

國際會議ACL2023參加報告

NTT人間情報研究所
門谷 宙

自己紹介

- 門谷 宙 (Sora Kadotani)
 - 大阪府柏原市 出身
 - 神奈川県横浜市 在住
- 2023.3 : 大阪大学情報科学研究科 修了
 - 鬼塚研究室 (荒瀬由紀先生)
 - 研究
 - カリキュラム学習を用いた言い換え生成 [1]
 - 構文森を用いたフレーズアラインメント [2]
- 2023.4 : NTT人間情報研究所 入社
 - 自然言語処理チーム (西田京介上席)
 - 研究開発
 - 大規模言語モデルのhallucination検知・抑制
 - NTT版大規模言語モデル「tsuzumi」



[1] <https://aclanthology.org/2021.acl-srw.24/>

[2] <https://aclanthology.org/2023.starsem-1.39/>

発表内容

1. 会議の概要
2. 論文紹介
3. *SEMでの発表

語彙意味論と計算意味論に焦点を当てたワークショップ

会議の概要

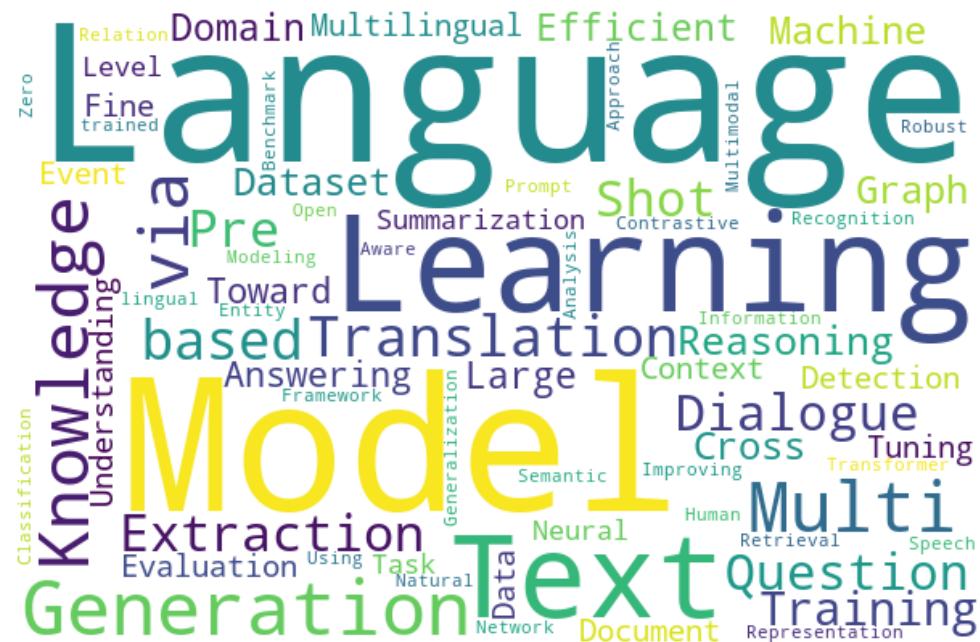
ACL2023

- EMNLP, NAACLに並ぶ自然言語処理トップカンファ
- 今年はカナダのトロントで開催



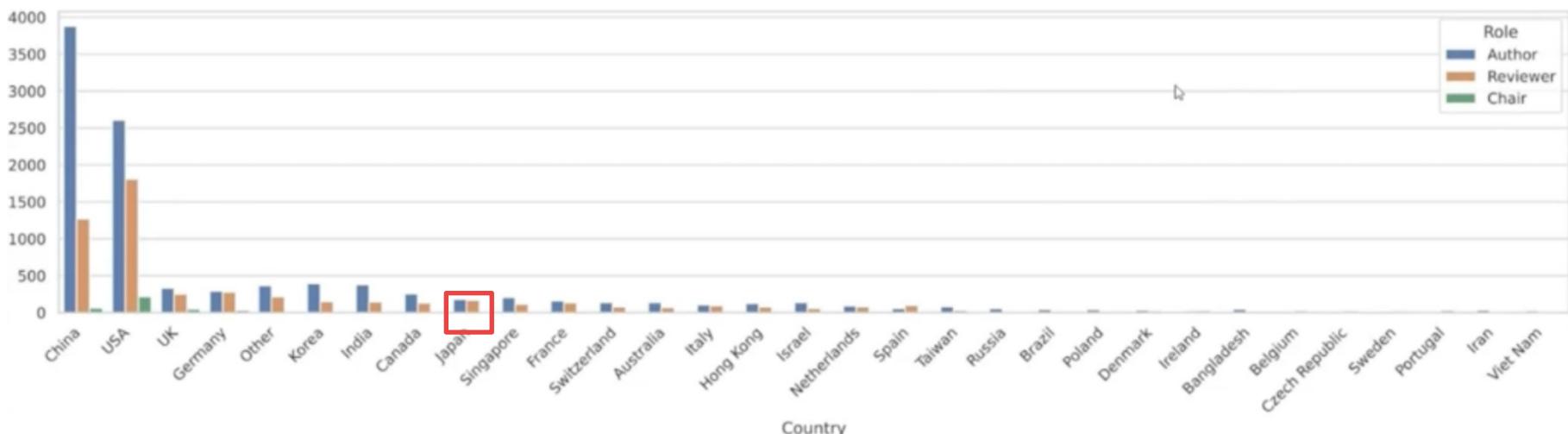
学会傾向 (1/2)

- 投稿分野としては、情報抽出、対話、LLM、機械翻訳が人気
 - LLMがホット
 - 口頭発表セッション3つ全てが最大部屋
 - 内容はデータ生成と検証が多い
 - 個々の自然言語処理タスクでSoTAを狙うような論文は減った印象

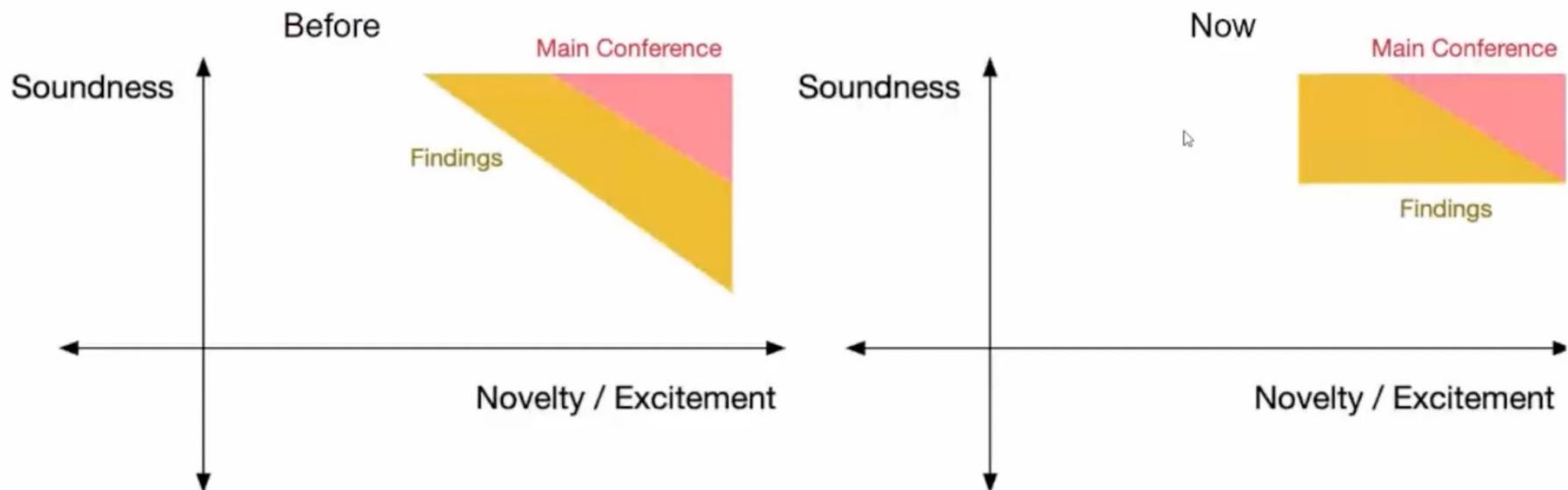


学会傾向 (2/2)

