

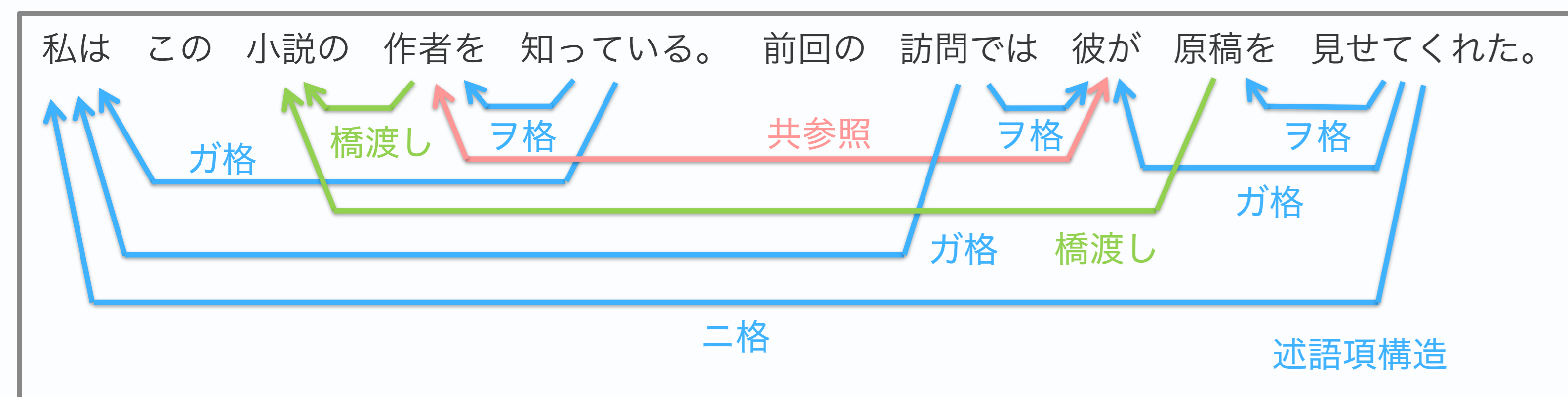
自然言語の意味理解に向けた名詞や述語間の関係解析

知能情報学専攻 植田 暢大, 黒橋 禎夫

{ueda,kuro}@nlp.ist.i.kyoto-u.ac.jp

研究背景：コンピュータによる自然言語の意味理解

- 自然言語テキストでは名詞や述語間に様々な意味的關係（結束性）が存在し、テキストの意味を支えている
- 結束性の認識はコンピュータによる自然言語の意味理解に不可欠である



