

レビュー文のカテゴリ分類のための重文自動分割

Compound Sentence Segmentation for Review Text Categorization

高師 遥 阿部 遥佳 鈴木 信太郎 大和矢 悠仁 延澤 志保 （東京都市大学）

Haruka Takashi, Haruka Abe, Shintaro Suzuki, Yuji Yamatoya, Shiho Hoshi Nobesawa (Tokyo City University)

研究背景・目的

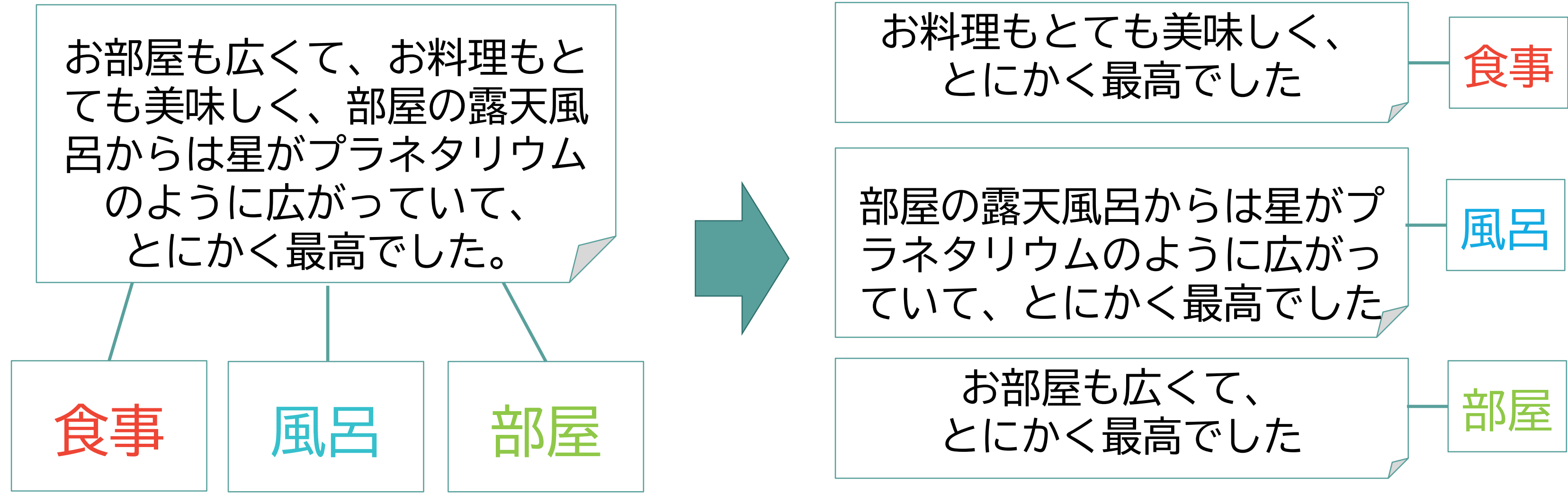
それぞれの宿泊施設について、レビュー文と6つの項目別5段階評価が投稿できる



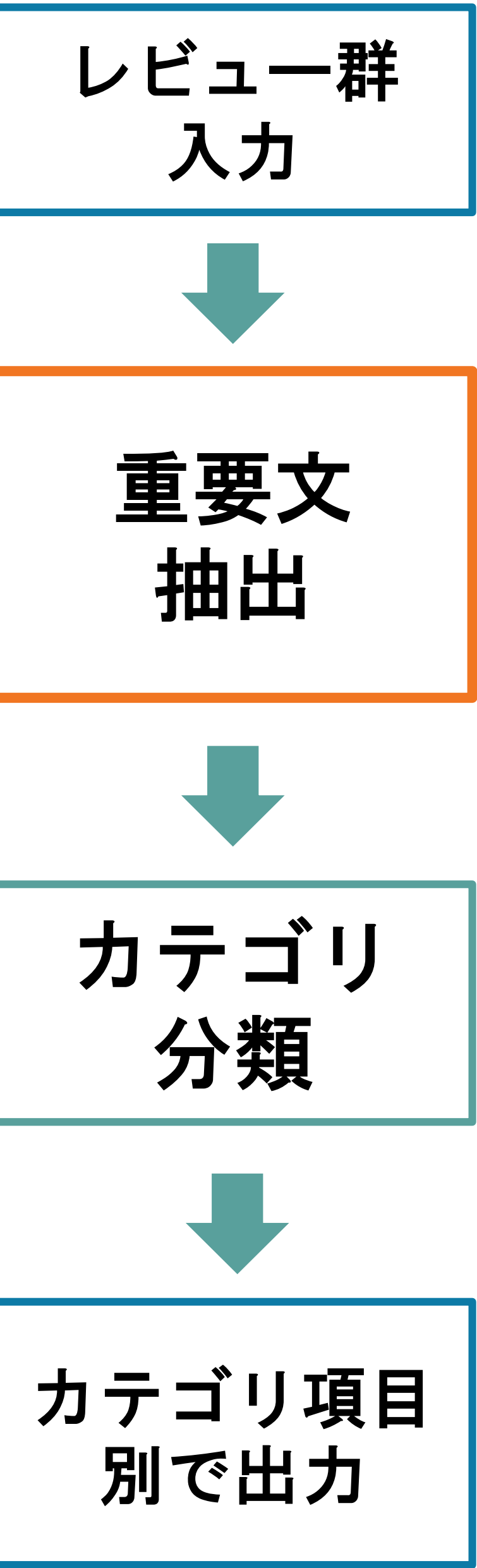