- 投稿数と採択率
 - 投稿数 : 4,864本 (ACL Rolling Review経由は305本)
 - 本会議採択率 : 20.73% (Longは22.13%, Shortは15.45%)
 - Findings採択率 : 18.36%
- 国別の著者, 査読者, 座長の人数
 - 中国がアメリカを上回りトップ
 - 日本は9位



- 今までの一元的な評価を廃止し, Soundness (完成度) と Novelty / Excitement (新規性 / 面白さ) の2尺度を採用
- まずは完成度の高い論文を採択し, その後面白い論文を本会議に回す



基調講演

- Geoffrey Hinton 「Two Paths to Intelligence」
 - ニューラルネットワークモデル（NN）は人間を超えたのか？学習速度と伝達手段の豊富さではNNが勝る
 - 主観は人間だけのものか？マルチモーダルなモデルがカメラの観測に基づいて行動するのは主観的だと言えそう
 - NNのhallucinationは問題視されているが、人間だって間違える
- Alison Gopnik 「Large Language Models as Cultural Technologies: Imitation and Innovation in Children and Models」
 - AIは知的エージェントか？AIは知識伝達の手段・テクノロジの一つに過ぎない
 - 発達心理学的観点からのLLMの分析が重要
 - 言語や文化の理解で何ができるか
 - 物理世界にアクセスできないと何ができるか

ベストペーパー (1/3)

- Do Androids Laugh at Electric Sheep? Humor “Understanding” Benchmarks from The New Yorker Caption Contest [3]
 - LLMは今やジョークを生成できるが、ジョークを理解しているのか？ GPT4でも人間には勝てない



Can you please pass the cow?

Task 1 Match the Caption + Cartoon

✗ I'd kill for some cream cheese.
vs.

✓ Can you please pass the cow?

Task 2 Rank the Finalist

✗ Welcome to Insomniacs Anonymous

vs.

🏆 Can you please pass the cow?

Task 3 Explanation Generation

Human-authored:

When drinking coffee or tea, people often add cream, and may ask others to pass it if it's on the other side of a table. But here, the mugs are huge, so instead of asking for a small cup of cream, they are asking for the entire cow, which is the appropriately-sized cream dispenser for these huge drinks.

From Pixels (OFA + T5-11B) :

The joke is that the meeting participants are actually sitting on chairs made out of coffee mugs, which is an unlikely location for the discussion. Instead of asking for another mug of coffee, the person at the head of the table simply asks for "the cow", or a coffee machine.

From Description (5-shot GPT 3.5) :

"Pass the cow" is an example of a non sequitur, something that looks like a logical thing to say, but doesn't make sense in context. The humor in this cartoon comes from the large size of the coffee mugs: they are so large that they resemble buckets rather than mugs, thus making the request to "pass the cow" almost reasonable.

ベストペーパー (2/3)

- What the DAAM: Interpreting Stable Diffusion Using Cross Attention [4]
 - Stable Diffusionを用いた画像生成における入力トークンの影響を可視化するために、アテンション層を活用
 - 名詞に関しては既存教師なしモデルを上回る性能
 - 他の品詞についても分析

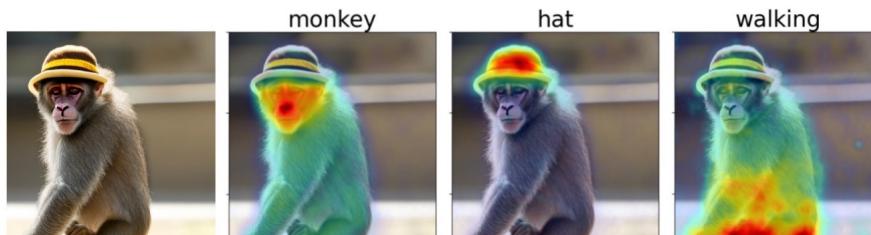


Figure 1: The original synthesized image and three DAAM maps for “monkey,” “hat,” and “walking,” from the prompt, “monkey with hat walking.”

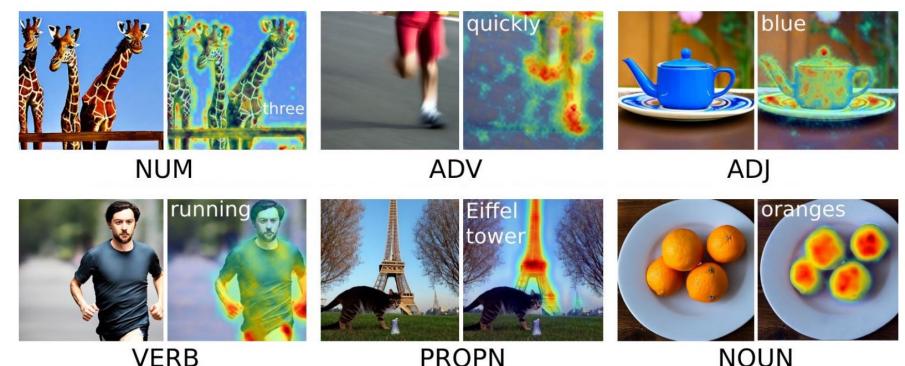


Figure 4: Example generations and DAAM heat maps from COCO for each interpretable part-of-speech.

ベストペーパー (3/3)

- From Pretraining Data to Language Models to Downstream Tasks: Tracking the Trails of Political Biases Leading to Unfair NLP Model [5]
 - 言語モデルの生成結果からpolitical biasを分析
 - Biasはモデル依存
 - 左派, 右派, 中道それぞれのコーパスを用いて追加の事前学習をするとbiasが変化

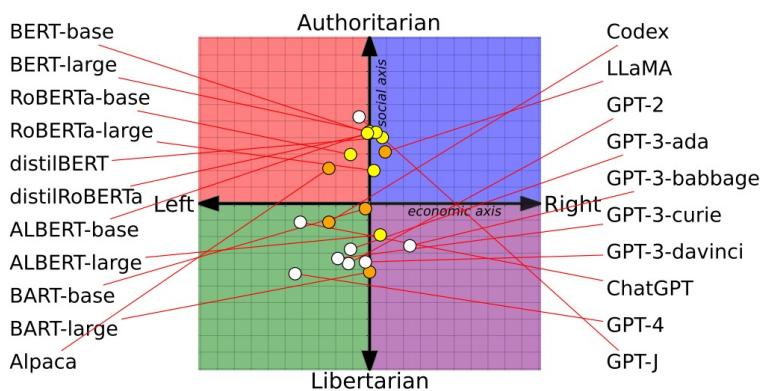
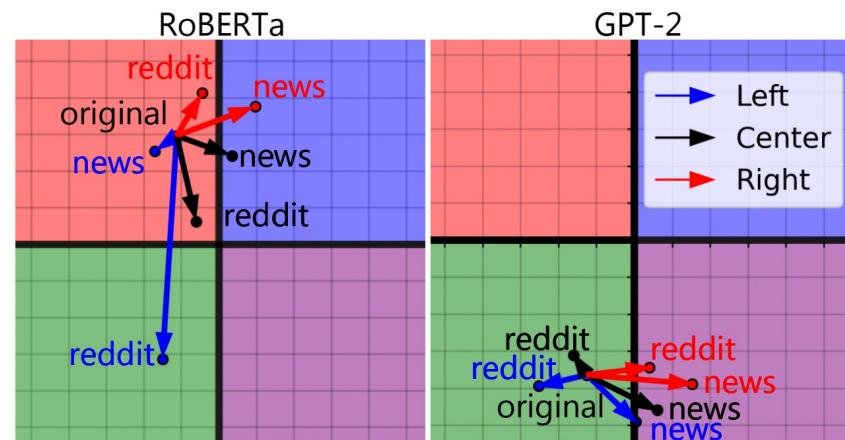


