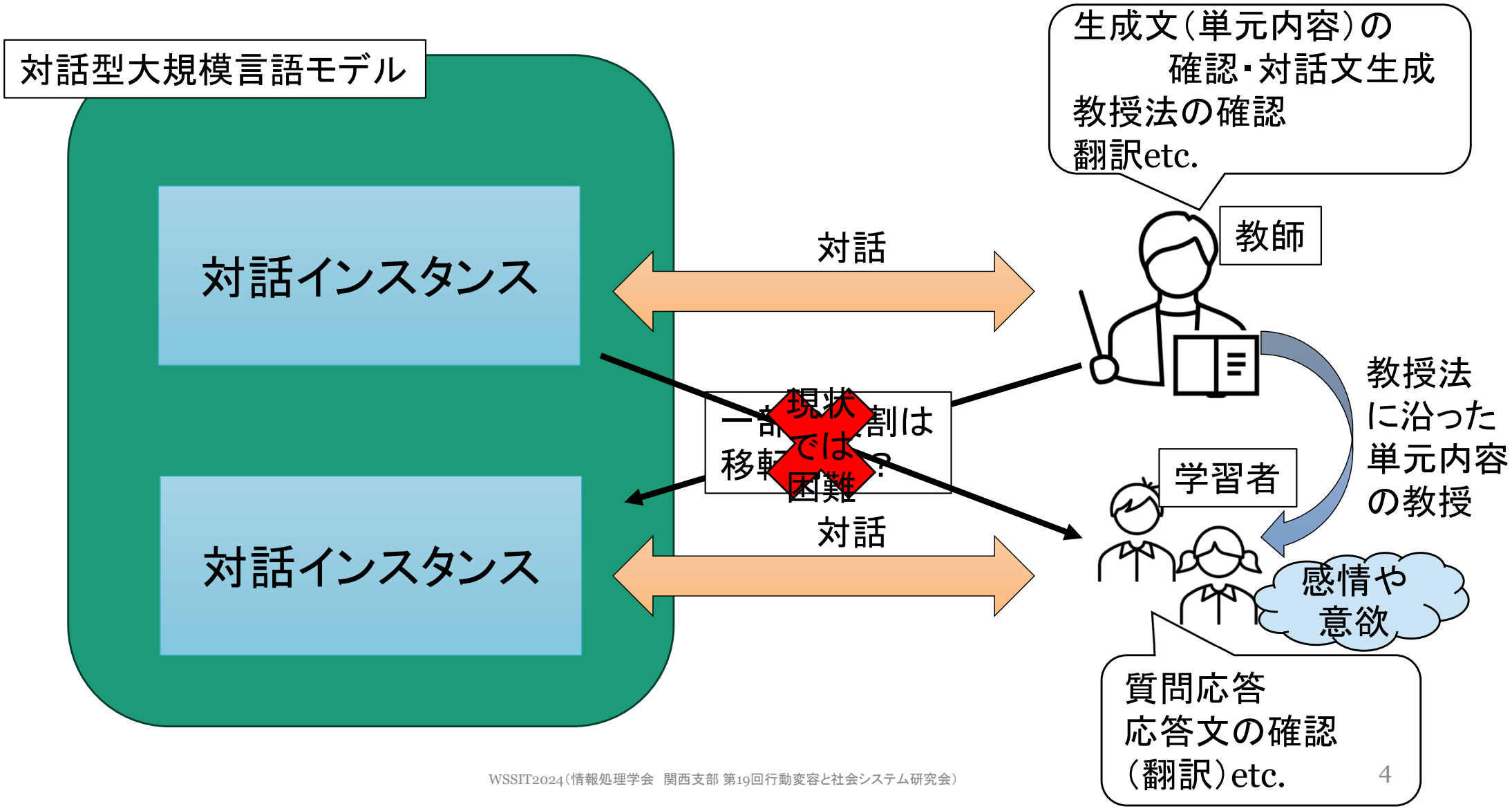
英語教育支援のための対話練習教材作成における大規模言語モデルの活用

阿部秀尚
文教大学情報学部情報システム学科
hidenao@bunkyo.ac.jp

研究背景

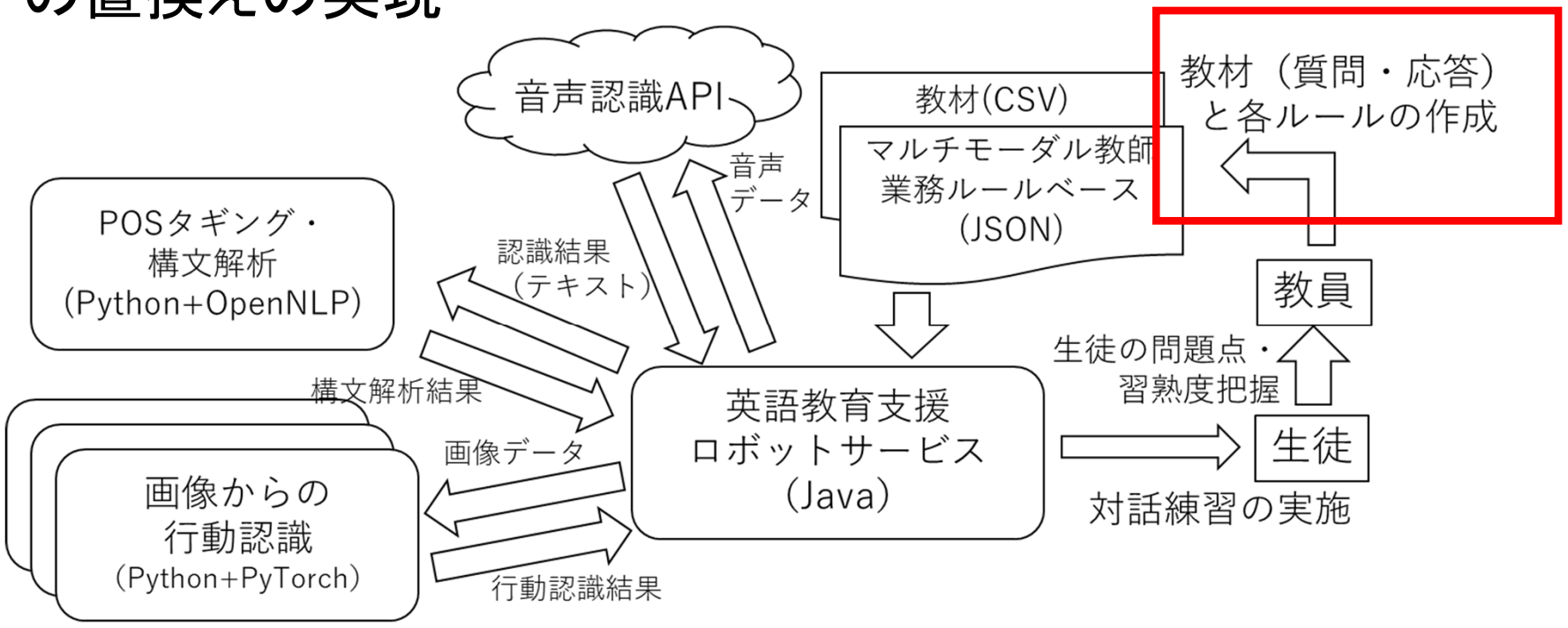
- クラウドサービスによる大規模言語資源や言語モデル利用環境の発展
  - →テキスト生成の性能が人による一般的なテキスト作成と同程度に向上
  - 言語モデルと画像生成（映像生成）を繋いだより大規模なモデルの公開と利用環境の実現
- 教師の業務知識を利用した授業支援システム（ロボットサービス＋状況把握）の開発
  - 業務知識に基づくロボットサービスの開発による個別の行動ルール適用知識（メタ知識）の明確化
  - →英語授業支援を目的としたマルチモーダルルールベースに基づくロボットサービスおよび授業支援システムの開発

