

# 論述構造解析における スパン分散表現

～背景について厚めに～

---

栗林樹生<sup>1,2</sup> 大内啓樹<sup>3,1</sup> 井之上直也<sup>1,3,4</sup> 鈴木潤<sup>1,3</sup>

Paul Reisert<sup>3</sup> 三好利昇<sup>5</sup> 乾健太郎<sup>1,3</sup>

1東北大学 2Langsmith株式会社 3理化学研究所

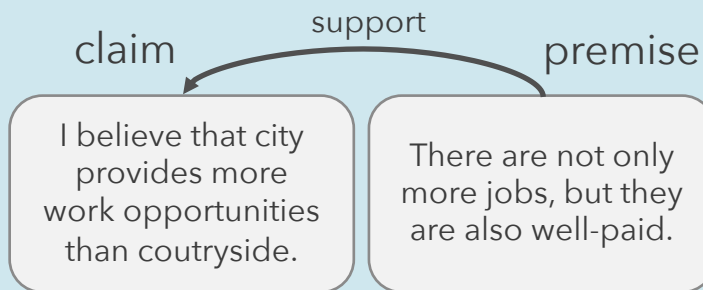
4出版時: Stony Brook University 5日立製作所

※本発表時点での所属とは一部異なる

# 概要

## 論述構造解析におけるスパン分散表現

論述的文章の談話構造を予測、議論マイニングにおける基礎解析技術。



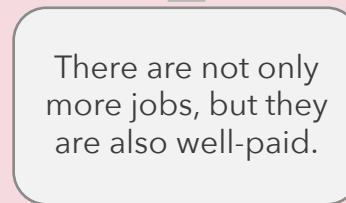
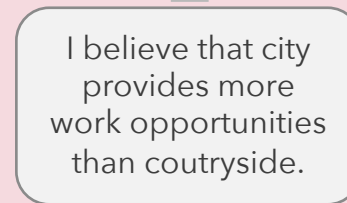
このタスクで

1つ以上のトークンから構成される言語単位に対する分散表現

スパン分散表現



↑ エンコード ↑



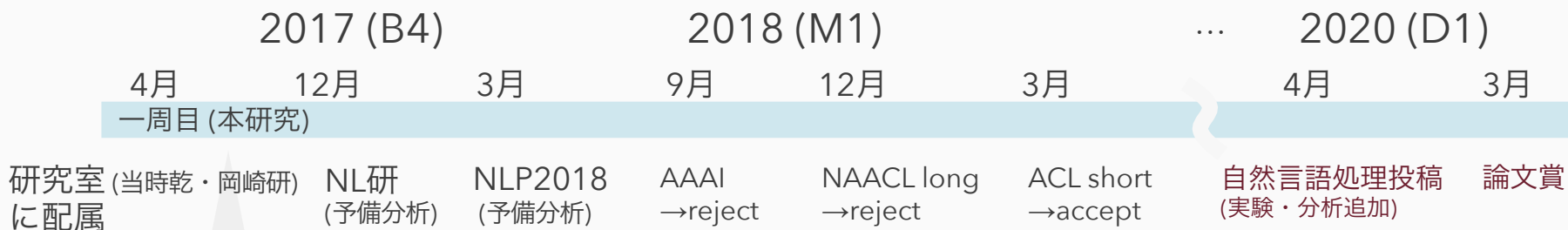
を工夫する

本研究では  
論述単位  
と呼ばれる  
節や文

# 前置き: 本論文の軌跡



# 前置き: 本論文の軌跡



二周目 (word order)

三周目 (cognitive modeling)

- 研究室に配属されて一周目の研究 (B4~M1)

- 文脈の解析 (談話処理) が難しいらしい @B4の基礎勉強会

- 現在も研究活動の関心の根は談話に (手前味噌ですが宣伝)

- 談話構造解析 (接続・論理的つながり): 本研究
- 主題構造・とりたて: **E7-1** 人と言語モデルが捉える文の主題 [藤原+, 第二著者], ACL2020 Kuribayashi+ (word order)
- 情報構造に関する分析を含む: **P4-1** 日本語の読みやすさに対する情報量に基づいた統一的な解釈 [筆頭]
- 物語中のイベントの顕現性: **E7-4** 物語におけるイベントの顕現性推定と物語類似性計算への応用 [大竹+, 共著]
- 文章処理の認知モデリング: **C2-3** 予測の正確な言語モデルがヒトらしいとは限らない [筆頭], **P4-1**

共著外ですが関連して宣伝:  
P-2 論述への説明性の高いフィード  
バック提示に向けたコーパスの試作  
[内藤+]

### NNモデルの中の言語を解像度高く分析・解釈するために

- A7-2** Transformerの文脈を混ぜる作用と混ぜない作用 [小林+, 第二著者], EMNLP2020 Kobayashi+ (norm, not weight),
- D3-4** 事例ベース依存構造解析のための依存関係表現学習 [大内+, 共著]

推敲支援・社会実装: **P9-15** Langsmith: 人とシステムの協働による論文執筆 [伊藤+, 第二著者], EMNLP2020 Ito+ (the Langsmith)

学振**DC1**: テクストの数理的モデリングと、数理モデルを通じたテキストらしさの解明への挑戦

# 前置き: 本論文の軌跡



## ● 国際会議で二度reject

- アップデートし続けることが大事? 諦めたらそこで試合終了?  
共著者陣のガッツに学んだ.
- META REVIEWER (ACL2019):  
I wonder if this should be a long paper, without the appendix...

# 背景: 議論マイニング (Argument mining)

---

議論 (論述文・ディベートなど) の解析を行うNLPの一分野

科学的興味: 議論や説得の仕組みをデータドリブンに分析


応用的興味: 議論の集約・意思決定の支援 ("why"のmining)

# 背景: 議論マイニング (Argument mining)

議論 (論述文・ディベートなど) の解析を行うNLPの一分野

科学的興味: 議論や説得の仕組みをデータドリブンに分析

応用的興味: 議論の集約・意思決定の支援 ("why"のmining)

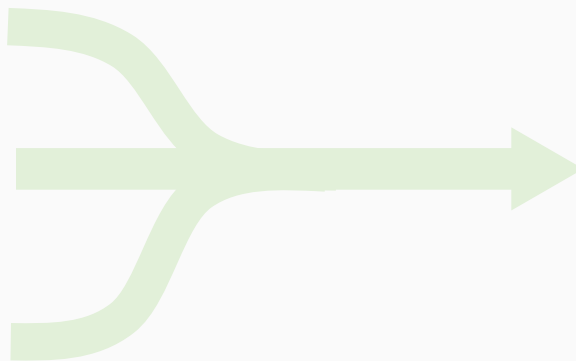
 推定100億ドルの市場があるらしい [Lawrence&Reed, 2019]

Argumentation  
theory

Discourse and  
pragmatics

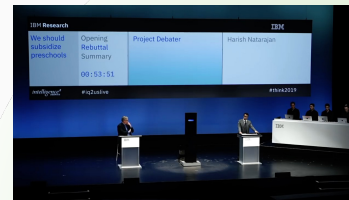
Opinion mining

⋮



## ディベートの自動化

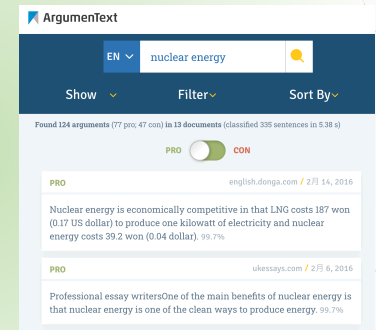
IBMのproject debater teamが52の関連論文



<https://www.research.ibm.com/artificial-intelligence/project-debater/how-it-works/>

## 意見の集約

[Reimers+, 2019]  
[Dexenberger+, 2020], ...



<https://www.argumentsearch.com/>

## 議論の質の評価

[Wachsmuth+, 2017][Gretz+, 2020][Paul+, 2019],...



# 背景: 議論マイニング (Argument mining)

議論 (論述文・ディベートなど) の解析を行うNLPの一分野

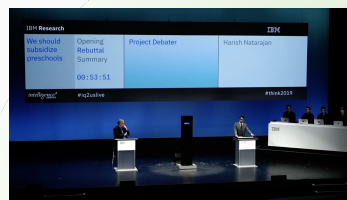
科学的興味: 議論や説得の仕組みをデータドリブンに分析

応用的興味: 議論の集約・意思決定の支援 ("why"のmining)

💡 推定100億ドルの市場があるらしい [Lawrence&Reed, 2019]

## ディベートの自動化

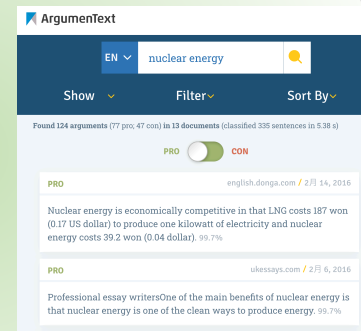
IBMのproject debater teamが52の関連論文



<https://www.research.ibm.com/artificial-intelligence/project-debater/how-it-works/>

## 意見の集約

[Reimers+, 2019]  
[Dexenberger+, 2020], ...



<https://www.argumentsearch.com/>

## 議論の質の評価

[Wachsmuth+, 2017][Gretz+, 2020][Paul+, 2019],...



