> 。 部屋：否定 料金：肯定
	<u>建物は古い</u> のですが、 <u>最寄り駅に近く非常に便利</u> でした。 施設：否定 立地：肯定
肯定表現から否定表現への遷移	
	<対象は同一>
	<u>駅から近い</u> のですが、 <u>場所がちょっと分かりづらい</u> です。 立地：肯定 立地：否定
	<u>バイキングも肉系は美味しかった</u> けど、 <u>品数が少ない</u> 感じでした。 食事：肯定 食事：否定
	<対象は別>
	<u>部屋は広い</u> のですが、 <u>建物が古い</u> 。 部屋：肯定 施設：否定
	<u>朝食は美味しかった</u> のですが、 <u>少し料金が高い</u> です。 食事：肯定 料金：否定

図5-1 評価極性の遷移する表現

5.3 評価極性の遷移する表現の抽出

前節で検討した手法を用いて，評価が遷移する評価表現を抽出する．宿泊予約サイト"じゃらんネット"の宿泊レビューを利用し，実験を行った．対象としたデータの内容を表 5-1 に示す．

表 5-1 感想コメント投稿数

ビジネスエリア		レジャーエリア	
東京駅・日本橋	1,000	箱根	1,000
大阪キタ	1,000	由布院	1,000
札幌	1,000	草津	1,000
福岡(博多)	1,000	登別	1,000
名古屋	1,000	道後温泉	1,000
総計			10,000

次に，図 5-2 にテキストマイニングによる分析のフローを示す．

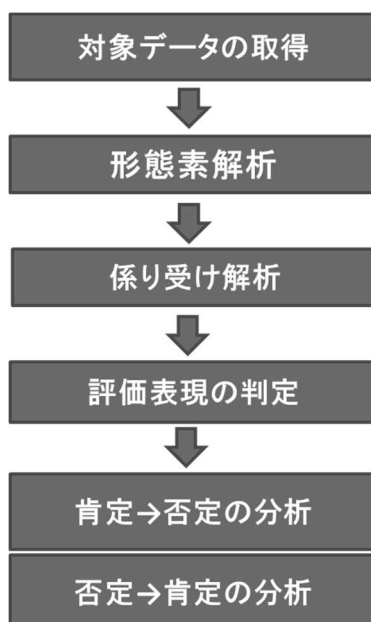


図 5-2 テキストマイニングによる分析フロー

最初に、宿泊レビューの感想コメントを、文単位で形態素解析を行い、次に係り受け解析を行う。そして評価表現辞書を用いて肯定表現、否定表現の判定を行い、対象ごとに分類を行う。その結果に基づいて、一文中の前半と後半において評価表現が1回ずつ出現する評価の遷移する文のみを抽出する。この抽出フローにより、数値評価項目に対応する評価表現は10,000件の感想コメントの中から14,347文の評価表現文を得られた。そしてその中から、肯定と否定がそれぞれ1回ずつ出現する文を選別したところ、肯定から否定に遷移する文を166文、また否定から肯定に遷移する文を104文得ることができ、合計で270文を得た。

次節ではこれらの文の利用し、その特徴の分析を行う。

5.4 評価の遷移する数値評価項目の抽出

本節では、数値評価項目と感想コメントの評価表現の極性との関連を検討する。数値評価項目は、宿泊レビューの評価項目の1つであり、宿泊利用者に関心の高い代表的な評価の対象への評価を数値で行うためのものである。数値評価項目は、宿泊検討者にとって重要な判断材料であるが、宿泊利用者が宿泊レビュー評価を行う際、関心が高い項目であるかないかにかかわらず数値評価項目を評価しなければならず、3.3節で示したように数値評価は集計すると偏りが生じてしまう。そのため宿泊検討者があとから閲覧する際、どの評価項目が重要であったかを判断するのが難しい。数値評価は、感想コメントとは独立した評価基準であるが、数値評価項目と感想コメントから抽出した評価表現の評価対象を関連付けることで分析に利用する。

そのために前節で示した手法により宿泊予約サイトの宿泊レビューの感想コメントから評価極性が遷移する表現を抽出し、代表的な数値評価項目8項目と関連付けを行う。そして、肯定評価から否定評価、否定評価から肯定評価への評価の遷移する頻度を集計した。その得られた集計結果のうち、否定評価から肯定評価に評価が遷移した例を表5-3に示す。

分析の結果、否定評価から肯定評価に遷移する表現は、評価対象の「部屋」と「食事」に関する評価の遷移が多く、また前半と後半の評価対象が同一のものである場合が多くみられた。それらの原文を参照したところ、「部屋」が対象となった場合、具体的には部屋は狭いがベッドが広いや、部屋が狭いが寝心地が良かったなど、部屋の狭さに対する不満はあるものの、部屋での寝心地が良かったので満足であることが記載されていた。また「食事」を対象とする評価の場合は、提供される食事の品数に関する評価と、提供される食事の内容に対する評価に分かれていた。

表 5-3 否定評価から肯定評価への遷移例

前半否定	後半肯定	出現回数
部屋	部屋	18
食事	食事	18
部屋	清潔度	17
風呂	風呂	16
部屋	風呂	10
立地	立地	10
部屋	立地	6
部屋	料金	6
施設	サービス	4
施設	食事	4
サービス	施設	4
サービス	サービス	3
風呂	清潔度	3
立地	部屋	3
風呂	立地	3
立地	サービス	3
部屋	施設	3
風呂	施設	3
施設	部屋	3
施設	清潔度	3

