



ねえAI この言葉の意味、わかる？

パラフレーズ現象の解明のための言語資源構築と
パラフレーズアラインメント技術の確立
荒瀬 由紀 arase@ist.osaka-u.ac.jp

パラフレーズ現象を解明し、AIによる意味理解を目指す 理解に基づく対話システム・質問応答・検索技術の実現に貢献

パラフレーズって何？

同じ意味内容だけど表面上は異なる言語表現だよ！
こういうのだね

The research group developed the antivirus vaccine.
The scientific team created a vaccine against the virus.

なぜ重要な？

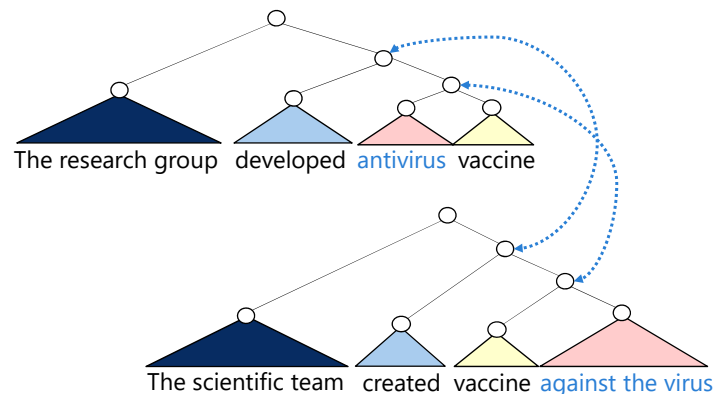
言語の意味構成メカニズムを解明する手がかりなんだよ！

それはすごい！
応用ではどう役立つの？

すばり言語を数値表現化する
深層学習でも有効なんだ！

世界最大！

パラフレーズ対応付けコーパスを構築



世界最高性能！

フレーズベクトル化技術の開発

Transfer fine-tuning [1]

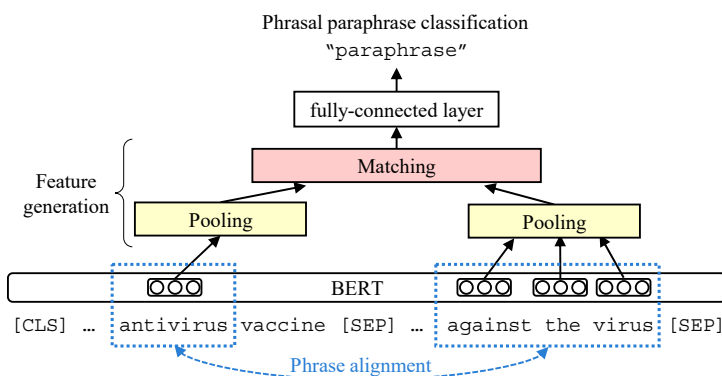
パラフレーズの学習でBERT
をパワーアップ！

Transfer fine-tuning on BERT

- Pre-trained BERT でパラフレーズ認識を学習 [1]
- パラフレーズ認識だけでなく、種々の文対モデリングタスクで高い性能を実現
- モデルサイズは維持したまま、小さな Fine-tuning データセットで大きな性能改善

Extended Syntactic Phrase Alignment Dataset (ESPADA)

- 言語学者による HPSG 構文木アノテーションと 3名によるフレーズ対応アノテーション
- パラフレーズの系統的分析が可能に！
- 近日公開予定



文対数	1,916
語彙サイズ	9,540
フレーズ総数	152,480
フレーズ対応の総数	251,972
2名以上が合意したフレーズ対応の総数	80,572
Non-compositional なパラフレーズの割合	7.93 %