Argumentation theory  
Discourse and pragmatics  
Opinion mining



### 研究方向の多様化:

論述の生成, 言語・ドメイン横断的研究, より高度なアノテーションなど.

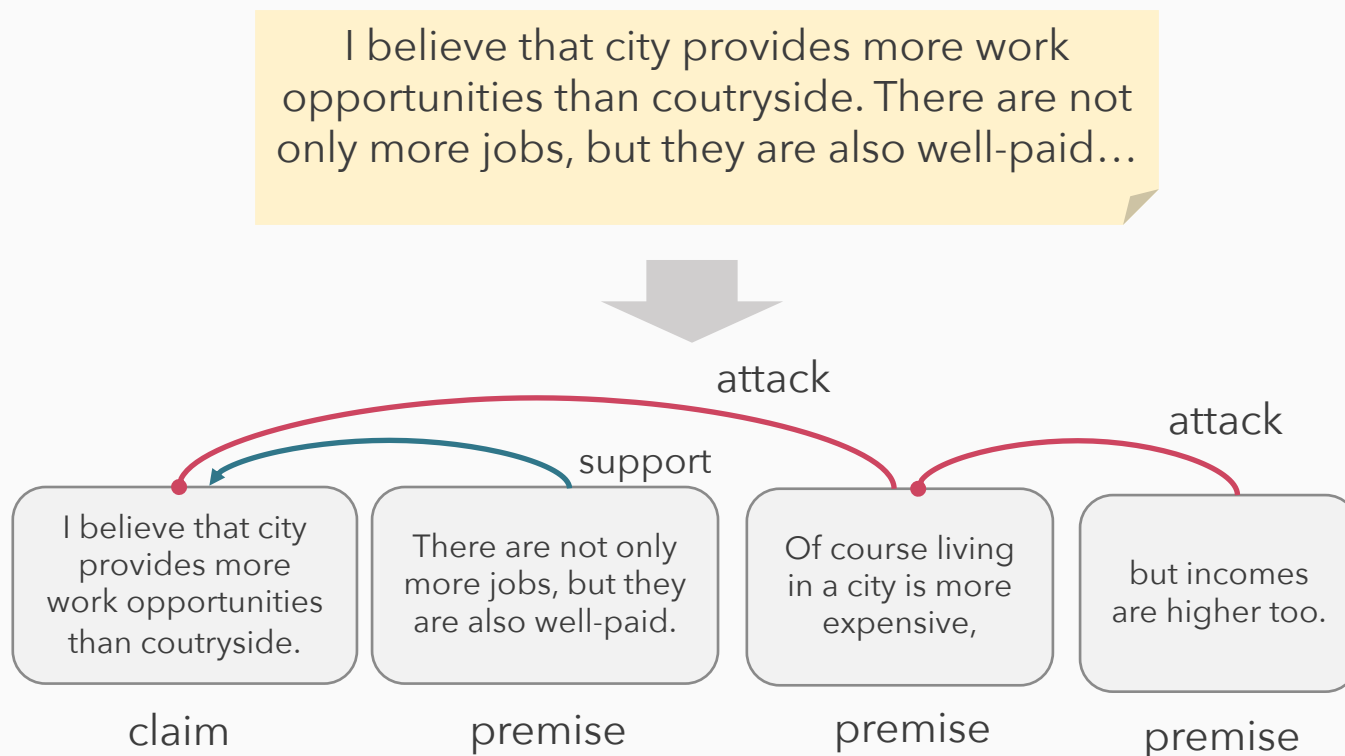
ちなみにsentence-BERT [Reimers&Grevych, 2019] も argument miningの文脈 [Reimers+, 2019] で登場



# 背景: (典型的な) 論述構造解析

## Argument miningにおける基礎解析タスク・技術

論述単位 (談話単位) の論述における役割, 単位間の関係を同定



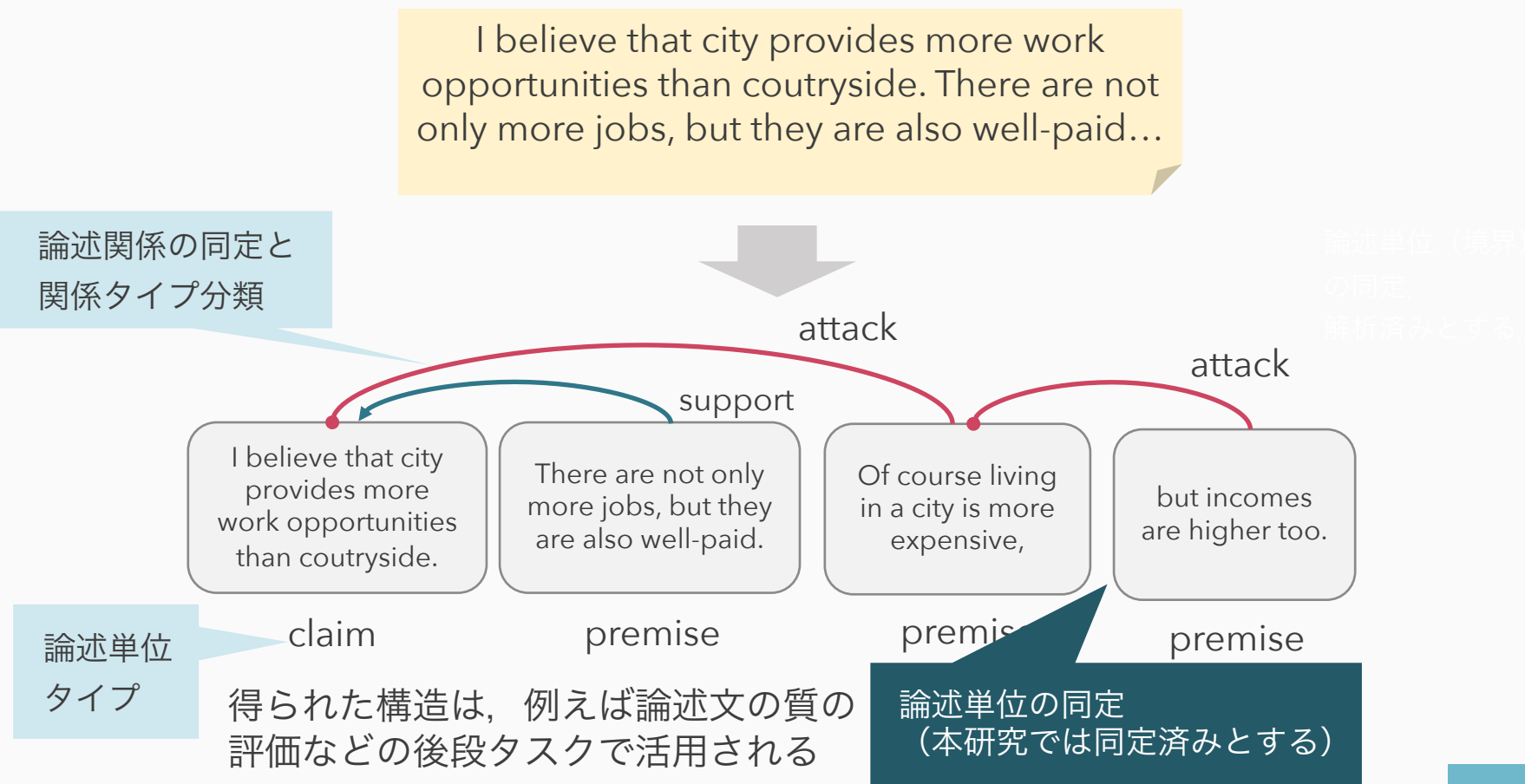
得られた構造は, 例えば論述文の質の評価などの後段タスクで活用される

[Stab&Gurevych, 2016] [Nguyen&Litman, 2018]

