S04

動画サイトにおける
時刻同期コメントを用いた単語変化の検出

西尾駿斗，武藤敦子，島孔介，森山甲一，犬塚信博 （名古屋工業大学）
松井藤五郎 （中部大学）

背景

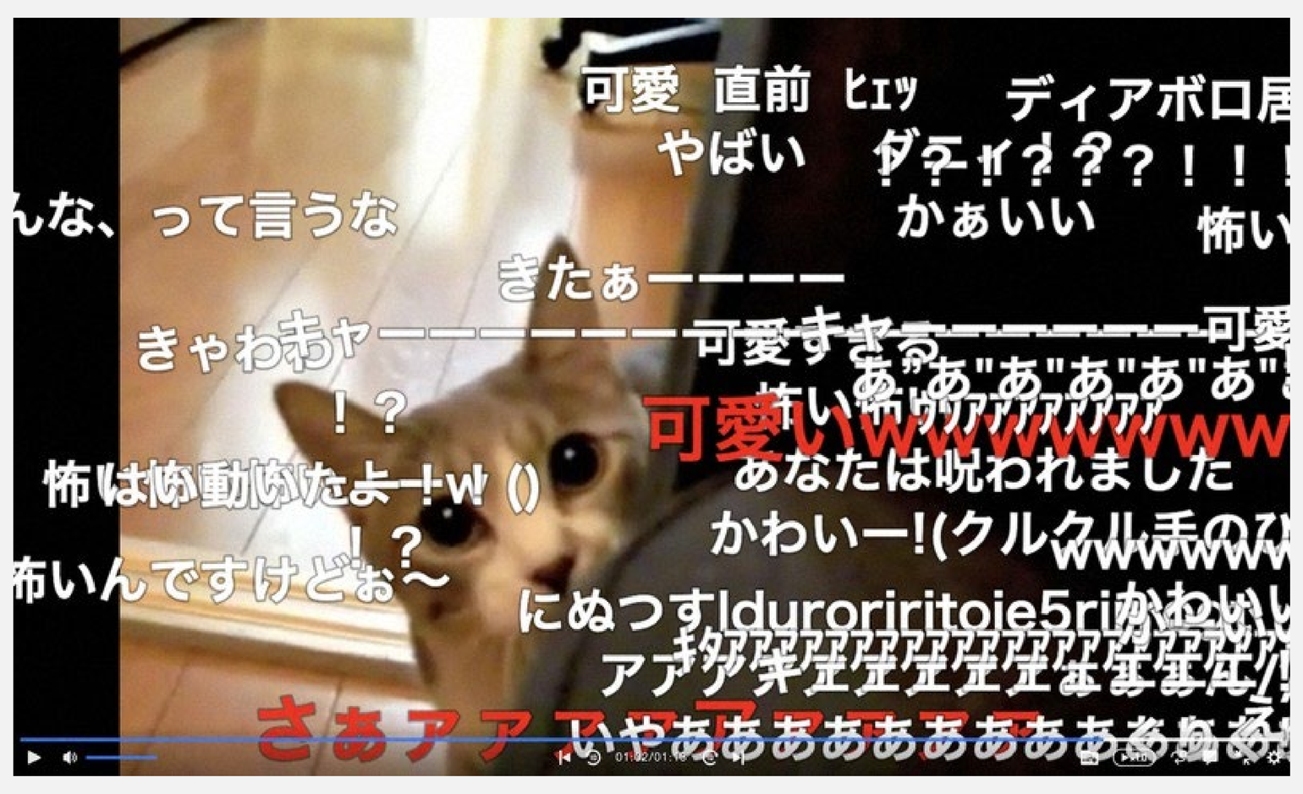
インターネットの普及により、SNSや動画共有プラットフォームの利用が一般化し、特有の言語文化やスラングが生まれているが、既存の言語変化の研究は主観的な分析が中心である。インターネット上で発生する言語変化を体系的に分析することで、現代の言語変化の理解に大きく役立つ。ニコニコ動画は**ユーザが任意の再生位置にコメントを投稿**でき、同じ場面のコメントを通時的に分析することで、言語変化の発見をするためのデータとして有用である。

