

研究スタッフ

教 授：乾 健太郎

助 教：井之上 直也、折田 奈甫(兼任)

特任助教：松林 優一郎、田 然、水本 智也

研究員：高橋 容市、松田 耕史、山口 健史

研究目的

情報伝達のための最も重要なメディアは、日本語や英語など、誰もが日常で使っている言語（自然言語）です。本研究室では、自然言語で表現され、伝達され、蓄積される情報や人の知識をコンピュータで処理するための基礎理論、基盤技術、応用技術に関する研究を行っています。

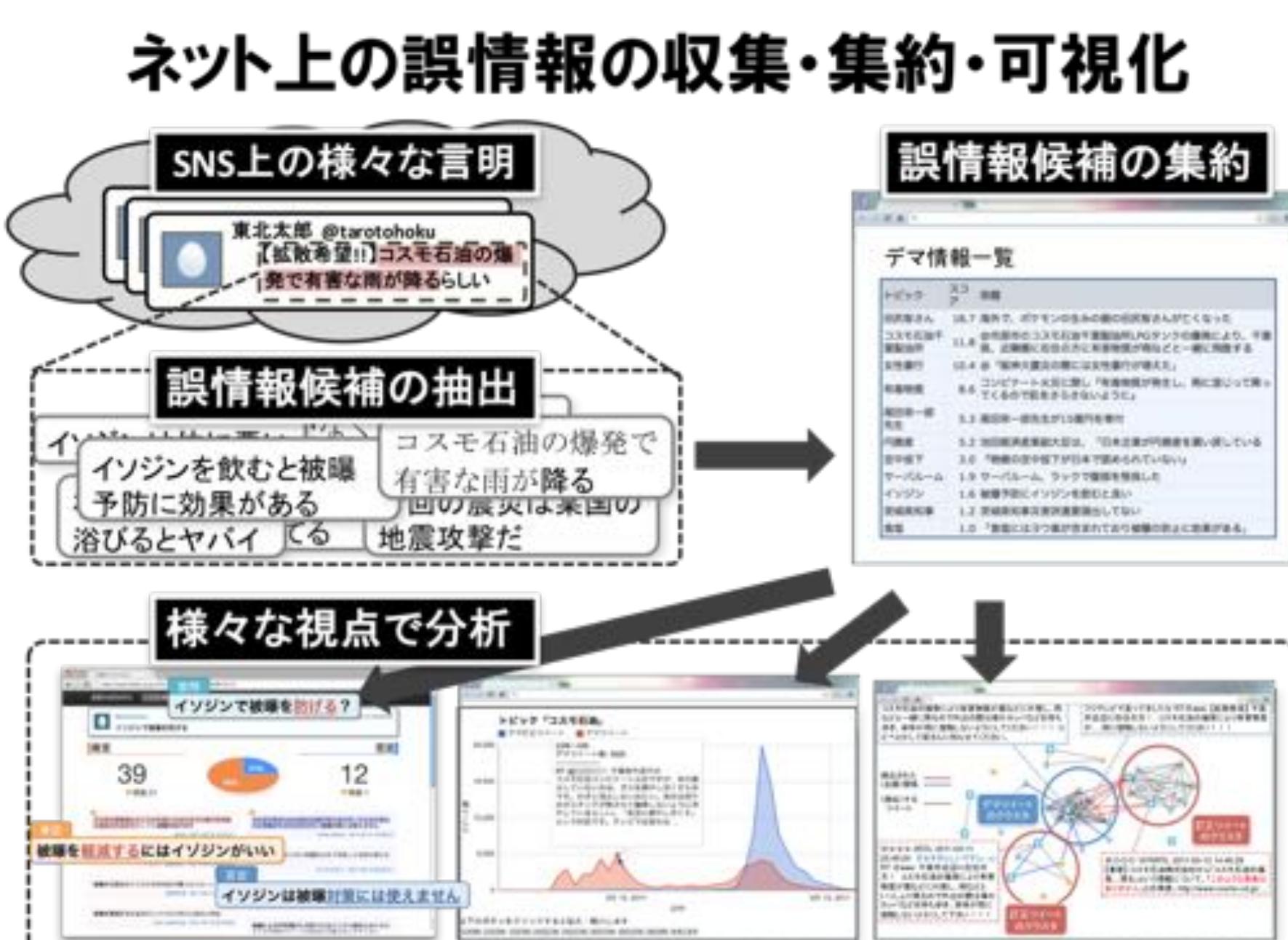