対話練習における大規模言語モデルと教師・学習者の関係



研究目的

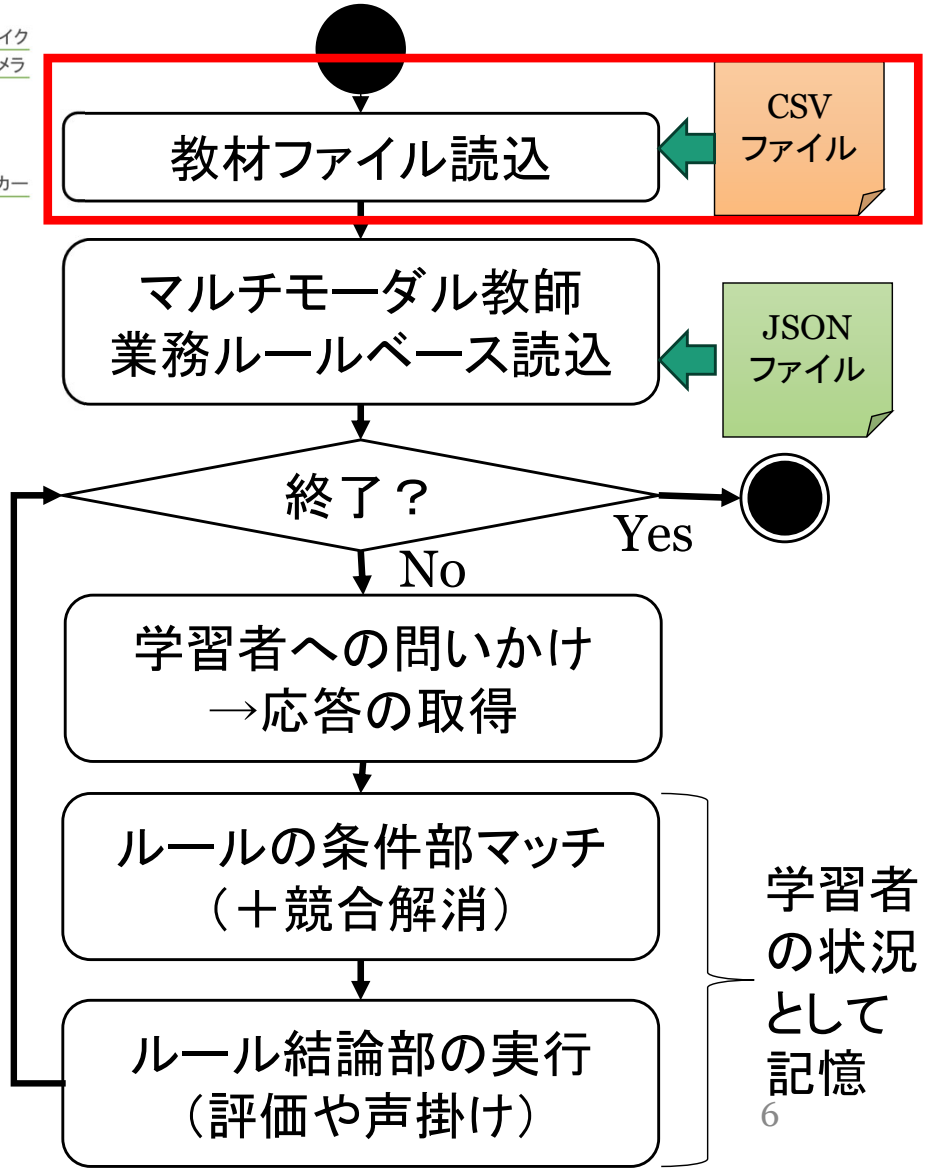
- 大規模言語モデルを利用した対話練習用教材の作成
  - 大規模言語モデルへのプロンプトによる対話練習用英文（質問・応答文）の作成の確認
  - 「中学生の学習範囲内」を担保するため，言語モデルによる単語の置換えの実現



マルチモーダル教師業務ルールベースによる英語教育支援(対話練習)ロボットサービス

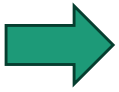
動作環境

- 対話ロボット:VStone社製Sota
  - プログラミング言語:Java
  - マイクやカメラの性能はSotaによる
- 音声認識:IBM Cloud Speech-to-Text
  - 音声認識精度や機能はSTTサービスの品質による
- 自然言語処理:Apache OpenNLP
- 画像認識:独自に作成したCNN
  - プログラミング言語:Python(PyTorch)



