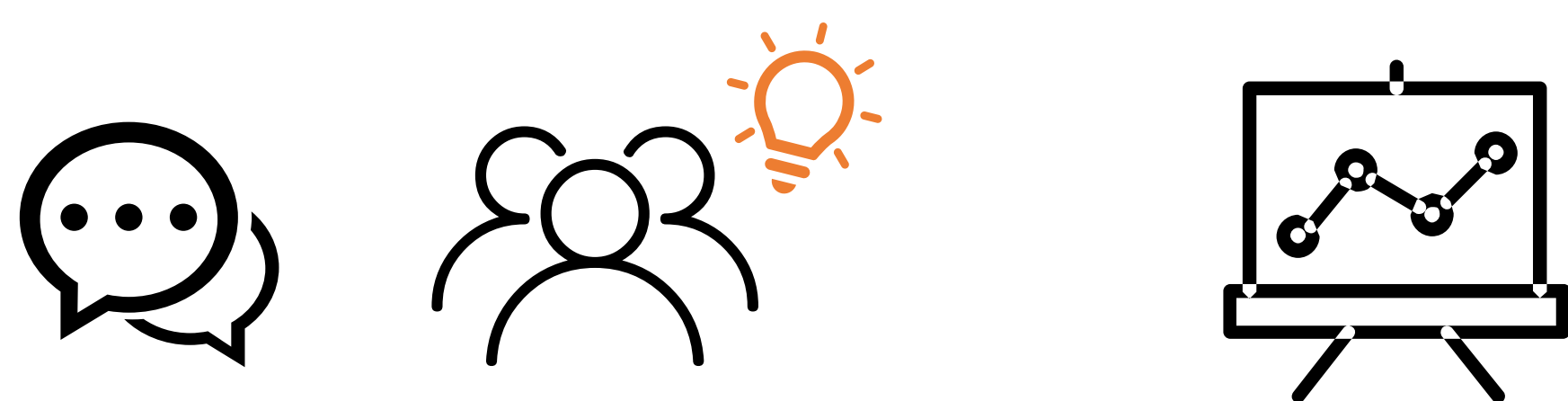


1. 研究背景

❖ ECサイトに投稿されるレビューは有用な**情報資源** [1]

例)