東北大 × 産学・学際連携



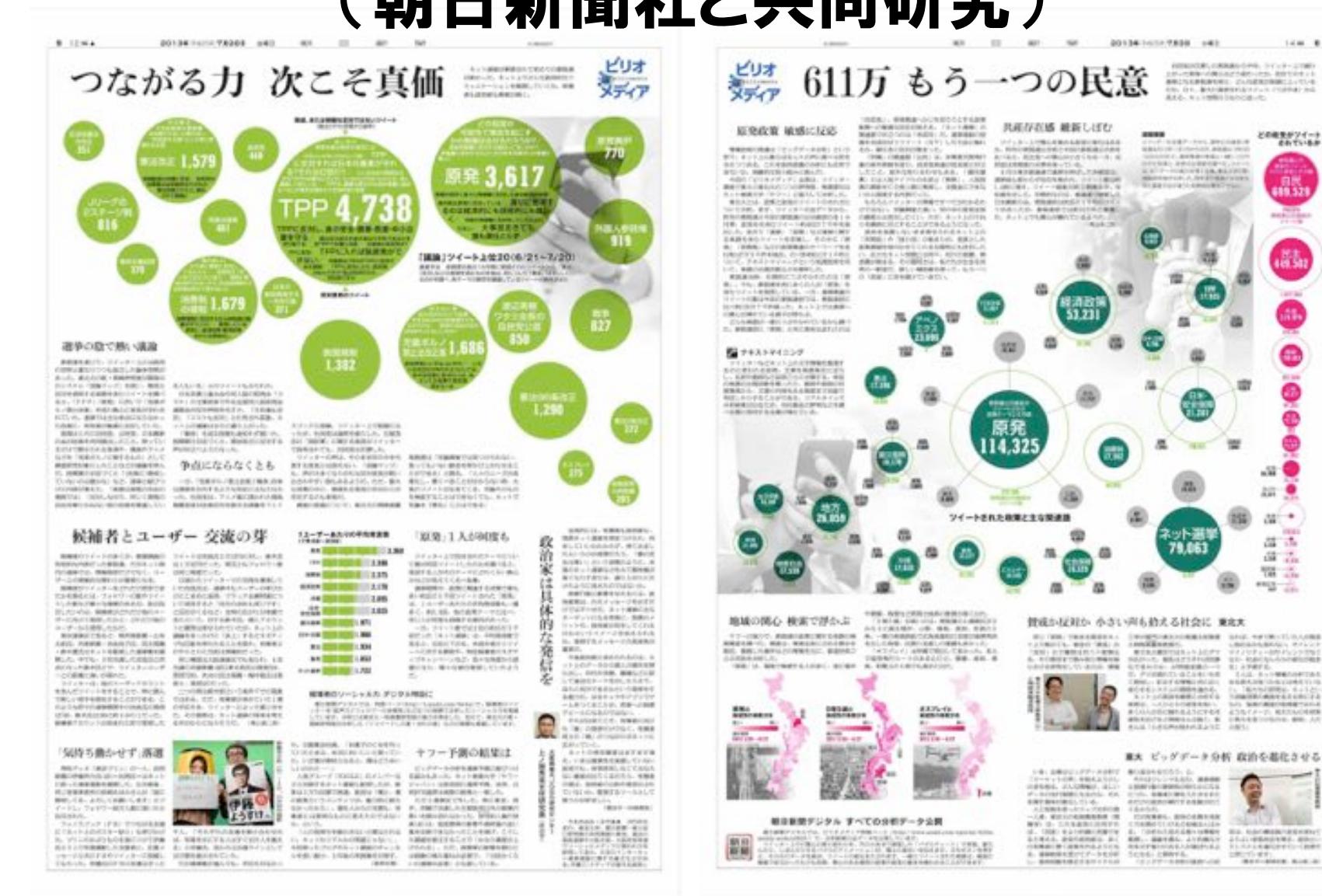
主な研究テーマ

1. Web情報分析

Webには多様な情報があり、検索エンジンによって容易に入手できるようになりました。しかし、一つの情報をただちに信じることは危険な場合があります。

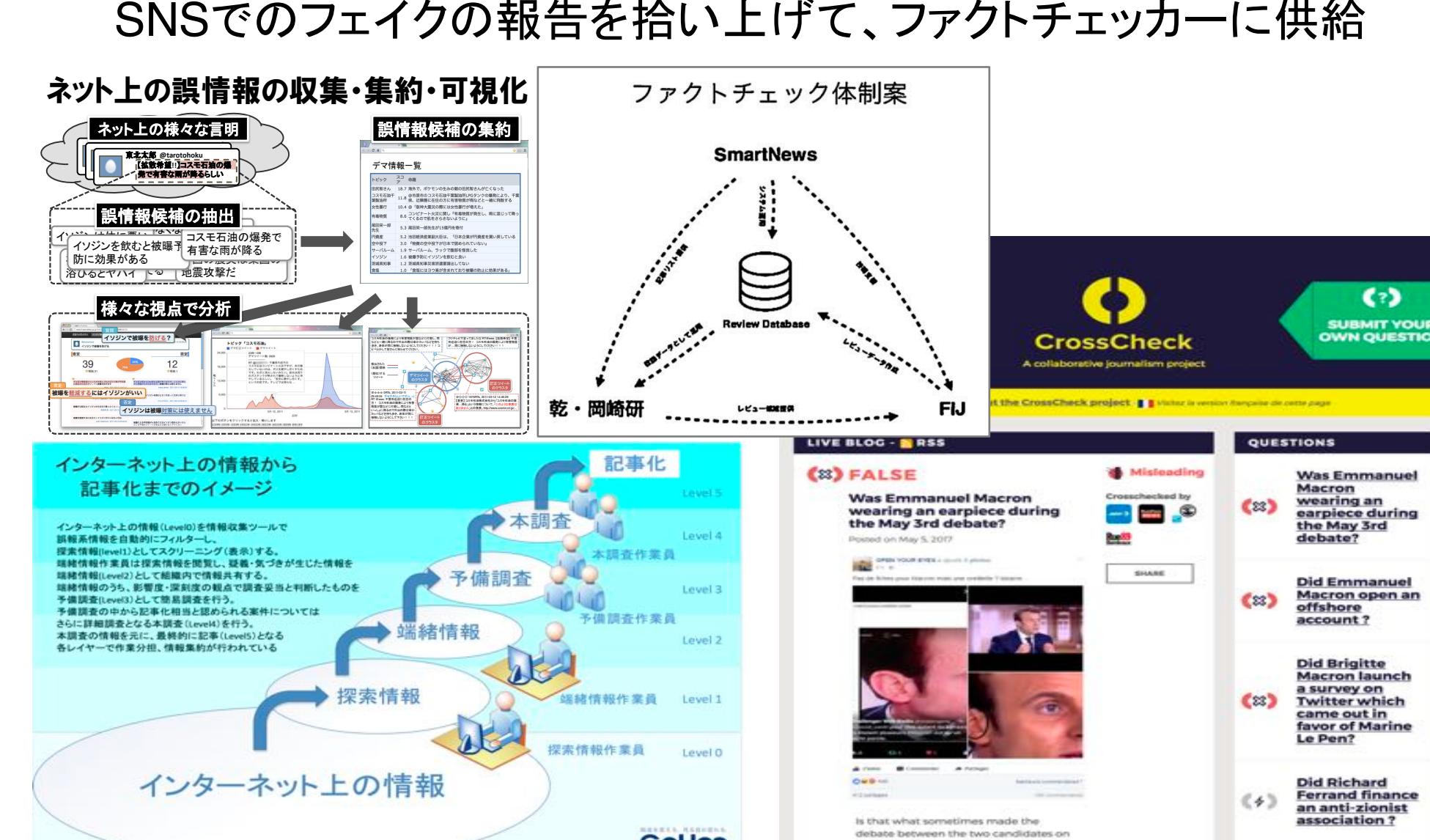


参議院選挙に関するツイートデータの分析 (朝日新聞社と共同研究)