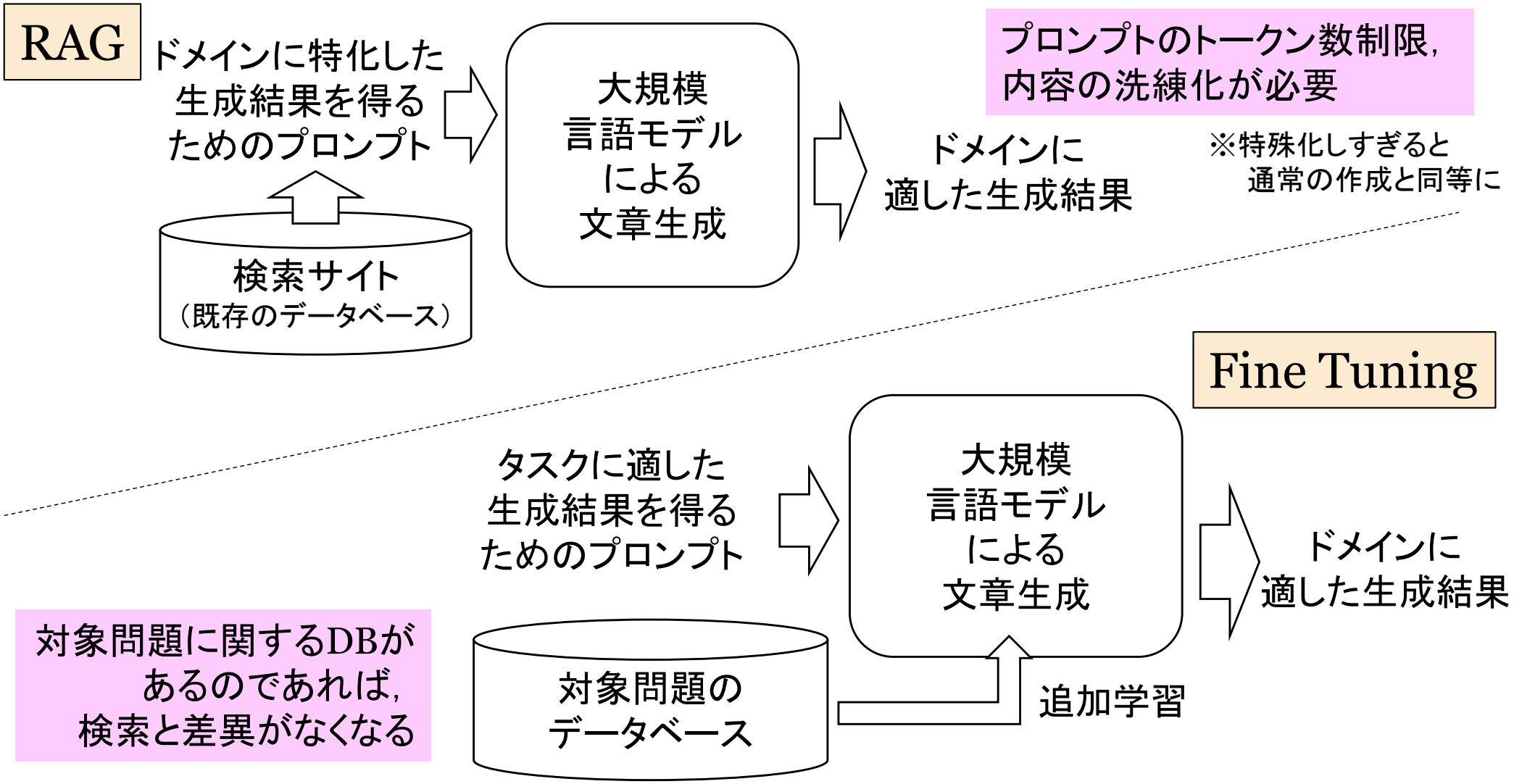
関連研究

- ロボットとの質問応答形式による学習効果の評価 [村本2023]
  - リピーティング学習（例：Audio-Lingual Method）よりも質問応答形式による学習で中期的な定着度が向上
- 大規模言語モデルにより生成される文章のドメイン特化手法
  - RAG: Retrieval Augmented Generation
  - Fine Tuning