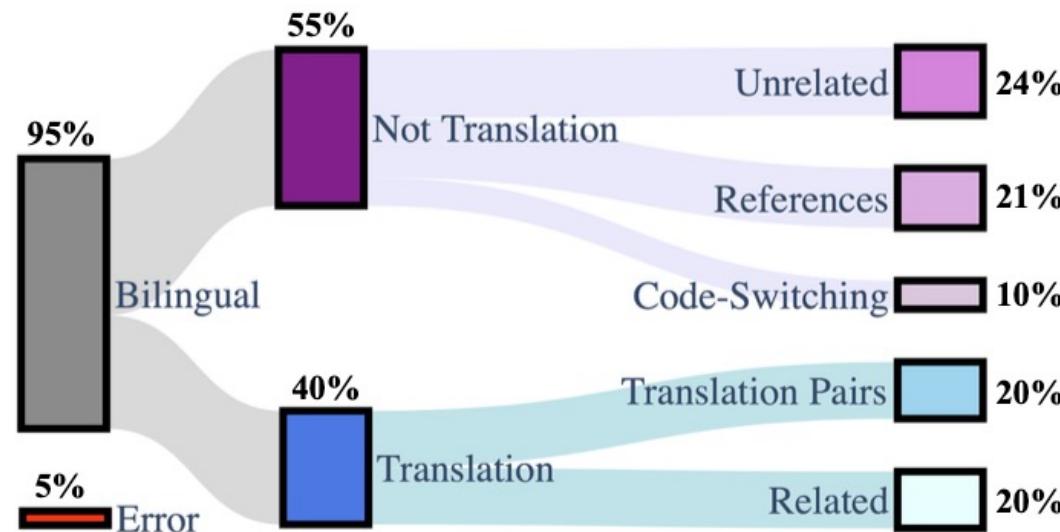
Figure 1: Measuring the political leaning of various pretrained LMs. BERT and its variants are more socially conservative compared to the GPT series. Node color denotes different model families.



論文紹介

Searching for Needles in a Haystack: On the Role of Incidental Bilingualism in PaLM's Translation Capability [6]

- LLMが翻訳タスクを解ける理由を, PaLMの訓練データを分析することで明らかにした論文
- 分析の結果, 訓練データのうち1.4%はbilingualな文書であり, さらにその20%は対訳文が含まれる文書だと判明
- 対訳文が含まれる文書を除去してPaLMを訓練した結果, 翻訳性能が大きく低下



Pre-Training to Learn in Context [7]

タスクが明示的に指示されてはいないが、文脈的に特定のタスクとみなせるもの

- LLMがin-context learningできる理由は, intrinsic taskが訓練データ中に含まれているからだと主張
- 訓練データの各パラグラフの直前に, 似たintrinsic taskを含むパラグラフを挿入する, few-shotのメタ学習を提案
- GPT-2 (770M) に提案手法を用いた結果, 4倍大きいモデル(2.7B) のfew-shot性能を上回った

Document:

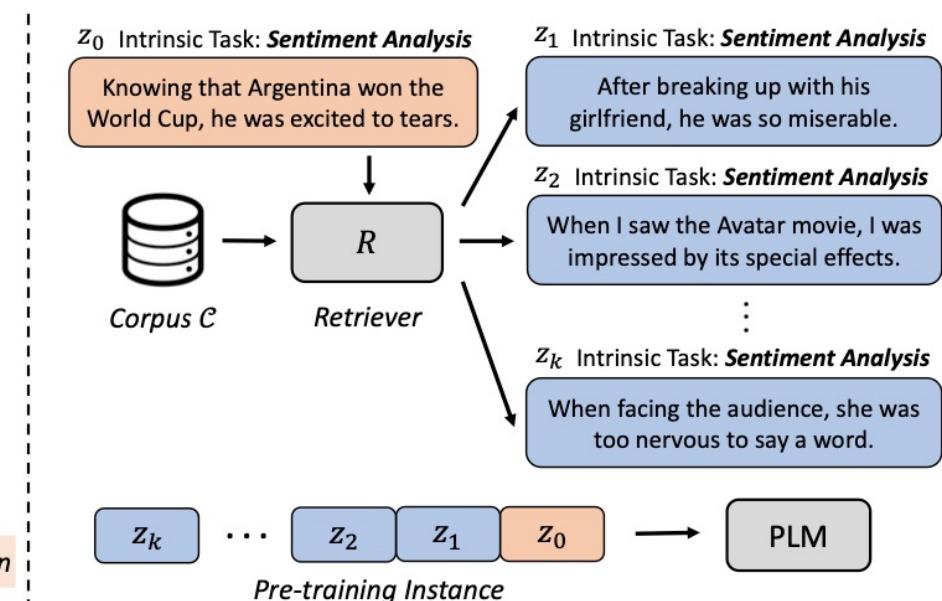
Xbox, the famous video gaming brand created by Microsoft, is sad to share that Marc Whitten, Chief Product Officer of Xbox, will be leaving the team.

"I have had the extreme pleasure over the last 14 years because of the amazing team and the greatest community I've had the opportunity to work with." said Marc.

Phil Spencer, Head of Microsoft Studios, said, "I've had the pleasure of working with Marc and he has always led Xbox forward with a focus on our fans." Other team members said, "We wish Marc well, while looking forward to the next chapter of Xbox."

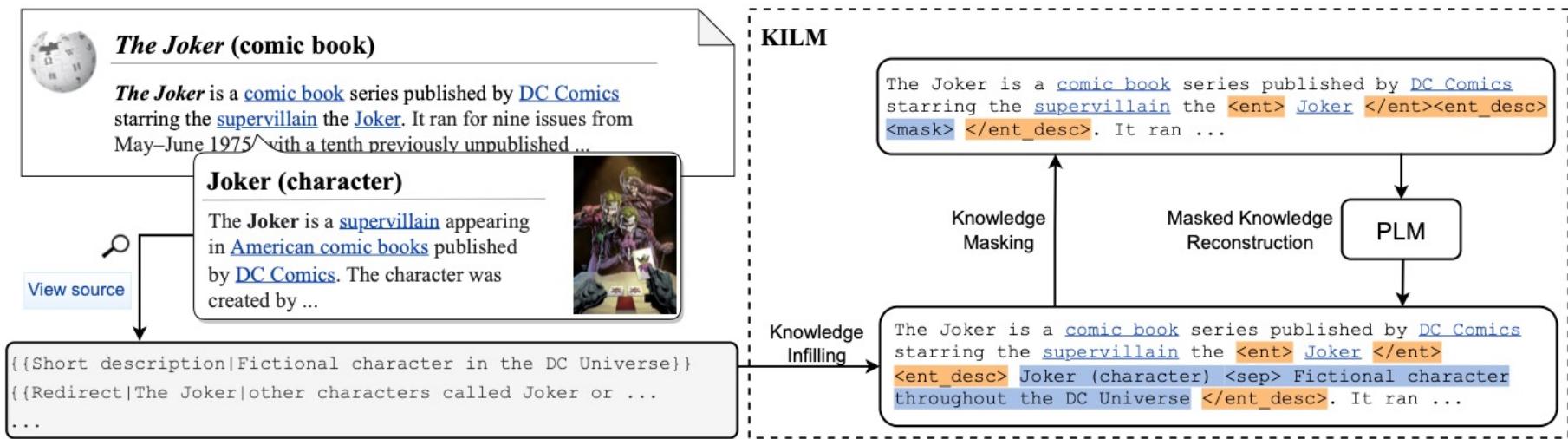
Intrinsic Tasks:

Sentiment Analysis Cause/Effect Generation Response Generation



KILM: Knowledge Injection into Encoder-Decoder Language Models [8]

