

テキストマイニングを用いた観光の言説分析

守屋 豊[†] 井出 明^{††}

近畿大学[†] 首都大学東京^{††}

観光という営みは言語情報と深く関係している。それゆえ、観光体験をより良いものにし、またそれを記憶として伝達するためには、観光に特化した言語シソーラスが必要であると筆者らは主張してきた。これまで、国内観光情報サイトの記事を素材として、図書館情報学におけるシソーラス構築のノウハウを応用し、語と語の関連性をはかるという手法を用いることで、日本語による観光情報シソーラスの設計の可能性について研究を重ねてきた。本研究では、世界的に最も人気を博している観光地であるパリを対象として、国内における海外旅行口コミサイトである「フォートラベル」から日本語で書かれた記事を取り出すと共に、海外における観光情報スペースである“travel blog”から英文記事を抽出した。そして、それぞれに対してテキストマイニングの手法を導入し、語をシソーラスの形でまとめ、比較を行った。最終的には、ビジャーの抱える文化の違いによって、パリという一地域における記述の言語表現にどのような変化が見られるかについて分析した。

A discourse analysis of tourism with text mining

Yutaka Moriya[†] Akira Ide^{††}

Kinki University[†] Tokyo Metropolitan University^{††}

The experiences of tourism have many strong relationships with linguistic information. In order to make the experience of tourism better and to inform other people of experiences as memories the thesaurus unique to tourism is needed. Therefore, in our research, using the data we collected in tourism information sites, written by Japanese tourists, we tried to make a tourism thesaurus by applying a method in LIS (Library Information Science), which analyze relationships among words to construct a thesaurus in Japanese. In this research, we extracted data about Paris from “4-travel”, which is written by Japanese people, and from “travel blog”, which is written by English speaking people. Then, we applied a method of text mining to each data, and compared them for creating a thesaurus. Finally, we analyzed whether linguistic expressions for Paris changed according to the cultural difference of visitors.

テキストマイニングを用いた 観光の言説分析

2008年1月31日(木) EIP39

近畿大学 経済学部経済学科
守屋 豊

首都大学東京大学院 都市環境科学研究所
井出 明

目次

はじめに

- 1. 前回までの研究
- 2. 研究の目的
- 3. 先行研究
- 4. 実験方法
 - 4.1. 日本語レビューのテキストマイニング
 - 4.2. 英語レビューのテキストマイニング
- 5. 考察
- 6. まとめ
- 7. 今後の課題

「パリとはどういう街か」

- パリのイメージ?
 - グルメ、ファッション、歴史、芸術文化…
 - 人種・民族の違いでパリのイメージは変わるのか
 - これまで、それを計量的にはかることが出来なかつた
- ↓
- 情報化社会の到来
 - テキストマイニング→言語を計量分析することが可能に

これまでの観光における計量分析

- TSA(Tourism Satellite Account)
 - ツーリズム部門会計
 - UNWTOの開発した、統計情報を元に観光がGDPにどれだけ貢献しているかを測定
 - 経済面での分析
- 観光地のイメージを含めた全体像を計量的に捉えることはできない^[1]

新しい計量分析の可能性

- テキストマイニングを用いる
 - インターネット上のブログや口コミ情報を用いて、その土地がもつ雰囲気や全般的な情報を言語的に計量化することで測定

1. 前回までの研究

- 観光という営みが言語情報と深く関係
 - 観光体験をより良いものにする記憶としてそれを伝達する
- 観光に特化した言語シソーラスの必要性
- 図書館情報学の手法を導入することで日本語による観光情報シソーラス構築が可能に^[2]

一つの概念に対して語がさまざまにある。



フリーワード検索 (楽天トラベル内の口コミ検索) [3]

夕日	357件
夕陽	122件
夕焼け	39件
夕やけ	8件

ヒット件数に大きな違い
求められる情報にアクセスできなくなる

シソーラスの一例

『日本語大シソーラス』^[4] の“夕日”に関する単語

- ・ 太陽
 - ・ 夕日
 - ・ 夕陽
 - ・ 斜陽
 - ・ 落日
- 観光情報サイトから抽出した“夕日”関連語
- 夕焼け
 - 灯台
 - 神秘的
 - 感動
 - オレンジ
 - 黄色
 - レストラン
 - 温泉
 - ⋮