次に肯定評価から否定評価に遷移した例を表 5-4 に示す。

表 5-4 肯定評価から否定評価への遷移例

前半肯定	後半否定	出現回数
食事	食事	20
風呂	風呂	12
部屋	部屋	7
立地	立地	7
風呂	清潔度	4
立地	部屋	3
部屋	立地	3
部屋	施設	3
部屋	食事	3
施設	部屋	3
立地	風呂	2
部屋	清潔度	2
サービス	部屋	2
施設	清潔度	2
サービス	清潔度	2
立地	料金	2

分析の結果、評価の遷移は評価対象が「食事」から「食事」へ、「風呂」から「風呂」へなど、表 5-3 と同様に同一評価対象で評価の遷移が多く起きることが分かった。食事に関する評価の場合は、原文を参照してみると食事内容自体は良いものの、食事メニューの品数が少ないことに不満があるものが多かった。また、風呂に関する評価の場合、風呂の内容自体には満足しつつも、露天風呂が狭かったことへの不満や、そもそも露天風呂がなかったことへの不満が見られた。そのほかにも他の利用客のマナーの悪さや、シャンプーなどの風呂場に設置される備品の不備を指摘する評価も見られた。

次に、これらの評価の遷移結果をどの評価項目からどの評価項目への遷移になっているかを理解するために、否定評価から肯定評価への遷移をマトリクス表にしたものを表 5-5 に示す。

表 5-5 否定評価から肯定評価への遷移

	部屋+	食事+	風呂+	サービス+	清潔度+	施設+	立地+	料金+	計
部屋-	18	2	10	1	17	3	6	6	63
食事-	1	18	2	1	1				23
風呂-	1	1	16	1	3	3	3	1	29
サービス-	1			3		4			8
清潔度-					1	1			2
施設-	3	4		4	3		1		15
立地-	3			3	2		10	1	19
料金-	1		2	2		1		1	7
計	28	25	30	15	27	12	20	9	166

横軸は前半の数値評価項目を示しており、縦軸は後半の評価に対する数値評価項目を示している。そして表中の数値は、評価の遷移の頻度を示している。

表 5-5 の結果は、否定評価から肯定評価への評価の遷移は、同一の数値評価項目で多く見られ、前半が「部屋」、「風呂」、「食事」に関する否定表現から肯定表現への遷移が多くみられることを示している。これは、宿泊利用者は感想コメントを書く際、まず前半の「部屋」の評価対象に含まれる属性に何らかの不満を持つ。しかし、後半の評価対象に含まれる属性への評価により、全体としては満足したものと考えられる。「風呂」や「食事」の場合、評価の遷移は同

一の数値評価項目である「風呂」や「食事」への遷移が多く見られた。一方、前半で「部屋」の否定評価を行った場合をみると、宿泊利用者は、前半では「部屋」を否定評価したのち、後半で肯定評価する評価対象は「部屋」が最も多く、その次に「清潔度」、「風呂」の順であった。

原文を参照してみたところ、評価対象が「部屋」の場合の前半の否定評価は、部屋が狭いまたは小さいことに不満を示す表現が多く見られた。その部屋への不満に対し、後半の評価対象項目への肯定評価で満足している。この場合、前半の「部屋」の評価対象を改善するには、部屋を大きくする事が必要になるわけだが、それは不可能に近いものであろう。表 5-6 に後半の評価対象である数値評価項目が肯定評価であったものおよび、実際の評価表現内容を示す。

宿泊施設の担当者は、ここに示す後半に出現する数値評価項目に対する評価に注目し、それら項目の評価を高める努力をすれば、宿泊利用者の満足感を高められることができる。

表 5-6 肯定評価の原因となる評価表現

評価表現	頻度
清潔度	20
清潔(類)-良い	13
清潔(類)-満足	5
清潔(類)-快適	2
風呂	18
風呂(類)-良い	6
お湯-良い	5
風呂(類)-広い	3
温泉(類)-良い	2
風呂(類)-楽しむ	2
食事	16
食事和食(類)-美味しい	4
味-良い	3
食事(類)-美味しい	3
品数-多い	2
食事(類)-満足	2
食事(類)-良い	2
部屋	11
部屋寝具(類)-広い	3
寝心地-良い	2
部屋寝具(類)-大きい	2
部屋(類)-満足	2
部屋寝具(類)-快適	2
立地	9
眺め(類)-良い	4
最寄駅(類)-近い	3
立地(類)-良い	2
サービス	7
接客(類)-良い	3
接客(類)-嬉しい	2
接客(類)-満足	2
料金	5
価格(類)-安い	3
コストパフォーマンス(類)-高い	2
施設	3
宿泊施設(類)-多い	3

次に、肯定評価から否定評価に遷移する場合を分析する。

表 5-7 は肯定表現から否定表現への遷移を示したものである。

表 5-7 肯定評価から否定評価への遷移

	部屋-	食事-	風呂-	サービス-	清潔度-	施設-	立地-	料金-	総計
部屋+	7	3	1		2	3	3	1	20
食事+	1	20		1	1	1		1	25
風呂+	1	1	12	1	4				19
サービス+	2	1	1	1	2				7
清潔度+	1		1			1			3
施設+	3	1		2	2		1		9
立地+	3		2		1	1	7	2	16
料金+	1	1	1		1			1	5
総計	19	27	18	5	13	6	11	5	104