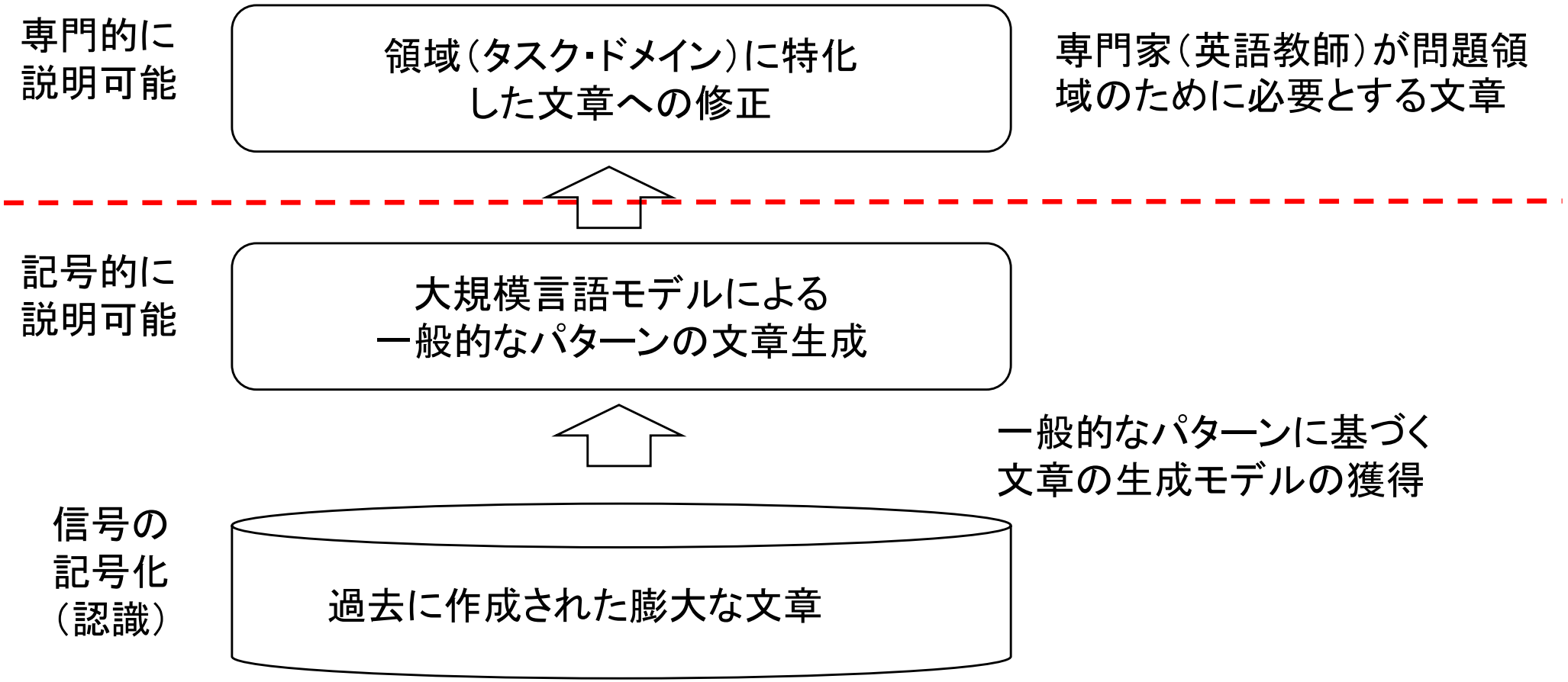


「中学生の学習範囲の単語・単元」のように膨大なテキスト量による前提を大規模言語モデルへのプロンプトに含めることは現状では困難

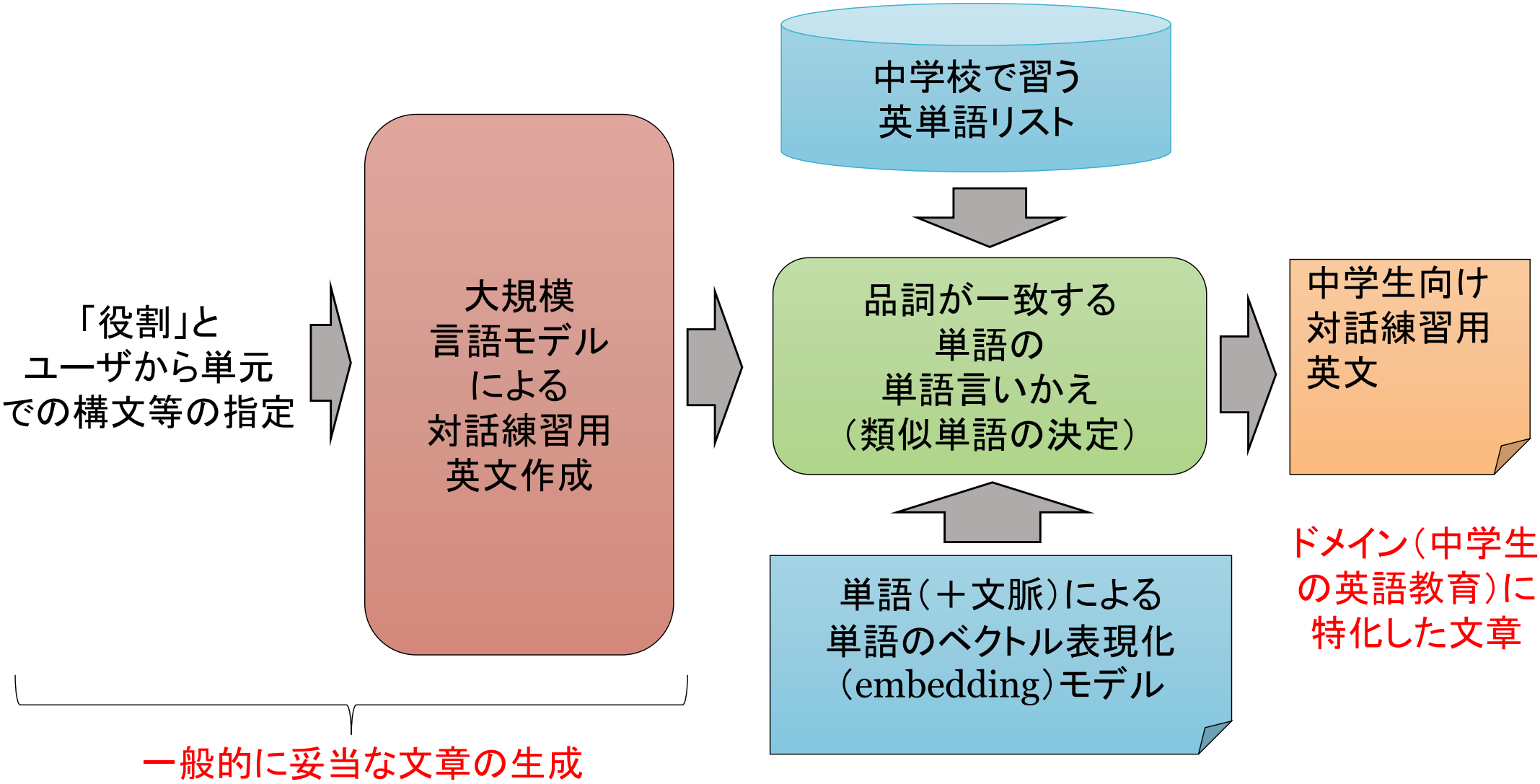
大規模言語モデルで生成される文章のドメイン（対象問題）特化手法の課題



XAIを参照した問題領域に特化したテキスト生成の概観



提案手法：
大規模言語モデルとドメイン特化型モデルの適用の統合による中学生向け対話練習教材の生成手法



日本の中学校で習う英単語

- 学習指導要領内では単元を指定
  →検定済み教科書に反映→単語リスト化
