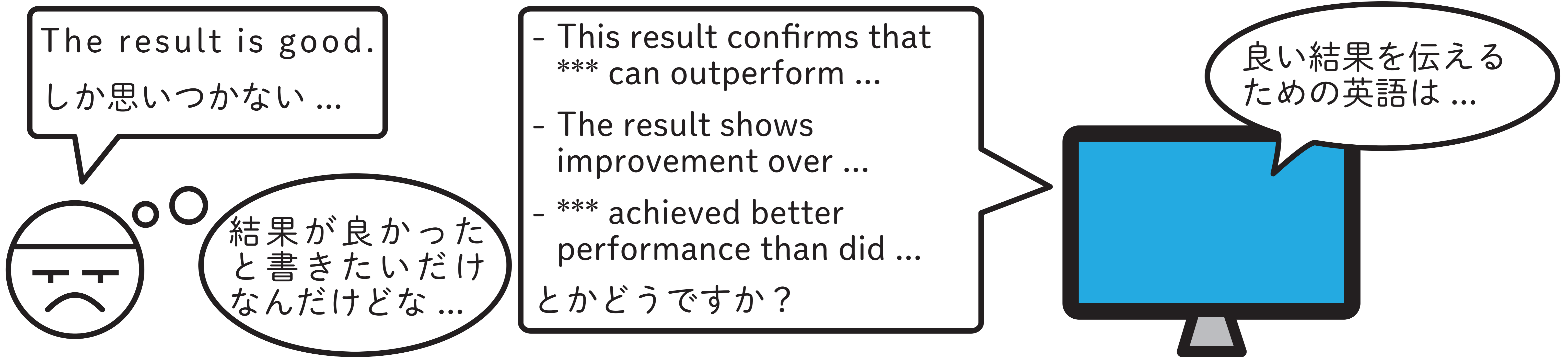


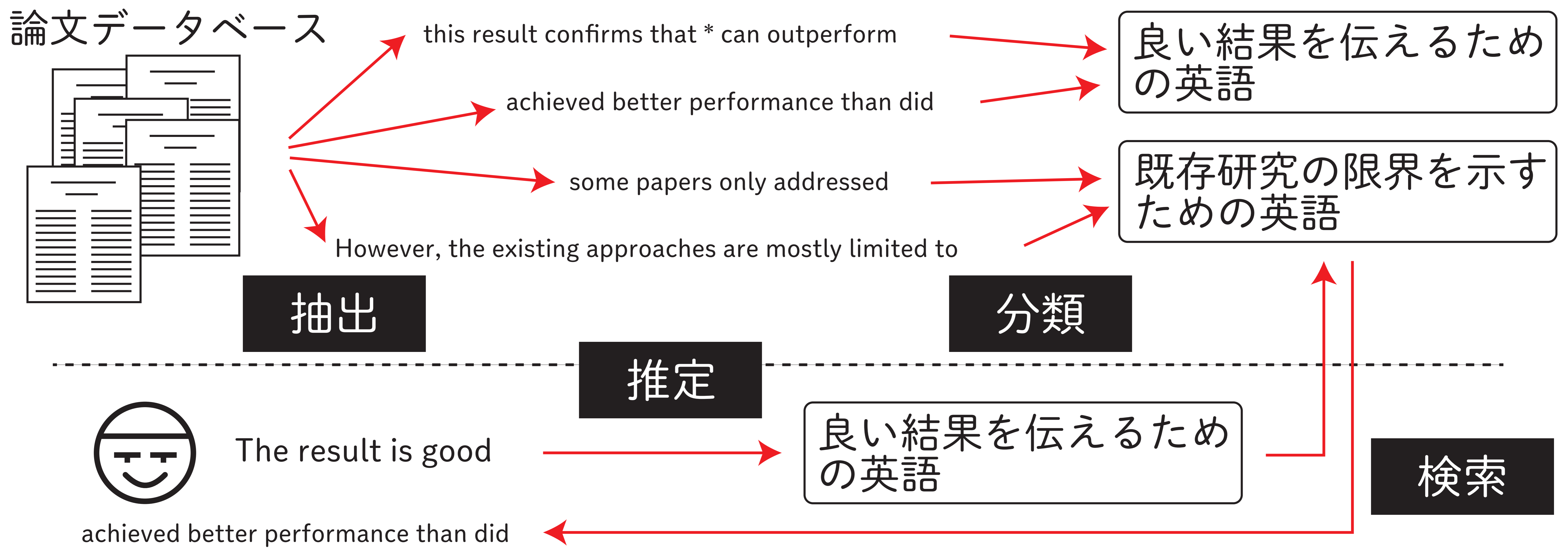
知らない英語表現を検索するには？ 英語論文執筆支援システム

岩月憲一（東京大学・博士後期課程）・相澤彰子（国立情報学研究所・教授）

どんな研究？



研究のゴール



これまでの研究成果

定型表現（よく使われる連語・構文）と、修辞構造に基づく機能（文の役割）を対応させたデータセットの構築に取り組んでいます。

① 定型表現の抽出

目的 英文を定型部分と非定型部分に分ける
 タスク 系列ラベリング問題として解く
 データ Academic Phrasebank という既存のリソースに対し、人手でアノテーションしたものを教師データとして利用
 抽出対象は、論文コーパス (ACL Anthology と PMC)
 手法 条件付き確率場 (CRF)
 結果 人手で抽出した定型表現を評価
 正解率：54% (ACL)；67% (PMC)
頻度のみ：14% (ACL)；14% (PMC)・MI：15% (ACL)；12% (PMC)

② 定型表現の分類

目的 定型表現に機能ラベルを付与する
 タスク 分類問題として解く
 データ Academic Phrasebank に人手でアノテーションした定型表現
 手法 入力に文ベクトル (単語 skip-gram + 内容語削除) + セクション情報
 分類器はランダムフォレスト
 結果 F 値：41.3% (ACL)；70.6% (PMC)
入力に定型表現のみの場合：24.1% (ACL)；32.0% (PMC)
 入力に例文コーパスを使わない場合：2.98% (ACL)；2.09% (PMC)

③ 定型表現機能の推定

目的 不完全な英語で検索する
 タスク 分類問題として解く
 データ ②と同じ
 手法 1. 定型表現部分不使用 + ②
 2. ランダムに単語を落とす + ②
 結果 1. の場合 (F 値)
 27.3% (ACL)；20.0% (PMC)
 2. の場合 (F 値)
 54.5% (ACL)；45.8% (PMC)

発表文献：Iwatsuki, K., & Aizawa, A. (2018). Using Formulaic Expressions in Writing Assistance Systems. In Proceedings of the 27th International Conference on Computational Linguistics, 2678-2689. 論文 URI: <https://www.aclweb.org/anthology/papers/C18/C18-1227/>



連絡先：岩月憲一 / 国立情報学研究所 相澤研究室
 Email: iwatsuki@nii.ac.jp