# 背景: (典型的な) 論述構造解析

## Argument miningにおける基礎解析タスク・技術

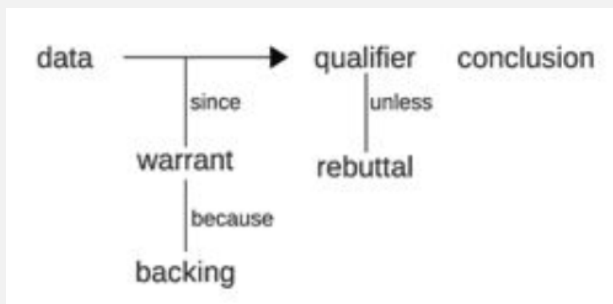
論述単位 (談話単位) の論述における役割, 単位間の関係を同定



# 背景: (典型的な) 論述構造解析

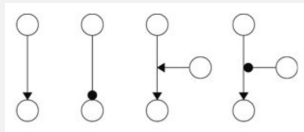
## 議論学の知見を踏まえたアノテーション方針

### Toulmin 1958

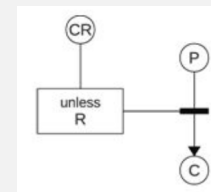


### Klein 1980 recursive nature

### Grewendorf 1980 representation of opponent (attack)



### Freeman 1991 counter-rebuttal

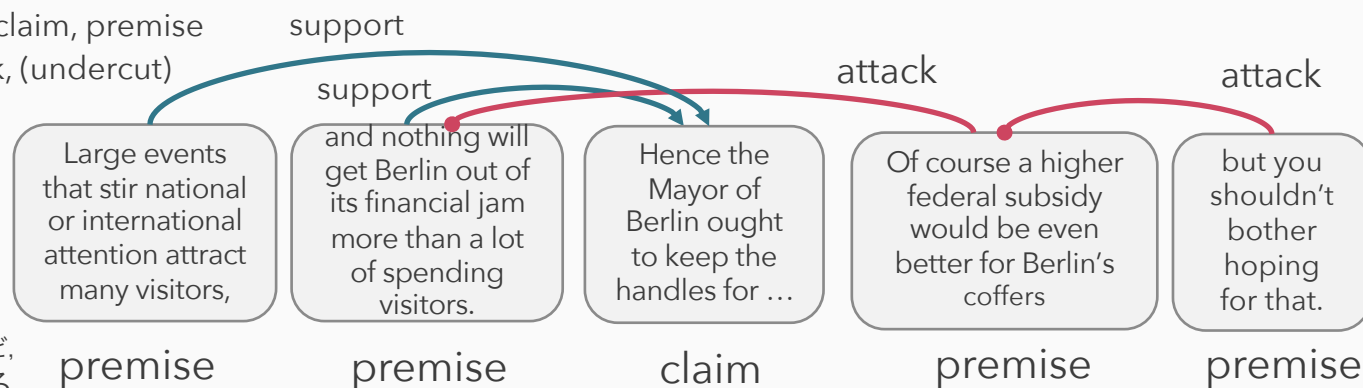


💡 RSTとの違いは？  
[Peldszus&Stede,2013][Stede+,2016]

## 交差を許す依存構造木

論述単位タイプ: (MajorClaim), claim, premise

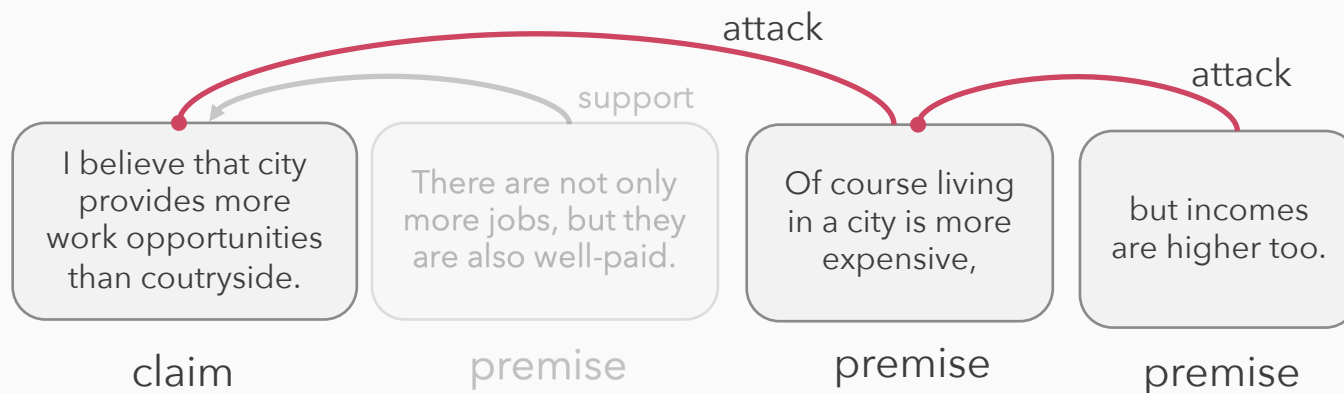
論述関係タイプ: support, attack, (undercut)



💡 よりリッチな情報を付与する研究 [Carlie+,2018][Egawa+,2019]など, 木構造を仮定しない研究もある [Niculae+,2017]など

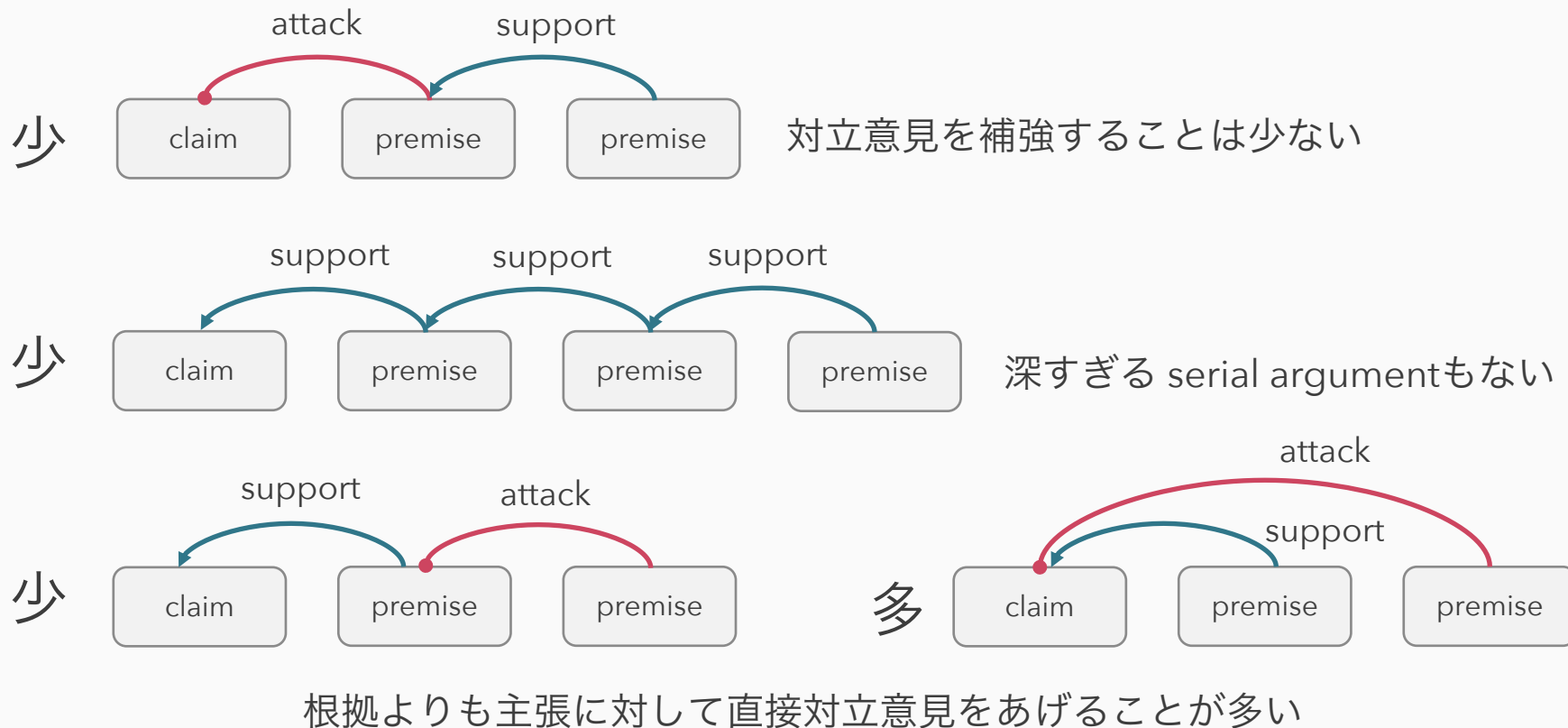
# 予備分析: マクロな構造 [Kuribayashi+, 2017]

関係予測の手がかり: 典型的なマクロな構造の存在



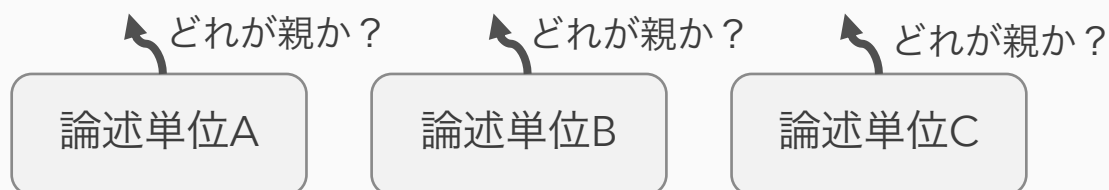
例えば, 主張に対する対立意見に言及した場合,  
その対立意見は反論し返されるのが自然  
(Freeman's concept of counter-rebuttal)