- 青森県版中学校英単語集 VERSION V

品詞	単語数
名詞	1,254
代名詞	60
動詞	343
助動詞	19
形容詞	361
副詞	148
前置詞	42
接続詞	42
冠詞	3
間投詞	14

言い換え対象の英単語の置換

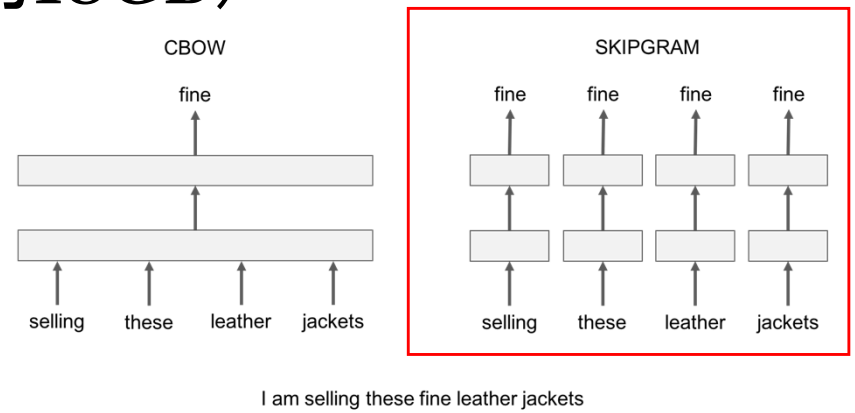
- 前提：英単語集に無い英単語は用いることができない
  - 「知らない単語」の利用は学習者の意欲低下につながる
  - （上級校への進学のための）入試ではアウト
- 適切な単語への置換手法
  - 置換対象に最も類似した英単語を見つける
    →fastTextによるSkipGramモデルにより単語のベクトル分散表現化

