



あなたの言葉、計算します。

マルチモダリティを可能にする テキストのベクトル化技術の創出

荒瀬 由紀 (大阪大学大学院情報科学研究科)
arase@ist.osaka-u.ac.jp

フレーズを表現するベクトルを生成 言語を「計算」できる世界へ

- ユーザの発言の意味を理解する対話システム
- 画像、センサデータとテキスト処理との融合

アプローチ

- 既存研究は単語の出現文脈に基づいてベクトル生成
「爽やかな天気」と「どんよりした空模様」が近い意味？
- 同一の意味を異なる言語表現で表すパラフレーズを考慮することで意味的な近さ・遠さを再現するベクトル生成を実現

Relying on team spirit, expedition members defeated difficulties.
 <-> *Members of the scientific team overcame challenges through teamwork.*

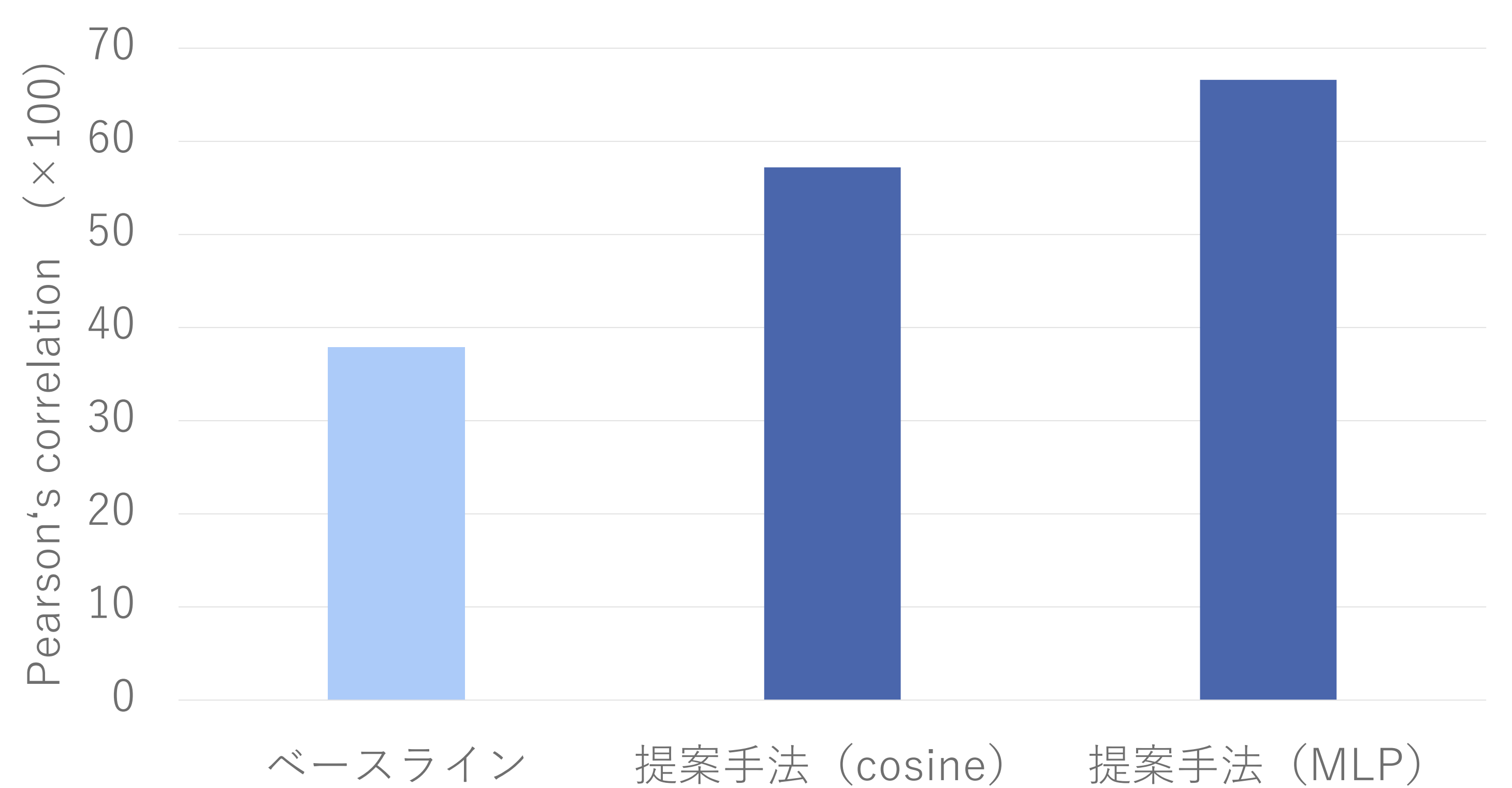
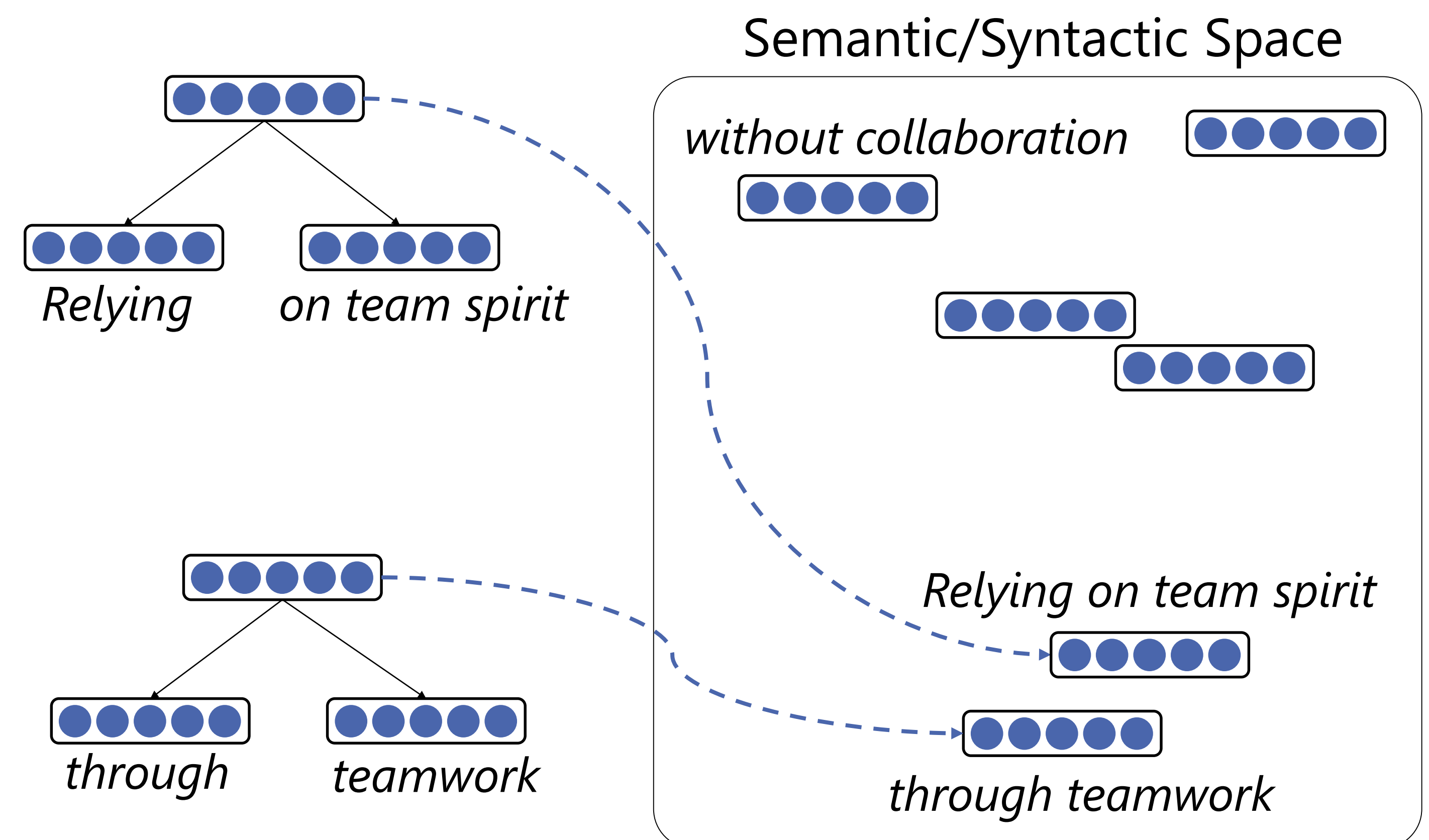
Semantic Textual Similarity (STS) Benchmark による評価実験