# 予備分析: マクロな構造 [Kuribayashi+, 2017]



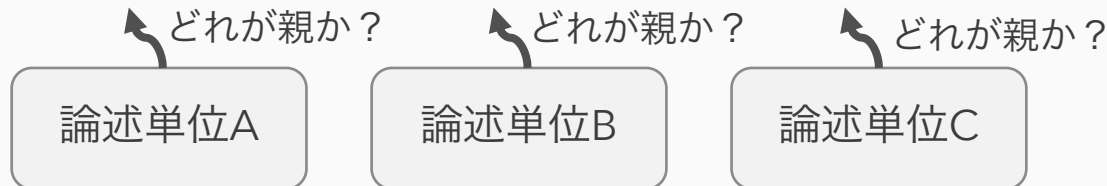
# 仮説：文章全体の（大域的な）情報を考慮する必要がある

- 前提: 各論述単位の親（関係）を当てる部分問題に分解して解く



# 仮説：文章全体の（大域的な）情報を考慮する必要がある

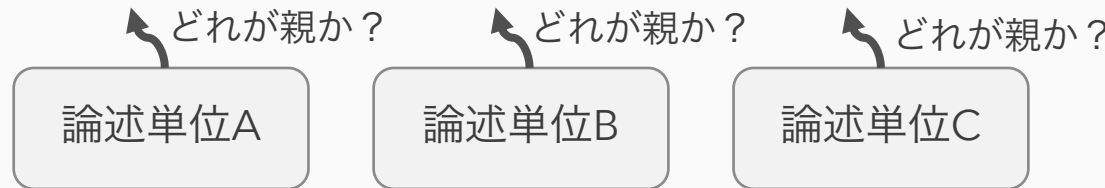
- 前提: 各論述単位の親（関係）を当てる部分問題に分解して解く



- 問題点: 文章全体の（大域的な）情報を加味すべき問題であるにも関わらず、既存手法ではスパン（論述単位）表現としてスパン内の局所的な情報のみが用いられる傾向

# 仮説：文章全体の（大域的な）情報を考慮する必要がある

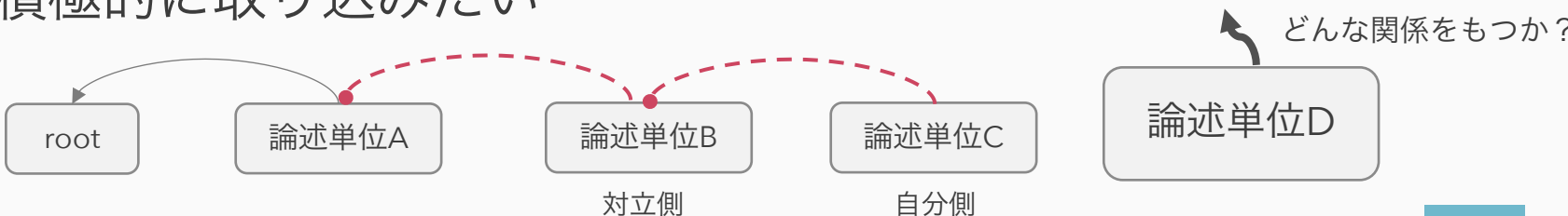
- 前提: 各論述単位の親（関係）を当てる部分問題に分解して解く



- 問題点: 文章全体の大域的な情報を加味すべきも関わらず、既存手法ではスパン（論）としてスパン内の局所的な情報のみ

特に本タスク（というか談話構造予測タスク全般）は学習データが少ないため、マクロな傾向をデータのみから捉えることに限界がある可能性

- スパン表現をエンコードする際、大域的、談話的な文脈を積極的に取り込みたい





# 関連研究: 高次の特徴量を考慮した交差を許す 依存構造解析 (high-order non-projective dependency parsing)

- 高次の情報（関係間の関係）を考慮する
  - second-order non-projective maximum spanning tree decoding (NP困難) を近似 [McDonald&Pereira, 2006],[Carreras, 2007]...
  - parsing historyを考慮 [Yamada&Matsumoto, 2003] [transition-based parsing全般] [pointer network系]...
  - ニューラルモデルに高次の特徴を考慮する機構を明に追加 [Pei+, 2015][Lee+, 2018]...



# 関連研究: 高次の特徴量を考慮した交差を許す 依存構造解析 (high-order non-projective dependency parsing)

- 高次の情報（関係間の関係）を考慮する
  - second-order non-projective maximum spanning tree decoding (NP困難) を近似 [McDonald&Pereira, 2006],[Carreras, 2007]...
  - parsing historyを考慮 [Yamada&Matsumoto, 2003] [transition-based parsing全般] [pointer network系]...
  - ニューラルモデルに高次の特徴を考慮する機構を明に追加 [Pei+, 2015][Lee+, 2018]...



# 関連研究: 高次の特徴量を考慮した交差を許す依存構造解析 (high-order non-projective dependency parsing)

- 高次の情報（関係間の関係）を考慮する

- second-order non-projective maximum spanning tree decoding (NP困難) を近似 [McDonald&Pereira, 2006],[Carreras, 2007]...
- parsing historyを考慮 [Yamada&Matsumoto, 2003] [transition-based parsing全般] [pointer network系]...
- ニューラルモデルに高次の特徴を考慮する機構を明に追加 [Pei+, 2015][Lee+, 2018]...
- LSTMで文脈化された表現 (LSTM-minus) を用いることで、高次の素性を考慮した解析器を凌ぐ性能がでる [Wang&Chang, 2016]
- 当時ELMoが注目され始めており、分散表現の文脈化への期待

decoding algorithmの工夫



encoder側で文脈を見る

# 関連研究: 高次の特徴量を考慮した交差を許す依存構造解析 (high-order non-projective dependency parsing)

- 高次の情報（関係間の関係）を考慮する

- second-order non-projective maximum spanning tree decoding (NP困難) を近似 [McDonald&Pereira, 2006],[Carreras, 2007]...
- parsing historyを考慮 [Yamada&Matsumoto, 2003] [transition-based parsing全般] [pointer network系]...
- ニューラルモデルに高次の特徴を考慮する機構を明に追加 [Pei+, 2015][Lee+, 2018]...
- LSTMで文脈化された表現 (LSTM-minus) を用いることで、高次の素性を考慮した解析器を凌ぐ性能がでる [Wang&Chang, 2016]
- 当時ELMoが注目され始めており、分散表現の文脈化への期待

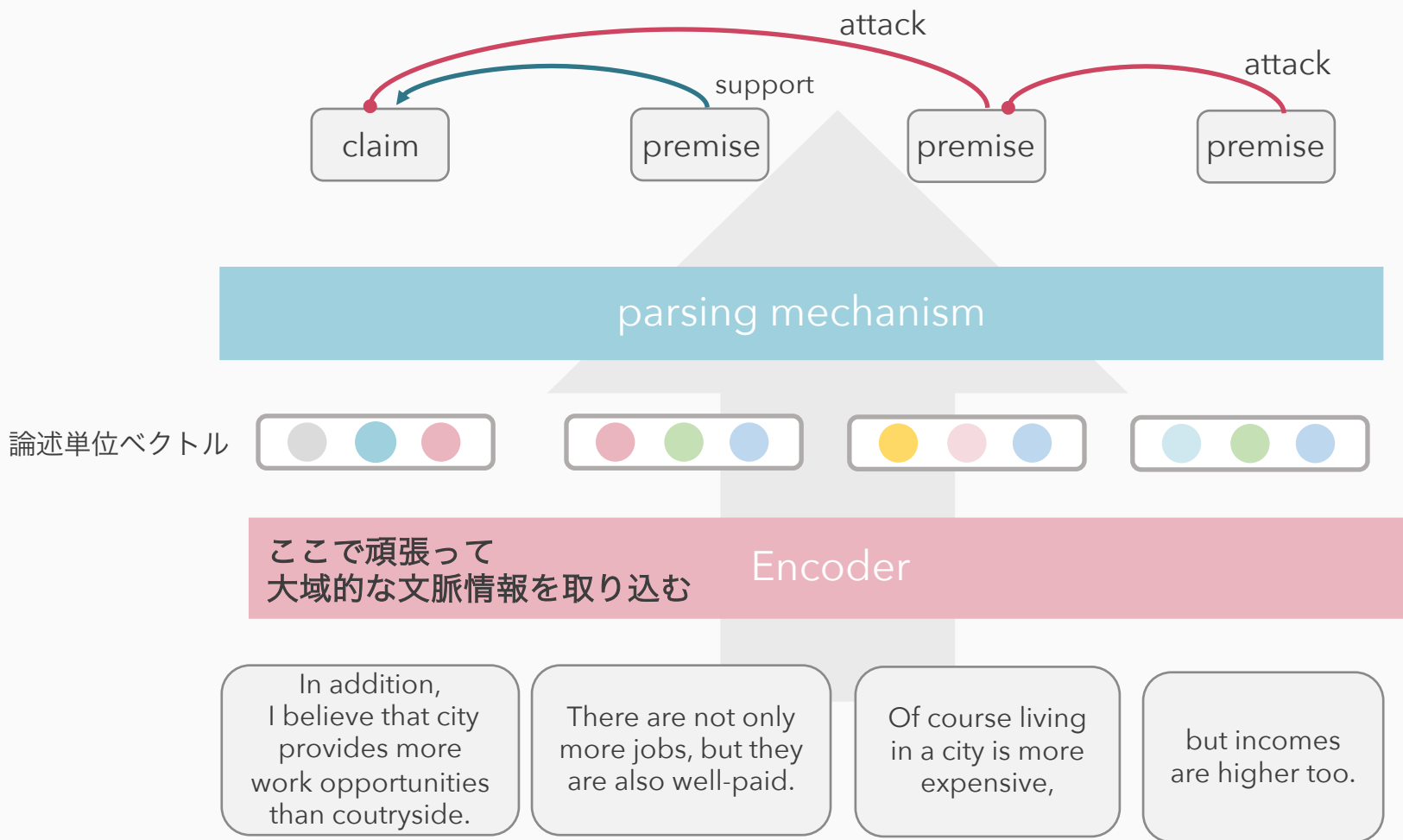
decoding algorithmの工夫

encoder側で文脈を見る

- 本研究のスタンス: 論述単位のエンコード時に、積極的に文脈を考慮するよう促せば大域的な（高次の）傾向を考慮できる

💡 文脈化された単語分散表現の活用と高次の特徴量の活用の関係については近年も分析が続いている [Fonseca&Martins,2020]