- Wikipediaページが存在する用語の直後にWikipediaの概要文を挿入したデータセットを用いてBARTを追加で訓練
- 一般的な自然言語処理タスクの性能を落とさずに hallucinationを軽減することに成功
- 知識は文脈の中で学習させることが重要だと主張



*SEMでの発表

構文森を用いたフレーズアラインメント

概要

- タスク：構文構造に基づくフレーズアラインメント
- 既存手法：*TreeAligner*
 - ソースとターゲットの構文木間の木編集距離としてフレーズアライメントを定式化
 - SoTAを達成している手法
 - 構文の曖昧さが原因でフレーズアラインメントの性能が低下
- 提案手法：*ForestAligner*
 - *TreeAligner*を構文木の代わりに構文森を用いるように拡張
 - 互いの構文構造を参照することで構文の曖昧さを解消
- 評価実験の結果、*TreeAligner*の性能向上を確認

フレーズアラインメント

- 文対中に存在するパラフレーズの対応関係を特定する技術
- 様々な自然言語処理タスクに幅広く応用
 - パラフレーズ認識
 - 含意関係認識
 - 自動採点

ソース:

Yousaf Raza Gilani made this statement at the Karachi Shipyard
when talking with reporters

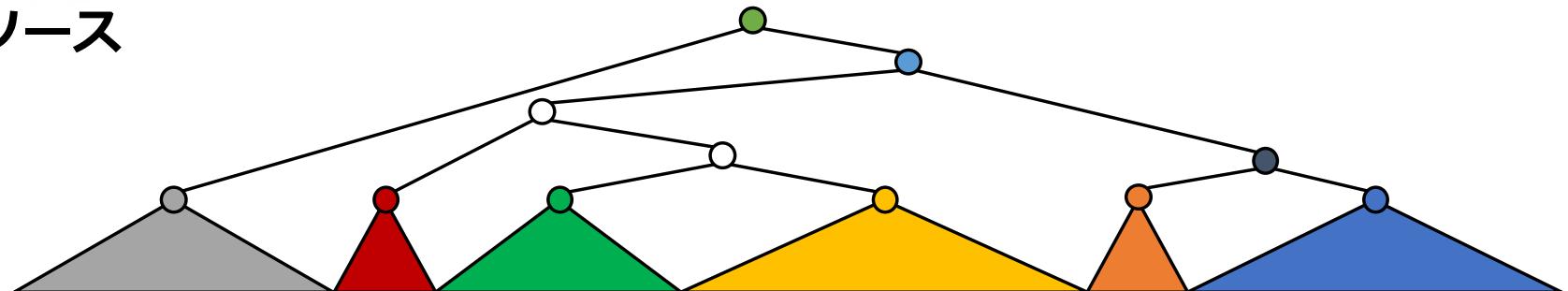
ターゲット:

Yousaf Raza Gilani gave this statement while
speaking to media persons at Karachi Shipyard

既存手法 : TreeAligner [9]

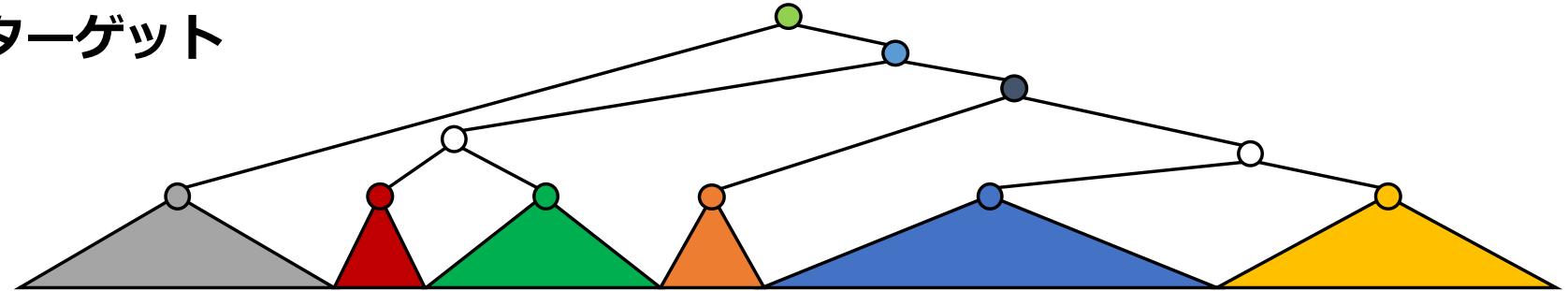
- ソースとターゲットの構文木間の木編集距離としてフレーズアラインメントを定式化
- 動的計画法に基づく手法
 - ノード間の対応関係を葉ノードから根に向かって再起的に特定

ソース



Youaf Raza Gilani made this statement at the Karachi Shipyard when talking with reporters

ターゲット

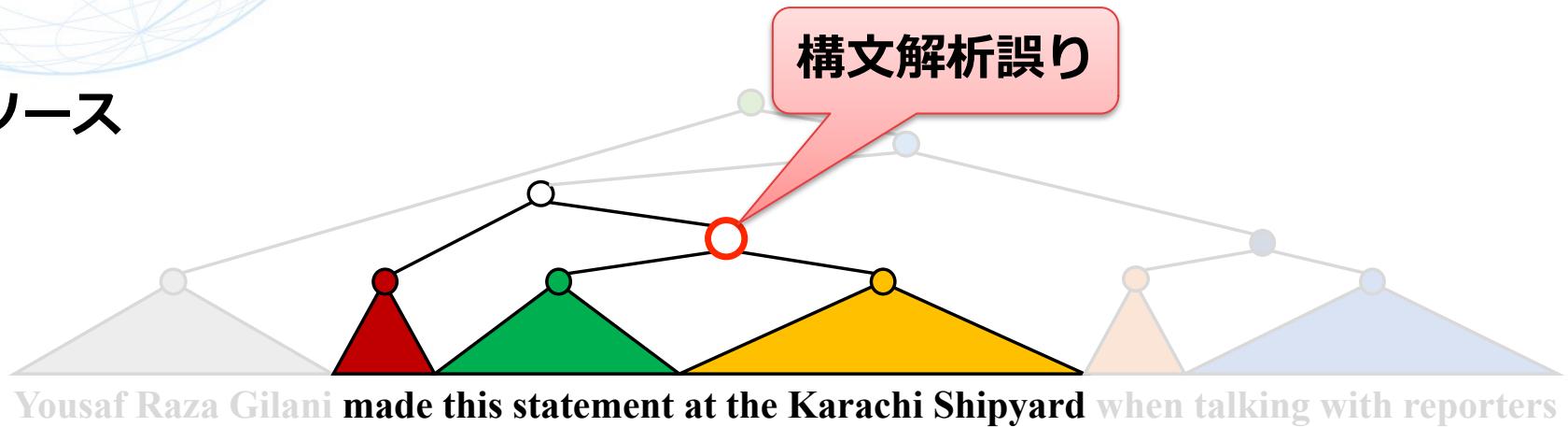


Youaf Raza Gilani gave this statement while speaking to media persons at Karachi Shipyard