情報の信頼性、ファクトチェック支援

世界中でファクトチェック体制の勃興
SNSでのフェイクの報告を拾い上げて、ファクトチェッカーに供給



記事単位でツイートをひとまとめにする

○正例

TwitterBot リツイート

×負例

性的表現
アフィリエイト

フェイクニュースの端緒情報をTwitterから抽出する作業を支援

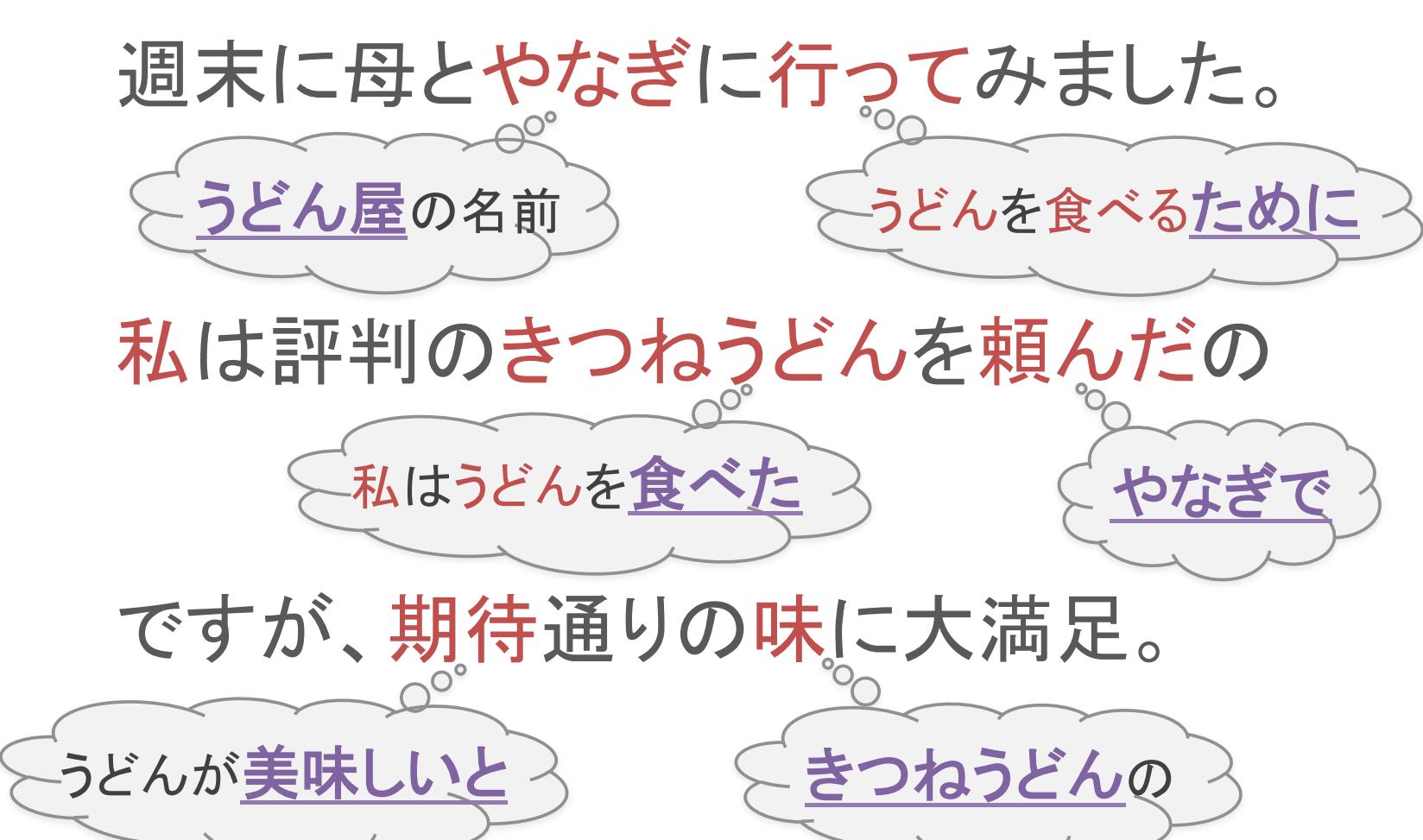
学習器で未知データをランキング付け

人手で検証

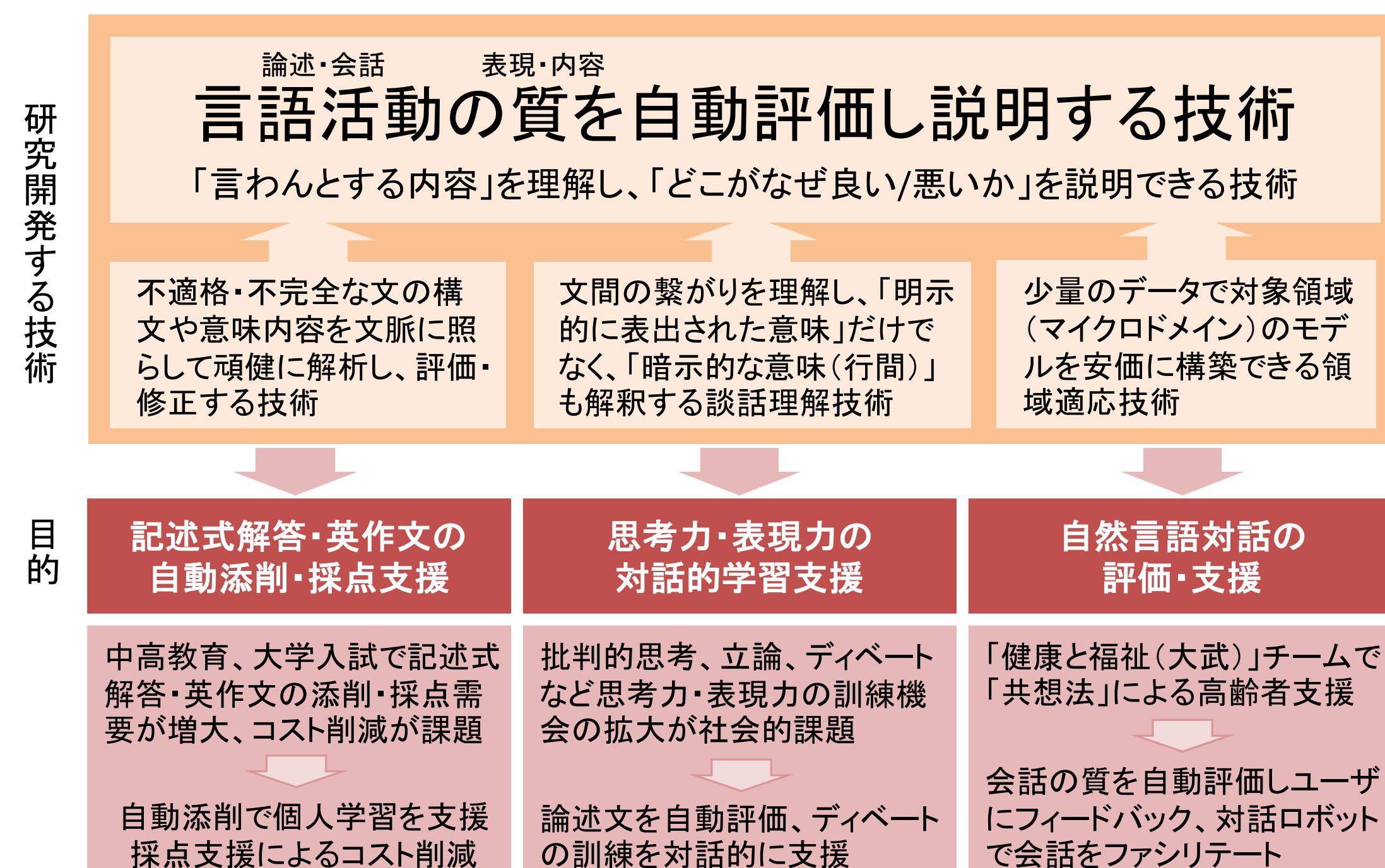