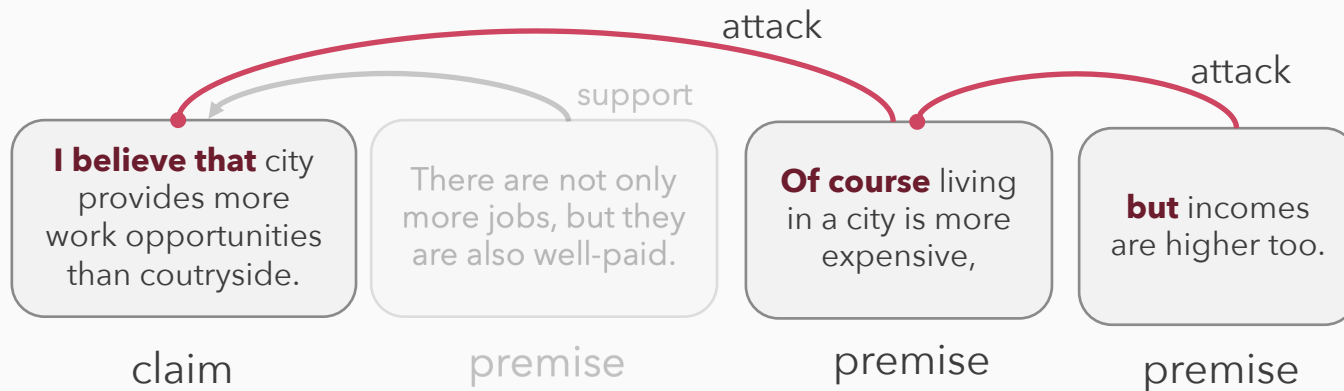
# エンコーダの強化





# 改良2: 命題と接続表現の分離・多粒度の文脈化

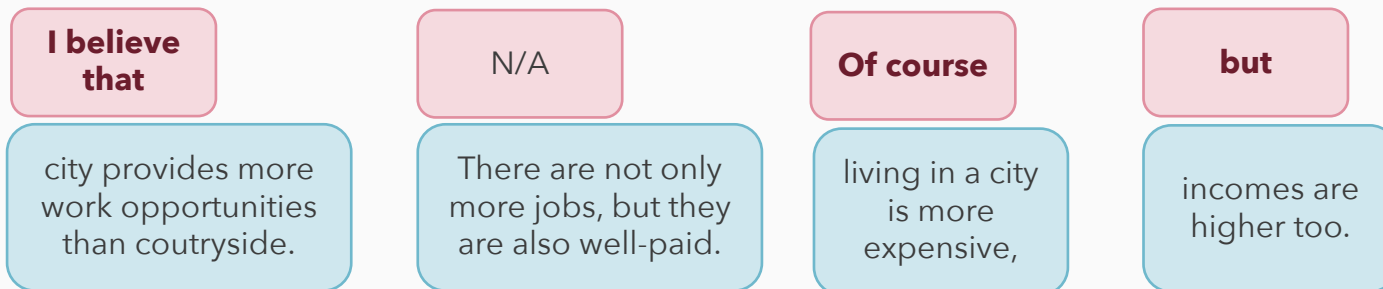
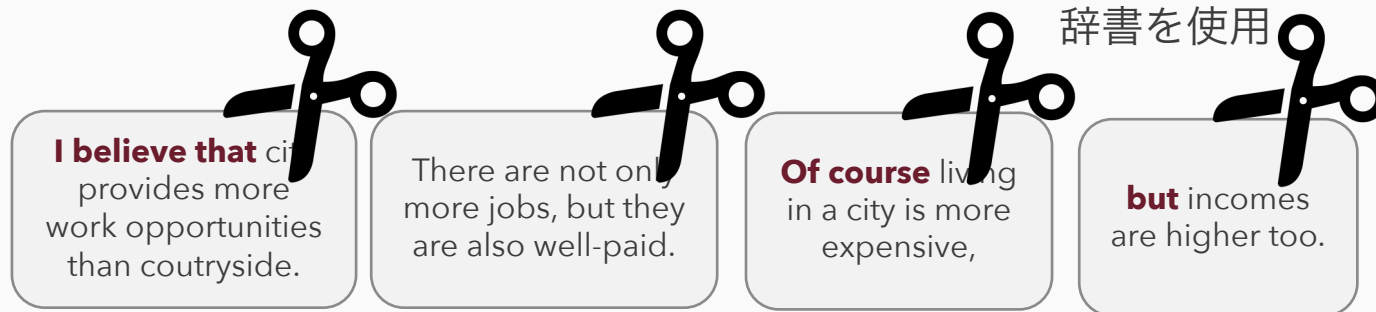
- 談話関係同士の大域的な傾向を捉えたい
  - 前提: 談話関係はしばしば接続表現（論述マーカ）によって示される
  - 仮説: 文章全体における接続表現の配置を考慮することで、  
論述の大域的なパターンを捉えやすくなる



# 改良2: 命題と接続表現の分離・多粒度の文脈化

- 談話関係同士の大域的な傾向を捉えたい
  - 前提: 談話関係はしばしば接続表現（論述マーカ）によって示される
  - 仮説: 文章全体における接続表現の配置を考慮することで、  
論述の大域的なパターンを捉えやすくなる

既存のアノテーションや辞書を使用

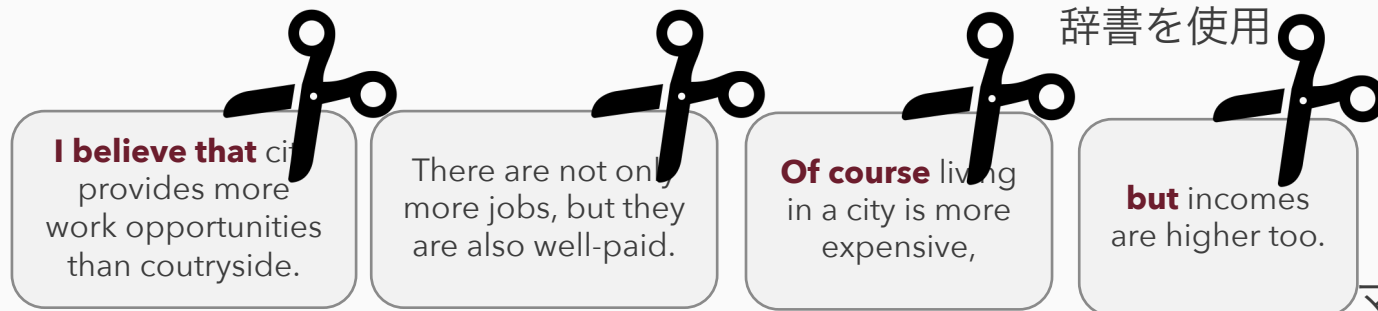




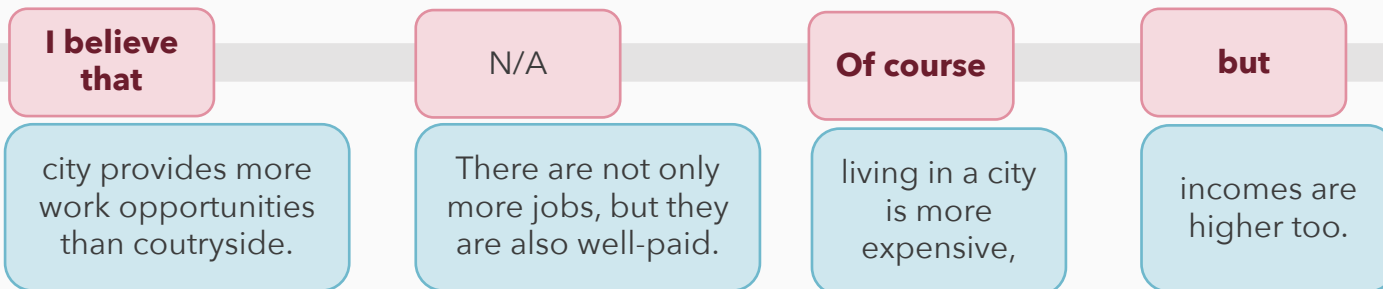
# 改良2: 命題と接続表現の分離・多粒度の文脈化

- 談話関係同士の大域的な傾向を捉えたい
  - 前提: 談話関係はしばしば接続表現（論述マーカ）によって示される
  - 仮説: 文章全体における接続表現の配置を考慮することで、  
論述の大域的なパターンを捉えやすくなる

既存のアノテーションや辞書を使用



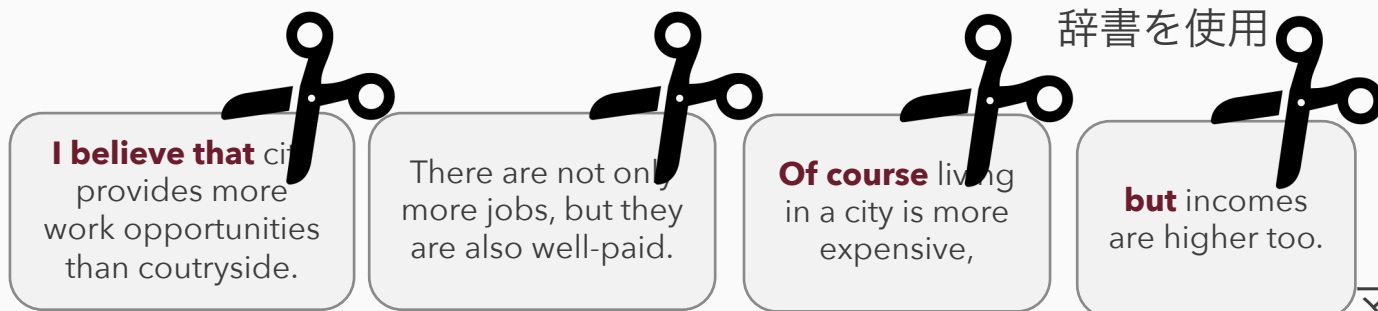
マクロな議論の流れ  
(議論の型)