夕日にに関する観光情報シソーラスの一部

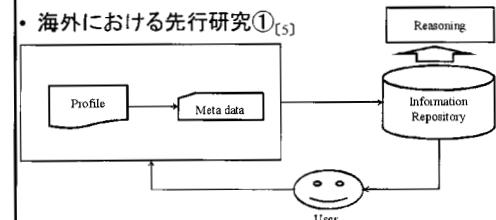


2. 研究の目的

- ・ 文化的な違いによる観光行動の比較
- 日本語と英語によるシソーラスの違いによって比較可能

3. 先行研究

・ 海外における先行研究①^[5]



- ・ ビジターのプロフィールを元にユーザーが興味をもつような情報を提供
(推論エンジンにオントロジーを使用)

3. 先行研究

- 海外における先行研究②^[6]

```
<TITLE> ※※※ </TITLE>
:
<KEYWORDS> ※※※ </KEYWORDS>
```

- 観光情報ウェブサイトのHTML構造からその内容を読み取る
・単語の共起関係、重要度による重み付けを利用

3. 先行研究

先行研究調査の結果

- 観光の口コミレビューに対して図書館情報学の手法を導入した試みは未だみられない



観光情報学の新たな分析ツールとして
テキストマイニングを用いることに十分な
合理性と新規性がある

4. 実験方法

- 対象観光地: パリ
 - 2005年 フランス 外国人訪問客数
 - 7910万人(うち日本人が約65万人)^[7]
 - 世界的な観光地
- 対象言語: 日本語、英語
- 対象情報: 観光ブログ記事(テキストと写真)
 - 日本語: フォートラベル^[8]
 - 英語: TravelBlog^[9]

4.1. 日本語レビューのテキストマイニング

「フォートラベル」

- パリに関する記述を直近300件を抽出(全2554件中)

- ①形態素解析システム“ChaSen”による形態素解析^[10]
- ②語の出現頻度の測定
- ③カテゴリ分類によるシソーラス構築

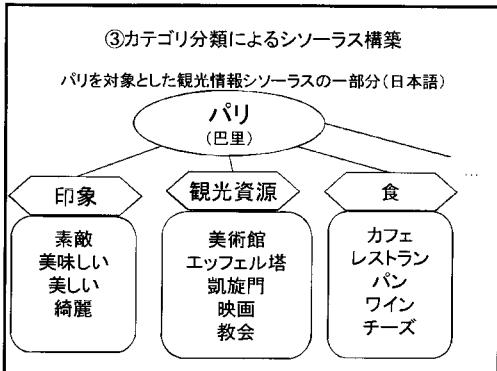
①形態素解析

(例) 「シャンゼリゼの夜景を満喫した。」

シャンゼリゼ	名詞-固有名詞-一般
の	助詞-連体化
夜景	名詞-一般
を	助詞-格助詞-一般
満喫	名詞-サ変接続
し	動詞-自立
た	助動詞
。	記号-句点

②語の出現頻度の測定(Excelを使用)

B	C	D	E	F
トピック	データの総数 / 頻数			
パリ	1415			
人	612			
東	615			
空氣	485			
晴れ	447			
オクタノル	450			
フランス	444			
晴れ	431			
パリ	405			
天候	380			
空氣	364			
晴れ	356			
天	354			
日	353			
風	340			
気温	323			
湿度	316			
中	314			
エントラード	306			
駅	297			
時間	279			
出	262			
セーヴ	260			
バス	260			
日本	254			
街	251			
パリタリダム	246			



4.2. 英語レビューのテキストマイニング

・“Travlblog”