単語ベクトル$\boldsymbol{a}, \boldsymbol{b}$間の類似度：（コサイン類似度）

$$\cos(\boldsymbol{a}, \boldsymbol{b}) = \frac{\boldsymbol{a} \cdot \boldsymbol{b}}{\|\boldsymbol{a}\|\|\boldsymbol{b}\|}$$

実装：
fastTextによるSkipGramモデルの作成

- 利用言語資源
  - 英語版Wikipedia2023年12月20日版ダンプ
    →全ページ記事（約1億6000万文；約18GB）
- 実行設定
  - Python＋fasttextライブラリ
  - fastTextにおけるSkipGramを利用
- 得られたモデル
  - 利用言語資源に対するSkipGramモデル（約14GB）
    →単語ごとの300次元のベクトル分散表現を獲得

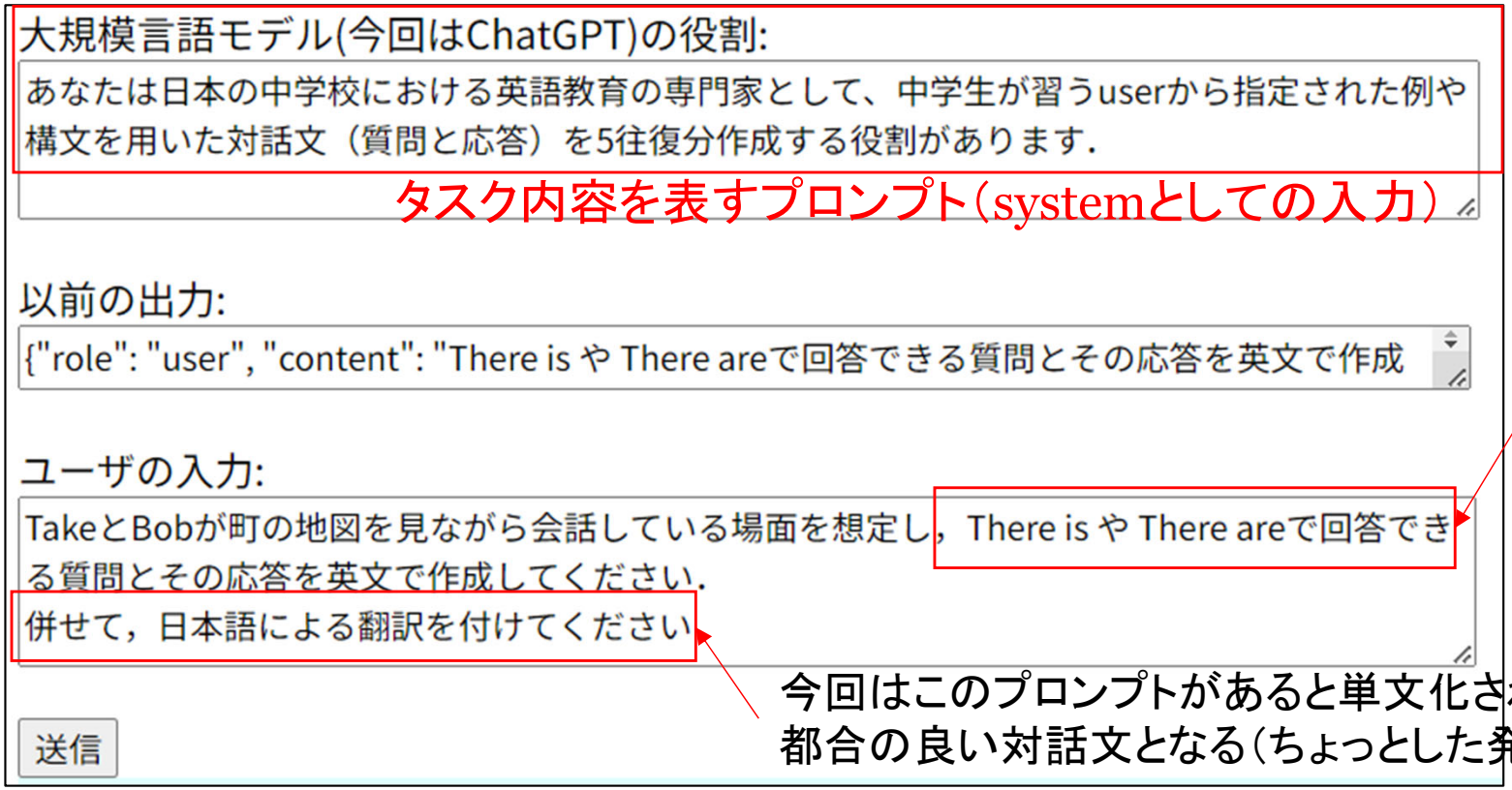


Python+FlaskによりWebAPI化したスクリプトにより実行

画像出典：https://fasttext.cc/docs/en/unsupervised-tutorial.html

大規模言語モデルによる英語対話練習用英文の作成（実行例）

- OpenAI APIからgpt-3.5-turboモデルを指定
  - 実装：HTML/JavaScriptによるインタフェース→PHPからcurlモジュールを通じてアクセス



中学校英単語集以外の単語の言い換え（実行例）

- 言い換え前の結果

ChatGPTの応答:
Takeshi: What can we find in the east side of the park according to the map?
（地図によれば、公園の東側に何がありますか？）

Bob: There are some tennis courts and a playground in the east side of the park.
（公園の東側にはテニスコートと遊び場があります。）

Takeshi: Is there a café near the museum?
（美術館の近くにカフェはありますか？）

Bob: Yes, there is a café just across the street from the museum.
（はい、美術館の向かいにカフェがあります。）

Takeshi: Do you know if there is a subway station close to the shopping mall?
（ショッピングモールの近くに地下鉄駅はあるか知っていますか？）

Bob: There is a subway station about a five-minute walk from the shopping mall.
（ショッピングモールから徒歩5分ほどのところに地下鉄駅があります。）

Takeshi: Could you tell me if there are any bus stops in front of the station?