# 改良2: 命題と接続表現の分離・多粒度の文脈化

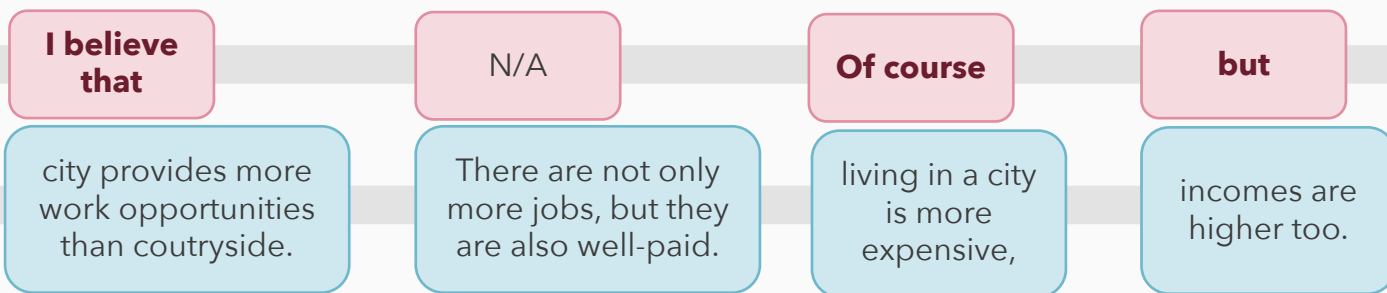
- 談話関係同士の大域的な傾向を捉えたい
  - 前提: 談話関係はしばしば接続表現（論述マーカ）によって示される
  - 仮説: 文章全体における接続表現の配置を考慮することで、  
論述の大域的なパターンを捉えやすくなる

既存のアノテーションや辞書を使用

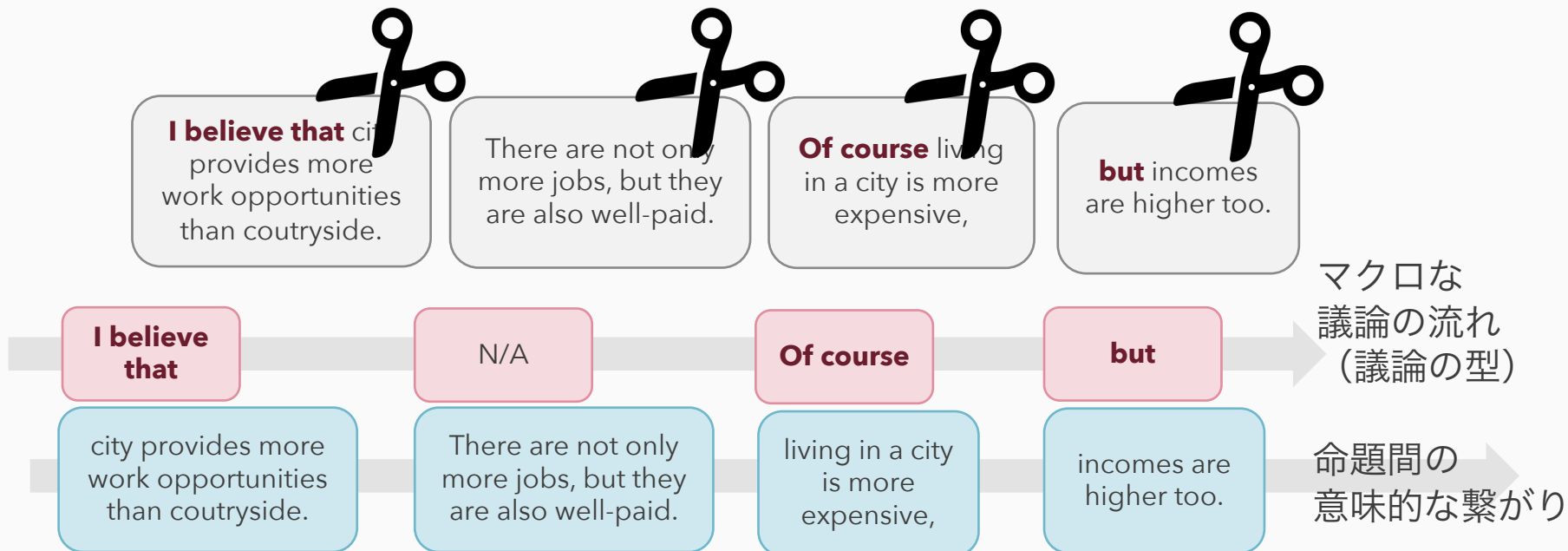


マクロな 相補的  
議論の流れ  
(議論の型)

命題間の  
意味的な繋がり



# 改良2: 命題と接続表現の分離・多粒度の文脈化



💡 様々な見方が可能

- 論題非依存の大域的構造（議論の型）と論題依存の内容を区別

テキスト言語学[Halliday&Hasan,1994]の視点から

- 結束性の範疇の区別（例えば「接続」と命題間の「語彙的結束性」など）をモデルアーキテクチャに取り込む

選択体系機能文法[Halliday,2014]の視点から

- テキスト形成的主題とそれ以外（題術含む）の区別

## 改良2: 命題と接続表現の分離・多粒度の文脈化

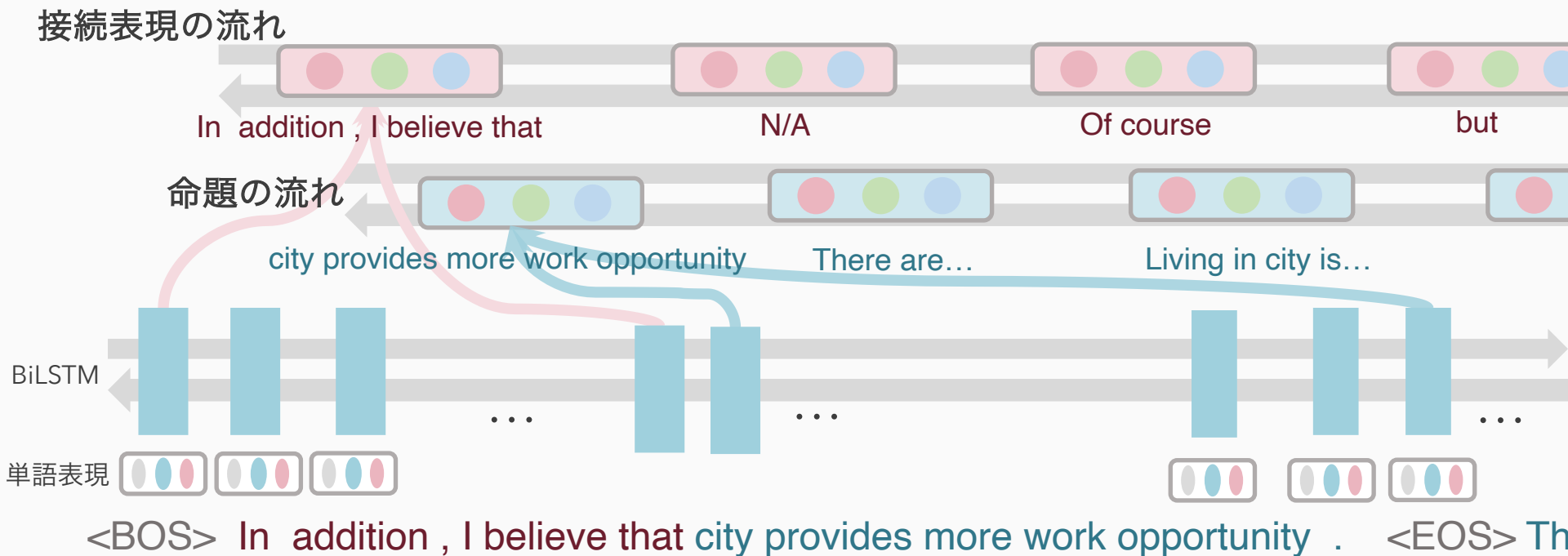
単語表現   

...