問題点

- ・評価の数字だけでは、具体的なレビュー内容がわからない
- ・投稿されたレビューではどこに何が書いてあるかわかりにくい
- ・レビュー数は膨大でありすべて読むことは困難である

施設ごとのレビュー文を句に分割し、評価視点である各カテゴリごとに出力することで、より簡潔な情報、必要な情報を得やすくする



提案手法



重要文抽出

重要語:固有名词、一般名词、サ変接続名词

レビュー群入力

句点を基に文に分割

係り受け解析

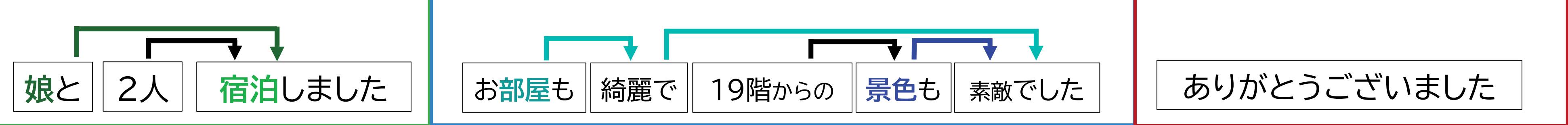
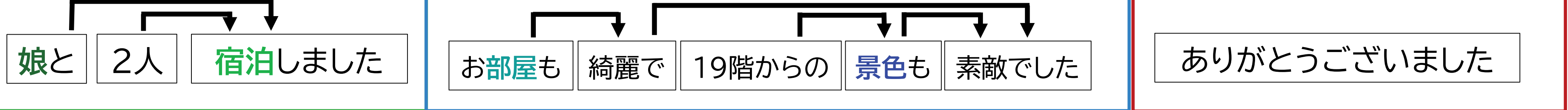
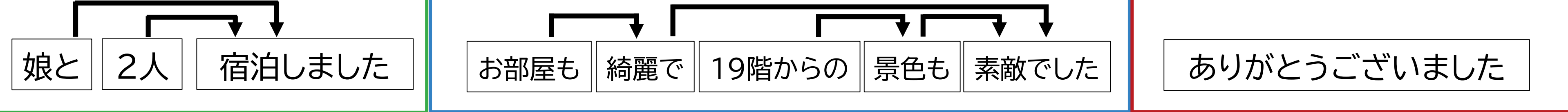
重要語認識