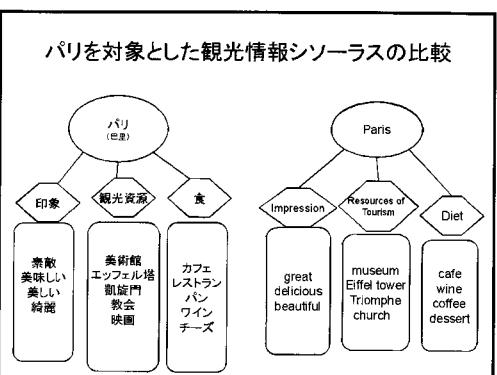
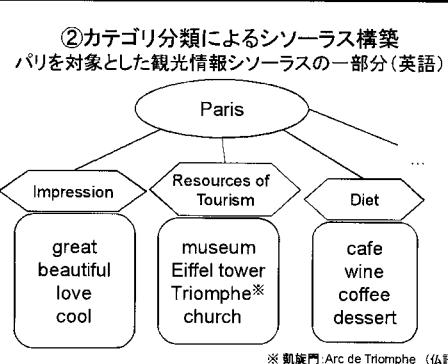
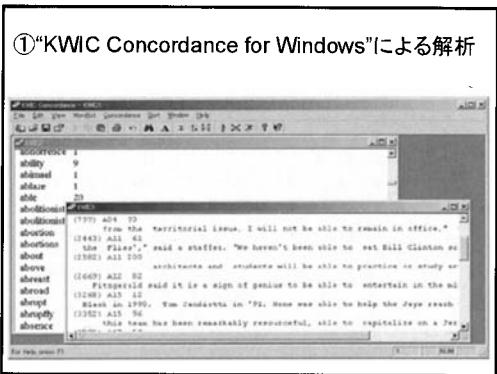
- “Paris”に関する記述を直近のもの300件抜き出し (1588件より)

“ChaSen”: 現時点では英語に非対応

①“KWIC Concordance for Windows”による解析^[11]

- 語の抽出
- 語の出現頻度の測定

②カテゴリ分類によるシソーラス構築



5. 考察

- ・観光資源(Resources of Tourism)に注目
- ・『ダ・ヴィンチ・コード』 ダン・ブラウン著
 - 2004年 日本翻訳版、出版
 - 2006年 劇場版、世界同時公開

日本語と英語の観光資源における語の出現頻度の比較

ダ・ヴィンチ・コード、 ダビンチコード…	41
Code	10
ルーブル美術館	228
louver	241
ピラミッド、…	93
pyramids	14
映画	111
movie, movies, film…	57

6.まとめ①

- テキストマイニングにより観光行動上の特徴の抽出
 - 歐米圏の人々と比較し、日本人の多くに映画のロケ地を巡るという特徴的な観光行動
- 漫然と記事を辿るだけでは気づき難い事象



サプライヤー側にとっての有用なマーケッティングデータ

6.まとめ②

- 各国のビジターの文化の違いから起こる観光行動の特性に応じたマイクロマーケティング展開が可能に
 - 検索漏れを防ぐシソーラスの本来の目的
 - テキストマイニングの過程で日本語、英語共に表記揺れを回収
- 口コミによる英語での観光情報シソーラスの構築

7.今後の課題

詳細なパラメータの付加(時間軸など)
他の観光地との比較



実用的なデータの収集

- 観光系学会
 - 学際的研究としての観光情報学の発展を目指す

参考文献①

- [1] 佐竹真一「観光資源評価の基礎概念」
『日本観光学会誌』 第48号No.48 June 2007 pp.68-80 2007年
- [2] 守屋豊・井出明「観光情報シソーラスの研究と設計」
『日本観光学会 第95回全国大会研究発表要旨』 pp.40-41 2006年
- [3] 美天トラベル(<http://travel.rakuten.co.jp/>)
2007年11月3日確認
- [4] 山口翼編『日本語大シソーラス—類語検索大辞典』
大修館書店 2003年
- [5] Chang, Choi, et al. "Travel Ontology for Recommendation System based on Semantic WebAdvanced Communication Technology," ICACT 2006
The 8th International Conference Volume 1, (2006): pp.624-627
- [6] Lobsan,Karoui, Et al. "Context-based Hierarchical Clustering for the Ontology Learning" Web Intelligence archive. Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence (2006) pp.420-427

参考文献②

- [7] 国際観光振興機構: 世界の国際観光の動向
(http://www.jnto.go.jp/jpn/tourism_data/global_tourism_trends.html)
2007年11月20日確認
- [8] フォートラベル (<http://4travel.jp/>) 2007年11月20日確認
- [9] "TravelBlog" (<http://www.travelblog.org/>) 2007年11月3日確認
- [10] "Chasen" (<http://chasen.naist.jp/niki/Chasen/>)
奈良先端科学技術大学院大学情報科学研究科 2007年11月3日確認
- [11] "KMIC Concordance for Windows"
(http://www.chs.nihon-u.ac.jp/eng_dpt/tukamoto/)
日本大学文理学部英文学科 2007年11月3日確認