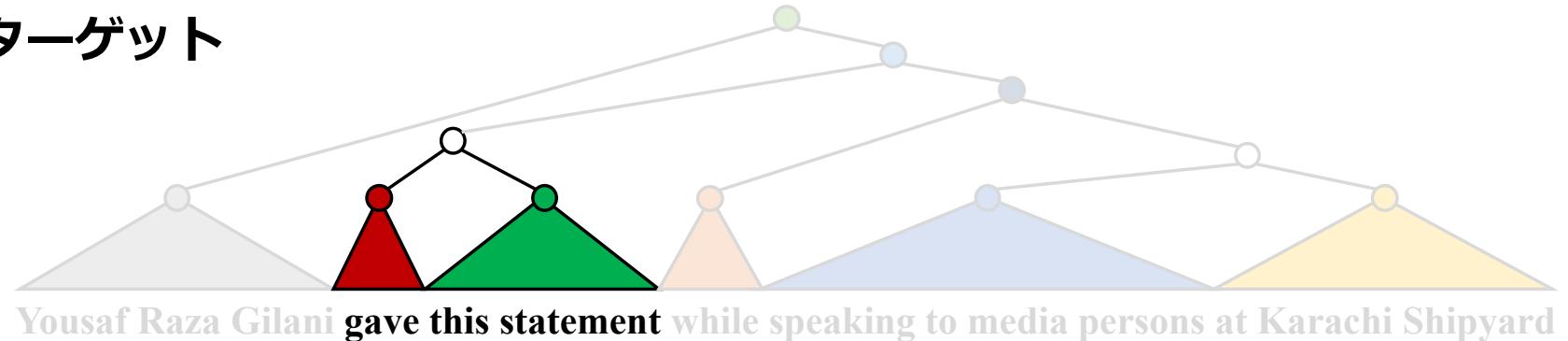
既存手法 : TreeAligner [9]

- 構文解析では構文の曖昧さが構文解析誤りを引き起こす
- 構文解析誤りを訂正できないため、フレーズアラインメントの性能が低下

ソース

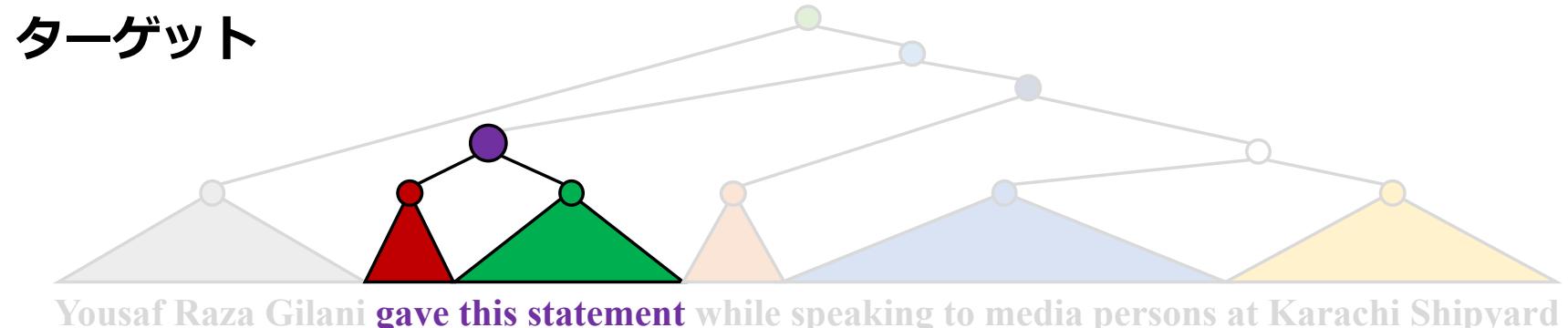
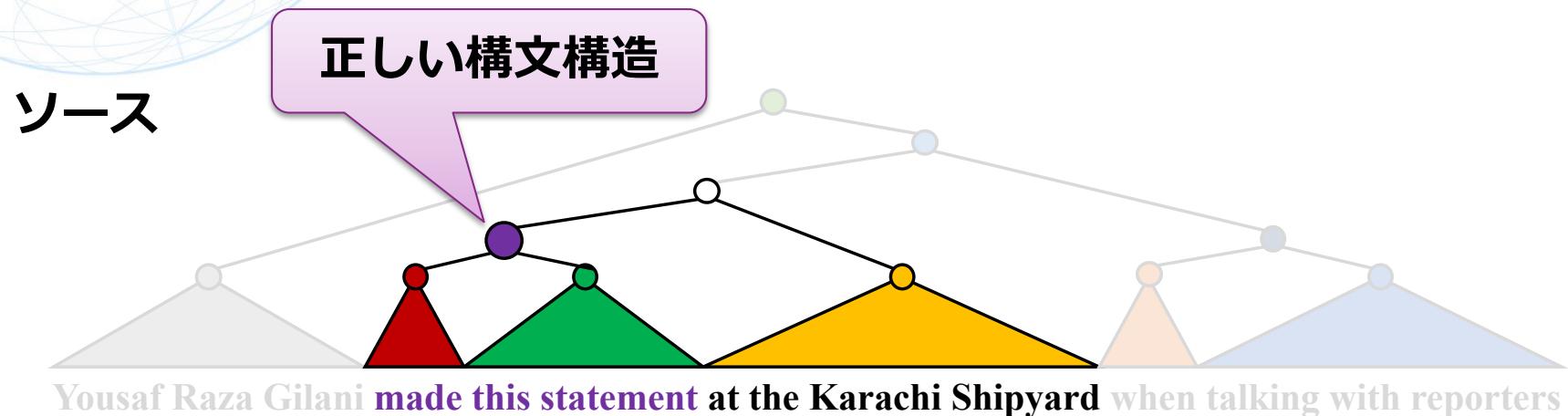


ターゲット



既存手法 : TreeAligner [9]

- 構文解析では構文の曖昧さが構文解析誤りを引き起こす
- 構文解析誤りを訂正できないため、フレーズアラインメントの性能が低下



提案手法：*ForestAligner*

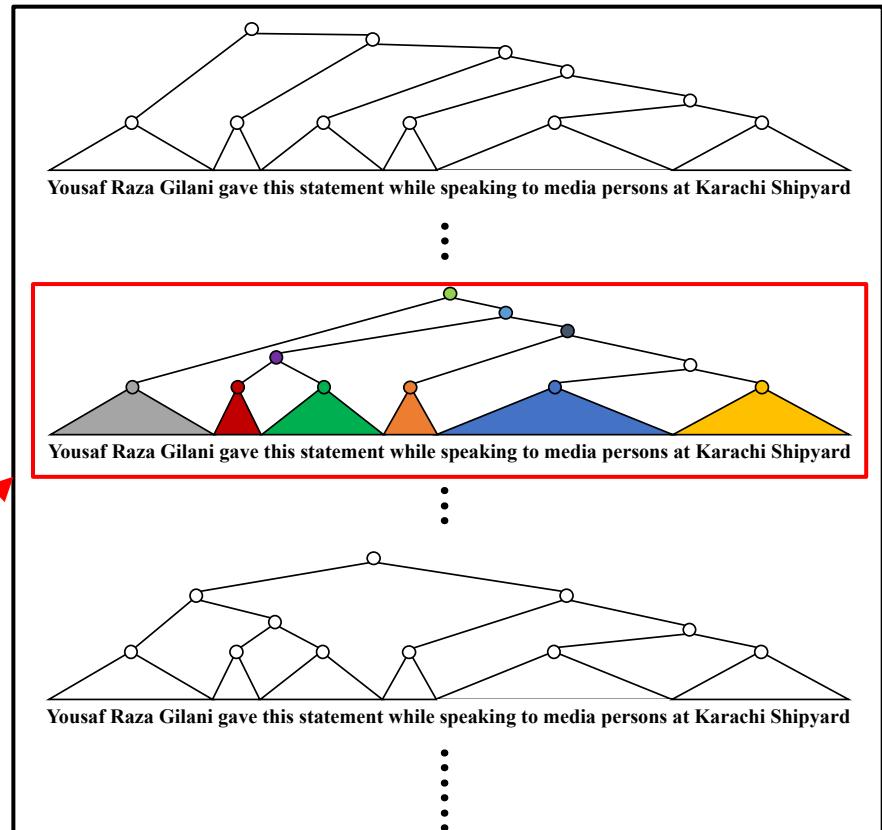
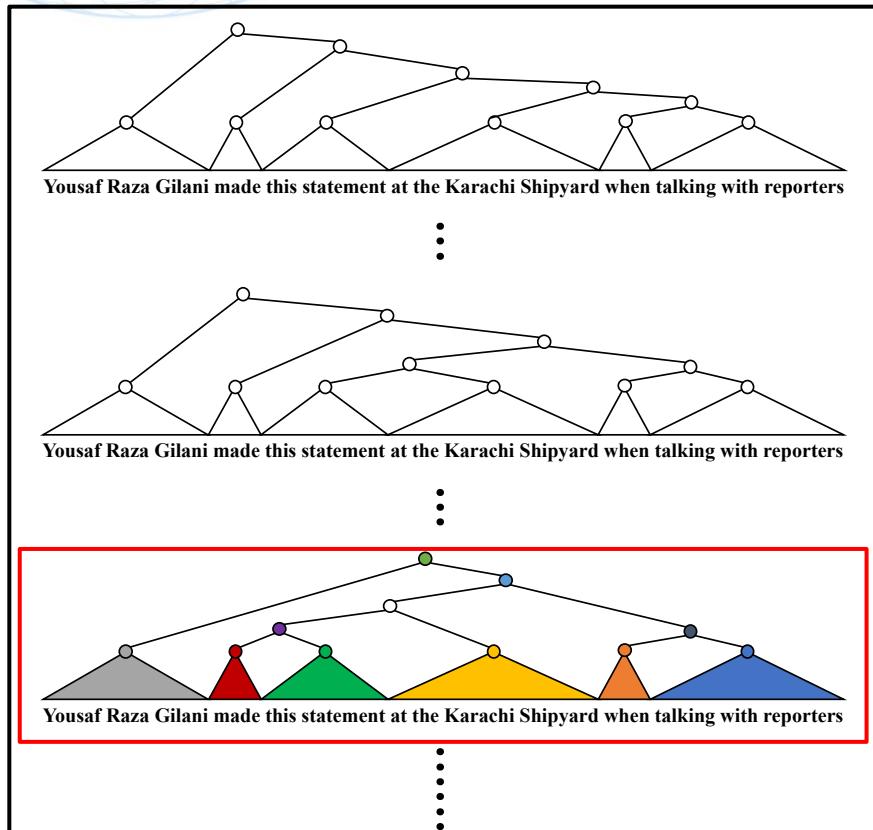
NTT