他のユーザの購買意思決定の補助 マーケティング分析

❖ Web上に存在するレビューは膨大であり、人手で欲しい情報を見つけ出すことは**負担**



❖ レビューを要約する際の**問題点**

• ECサイトのレビューには**商品とは関係ない内容**が含有



商品とは関係のない内容が**ノイズ**となる可能性

➤ 研究目的

商品との関連性を考慮した教師なしキーフレーズ抽出手法の提案

➤ 研究意義

商品との関連性を考慮することで、より精度の高いキーフレーズ抽出が可能

3. 提案手法

❖ 提案手法

- 商品と関連度の高いキーフレーズを抽出するために候補語と商品説明文との CrossAttention を使用

$$Key_Score = d \times PC + (1 - d) \times CS$$

d : ハイパーパラメータ

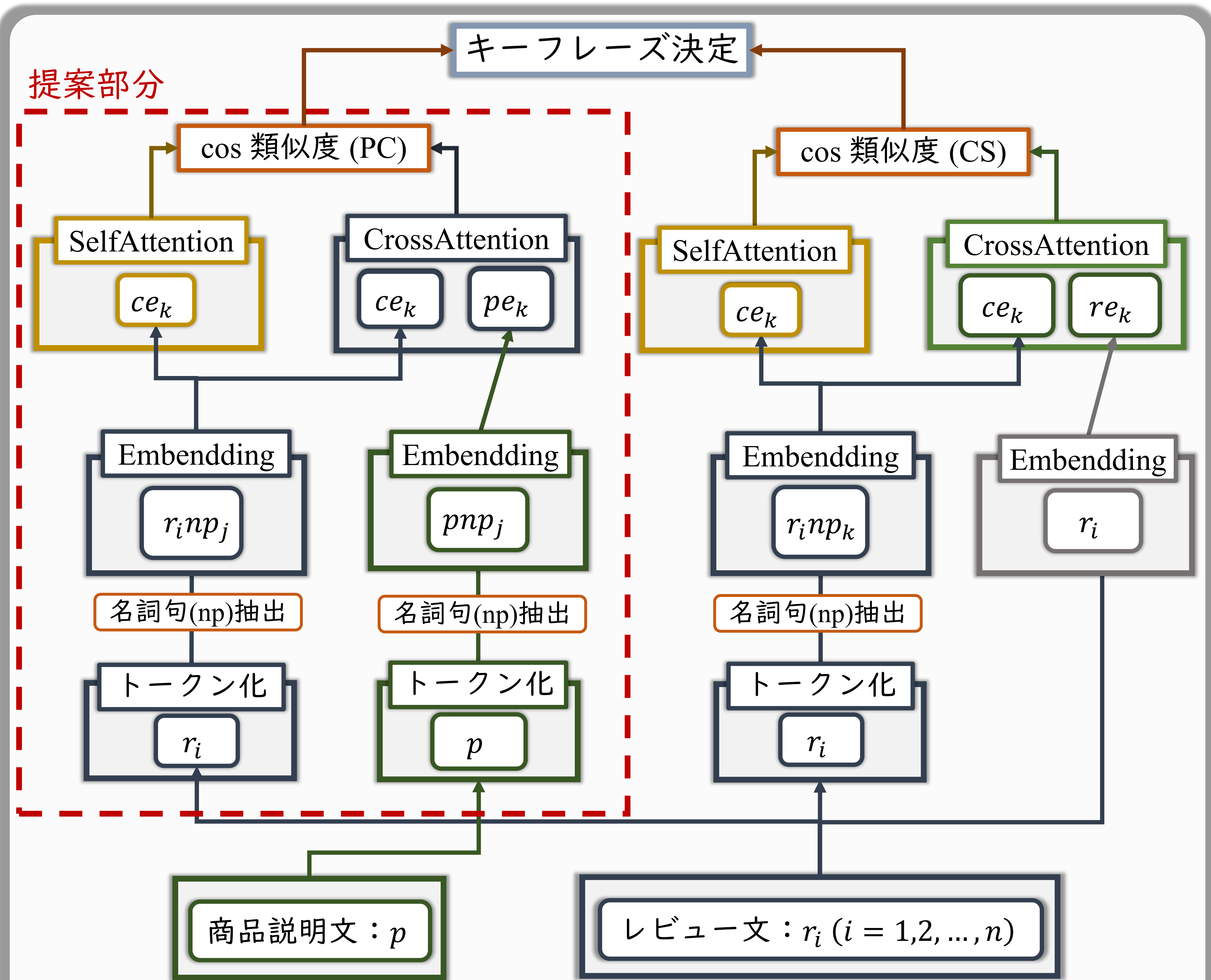


図2 提案手法のアーキテクチャ

❖ 提案手法の動作イメージ

➤ 商品レビュー

“発送元、配達会社、どちらのせいかわかりませんが、破れたダンボールで・・・商品自体の良さは知っていますが、気分が悪いのでもうここでは買いません。”

➤ 抽出した名詞句

「発送元、配達会社、破れたダンボール、中身が飛び出しそうな状態、商品自体の良さ」

商品と関係ある内容
(品質, 価格など)

“商品自体は良い”

商品と関係ない内容
(配送, 対応など)

“発送元、配達会社、どちらのせいかわかりませんが、破れたダンボールで潰れたり中身が飛び出しそうな状態で届きました。”

4. 評価実験

❖ 使用するデータセット

- 楽天市場データ (210件) [3] をランダム抽出し、人手によるアノテーションを実施

⇒ 商品説明文を読み、各レビュー文を代表してと思われる単語をキーフレーズとしてラベリング

表1 ラベリングの例

レビュー文	ラベルの例
届いてまだ雑巾しか縫ってませんが、音が静かで驚きました。・・・このミシンを使いこなせるようになりたいです。	「静かだ, ミシン」

❖ 評価方法

- 抽出されたキーフレーズがラベルに含まれているかを確認 → 再現率・適合率・F値を計算

❖ 使用したモデル

- 形態素解析・tokenizer : GiNZA [4]
- RoBERTa [5] : 早稲田大学河原研究室のモデル [6]

❖ 実験結果

表2 実験結果

比較項目		再現率	適合率	F 値
(TOP 3)	AttentionRank	0.512	0.127	0.203
	提案手法	0.536	0.133	0.213
(TOP 5)	AttentionRank	0.840	0.208	0.334
	提案手法	0.856	0.212	0.340