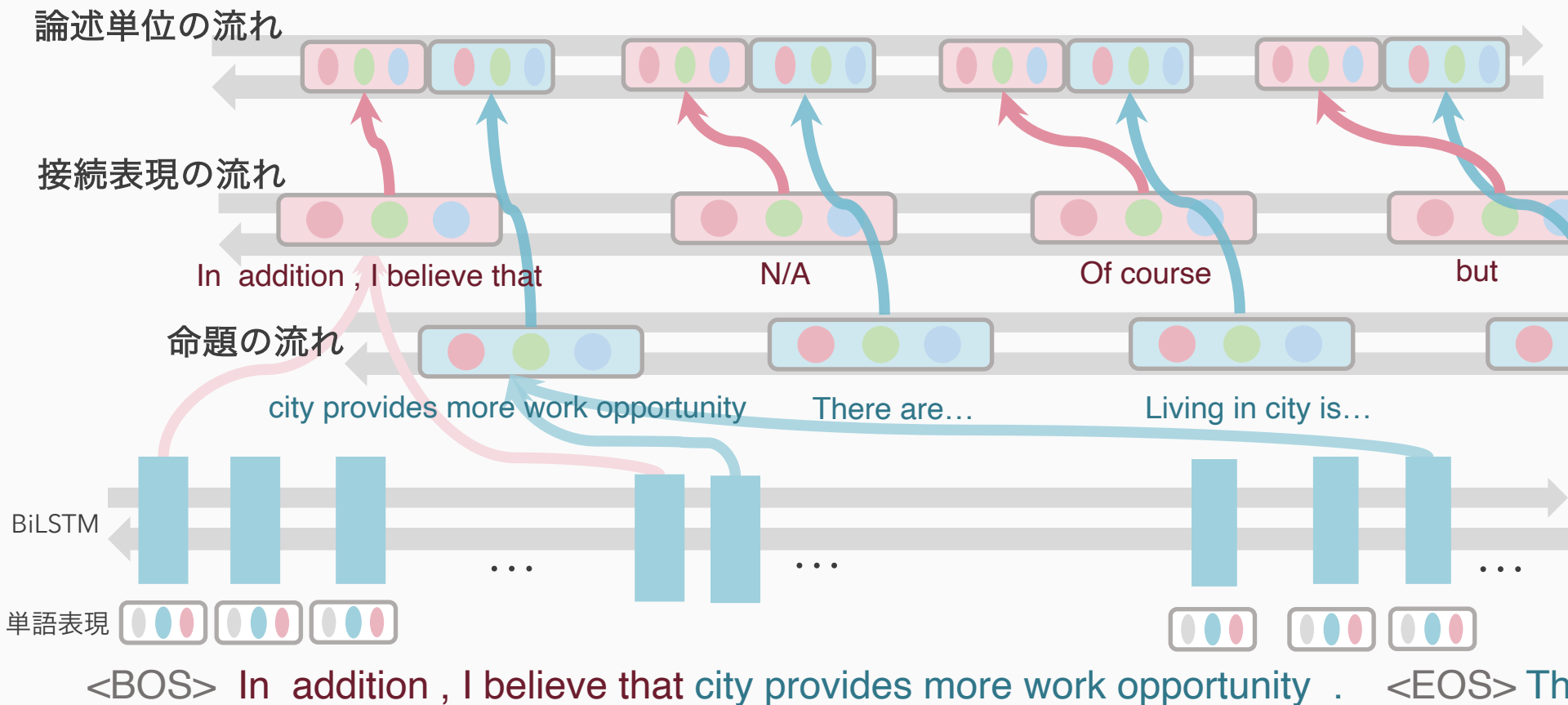
  

<BOS> In addition , I believe that city provides more work opportunity . <EOS> Th