■チャットボットへの応用例

このお店、料理がおいしそうだ。近くにあるか調べてくれない？

ユーザー

(お店) 料理
(料理が) おいしそう
(お店が) (近く)にある
(ボットが)(お店を)調べる

ボット

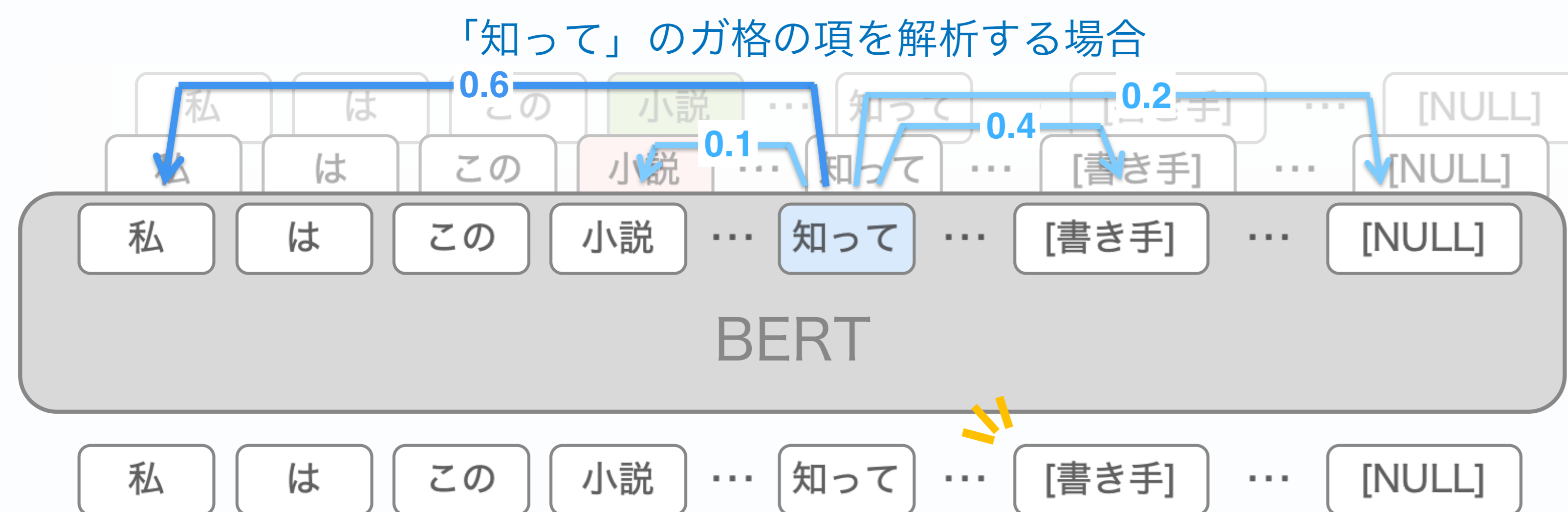
四条通にあるようです

■結束性解析タスクの種類

- **述語項構造解析**: 述語に対して「誰が」や「何を」にあたる単語を発見
- **共参照解析**: 名詞に対して実世界の同一のエンティティを指す別の名詞を発見
- **橋渡し照応解析**: 名詞に対して欠けている必須的な情報を補う名詞を発見

BERTに基づく結束性の同時解析 [Ueda+20]

- 提案手法: それぞれの単語ペアに対してガ格らしさ・ヲ格らしさのスコアを計算
正しいペアのスコアが高くなるようにニューラルネットワークを訓練



BERT [devlin+19]

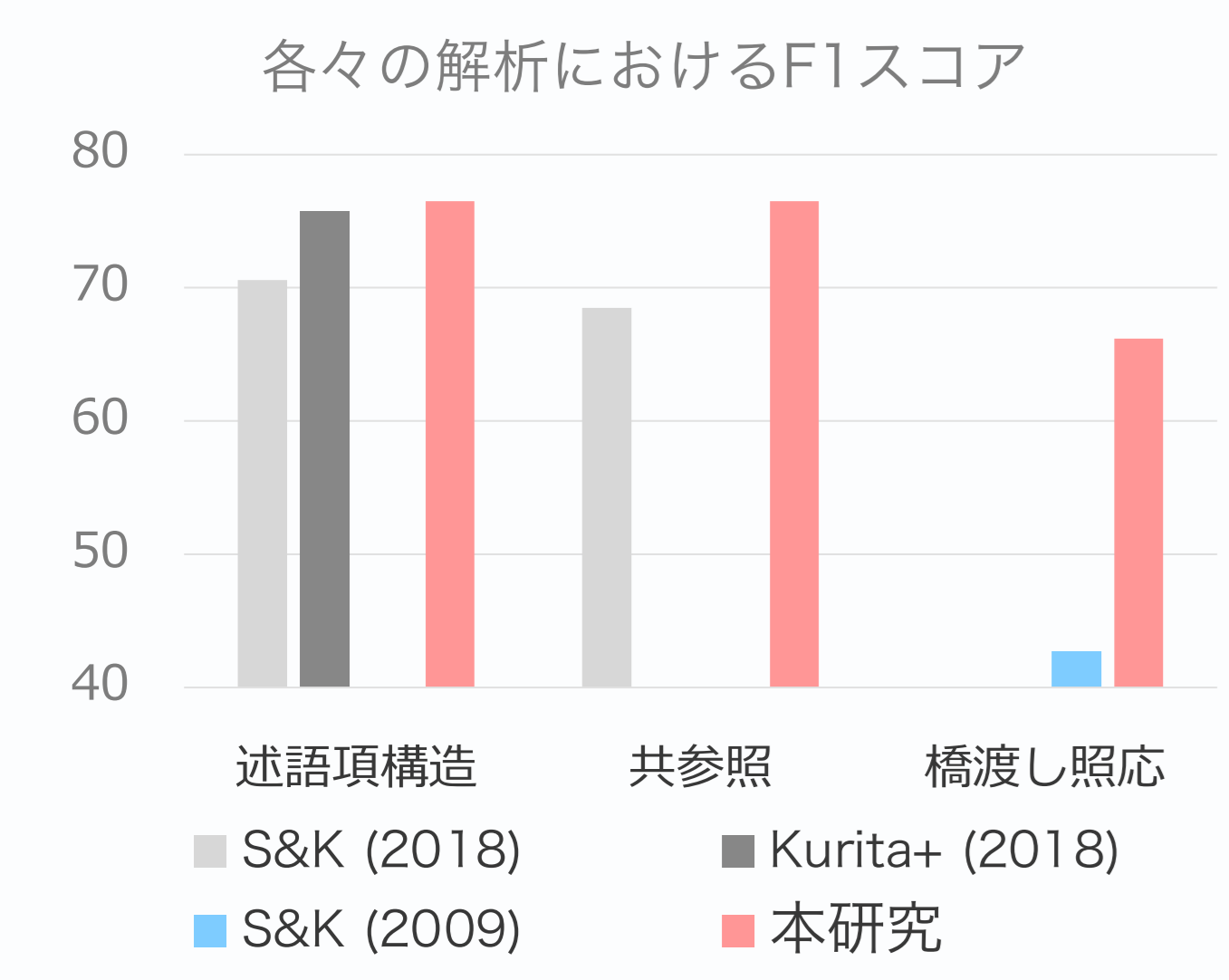
- ✓ 大規模なテキストで穴埋め問題を解く（事前学習）
- ✓ 事前学習から言語の汎用的な特徴を獲得