目的

本研究では、**（1）語彙の散らばりを用いた、単語変化における注目時間帯の特定**を行い、
その時間帯における、**（2）似た意味を表す単語の変化を検出**することを目的とする。

使用データ

- **ニコニコ動画コメントデータ**（提供：ドワンゴ）
  投稿時間,本文,書込再生位置,コマンド等を保持

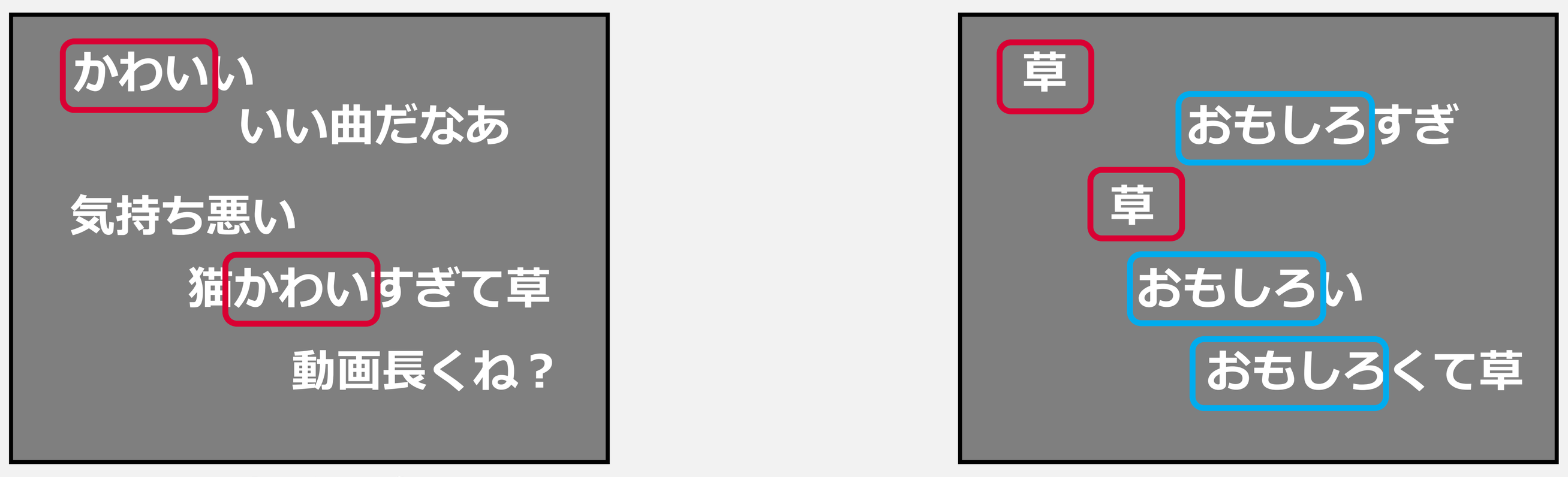


データの前処理

- **繰り返しの削除**
  繰り返し文字列は3回,長音は1回に短縮
  “キターーーー!!!!!!” → “キター!!!”
- **不要な記号の削除**
  顔文字などに使用される語や、文字間の空白を削除
- **表記揺れの統一**
  様々な長音記号や半角カナなどを全角に統一

