

[S05] 料理レシピにおける可読性指標の検討

上田茜, 岡本一志, 軽部幸起, 原田慧, 柴田淳司

電気通信大学

はじめに

背景と目的

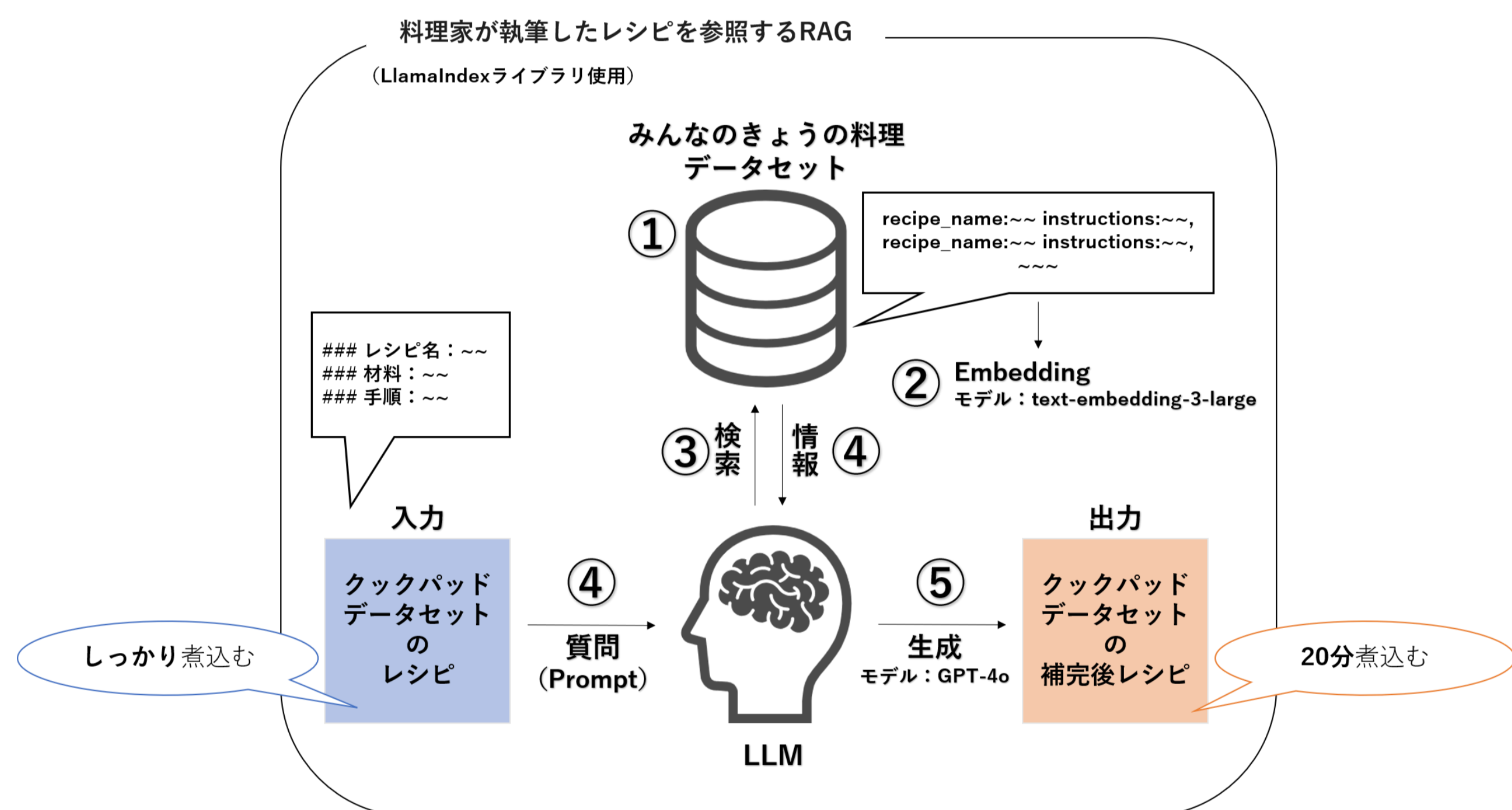
- 料理レシピ投稿サイトでは、閲覧者によって解釈が異なる**曖昧表現**が、レシピの理解度に影響
→ 曖昧表現を**自動補完**するシステムを提案し、可読性が向上 [上田+, 2025]
- レシピの可読性には、曖昧表現に限らず文章の構成など幅広い要素が影響
→ 曖昧度を包含し、レシピの**可読性を表す複合的な指標**の提案