共参照や橋渡し照応も単語間の関係なので同様に解析が可能

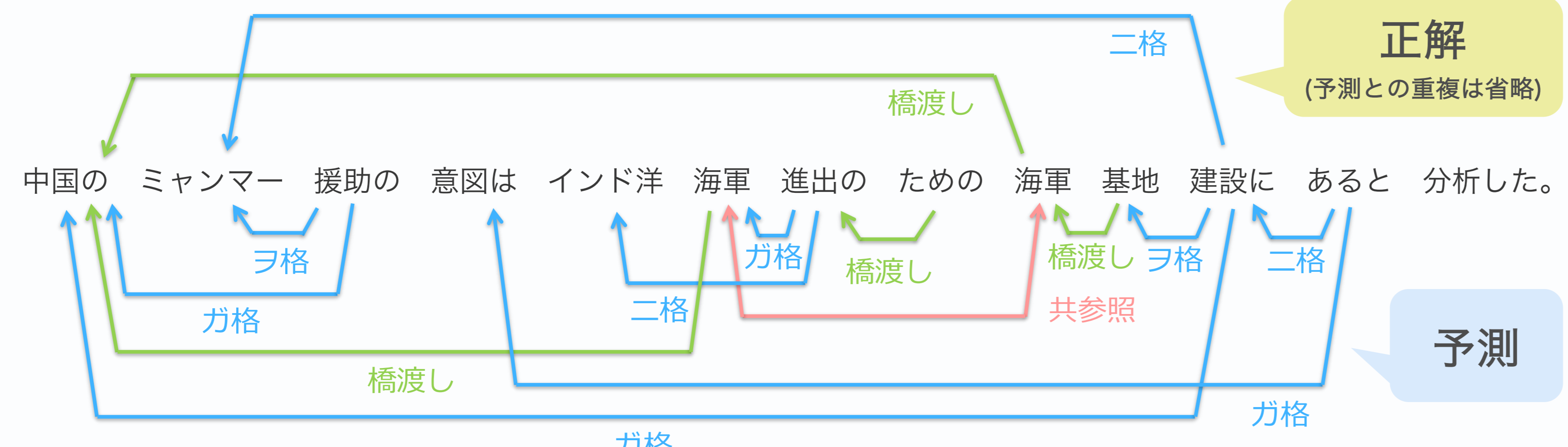
特殊トークン

- [書き手] • [その他 (人)] } 外界照応
- [読み手] • [その他 (物)] } に対応
- [NULL]