評価表現が否定評価から肯定評価に遷移する表 5-5 と比較して、評価表現の遷移の出現頻度は少ない。その中で部屋や風呂や料理についての評価の遷移が多くみられた。

このような評価の遷移が起こる場合、宿泊利用者は前半の数値評価項目に満足しているにもかかわらず、後半の評価項目が宿泊利用者を不満足にしたことになる。日本語における結論は後半にあることを考慮すると、表 5-7 における後半の不満足の要因は、宿泊サービス提供における重要な改善課題である。しかし、表 5-7 の結果は大分類を示す数値評価項目を集計したものであるため、具体的な改善項目が示されていない。そこで具体的な改善項目を調べるためには、評価対象物を検証する必要がある。評価対象を評価属性に分類し、頻度が高い否定評価の原因となる評価表現を表 5-8 に示す。「食事」に関しては品数についての課題が、「風呂」、「部屋」の場合には広さが課題となることが多いことがわかる。

表 5-8 否定評価の原因となる評価表現

評価表現	頻度
食事	16
品数-少い	5
夕食(類)-食べる+ない	3
味-薄い	2
メニュー(類)-少い	2
バイキング(類)-残念	2
朝食(類)-少い	2
風呂	10
特徴風呂(類)-狭い	5
風呂(類)-狭い	3
風呂(類)-不便	2
清潔度	7
清潔(類)-残念	7
部屋	6
部屋(類)-狭い	6
料金	2
価格(類)-狭い	2
施設	2
建物-古い	2

5.5 宿泊利用目的の違いによる特徴分析

本節では、宿泊利用者の利用目的による違いを検討する。宿泊施設をビジネスで利用する時と、レジャーで利用する時では、宿泊利用者の目的が異なる。そのため、宿泊利用者が注目する項目には違いがあると考えられる。そこで、表 5-5 及び表 5-7 の評価の遷移の結果を参照し、ビジネス利用とレジャー利用の宿泊施設に分類し、注目される項目の特徴の分析を行う。

ビジネスエリアにおける否定評価から肯定評価への遷移の集計結果を表 5-9 に示し、レジャーエリアにおける集計結果を表 5-10 に示す。

表 5-9 否定表現から肯定表現への遷移（ビジネスエリア）

	部屋+	食事+	風呂+	サービス+	清潔度+	施設+	立地+	料金+	総計
部屋-	12		6	1	14	1	4	6	44
食事-		3							3
風呂-	1	1	2		1	1			6
サービス-				1		1			2
清潔度-					1				1
施設-	2	1		3	2		1		9
立地-	3			1	1		7	1	13
料金-			1			1		1	3
総計	18	5	9	6	19	4	12	8	81

表 5-10 否定表現から肯定表現への遷移（レジャーエリア）

	部屋+	食事+	風呂+	サービス+	清潔度+	施設+	立地+	料金+	計
部屋-	6	2	4		3	2	2		19
食事-	1	15	2	1	1				20
風呂-			14	1	2	2	3	1	23
サービス-	1			2		3			6
清潔度-						1			1
施設-	1	3		1	1				6
立地-				2	1		3		6
料金-	1		1	2					4
総計	10	20	21	9	8	8	8	1	85

表 5-9 に示すように、ビジネス利用の宿泊施設では、評価表現の遷移が「部屋」に集中した。部屋の広さに不満があるが、その他の要素への肯定極性を持つ評価表現により満足に変わっている。ビジネスで宿泊施設を利用する際は、宿泊利用者は「部屋」に最も関心があることがわかる。そして「部屋」への不満を部屋自体の評価と、「清潔度」、「料金」とともに評価し、納得していることがわかる。

表 5-10 に示すように、レジャー利用の宿泊施設では評価対象となる数値評価項目は分散し、「風呂」と「料理」および部屋の評価対象において、評価表現の遷移が多く見られた。その際、同一の数値評価対象内の別な属性が評価されており、たとえば食事の場合であれば評価の遷移は食事の品数は少なかったが、味は美味しいといった評価が示されている。これらの傾向はビジネスエリアとレジャーエリアにおける宿泊利用者の注目する項目の違いを示している。

次に、ビジネスエリアにおける肯定評価から否定評価への遷移の集計と、レジャーエリアにおける評価の遷移をそれぞれ表 5-11 および表 5-12 に示す。

表 5-11 肯定表現から否定表現への遷移（ビジネスエリア）

	部屋-	食事-	風呂-	サービス-	清潔度-	施設-	立地-	料金-	総計
部屋+	5					3	2	1	11
食事+		3						1	4
風呂+			3	1					4
サービス+	2		1						3
清潔度+	1		1			1			3
施設+	3				2		1		6
立地+			2			1	5	2	10
料金+	1				1			1	3
総計	12	3	7	1	3	5	8	5	44

表 5-12 肯定表現から否定表現への遷移（レジャーエリア）

	部屋-	食事-	風呂-	サービス-	清潔度-	施設-	立地-	料金-	総計
部屋+	2	3	1		2		1		9
食事+	1	17		1	1	1			21
風呂+	1	1	9		4				15
サービス+		1		1	2				4
清潔度+									
施設+		1		2					3
立地+	3				1		2		6
料金+		1	1						2
総計	7	24	11	4	10	1	3		60

否定極性から肯定極性への評価表現の遷移は、その逆と比べて少ない。表 5-11 よりビジネスエリアの宿泊施設の場合、部屋と立地に関する評価の遷移が多いことが分かる。また、表 5-12 よりレジャー利用の場合には、食事に関する評価の遷移が多く、続いて風呂に関する評価の遷移が多い。この結果より、宿泊施設がサービス改善を実施する場合には、宿泊施設の利用目的に合わせて、対象とする項目を絞り、実施すべきである事が示される。