2. 「行間を読む」言語理解への挑戦

言葉を理解するには、言葉が伝える情景をイメージしたり、常識的知識を使いこなして何がなぜ起こったのかを推論したりと、高度な知能が求められます。

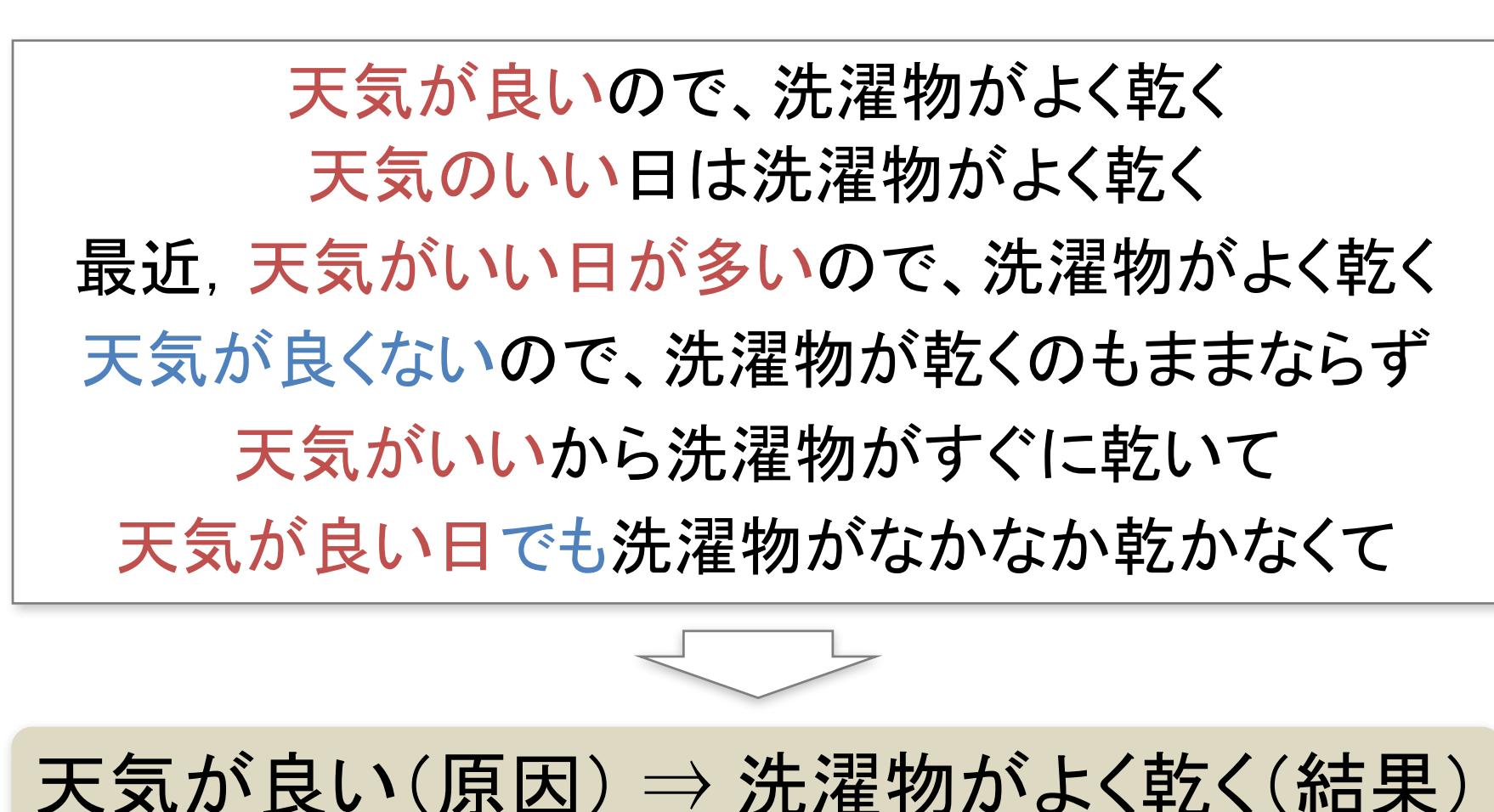