可読性指標の構築による予想される効果

- 閲覧者：可読性の高いレシピの選択が可能
- 投稿者：可読性の高いレシピの執筆の支援が可能

補完システムの開発 [上田+, 2025]

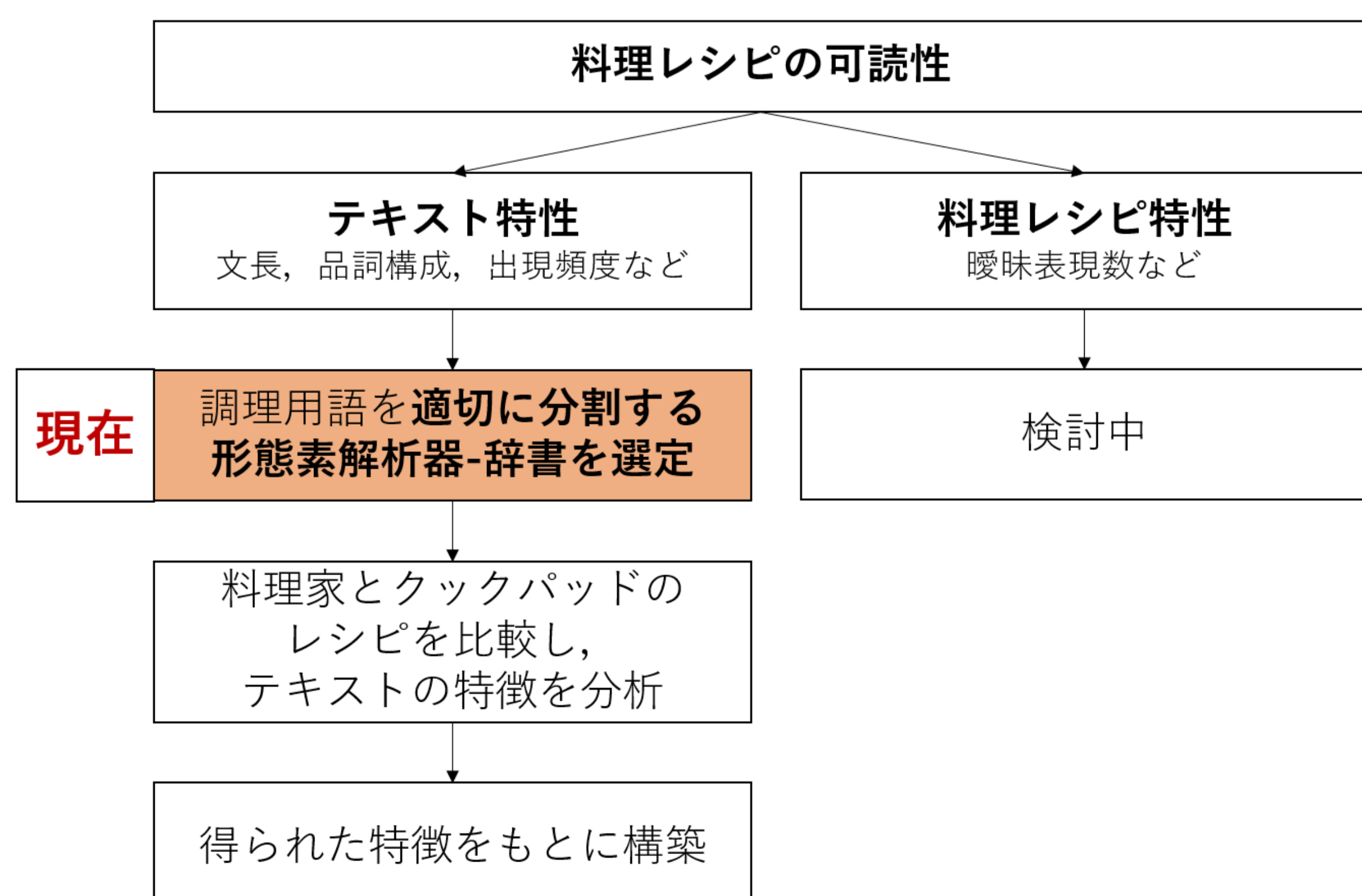
- 曖昧表現を補完するため、クックパッドのレシピに料理家が執筆したレシピを外部知識として参照する**検索拡張生成 (RAG)**を構築
- 補完後レシピの曖昧表現数は、補完前の**27%-50%に減少**