重要語を含む句の推定

選定した名詞を含む句を全て出力

娘と2人宿泊しました。お部屋も綺麗で19階からの景色も素敵でした。ありがとうございました。

娘と2人宿泊しました お部屋も綺麗で19階からの景色も素敵でした ありがとうございました



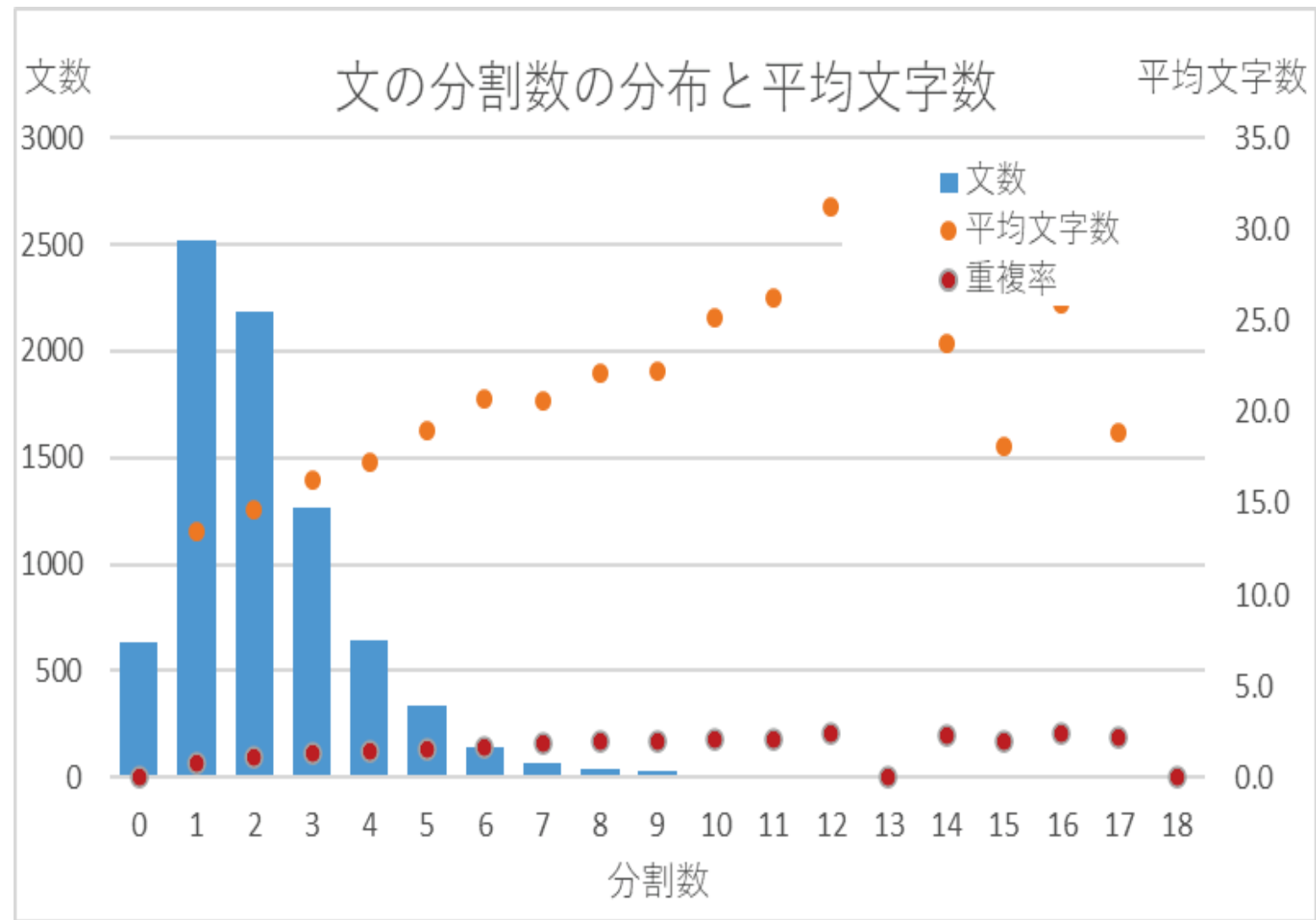
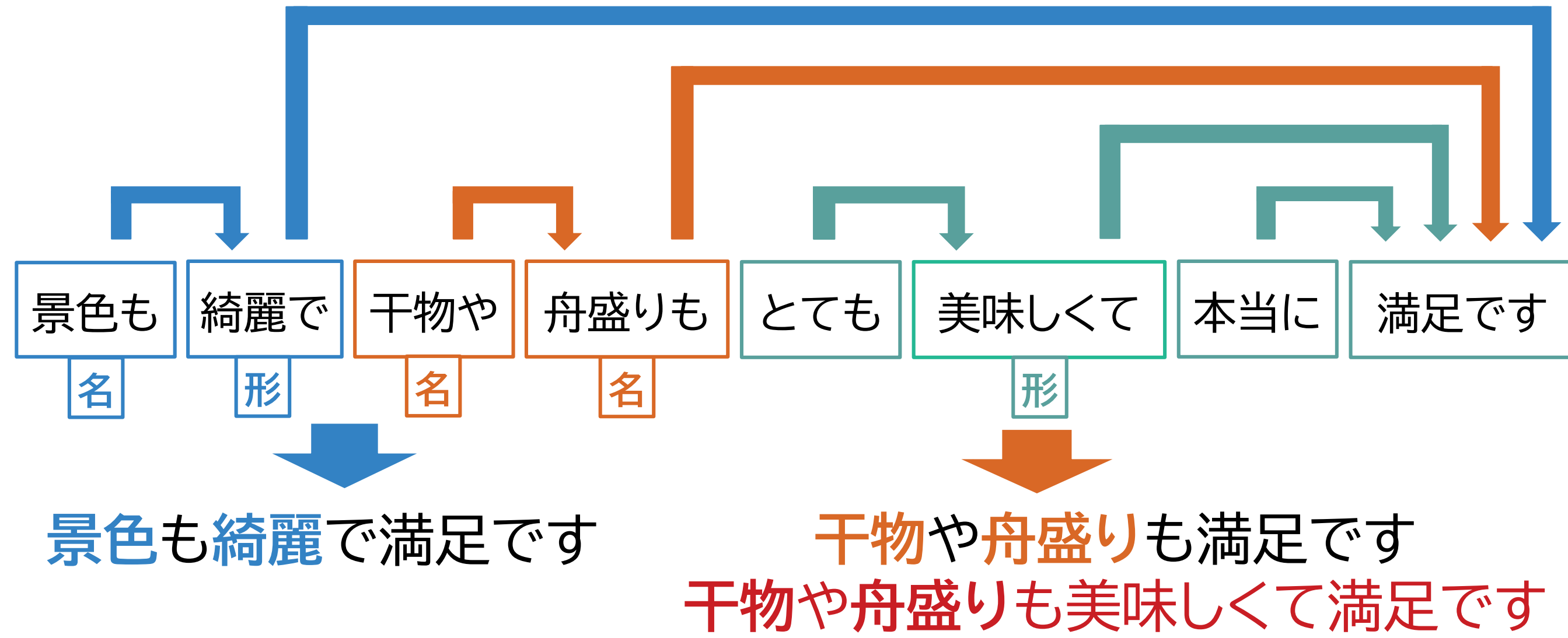
娘と宿泊しました 宿泊しました お部屋も綺麗で素敵でした 景色も素敵でした

実験結果・考察

過不足が多く見られた
○新語等、名詞だと認識されず削除された句が多く見られる
→新語への対処として辞書を使用する
→名詞の種類を特定せず、再度実験を行う

○口語のため、係り受けが正しく行われていない

○過不足の例



解析(7,900文)
分割数0:
削除された文 632文(8.57%)
○削除された文の内、目立ったレビュー文
・ありがとう 164文(25.95%)
・また来ます 19文(3.00%)
・美味しい 19文(3.00%)

謝辞

本研究では、国立情報学研究所のIDRデータセット提供サービスにより楽天グループ株式会社から提供を受けた「楽天データセット」 (https://rit.rakuten.com/data_release/) を利用しました。心から感謝申し上げます。

参考文献

楽天グループ株式会社 (2020): 楽天トラベルデータ. 国立情報学研究所情報学研究データリポジトリ. (データセット). <https://doi.org/10.32130/idr.2.3>
楽天グループ株式会社 (2021): 楽天トラベルレビュー: アспект・センチメントタグ付きコーパス. 国立情報学研究所情報学研究データリポジトリ. (データセット). <https://doi.org/10.32130/idr.2.14>