「行間を読む」



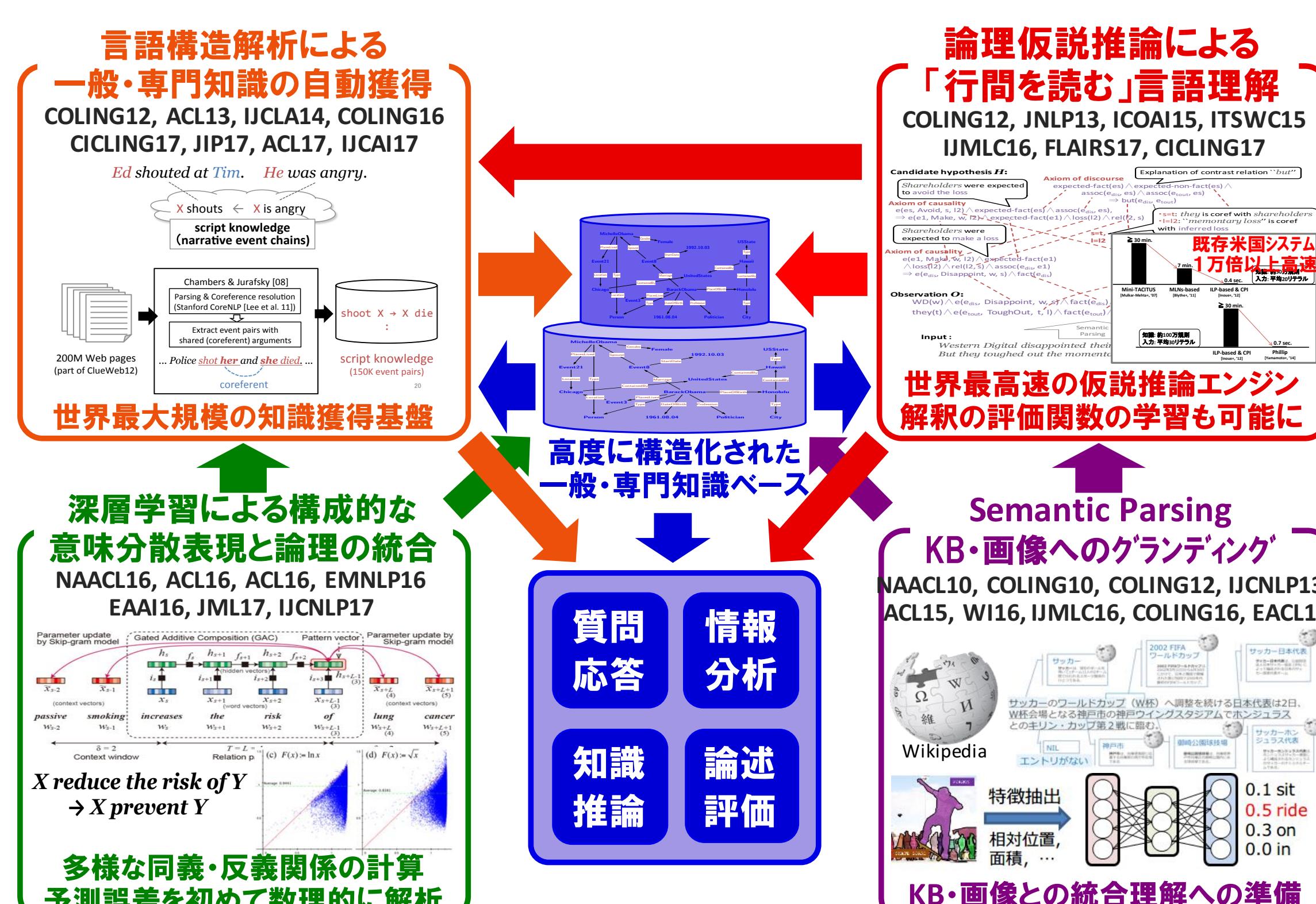
自然言語理解チーム 研究計画の概要



常識もビッグデータから集める



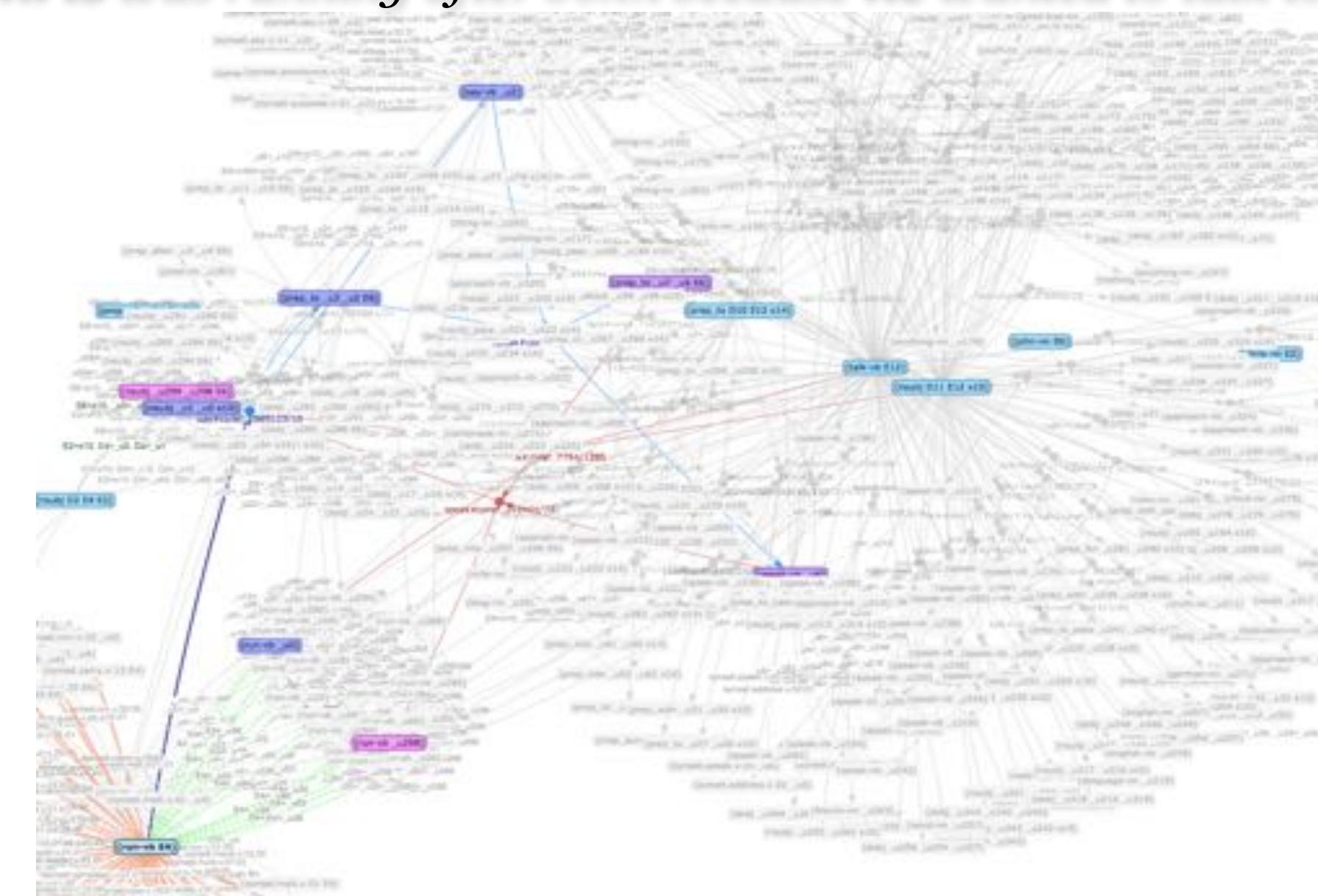