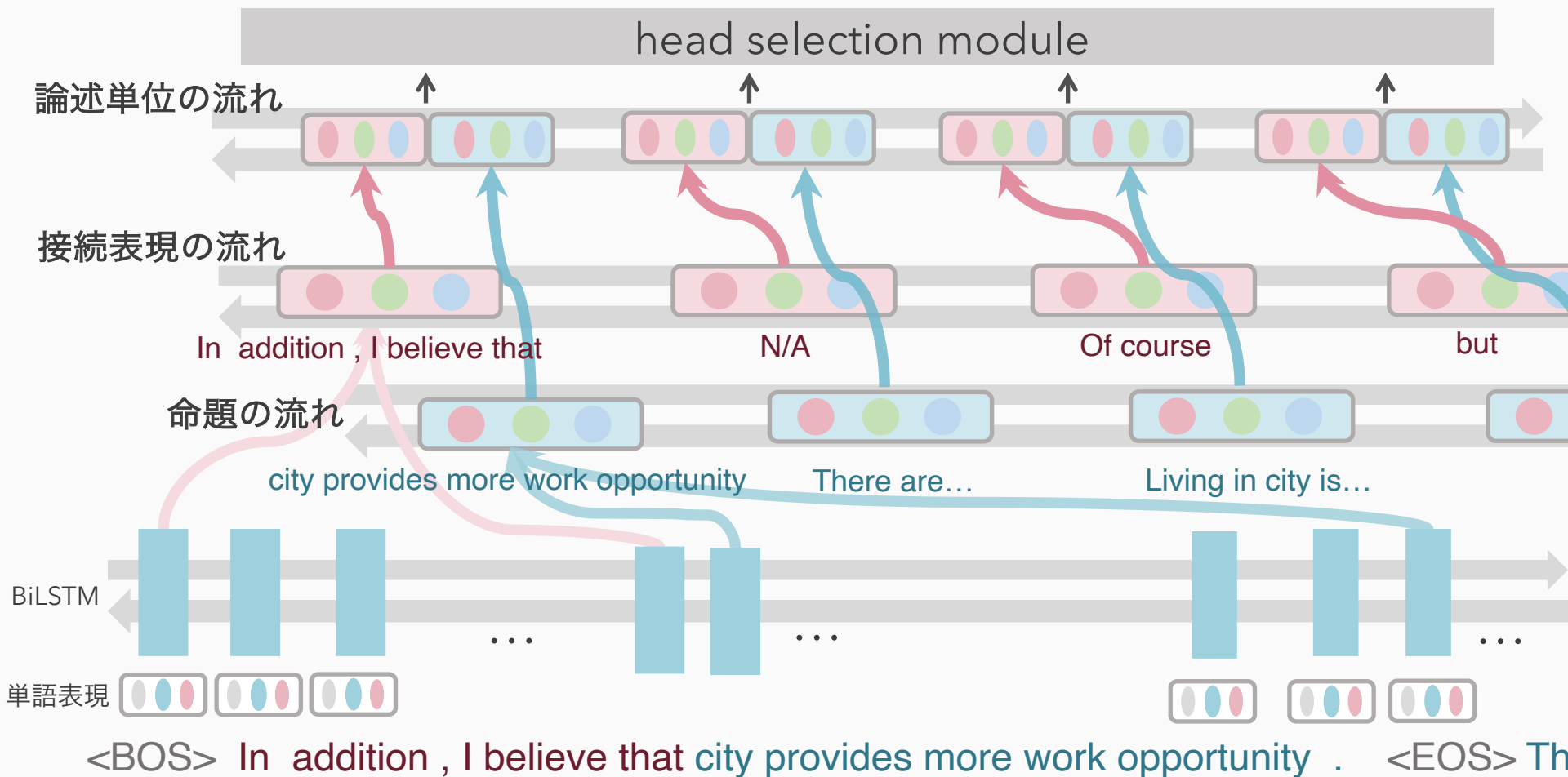
# 改良2: 命題と接続表現の分離・多粒度の文脈化



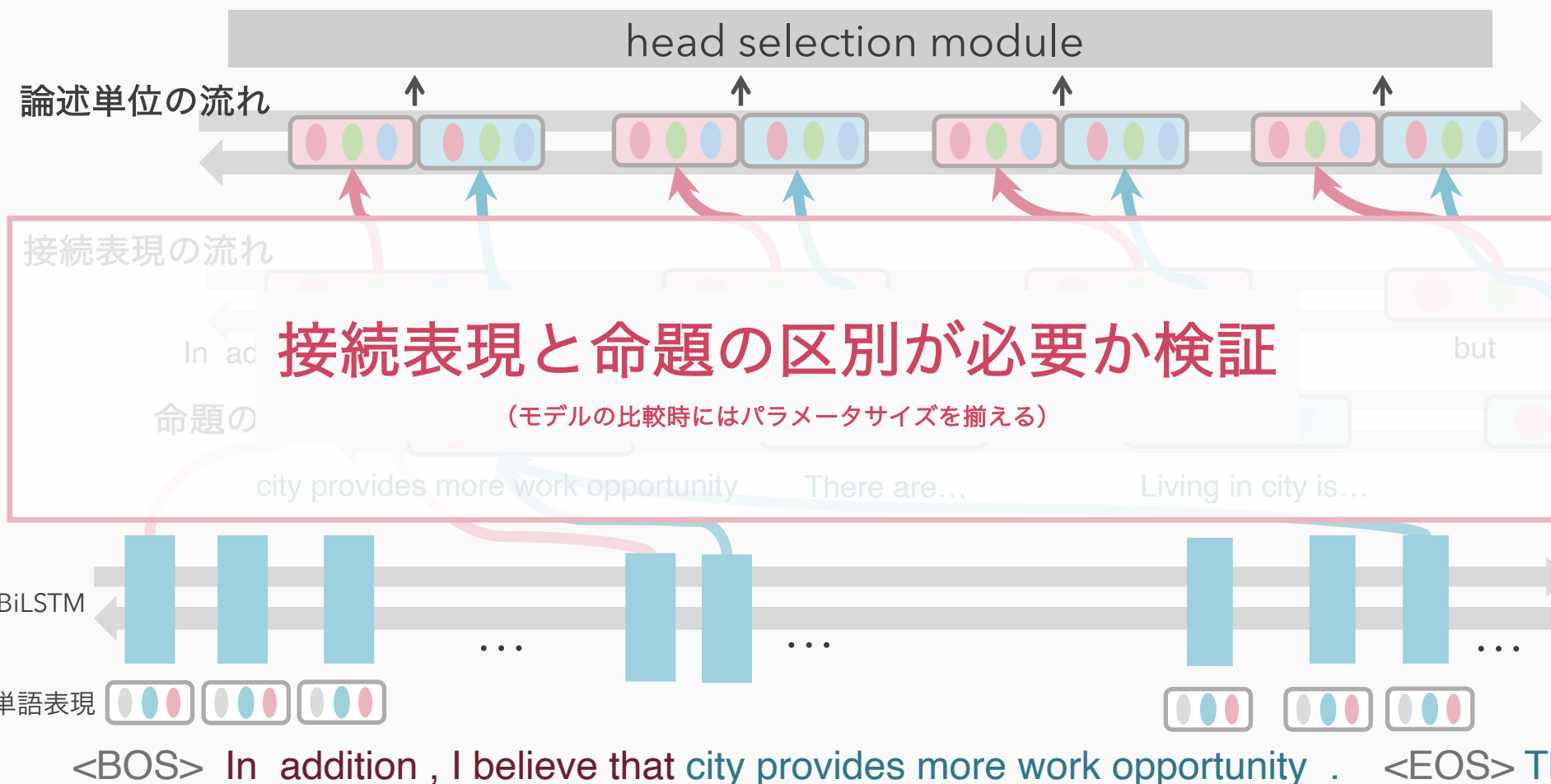
# 改良2: 命題と接続表現の分離・多粒度の文脈化



# 改良2: 命題と接続表現の分離・多粒度の文脈化



# 実験: 時間の都合上, 接続表現と命題の区別を行うこと (改良2) による性能向上についてのみ言及





# 結果: 10の実験設定 (2種類のコーパス×5種類の単語埋め込み) で性能に良い影響

コーパス	単語埋め込み	区別	関係同定 (macro-F)	関係タイプ予測 (macro-F)	論述単位タイプ (macro-F)
Persuasive Essay Corpus [Stab&Gurevych, 2017]	XLNET	○	<b>82.2</b>	<b>78.5</b>	<b>89.4</b>
		×	80.6	77.8	89.1
	RoBERTa	○	80.9	<b>79.4</b>	88.4
		×	<b>81.6</b>	77.9	<b>89.1</b>
	BERT	○	<b>80.9</b>	<b>78.0</b>	86.4
		×	80.4	74.9	<b>86.6</b>
	ELMo	○	<b>80.7</b>	<b>79.0</b>	85.7
		×	80.3	78.2	<b>86.9</b>
	GloVe	○	<b>78.8</b>	<b>76.5</b>	<b>83.9</b>
		×	77.7	75.0	83.7
Arg-microtext corpus [Peldszus&Stede, 2016]	XLNET	○	<b>75.8</b>	<b>73.7</b>	<b>88.3</b>
		×	73.6	69.3	85.2
	RoBERTa	○	<b>74.0</b>	<b>71.6</b>	<b>85.9</b>
		×	71.8	66.4	80.9
	BERT	○	<b>73.0</b>	<b>71.1</b>	<b>84.0</b>
		×	69.9	64.2	77.9
	ELMo	○	<b>73.9</b>	<b>77.2</b>	<b>83.4</b>
		×	73.2	71.3	80.5
	GloVe	○	<b>72.6</b>	<b>75.4</b>	<b>81.5</b>
		×	70.4	64.1	76.9



# 分析: 典型的な部分構造の予測

attackの連鎖をなす部分構造における予測性能



コーパス	単語埋め込み	区別	関係予測精度 (acc.)
Persuasive Essay Corpus [Stab&Gurevych, 2017]	XLNET	○	71.6
		×	67.8
	RoBERTa	○	69.9
		×	67.2
	BERT	○	68.8
		×	69.3
	ELMo	○	71.0
		×	68.9
	GloVe	○	63.9
		×	59.0

next sentence predictionで、既に談話的な情報が少なからず入っている??

# まとめ

論述構造解析において、

- 論述単位の表現を積極的に文脈化することにより、特に典型的な部分構造や複雑な構造 (詳細は論文) における解析性能が向上
- 接続表現と命題を区別し、複数の粒度で文章全体の流れを捉えることの有効性を示した
- BERTやXLNETなどの強力な言語モデルから得られた単語表現を用いた場合でも、後段タスクへの転用時にスパン表現を洗練させることは有効
- 良くも悪くも文章の表層的なパターンを捉えにいった研究
  - 💡 接続表現に過度に依拠した予測が行われているという分析も [Optiz&Frank, 2019]
  - 💡 (より深い理解について問うために) モデルの批判的思考能力を評価・分析する研究も近年でてきている [Betz+,2020][Yu+,2019]

# References

---

Lawrence, J. and Reed, C. (2019). "Argument Mining: A Survey." *Computational Linguistics*, 45, pp. 765-818

Wachsmuth, H., Naderi, N., Hou, Y., Bilu, Y., Prabhakaran, V., Thijmm, T. A., Hirst, G., and Stein, B. (2017). "Computational Argumentation Quality Assessment in Natural Language." *Proceedings of the EACL2017*, pp. 176-187.

Gretz, S., Friedman, R., Cohen-Karlik, E., Toledo, A., Lahav, D., Aharonov, R., & Slonim, N. (2020). A Large-Scale Dataset for Argument Quality Ranking: Construction and Analysis. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05), 7805-7813.

Reisert, P., Heinzerling, B., Inoue, N., Kiyono, S., & Inui, K. (2019). Riposte! A Large Corpus of Counter-Arguments. *arXiv e-prints*, arXiv-1910.

Reimers, N., Gurevych, I., Reimers, N., Gurevych, I., Thakur, N., Reimers, N., Daxenberger, J. and Gurevych, I., 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the EMNLP2019*.

Reimers, N., Schiller, B., Beck, T., Daxenberger, J., Stab, C., & Gurevych, I. (2019, July). Classification and Clustering of Arguments with Contextualized Word Embeddings. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 567-578).

Daxenberger, J., Schiller, B., Stahlhut, C. et al. ArgumenText: Argument Classification and Clustering in a Generalized Search Scenario. *Datenbank Spektrum* **20**, 115-121 (2020). <https://doi.org/10.1007/s13222-020-00347-7>

Nguyen, H. and Litman, D. (2018). "Argument Mining for Improving the Automated Scoring of Persuasive Essays.", *AAAI 2018*, pp. 5892-5899

Stab, C. and Gurevych, I. (2016). "Recognizing Insufficiently Supported Arguments in Argumentative Essay." *Proceedings of the Argmining 2016*, pp. 113-118

# References

---

Carlile, W., Gurrupadi, N., Ke, Z., & Ng, V. (2018, July). Give me more feedback: Annotating argument persuasiveness and related attributes in student essays. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 621-631).

Egawa, R., Morio, G., & Fujita, K. (2019, July). Annotating and analyzing semantic role of elementary units and relations in online persuasive arguments. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop* (pp. 422-428).

Niculescu, V., Park, J., & Cardie, C. (2017, July). Argument Mining with Structured SVMs and RNNs. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 985-995).

KURIBAYASHI, T., REISERT, P., INOUE, N., & INUI, K. (2017). Examining Macro-level Argumentative Structure Features for Argumentative Relation Identification (言語理解とコミュニケーション)-(第4回自然言語処理シンポジウム). *電子情報通信学会技術研究報告=IEICE technical report: 信学技報*, 117(367), 37-42.

McDonald, R., & Pereira, F. (2006, April). Online learning of approximate dependency parsing algorithms. In *11th Conference of the European Chapter of the Association for Computational Linguistics*.

Carreras, X. (2007, June). Experiments with a higher-order projective dependency parser. In *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (emnlp-conll)* (pp. 957-961).

Yamada, H., & Matsumoto, Y. (2003, April). Statistical dependency analysis with support vector machines. In *Proceedings of the eighth international conference on parsing technologies* (pp. 195-206).

# References

---

Pei, W., Ge, T., & Chang, B. (2015, July). An effective neural network model for graph-based dependency parsing. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 313-322).

Lee, K., He, L., & Zettlemoyer, L. (2018, June). Higher-Order Coreference Resolution with Coarse-to-Fine Inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)* (pp. 687-692).

Wang, W., & Chang, B. (2016, August). Graph-based dependency parsing with bidirectional LSTM. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 2306-2315).

Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018, June). Deep Contextualized Word Representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (pp. 2227-2237).

Fonseca, E., & Martins, A. F. (2020, July). Revisiting higher-order dependency parsers. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 8795-8800).

Toshniwal, S., Shi, H., Shi, B., Gao, L., Livescu, K., & Gimpel, K. (2020, July). A Cross-Task Analysis of Text Span Representations. In *Proceedings of the 5th Workshop on Representation Learning for NLP* (pp. 166-176).

HALLIDAY, M. (1976). Cohesion in English. *Longman*.

Halliday, M., Matthiessen, C. M., & Matthiessen, C. (2014). *An introduction to functional grammar*. Routledge.

# References

---

- Opitz, J., & Frank, A. (2019, August). Dissecting Content and Context in Argumentative Relation Analysis. In *Proceedings of the 6th Workshop on Argument Mining* (pp. 25-34).
- Betz, G. (2020). Critical Thinking for Language Models. *arXiv preprint arXiv:2009.07185*.
- Yu, W., Jiang, Z., Dong, Y., & Feng, J. (2019, September). ReClor: A Reading Comprehension Dataset Requiring Logical Reasoning. In *International Conference on Learning Representations*.