ビジネスエリア利用の場合とレジャーエリア利用の場合では、評価の遷移する対象として注目される数値評価項目が異なることがわかった。一方で、宿泊利用目的にかかわらず評価表現の遷移は数値評価項目が同一で、その数値評価項目の別の属性要素への評価の遷移が多く起こる傾向がみられた。この傾向は、肯定極性から否定極性への遷移の場合にも、否定極性から肯定極性への遷移のいずれの場合にもみられた。藤村の「変化の方向性」のサービス分類にはサービスには常態を向上させるものと状態を維持するためのものがあることを示しており [Fujimura 2009]、この尺度を用いると、レジャーエリアでは料理、風呂といった宿泊施設滞在中の「常態を向上」させるサービス内容に対し、より注目していることが分かる。反面、部屋での滞在といった、普段の「常態を維持」するための提供内容に対しての評価は低い。また、ビジネスエリアでは、出張先での普段の生活の「常態を維持」することが目的となるため、普段の生活を担保する項目である、部屋の快適さや利便性の良さといった項目に注目が

集まることが分かる。宿泊施設が様々な評価対象を多面的に改善するのは困難を要すると考えられるため、1つの評価対象の改善に注力すべきである。それにより効率的に宿泊利用者の評価を高められるはずである。そのため、宿泊施設がサービス改善を実施するにあたっては、後半で肯定評価されている評価対象である数値評価項目とその属性項目を強みとして生かしつつ、すべての数値評価項目を対象に実施するのではなく、宿泊利用者が特に強調したかったと思われる後半に否定評価された数値評価項目と同じ評価対象の属性となる項目の改善に注力すべきである。

5.6 結言

本章では、感想コメントのなかに現れる評価表現の遷移する文に注目し、宿泊エリアの評価遷移の傾向とサービスを改善するための情報抽出を目的に、感想コメントの分析を行った。具体的には宿泊利用者が投稿する宿泊レビューの感想コメントには文中に複数の評価表現が記載されることがあることに着目し、その評価の遷移傾向の分析を行なった。そして肯定評価と否定評価がそれぞれ1回ずつ出現し、評価極性が遷移する表現の抽出方法を提示した。

その結果、評価極性が遷移する表現は、同一の評価対象となる数値評価項目どうしで遷移する傾向があることがわかった。また、宿泊利用目的の異なるビジネスエリアとレジャーエリアに分類し、評価表現の遷移の分析を行ったところ、宿泊施設の利用目的によって評価対象が異なることを示した。後半の評価が肯定的な場合、それは宿泊施設の強みとなり、後半の評価が否定的な場合は改善すべき項目となる。

評価表現の遷移から不満の要因となる否定極性の評価表現とその不満を解消する満足の要因となる肯定極性の評価表現を抽出することで、宿泊施設の利用目的に合ったサービスの改善を効率的に実施するための評価対象を抽出できることを示した。

第6章

結論

本論文では、インターネット宿泊予約サイトの利用者から寄せられる宿泊レビューの感想コメント情報に着目し、テキストマイニングによる宿泊利用者が注目する評価表現の抽出を行った。そしてその評価表現にみられる宿泊用途の異なるビジネスエリアとレジャーエリアでの特徴の違い、宿泊施設単位での特徴の違いを分析し、宿泊レビューの数値評価と感想コメントを結合した情報掲載方法への提案を行った。

21世紀の産業として期待を集める旅行産業は、地域社会への経済的貢献も大きく、これを拡大するために国内外の利用者の増加が著しいインターネットでの旅行販売が期待されている。インターネット宿泊予約サイトは、宿泊検討者の希望条件に従い日本全国各地の宿泊施設を24時間いつでも検索し、予約することが可能である。宿泊施設の担当者や旅行会社の社員のアドバイスを受けことなく、宿泊検討者自身の判断によって宿泊施設を予約する必要があるが、その参考情報の1つに宿泊施設を利用した宿泊利用者が書き込む宿泊レビューが存在する。宿泊レビューには宿泊利用者の数値評価と感想コメントに書かれており、これらを宿泊検討者は予約時の判断材料に利用することが出来る。宿泊利用者が投稿する宿泊レビューは利用者目線の生の評価情報であり、この情報を活用する事はインターネット宿泊予約サイトにおいて必要不可欠な存在である。

しかし現在のインターネット宿泊予約サイトは、大量に寄せられる宿泊レビューを、宿泊利用者が採点した数値評価を集計した平均点と、個々の宿泊レビューを画面に羅列するだけになっており、宿泊検討者にとって使い勝手が良いとは言えない。そこで宿泊レビューから宿泊施設選択の参考となる評価情報を抽出し、提示する方法を改善する必要がある。