理研AIPセンター 言語情報アクセス技術チーム & 自然言語理解チーム



東北大で世界最高速の仮説推論システムを開発

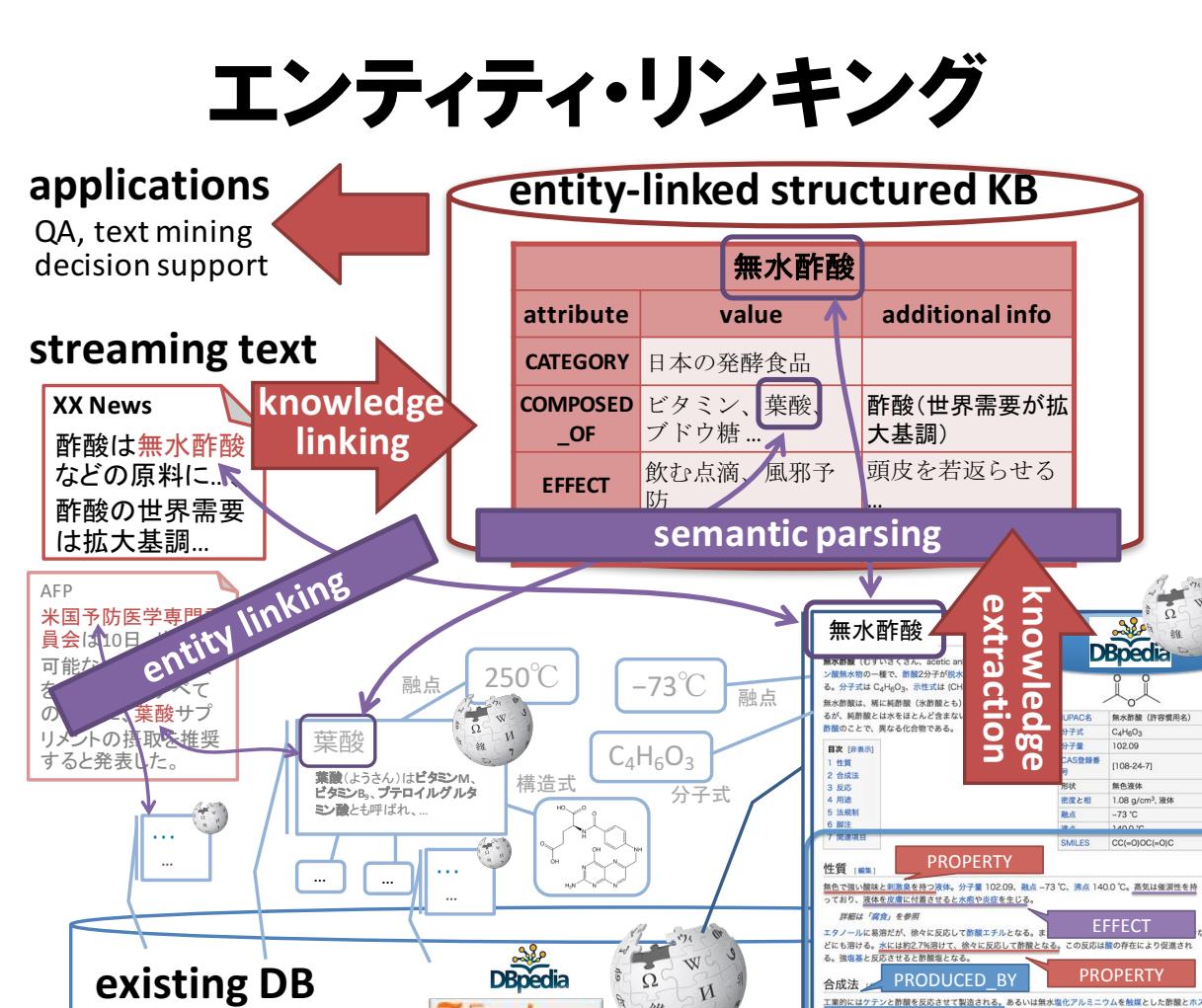
(米大システムの1万倍)