■既存研究との比較



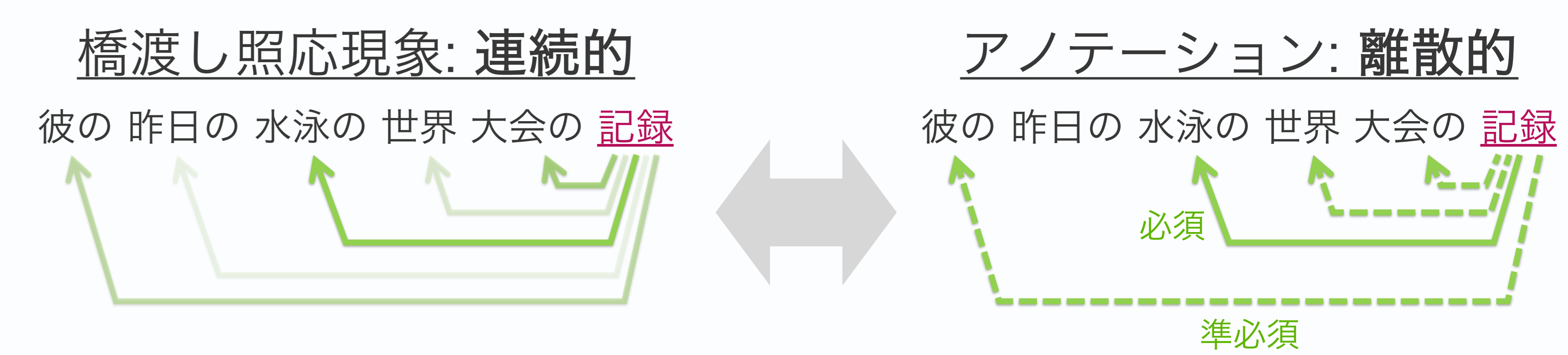
■システムの解析例



ソースコード: <https://github.com/nobu-g/cohesion-analysis>
デモ: <https://lotus.kuee.kyoto-u.ac.jp/cohesion-analysis/public/>

クラウドソーシングによる橋渡し照応解析の改善

- 背景: 述語項構造・共参照解析に比べ橋渡し照応解析の精度は低い
- 原因: 橋渡し照応現象とコーパスアノテーションの間に乖離が存在



- 提案手法: クラウドソーシングで多人数のデータを集め、より連続性を反映

8人のクラウドワーカー

- 「〇〇の種類」に当てはまる単語を全て選択
- 同時に最も必須的な単語を選択

問題8

× 38,840問

【該当なし】 【書き手】 【読み手】 【その他 (人)】 【その他 (物)】

イヌワラビは市街地の花壇の中に生えることもあります。このなかまも種類が多く、しかも雑種が多く見分けも困難です。でも、ヘビネコザというシダはある鉱物を含んだ土地に生えるのです。

ワーカーの回答を集約

問題8

【該当なし】 【書き手】 【読み手】 【その他 (人)】 【その他 (物)】

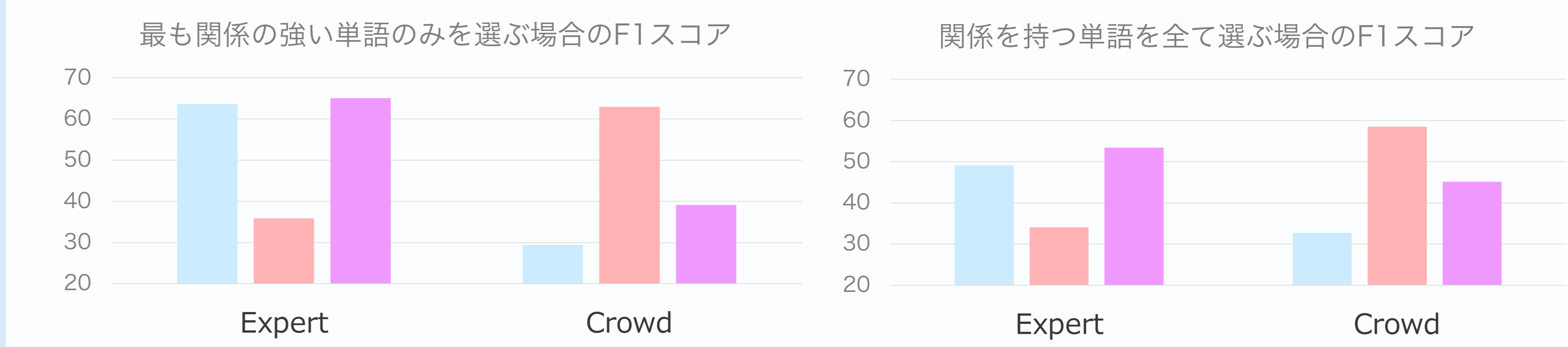
イヌワラビは市街地の花壇の中に生えることもあります。このなかまも種類が多く、しかも雑種が多く見分けも困難です。でも、ヘビネコザというシダはある鉱物を含んだ土地に生えるのです。

5/16 6/16

- 実験: 両コーパスで橋渡し照応解析モデルを訓練

- データセット
 - Expert (既存コーパス、5,124 文書)
 - Crowd (本研究で作成、3,923 文書)

- モデル:
 - Expertで訓練
 - Crowdで訓練
 - Expert+Crowdで訓練



- 結果: いずれの評価においてもCrowdまたはExpert+Crowdが最も良い性能

クラウドソーシングタスク画面: <https://tulip.kuee.kyoto-u.ac.jp/bridging-crowd/public/tutorial/220112-3588952734/0>