本論文の第3章では、国内宿泊予約サイトの宿泊レビュー情報を用いて、宿泊利用者の感想コメントからビジネス利用と、レジャー利用の場合でのエリアによる評価の特徴の違いと、宿泊検討者に向けた情報提示方法を提案した。感想コメントをテキストマイニングにより肯定、否定表現に分類し、数値評価項目への対応を行う事で、エリアによる特徴語、特徴表現の頻度分布の違いが存在し、利用者の数値評価項目への関心の強さに強弱が存在することを明らかにした。宿泊レビューを投稿する際、数値評価はあまり深く吟味することなく同じ数値で投稿してしまいがちだが、感想コメント情報を用いることで、ビジネス、レジャー各エリアの利用者が注目する数値評価項目の優先順位を明らかにした。この評価手法を用いて、宿泊検討者に対し、注目すべき数値評価項目を知らせることで、満足いく判断を行う手助けになると考えている。またビジネスエリアでの宿泊においては、立地条件が一番重要な項目であり、また部屋の評価では、部屋の広さが肯定、否定の評価にかかわる重要な評価属性であった。また、レジャーエリアの宿泊では料理と風呂と接客・サービスが重要な評価項目であった。これらの数値評価項目に対して、宿泊サービス提供者側が自ら発信する情報提供を強化する事で、利用者はさらに満足のいく判断につながるであろう。そしてビジネスエリア、レジャーエリアを構成するそれぞれの地区に分割してさらに検証すると、地区による特徴の違いが存在した。また、これらの研究成果の活用方法として、数値評価の項目へ評価マークの表示、書き込みの多い評価項目の追加表示、抽出した特徴表現数の表示、高頻度で出現した特徴表現〈属性〉の表示と、宿泊検討者にとって有益となる情報を宿泊レビューに掲載する手法を提案した。宿泊検討者にとって、さらに宿泊施設選択を納得したもとする支援情報となることを期待している。

第4章では、日本を訪れる旅行者である訪日旅行者によって、英語で書かれた宿泊レビューから感想コメントのテキストマイニング分析を行い、注目する評価項目とその評価表現の抽出を行った。第3章で用いた手法と同様に、感想コメントの評価表現を数値評価項目に対応させることで、どの数値評価項目において利用者の関心が高いかを検証した。訪日旅行者に人気の京都エリアを対象とし分析を行い、注目される数値評価項目に優先順位があることと、エリア内のホテルごとに評価される数値評価項目に、違いがあることを示し、英語で

書かれた感想コメントにおいても提案手法が適用できることが確認できた。この成果を生かし、訪日旅行者の旅行目的に応じて注目すべき数値評価項目を強調して表示させる方法や、感想テキストから得た評価表現を注目情報として宿泊レビューに表示させる方法をするなど訪日旅行者の施設選択時の判断に役立つ情報提供の実現に向けた対応が可能となる。

第5章では、宿泊利用者が発信する宿泊レビューの感想コメントには、1つの文中に複数の評価表現が記載されることに着目し、その評価の傾向の分析を行った。感想コメントの中で肯定評価と否定評価の評価極性が途中で反転遷移する文に着目し、評価極性が遷移する表現の抽出方法を提示した。結果を数値評価項目に対応させて分析したところ、評価極性が遷移する表現は、同一の数値評価項目間で遷移する傾向があることがわかった。また、宿泊利用目的の異なるビジネスエリアとレジャーエリアに分類し、評価表現の遷移の分析を行ったところ、宿泊施設の利用目的によって評価対象が異なることを示した。後半の評価が肯定的な場合、それは宿泊施設の強みとなり、後半の評価が否定的な場合は改善すべき項目となる。これらの評価表現の遷移から不満の要因となる評価表現を抽出することで、宿泊施設の利用目的に合わせたサービスを改善するための評価対象項目を特定し、宿泊施設のサービス改善に貢献するための情報取得方法を提案した。

以上、本研究で提案する宿泊レビューの感想コメントを利用した特徴となる評価表現の抽出と、数値評価と対応させた情報提示方法について報告した。今後は、検討した手法を用いて、実際に宿泊施設を予約しようとする宿泊検討者、また宿泊者により良いサービスを提供したいと願う宿泊施設サービス提供者に対し、双方に有益な情報提供ができる宿泊レビュー提示システムの構築を目指す予定である。

謝辞

博士課程の在学中，公私にわたってたいへんお世話になった筑波大学大学院システム情報工学研究科の 津田 和彦 教授に深く感謝いたします。

論文審査を快くお引き受け頂き，的確なアドバイスを頂戴した筑波大学大学院システム情報工学研究科の 吉田 健一 教授，佐藤 美佳 教授，倉橋 節也 准教授に深く感謝いたします。

論文の作成にあたって，日頃より貴重なご助言を頂いた帝京大学 文学部 社会学科の 藤田 昌克 准教授ならびに山口大学大学院 技術経営研究科 高橋 雅和 准教授に深く感謝いたします。

また，津田研究室の方々には，日頃より研究の内容および進め方についての貴重な示唆やご意見を頂戴いたしました。ここに深く感謝いたします。

著者が在籍する株式会社日本旅行で本研究の意義を理解し，温かく見守り励まして下さった皆様にも深く感謝いたします。また旅行業界に在籍する諸先輩からいただいた感想および意見は，研究を進めるにあたって重要なヒントになりました。

最後に，休日のほとんどの時間を研究に集中でき，元気づけられ，ここまで辿り着けたのは，私の家族，妻 亜希子，息子の聡磨の理解の賜物です。

ここに，記して以上の方々に深く感謝いたします。

参考文献

[Adachi 2011]

安達 清治: “変革期を迎えた旅行業 —旅行形態の変化とネット時代の展望—”, 大阪観光大学紀要 11, pp.101-104 (2011)

[ADK 2012]

“ADK 旅行クチコミ調査”, アサツー ディ・ケイ (2012)

http://www.adk.jp/html/news/2012/20120820_001614.html (2013/11/28 accessed)

[Asahara 03a]

浅原正幸, 松本裕治: “形態素解析システム「茶釜」 version 2.3.3 使用説明書”, 奈良先端科学技術大学院大学 (2003)