Chris was running after John because he wanted to talk to him.

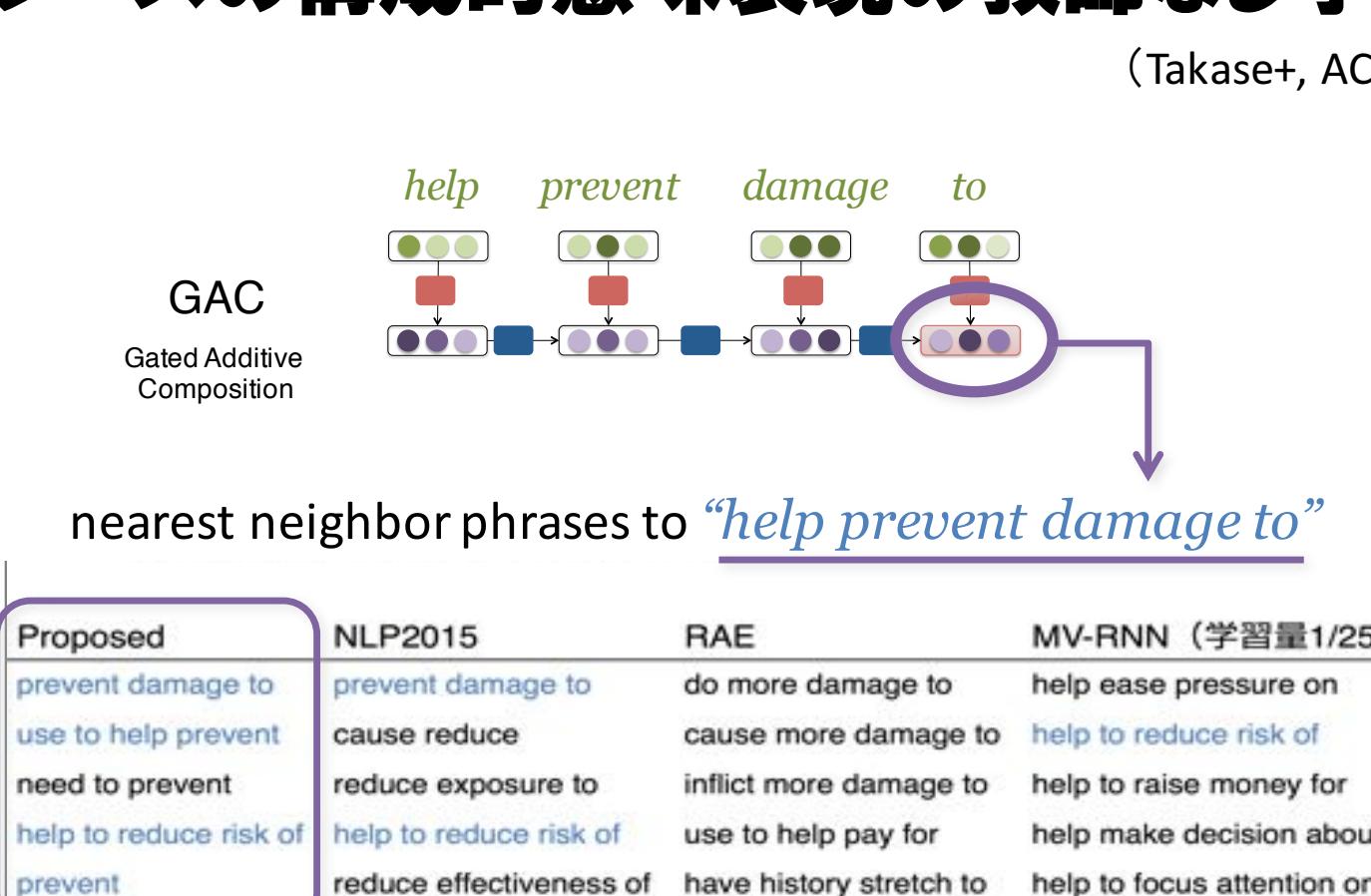


3. 頑健な言語解析技術を支える基礎技術の研究

言葉の意味を精緻に捉えるには、言葉と実世界の対応付け、言語表現の多様性を捉える数理モデルの構築が必要です。



フレーズの構成的意味表現の教師なし学習



Vector-based DCS