（駅の前にバス停があるか教えてもらえますか？）

Bob: The
（駅の真

- 言い換え後の結果

単語置換チェック後の結果:
Takeshi: What can we find in the east side of the park although to the map?
（地図によれば、公園の東側に何がありますか？）

Bob: There are some tennis courts and a picnic in the east side of the park.
（公園の東側にはテニスコートと遊び場があります。）

Takeshi: Is there a restaurant near the museum?
（美術館の近くにカフェはありますか？）

Bob: Yes, there is a restaurant just across the street from the museum.
（はい、美術館の向かいにカフェがあります。）

Takeshi: Do you know if there is a subway station close to the shopping shopping?
（ショッピングモールの近くに地下鉄駅はあるか知っていますか？）

Bob: There is a subway station about a hour walk from the shopping shopping.
（ショッピングモールから徒歩5分ほどのところに地下鉄駅があります。）

Takeshi: Could you tell me if there are any bus stops in front of the station?
（駅の前にバス停があるか教えてもらえますか？）

問題の無い言い換え：café→restaurant, five-minute→hour
適切と言えない言い換え：according→although, playground→picnic, mall→shopping

考察

- fastText+SkipGramモデルによる言い換えが可能であることを確認
  - 複合語への対処や置換候補の提示に留めるなどの改善が必要
- 言い換え後の英文の“自然さ”の改善
  - 最も類似する単語への言い換えにより不自然な英文になる
    →より自然な言い換え文に
  - 英文自体をベクトル分散表現化するBERTや大規模言語モデルによるEmbeddingの実施
    →元の文に（感覚的に）最も近い言い換え文の作成
- 視覚的な教材の同時作成
  - 画像生成モデルによる動的な画像生成

おわりに

- 大規模言語モデルにより生成された文章のドメイン特化の文章への修正手法の提案
  - 一般的なパターンに基づく文章
    →ドメイン特化型の言語モデルにより修正が可能

- 今後の課題
  - “自然な”修正後英文の選定手法の導入と専門家による提示内容の質的評価
  - 英語教育支援システムへの統合
    - 教師業務支援入力インタフェースとの連携
    - ロボットサービスとの連携