[Asahara 03b]

浅原正幸, 松本裕治: “ipadic version 2.7.0 ユーザーズマニュアル”, 奈良先端科学技術大学院大学 (2003)

[Bjørkelund 2012]

E. Bjørkelund, T.H. Burnett, K. Nørvåg, “A study of opinion mining and visualization of hotel reviews”, Proc. 14th Int. Conf. Inf. Integr. Web-Based Appl. Serv. - IIWAS '2012, pp. 229-238 (2012)

[DBJ 2013]

“地域のビジネスとして発展するインバウンド観光—日本型 DMO による「マーケティング」と「観光品質向上」に向けて—”, 日本政策投資銀行, 株式会社日本経済研究所 (2013)

[Fujii 02]

藤井敦, 石川徹也: “World Wide Web を用いた事典知識情報の抽出と組織化”, 電子情報通信学会論文誌, Vol.J85-D-II, No.2, pp.300 (2002)

[Fujimura 2003]

藤村滋, 豊田正史, 喜連川優, “電子掲示板からの評価表現および評判情報の抽出”, 第18回人工知能学会全国大会, 3F1-03 (2003)

[Fujimura 2009]

藤村 和宏: “医療サービスと顧客満足”, 医療文化社, pp6-8 (2009)

[goo 2013]

goo リサーチ, “「国内旅行と情報メディア」に関する調査”, NTT コム オンライン・マーケティング・ソリューション株式会社, (2013)

<http://research.goo.ne.jp/database/data/001529/> (2013年11月25日 accessed)

[Gretzel 2008]

Gretzel, U. Kyung, H, Y: “Use and Impact of Online Travel Reviews” Information and Communication Technologies in Tourism 2008, pp35-46 (2008)

[Grishman 86]

Ralph Grishman: “Computational Linguistics: An Introduction”, Cambridge Univ Press (1986)

[Hamaoka 2006]

濱岡 豊: “消費者間の相互作用”, 田中洋, 清水聰 (編) 『消費者・コミュニケーション戦略』, 有斐閣アルマ, pp57-93 (2006)

[Harada 01]

原田実, 水野高宏: “EDR を用いた日本語意味解析システム SAGE”, 人工知能学会論文誌 Vol.16 No.1 (2001)

[Hase 1999]

長谷 政弘: “観光ビジネス論”, 同友館, (1999)

[Hatakama 2004]

幡鎌博: “宿泊予約サイトの裏側に見る BtoB の進展と課題”, 経営情報学会 秋季全国大会, 6E-3 (2004)

[ICHSA 2012]

社団法人 国際交流サービス協会：“外国人旅行者受け入れについての調査”，
(2010).

[Inui 2006]

乾 孝司,奥村 学：“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 Vol.13 No.3, pp201-241 (2006)

[JTA 2012]

“観光立国推進基本計画”，観光庁 (2012)

[JTA 2013a]

“観光白書 平成 25 年版”，観光庁 (2013)

[JTA 2013b]

“宿泊旅行統計 平成 24 年”，観光庁 (2013)

[JTA 2013c]

“観光立国推進基本計画参考資料”，観光庁 (2013)

[JTA 2013d]

“ビジットジャパン事業について”，観光庁 (2013)

<http://www.mlit.go.jp/kankocho/shisaku/kokusai/vjc.html> (2013/11/25
accessed)

[JTA 2013e]

“史上初の訪日外国人旅行者数 1000 万人達成！！”，観光庁， (2013)

http://www.mlit.go.jp/kankocho/topics08_000111.html (2013/12/21
accessed)

[JTB 2013]

“旅行年報 2013”，公益財団法人日本交通公社 (2013)

[JATA 2004]

日本旅行業協会国内旅行委員会：“更なる国内旅行にむけて —新時代の旅
行業の役割—”，日本旅行業協会 (2013)

[JATA 2013]

“数字が語る旅行業 2013”，日本旅行業協会 (2013)

[JFC 2013]

“国内宿泊施設の利用に関する消費者意識と旅館業の経営実態調査”, 日本政策金融公庫 (2013)

[JNTO 2013]

“TIC 利用外国人旅行者調査報告書”, 日本政府観光局 (2013)

[JRC 2013]

“じゃらん宿泊調査 2013”, じゃらんリサーチセンター (2013)

[Kamichi 2012]

上地 恵龍: “沖縄県宿泊施設における格付け制度の考察 — 特色ある宿泊施設づくりを事例として —”, 琉球大学観光科学 4, pp.43-53 (2012)

[Kano 2008]

狩野 美智子: “旅行業者の競争戦略の分析: 大手 4 社の経営行動に着目して”, 静岡大学経済研究. 13(3), pp.51-83 (2008)

[Kasper 2011]

W. Kasper, M. Vela, “Sentiment Analysis for Hotel Reviews”, Proc. Comput. Linguist. Conf. 2011, pp.45–52 (2011)

[Kawano 02]

河野浩之, 山田誠二, 北村泰彦, 高橋克己: “インターネットの知的情報技術 情報検索とエージェント”, 東京電機大学出版局 (2002)

[Kin 03]

金明哲, 村上征勝, 永田昌明, 大津起夫, 山西健司: “言語と心理の統計, 統計科学のフロンティア 10”, 岩波書店 (2003)

[Kita 99]

北研二: “確率的言語モデル”, 東京大学出版会 (1999)

[Kobayashi 2003]