- 文間類似度を6段階でラベル付け
- 提案手法
 - フレーズアラインメントを行った35kのパラフレーズ文ペアによるモデル学習
 - コサイン類似度による類似度推定
 - Multi-layer perceptron (MLP) による類似度推定
 - STS Benchmarkの学習データを使用
- ベースライン：単語ベクトル[5]の平均を計算し、コサイン類似度で類似度推定

類似度推定性能をそれぞれ **51%, 76%** 向上

パラフレーズによる制約付きベクトル生成

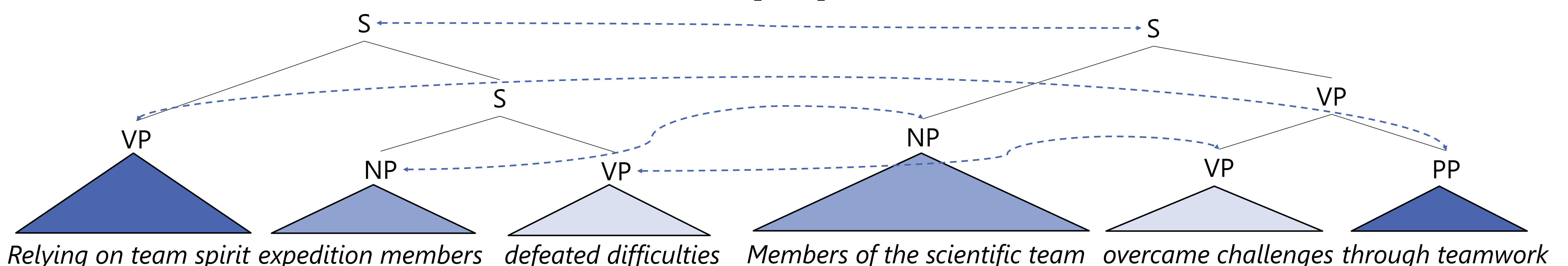
- 構文構造に基づくベクトル化 [4]
- 文単位のパラフレーズからフレーズ対応を特定
- パラフレーズのベクトル間類似度が高くなるように制約を付加



加速フェーズへの展開

- 多様なパラフレーズ現象を扱えるよう、フレーズアラインメント手法の改善
 - アノテーションデータの構築
 - フレーズベクトル生成モデルの学習データの拡大
- 構文構造の情報をより積極的に利用するモデルの設計
- フレーズアラインメントによるベクトル品質の評価

パラフレーズにおけるフレーズアラインメント [1-3]



参考文献

- [1] Y. Arase and J. Tsujii. 2017. Monolingual Phrase Alignment on Parse Forests, in Proc. of EMNLP, pp. 1-11.
- [2] Y. Arase and J. Tsujii. 2018. SPADE: Evaluation Dataset for Monolingual Phrase Alignment, in Proc. of LREC (to appear).
- [3] SPADE: <https://catalog.ldc.upenn.edu/LDC2018T09>
- [4] K. S. Tai et al. 2015. Improved Semantic Representations from Tree-Structured Long Short-Term Memory Networks, in Proc. of ACL, pp. 1556-1566.
- [5] J. Pennington, R. Socher, and C. D. Manning. 2014. GloVe: Global Vectors for Word Representation, in Proc. of EMNLP, pp. 1532-1543.