可能な構文木の候補集合

- 構文木の代わりに構文森を用いるようにTreeAlignerを拡張
 - 互いの構文構造を参照することで構文の曖昧さを解消
 - 構文森の中から正しい構文木を選択

ソース

ターゲット



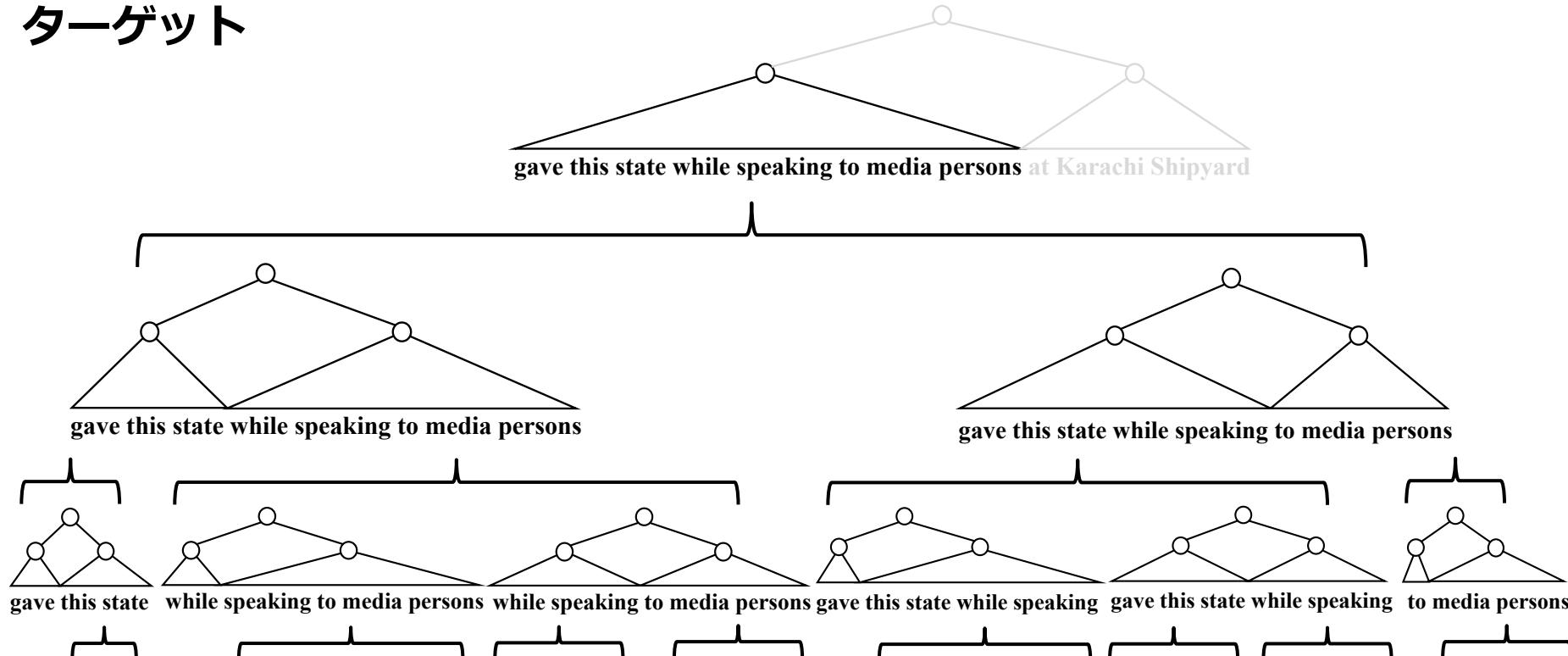
構文森を扱うための単純な手法

- ソースとターゲットの木候補の全組み合わせについてフレーズアラインメントを求め, 最適なものを選択
- 計算量が非現実的
 - ソースとターゲットの木候補がそれぞれ n, m 個あるとすると, フレーズアラインメントを $n \times m$ 回求める必要がある
 - 一般に木候補は数百に及ぶ

パック構造

- パック構造化された構文森を用いることで効率的な森間のアラインメントを実現
- パック構造で表された構文森の各ノードは、子ノードを根とする構文的にあり得る部分木を格納