可読性指標の検討

- 料理レシピの可読性について、曖昧表現数を基準として人手評価したが、曖昧表現に限らず文章の構成など幅広い要素が影響
 - 指標を用いて可読性を自動評価したい
- 曖昧表現だけでなく、文章構成などを含めた**料理レシピの可読性を複合的に測る指標**が必要
- 本研究では、テキスト特性に関して分析するため、**調理用語を適切に分割する解析器-辞書**を選定することを目的とし、形態素解析を実施

可読性指標構築のプロセス



トークン分割の比較

- 目的：調理用語を1トークンとして分割できるか、サンプリングしたデータで調査
- 使用したデータ：クックパッドデータセットのレシピ10件
- 使用した形態素解析器と辞書：Mecab: IPadic, Unidic, NEologd / Sudachi: SudachiDict
- 全解析器-辞書にかけた結果、正しく分割されていない単語 (名詞に限定) を35件人手で抽出し、解析器-辞書別に正答単語数を整理

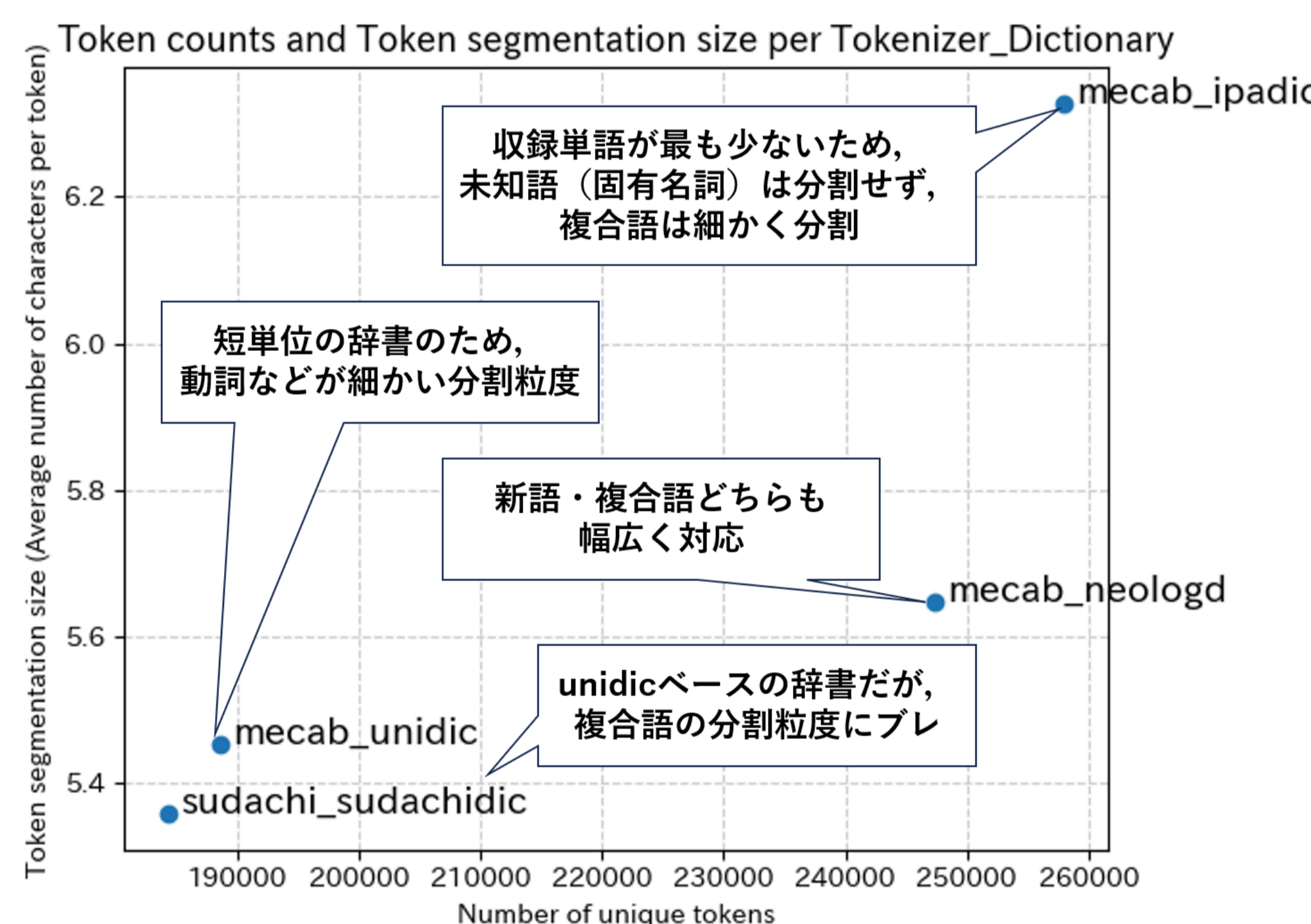
単語例	ipadic	unidic	neologd	sudachidict
粉チーズ	'粉' 'チーズ'	'粉' 'チーズ'	'粉' 'チーズ'	'粉' 'チーズ'
ほんだし	'ほん' 'だし'	'ほん' 'だし'	'ほんだし'	'ほん' 'だ' 'し'
ピザ生地	'ピザ' '生地'	'ピザ' '生地'	'ピザ' '生地'	'ピザ' '生地'

解析器-辞書	正答単語数	相対正答率