小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: “テキストマイニングによる評価現象の収集”, 情報処理学会研究報告. 自然言語処理研究会報告. 2003, pp.77–84 (2003)

[Kobayashi 2005]

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 (2005)

[Kobayashi 2006]

小林 のぞみ, 飯田 流, 乾 健太郎, 松本 裕治: “照応解析手法を利用した属性-評価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, C2-6 (2005)

[Kobayashi 2010]

小林 裕和: “旅行業における商品イノベーションを引き起こす旅行商品の特性について”, 国際広報メディア・観光学ジャーナル 10, pp61-72 (2010)

[Kokuryo 1999]

国領 二郎, “オープン・アーキテクチャ戦略”, ダイヤモンド社 (1999)

[Kotler 2003]

P. Kotler, J.T. Bowen, J.C. Makens: “Marketing for Hospitality and Tourism”, Pearson Education Limited (平林祥, 白井義男 訳: “コトラーのホスピタリティ&ツーリズム・マーケティング”, ピアソン・エデュケーション (2003)) .

[Lovelock 1999]

Christopher Lovelock, Lauren Wright, “Principles of Service Marketing and Management”, Prentice-hall (小宮路 雅博 訳: “サービス・マーケティング原理”, 白桃書房, 2002)

[Mani 01]

Inderjeet Mani: “Automatic Summarization”, John Benjamins Publish (2001)

[Matsuoka 00]

松岡弘, 庵功雄, 高梨信乃, 中西久美子, 山田敏弘: “日本語文法ハンドブック”, スリーエーネットワーク (2000)

[Masuoka 92]

益岡隆志, 田窪行則: “基礎日本語文法 (改訂版)”, くろしお出版 (1992)

[Masuoka 97]

益岡隆志, 仁田義雄, 郡司隆男, 金水敏: “文法, 言語の科学 5”, 岩波書店 (1997)

[Matsumoto 01]

松本祐治, 伝康晴: “話し言葉の形態素解析”, 情報処理学会 自然言語処理研究会 NL- 143-7 (2001)

[Matsumoto 04]

松本祐治, 高岡一馬, 浅原正幸, 工藤拓: “茶釜と南瓜による日本語解析一構文情報を用いた文の役割分類一”, 人工知能学会誌 Vol.19 No.3 (2004)

[MIC 2012]

“平成 24 年版 情報通信白書”, 総務省 (2012)

[Mikasa 2013]

三笠弘貴, 奥野拓: “観光記事における概要文章の個人適応に向けた文章分類手法の提案”, 情報処理学会 全国大会講演論文集 2013, pp621-623 (2013)

[Misono 2013]

御園 謙吉: “地方自治体の観光政策と観光統計”, 法政大学日本統計研究所 研究所報 42 観光統計, (2013)

[Morii 2011]

森井 庸介, 櫻 惇志, 波多野賢治, “複数評価文書を用いた極性判定とその判定理由の提示による消費者購買行動システムの提案”, DEIM Forum, C2-1, (2011)

[Morikawa 2001]

森川 毅: “ホテル・旅館の評価: 心理学的側面からの分析“, 長崎国際大学論叢 1, pp273-280 (2001)

[Nagao 96]

長尾真編: “自然言語処理, 岩波講座ソフトウェア科学 15”, 岩波書店 (1996)

[Nagao 99]

長尾真, 中川裕志, 松本裕治, 橋田浩一, John Bateman: “言語の数理, 言語の科学 8”, 岩波書店 (1999)

[Nagashima 2009]

長島 健太: “サービス商品の選択・評価におけるクチコミの機能—ホテルサービスの場合—”, 東洋大学経営学部 経営論集 74 号, pp91-108 (2009)

[Nasugawa 2006]

那須川 哲也: “テキストマイニングを使う技術／作る技術”, 東京電機大学出版局 (2006)

[Nielsen 2012]

“Consumer Trust in Online, Social and Mobile Advertising Grows”, Nielsen

<http://www.nielsen.com/us/en/newswire/2012/consumer-trust-in-online-social-and-mobile-advertising-grows.html> (2013/11/25 accessed)

[Nomura 88]

野村浩郷: “自然言語処理の基礎技術”, コロナ社 (1988)

[NTA 2006]

“日本旅行 100 年史”, 日本旅行, (2006)

[O'Connor 2008]

P. O'Connor: “User-Generated Content and Travel: A Case Study on Tripadvisor.Com”, Information and Communication Technologies in Tourism 2008.2, pp 47–58 (2008)

[Oi 2013]

大井 達雄: “観光地ブランドの評価に関する一考察”, 法政大学日本統計研究所 研究所報 42 観光統計 (2013)

[Okumura 2010]

奥村 学, 高村 大也: “自然言語処理シリーズ 1 言語処理のための機械学習入門”, コロナ社 (2010)

[Okumura 2013]

奥村 学：“ソーシャルメディアを対象としたテキストマイニング”，電子情報通信学会 基礎・境界ソサエティ：Fundamentals Review Vol.6 No.4 (2013)

[O'Mahony 2009]

M.P. O'Mahony, B. Smyth: “Learning to recommend helpful hotel reviews”, Proc. Third ACM Conf. Recomm. Syst. - RecSys '09, ACM Press 2009, pp. 305-308 (2009)

[Ouchi 2008]

大内 東,大柳 幸彦：“日本のホテル評価システムに関する研究”，開発こうほう (539),pp.27-31 (2008)

[Rosen 2002]