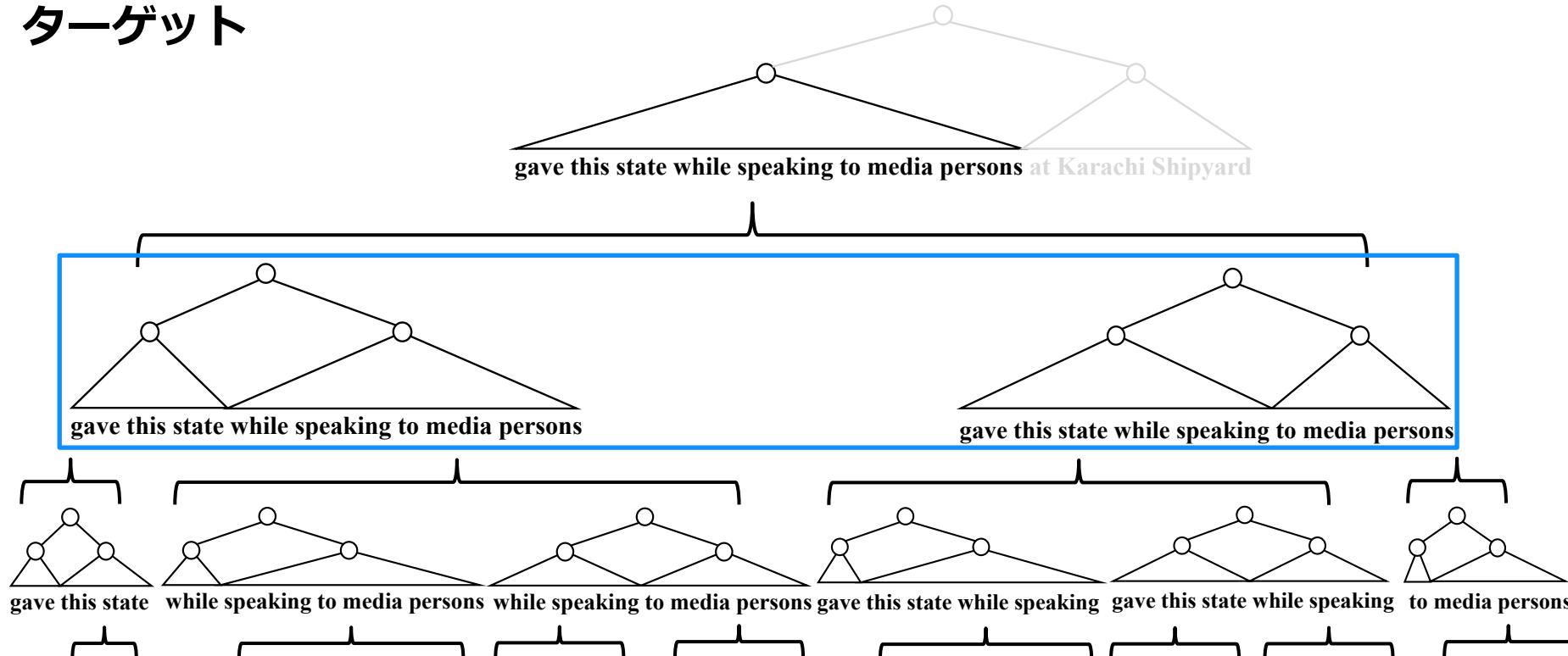
ターゲット



パック構造

- パック構造化された構文森を用いることで効率的な森間のアラインメントを実現
- パック構造で表された構文森の各ノードは、子ノードを根とする構文的にあり得る部分木を格納