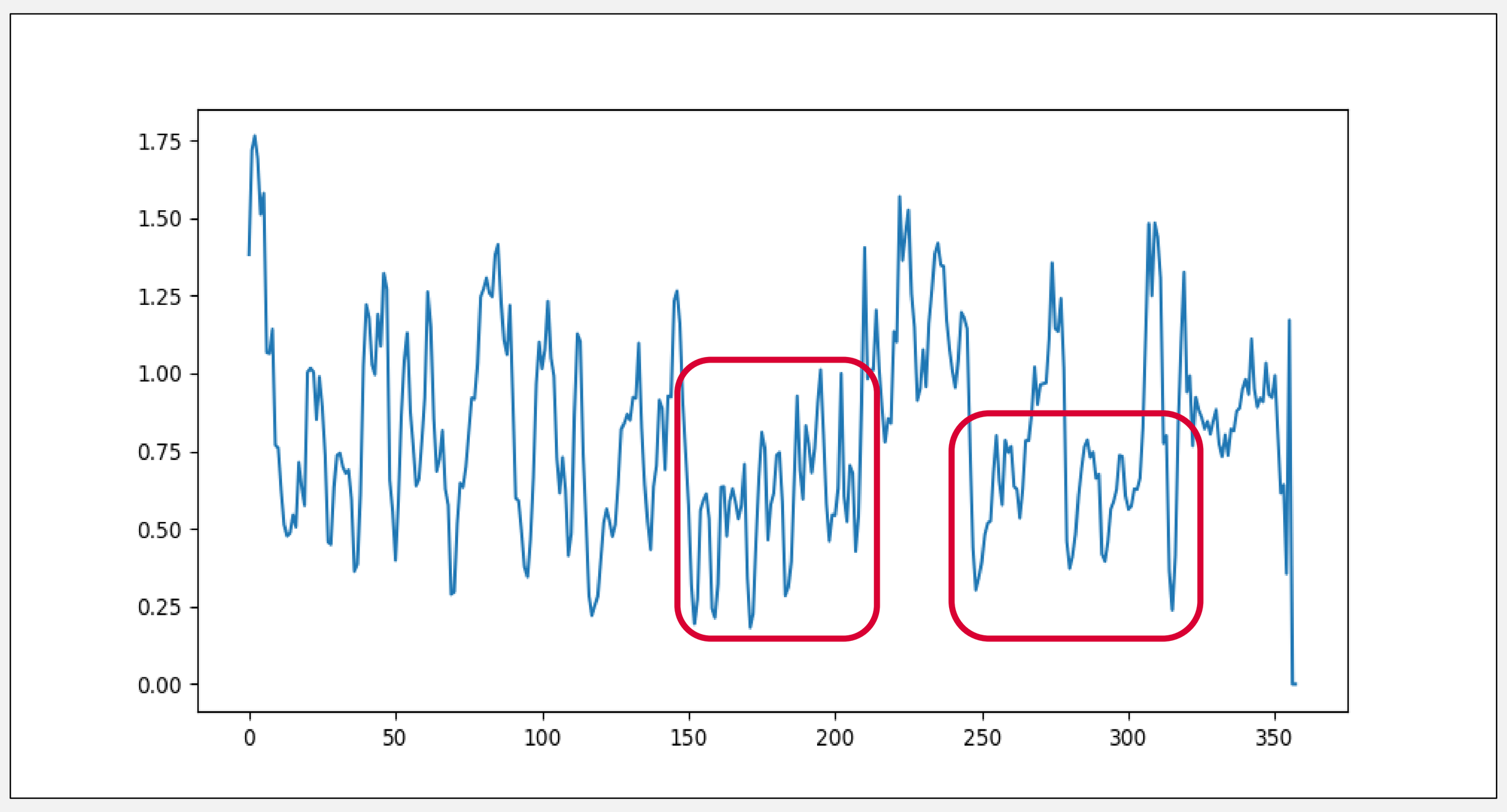
1. 単語変化における注目時間帯の特定

- **単語変化の検出がしやすい場面とは？**
  多くの人が同じような感想を述べている場面は、似た意味のコメントが投稿されやすく、単語の変化の検出を行うのに適したタイミングであると仮定