E. Rosen: “The Anatomy of Buzz: How to Create Word of Mouth Marketing”, Random House (2002)

[Ryoken 2010]

リョケン調査隊アンケート：“ネット予約・検索対策等の状況調査”，株式会社リョケン (2010)

[Sernovitz 2009]

Andy Sernovitz: “Word of Mouth Marketing”, Kaplan Publishing, (花塚 恵 訳, “WOM マーケティング入門”，海と月社, (2010))

[Shimakawa 2010]

島川 崇：“IT 普及と旅行業の変革”，運輸と経済 Vol70,No5 (2010)

[Shimooka 95]

下岡和也, 内元清貴, 河原達也, 井佐原均: “日本語話し言葉の係り受け解析と文境界推定の相互作用による高精度化”，自然言語処理 Vol.2 No.3 (1995)

[Sparks 2011]

Sparks, B.A. and Browning, V. “The Impact of Online Reviews on Hotel Booking Intentions and Perception of Trust”, Tourism Management 32(6), pp.1310-1323 (2011)

[Suzuki 2008]

Nobuo Suzuki, Kazuhiko Tsuda: “An efficient method of intention understanding and learning for the ambiguous expressions of sentence end” IJCAT 33(2/3), pp.99-108 (2008).

[Suzuki 2010]

鈴木 勝: “我が国の旅行会社における流通チャネルと経営の構造変化”, 運輸と経済 Vol70, No5 (2010)

[Tanabe 2008]

田邊 亘, 後藤 正幸: “宿泊施設の戦略構築を支援するユーザーレビュー分析に関する一考察”, 武蔵工業大学環境情報学部情報メディアセンタージャーナル Vol.9, pp91-101 (2008)

[Tateishi 2001]

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”, 自然言語処理. 144-11 (2001).

[Tokunaga 1999]

徳永 健伸: “言語と計算 5 情報検索と言語処理”, 東京大学出版会 (1999)

[Tsujii 1999]

辻井潤一, 北研二: “確率的言語モデル, 言語と計算 4”, 東京大学出版局 (1999)

[Tsujii 2011]

辻井潤一, “英語構文解析器 Enju 2.4.2 マニュアル”, 東京大学 (2011)

[Tsujii 2012]

K. Tsujii, T. Ikoma, K. Tsuda: “An Evaluation Method for Segmental Accommodation Reviews with Text Mining”, Advances in Knowledge-Based and Intelligent Information and Engineering Systems Vol.243, pp. 893–900 (2012)

[Uetake 2011]

植竹 朋文, 青木 章通: “繁閑格差の大きい地区のホテルにおけるネット直販に関する意識調査”, 専修大学 情報科学研究所所報 No.76, pp.1-26 (2011)

[Vermeulen 2009]

Ivar E. Vermeulen, Daphne Seegers: “Tried and Tested: The Impact of Online Hotel Reviews on Consumer Consideration”, *Tourism management* Vol.30-1, pp.123-127 (2009)

[Wu 2010]

Y. Wu, F. Wei, S. Liu, N. Au, W. Cui, H. Zhou: “OpinionSeer: interactive visualization of hotel customer feedback”, *IEEE Transactions on Visualization and Computer Graphics*. 16, pp.1109–1118 (2010)

[Yamaguchi 03]

山口 翼: “日本語大シソーラス 類語検索大辞典”, 大修館書店 (2003)

[Yamamoto 2008]

山本 晶: “インフルエンサーの発見とクチコミの効果～AIDEES モデルの実証分析～”, *季刊マーケティングジャーナル* 28(1), pp4-18 (2008)

[Yoshimura 00]

吉村賢治: “自然言語処理の基礎”, サイエンス社 (2000)

[Zhang 2009]

L. Zhang, B. Pan, W. Smith, X. (Robert) Li: “An Exploratory Study of Travellers’ Use of Online Reviews and Recommendations”, *Information Technology & Tourism*. 11, pp157–167 (2009)

関連業績リスト

【参考論文】

・公表済み論文

- (1) 辻井康一, 津田和彦: “テキストマイニングを用いた宿泊レビューからの注目情報抽出方法”, Journal of Digital Practices. 3 pp.289-296 (2012)
- (2) K. Tsujii, Y. Fujita, K. Tsuda: “Attention information extraction of the foreign visitors using Text Mining”, International Journal of Intelligent Systems Technologies and Applications. pp.194-206 (2013)
- (3) K. Tsujii, K. Tsuda, Y. Fujita: “The extraction method of the service improvement information from Guests’ review”, Advances in Knowledge-Based and Intelligent Information and Engineering Systems, pp. 250–257 (2013)

【その他の論文】

・公表済み論文

- (1) K. Tsujii, T. Ikoma, K. Tsuda: “An Evaluation Method for Segmental Accommodation Reviews with Text Mining”, Advances in Knowledge-Based and Intelligent Information and Engineering Systems, pp. 893–900 (2012)

・採録決定論文

- (2) K. Tsujii, M. Takahashi, Y. Fujita, K. Tsuda: “Extraction of Accommodation Evaluation by Foreign Tourists with Text Mining”, International Journal of Trade, Economics and Finance. (Vol.5 No.2 に掲載予定)

・査読のない発表論文

- (3)辻井康一, 津田和彦: “宿泊レビューを用いた宿泊施設サービス改善のた

めの情報抽出”，人工知能学会全国大会（2013）

(4)辻井康一，津田和彦：“ユーザーレビューを用いた宿泊施設の情報補完手法，情報科学技術フォーラム 2011, pp.201-202 (2011):