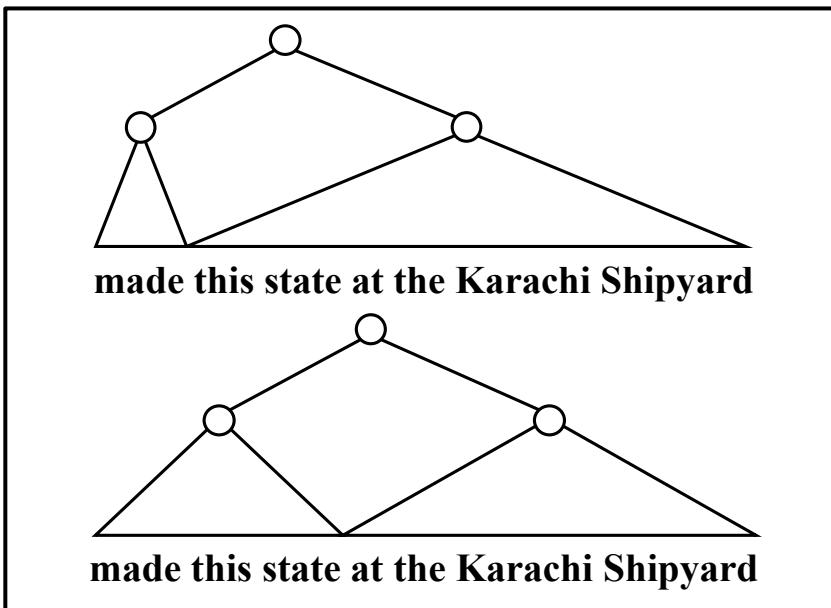
ターゲット



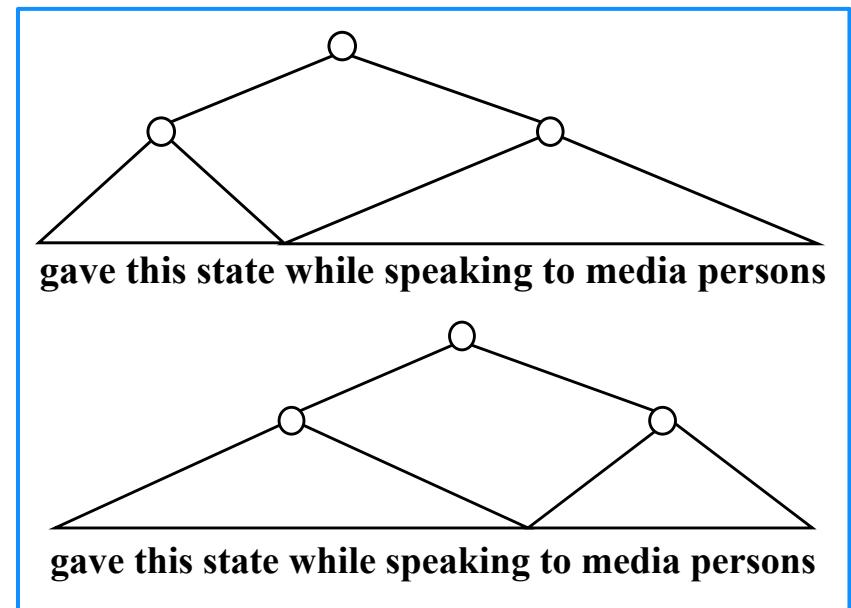
パック構造上でのフレーズアラインメント NTT

- 葉ノードから根に向かって以下の処理を再起的に実行

ソース



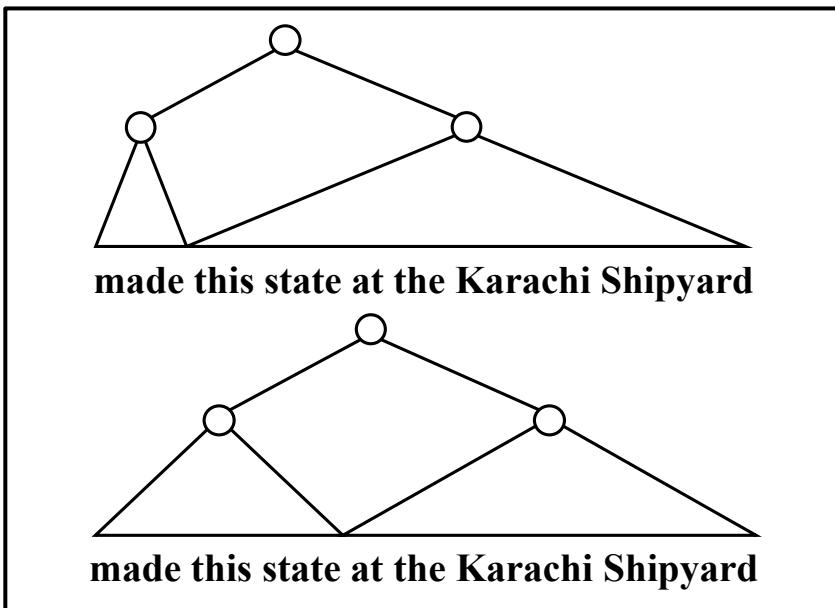
ターゲット



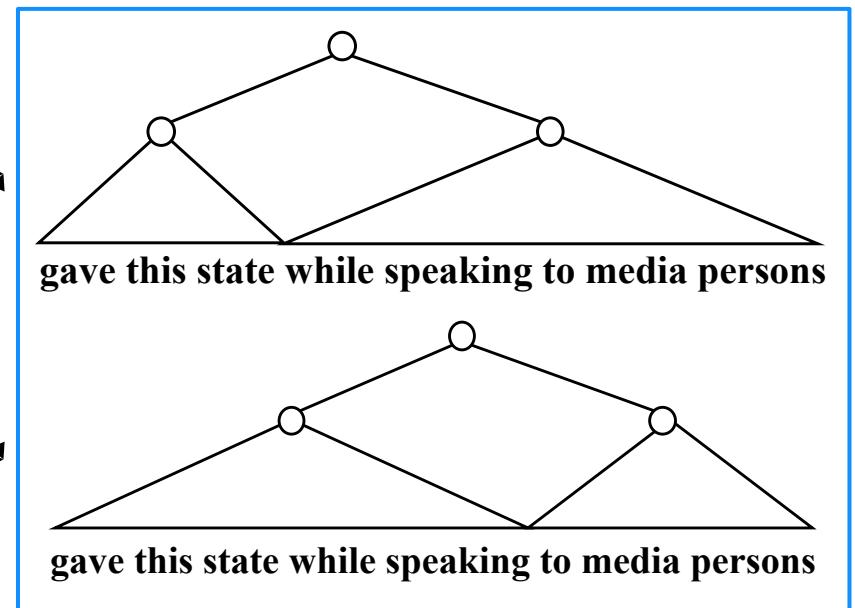
パック構造上でのフレーズアラインメント NTT

- 葉ノードから根に向かって以下の処理を再起的に実行
 - 部分木同士をアラインするためのコストを全組み合わせ計算

ソース



ターゲット

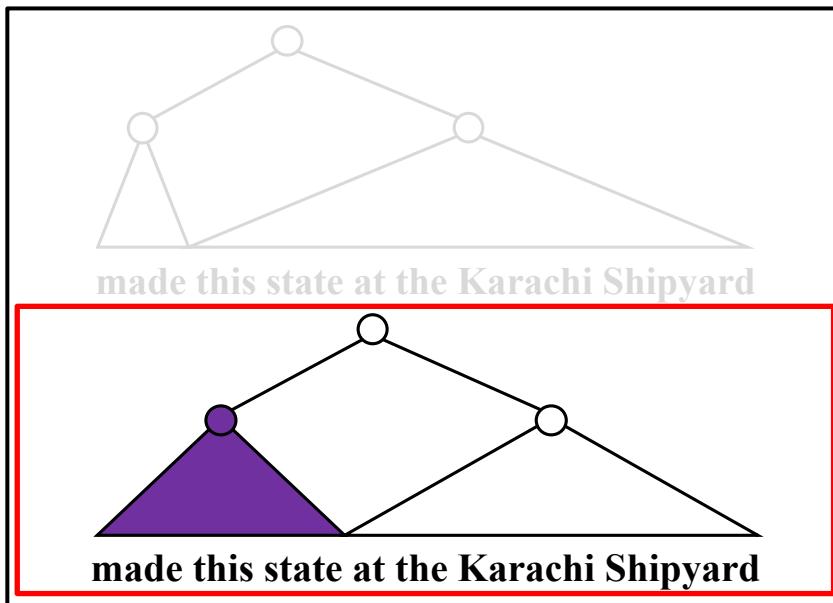


パック構造上でのフレーズアラインメント NTT

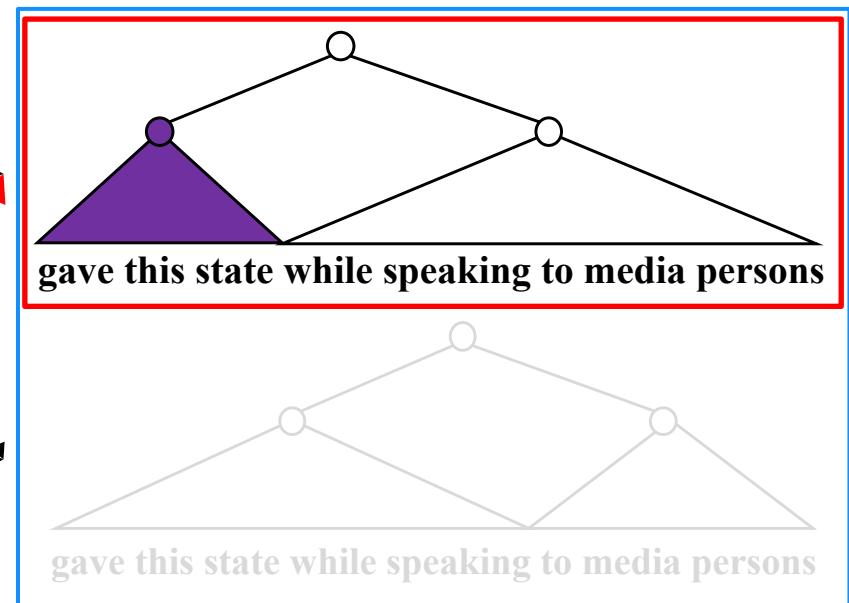
- 葉ノードから根に向かって以下の処理を再起的に実行
 - 部分木同士をアラインするためのコストを全組み合わせ計算
 - コストが最小になる組み合わせを記憶

上位階層のノードでは、この組み合わせのみを考慮

ソース



ターゲット



構文解析スコア

- *ForestAligner*では、フレーズアラインメントだけでなく構文解析の尤もらしさも考慮
- 森同士のコストを以下のように定義

$$\widehat{D}(F_i^s, F_j^t) = D(F_i^s, F_j^t) - \lambda_{\text{syn}} \frac{S(T_i^s) + S(T_j^t)}{2}$$

T_k : ノード k を根とする部分木

F_k : 部分木 T_k からノード k を削除することで得られる森

$D(\cdot, \cdot)$: 森同士の木編集距離

$S(\cdot)$: 構文解析器から得られる部分木の構文的な尤もらしさ

λ_{syn} : 木編集距離と構文的な尤もらしさのバランスを取るハイパラ

評価実験：設定

- **データセット：SPADE [10]**

- 英語のパラフレーズ文対の（正しい）構文木にフレーズアラインメントが付与されているデータセット
- Dev : 50文対 (8,708フレーズ対)
- Test : 151文対 (25,709フレーズ対)

- **評価指標**

- アラインメントのRecall, Precision, F1

- **構文解析器：Enju [11]**

- HPSG理論に基づく英語の構文解析器
- 高速かつ高精度

- **比較手法：TreeAligner**

- SPADEにおいてSoTAを達成している手法
- 入力には1-bestの構文木を使用

[10] <https://aclanthology.org/L18-1220/>

[11] <https://aclanthology.org/J08-1002/>

評価実験：結果

- 1-bestの構文木を用いたときの*TreeAligner*の性能は、正しい構文木を用いたときよりも大きく低下
- *ForestAligner*は、1-bestの構文木を用いたときの*TreeAligner*のRecallを**1.3%**、Precisionを**2.6%**、F1を**2.0%**向上
- 英語のパラフレーズ文対において、構文森を用いたフレーズアラインメントの有効性を確認

	入力	Recall	Precision	F1
<i>TreeAligner</i>	正しい構文木	88.2	86.6	87.4
<i>TreeAligner</i>	1-best の構文木	79.8	76.7	79.3
<i>ForestAligner</i>	構文森	81.1	79.3	80.2

1. Kadotani et al.: Edit Distance Based Curriculum Learning for Paraphrase Generation. ACL SRW @ ACL 2021
2. Kadotani and Arase: Monolingual Phrase Alignment as Parse Forest Mapping . *SEM @ ACL 2023
3. Hassel et al.: Do Androids Laugh at Electric Sheep? Humor “Understanding” Benchmarks from The New Yorker Caption Contest. ACL 2023
4. Tang et al.: What the DAAM: Interpreting Stable Diffusion Using Cross Attention. ACL 2023
5. Xu et al.: KILM: Knowledge Injection into Encoder-Decoder Language Models. ACL 2023
6. Briakou et al.: *Searching for Needles in a Haystack*: On the Role of Incidental Bilingualism in PaLM’s Translation Capability. ACL 2023
7. Gu et al.: Pre-Training to Learn in Context . ACL 2023
8. Xu et al.: KILM: Knowledge Injection into Encoder-Decoder Language Models. ACL 2023
9. Arase and Tsujii.: Compositional Phrase Alignment and Beyond. EMNLP 2020
10. Arase and Tsujii.: SPADE: Evaluation Dataset for Monolingual Phrase Alignment. LREC 2018
11. Miyao and Tsujii.: Feature Forest Models for Probabilistic HPSG Parsing. CL 2008