語彙数：多　　語彙数：少

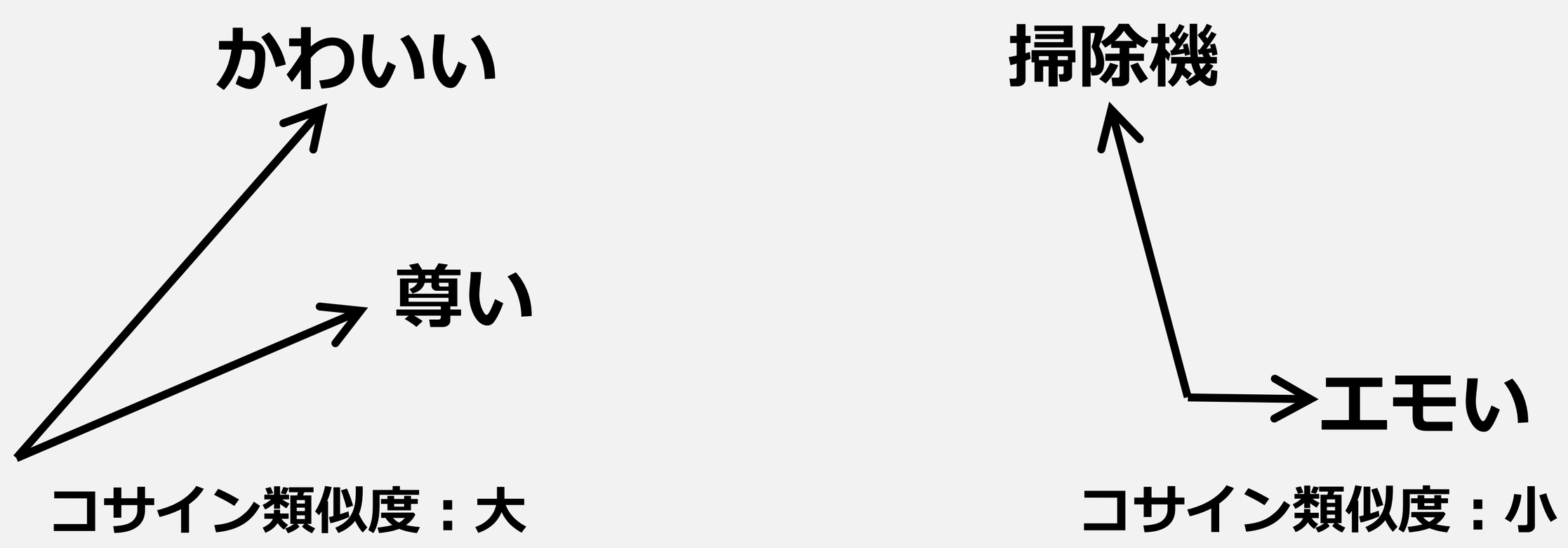
- **手法**
  再生時間帯ごとに、**「語彙数/コメント数」**を算出し、少ないタイミングを注目時間帯とする
  ※語彙数：重複を除いたユニークな単語数
- **実験**
  猫の動画集(sm1736482)を用いて、単語の種類数を計算
  いくつかの時間帯において単語の種類が
  減少するタイミングを発見



→注目時間帯に設定し、単語の変化を検出する

2. 似た意味を表す単語の変化の検出

- **類似語の特定（未定）**
  学習済みWord2vecを用いて、類似語を検出する。使用頻度の高い語について**コサイン類似度**を計算し、類似語のセットを特定する



Word2vec
単語をベクトルで表現し、意味の似た単語を近い位置に配置する技術。単語間の意味的な関係を捉えるため、
自然言語処理で広く活用されている

- **通時的な言語変化の検出（未定）**
  類似語のセット内の語の**出現頻度を時系列ごとに比較**し、使用回数の入れ替えを発見し、通時的言語変化を検出する

今後の展望

- Word2vecを用いた類似語の特定と、通時的な変化の検出
- 「注目時間帯において単語変化が発生しやすい」という仮定の妥当性の検証
- 他のプラットフォーム（X,YouTube)への手法の一般化
- 単語変化の自動検出精度の向上

連絡先：名古屋工業大学 犬塚・武藤研究室M2 西尾駿斗
Email：h.nishio.869@stn.nitech.ac.jp

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、株式会社ドワンゴのデータを利用させていただきました。ここに厚く御礼申し上げます。