❖ 考察

- Top 3, Top 5 の両方で精度の改善が見られた
- 商品レビューを閲覧する際に商品と関係ある内容に注目していると推測される

[3] NII, 情報学研究所 データポジトリ 楽天データセット, <https://www.nii.ac.jp/dsc/idr/rakuten/>, アクセス日: 2022/12/1.
 [4] 松田寛 et al.; (2019) "短単位品詞の用法曖昧性解決と依存関係ラベリングの同時学習." 言語処理学会第 25 回年次大会発表論文集, pp. 201-204.
 [5] Yinhan Liu et al.; (2019) "Roberta: A robustly optimized bert pretraining approach." arXiv: 1907.11692.
 [6] nlp-waseda, roberta-base-japanese, <https://huggingface.co/nlp-waseda/roberta-large-japanese-scg512/tree/main>, アクセス日: 2022/12/1.

2. 先行研究

❖ AttentionRank [2]

- 教師なしのキーフレーズ抽出手法
- CrossAttention を用いて文章の中での関係性を考慮
候補語: レビュー文から取得した名詞句

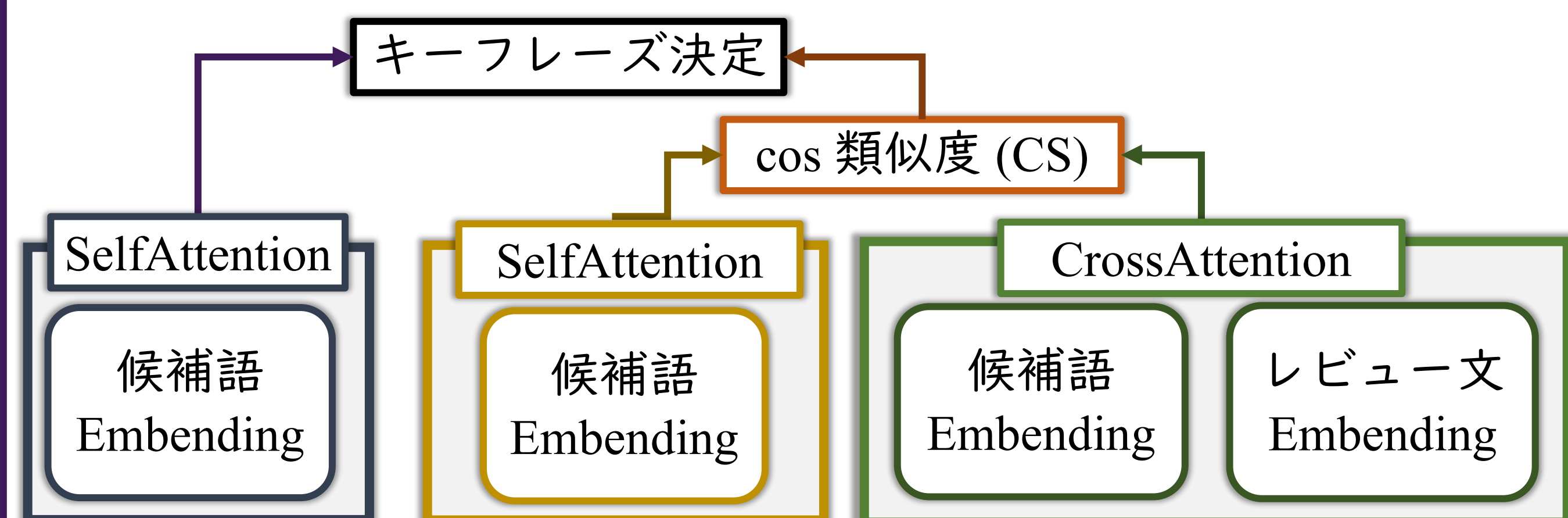


図1 AttentionRankのアーキテクチャ

5. まとめ・今後の課題

まとめ

- ECサイトにおける商品レビュー分析を行う際に商品との関連性を考慮する必要

今後の課題

- 名詞句を取る方法が精度に多く影響を与えるので、名詞句の取得方法を考案する必要

[2] Haoran Ding, Xiao Luo; (2021), "AttentionRank: Unsupervised Keyphrase Extraction using Self and Cross Attention", In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pp. 1919-1928.