mecab_ipadic	7	0.20
mecab_unidic	7	0.20
mecab_neologd	25	0.71
sudachi_sudachidict	19	0.54

- 35単語中4単語ほどの解析器-辞書でも正答ならず
- 予備調査として、GPT-4o で35単語解析した結果、全て正答
→ LLMが上記解析器-辞書以上の性能を保有する可能性

トークン数・トークンサイズの比較

- 目的：レシピ全体で、トークン分割傾向を調査
- 使用したデータ：クックパッドデータセットのレシピ166万件
- トークン数とトークンサイズを解析器-辞書ごとに調査し、サンプルデータでの結果をもとに分割傾向を考察



おわりに

本研究のまとめ

- 解析器-辞書を比較した結果、mecab_neologd が最良
- LLMが既存の形態素解析器以上の性能を保有する可能性が示唆

今後の予定

- プロのレシピとクックパッドのレシピの品詞構成と内訳を調査
- 可読性指標の詳細設計と構築

[上田+, 2025] 上田茜, 岡本一志, 原田慧, 柴田淳司, 軽部幸起: 大規模言語モデルを用いた料理レシピの曖昧表現補完, 人工知能学会全国大会論文集, 2025.

クックパッド株式会社: クックパッドデータセット, 国立情報学研究所情報学研究データリポジトリ, 2015. [みんなのきょうの料理] NHKエデュケーショナル, "プロの料理家レシピ満載【みんなのきょうの料理】", <https://www.kyounoryouri.jp/>, 2024/11/09 アクセス.