翻訳元言語における構文構造を利用したニューラル機械翻訳

2017/02/10
東京大学 鶴岡研究室 D2 江里口 瑛子

自己紹介

- 江里口 瑛子 (えりぐち あきこ)
  - 2015年3月　　お茶の水女子大学 修士課程 修了
  - 2015年4月~　東京大学 工学系研究科 博士課程 在籍
  - 2016年8~11月　ニューヨーク大学に研究滞在
  - 研究テーマ: ニューラル機械翻訳

2016年11月某日のこと

コンニチハ。
ゴキゲン ハ イカガ デスカ。

「…!?」

機械翻訳技術は身近なもの

一緒に住んでいたアメリカ人の11歳の女の子が

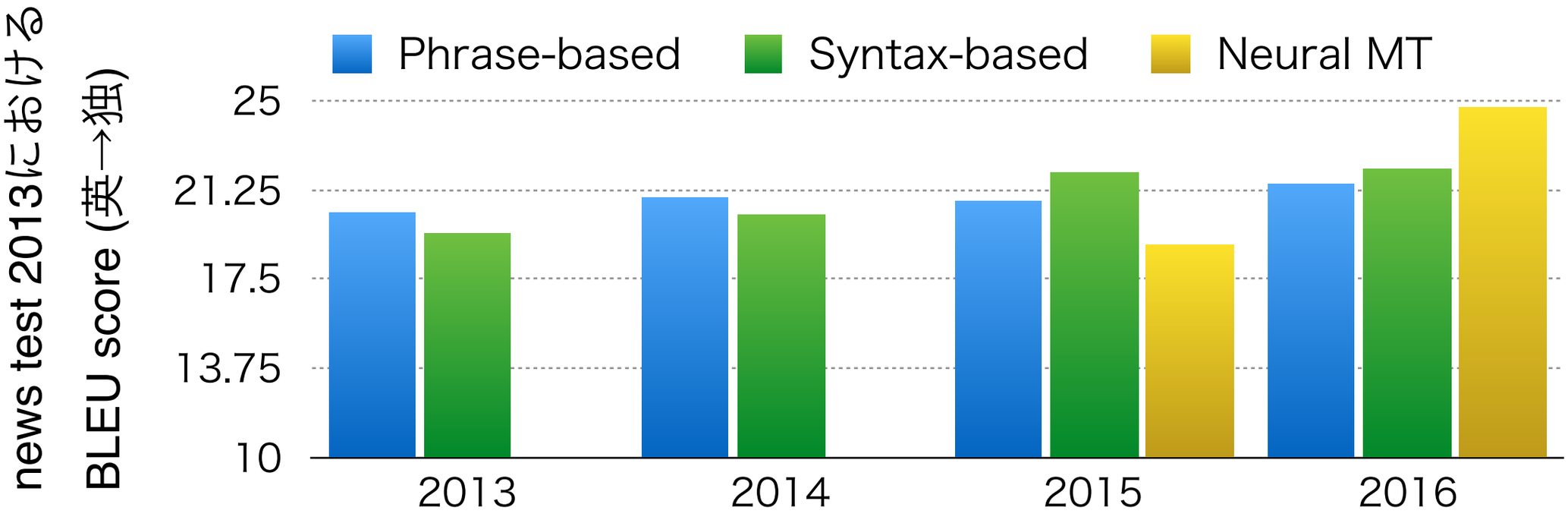
Google翻訳を使って英日翻訳+出力文の音声読み上げを行っていた

多言語処理

- 英日に限らず、世界には多種多様な言語が存在
- 複数言語間の処理がしたい
  - 翻訳
  - 多言語間文書分類
- 人間による翻訳作業、多言語処理はコストが高い
- 計算機による自動処理に注目
  - 翻訳 = 言語間テキスト変換タスク
  - 多言語間文書検索 = (機械翻訳 + 文書分類) タスク

機械翻訳研究の動向

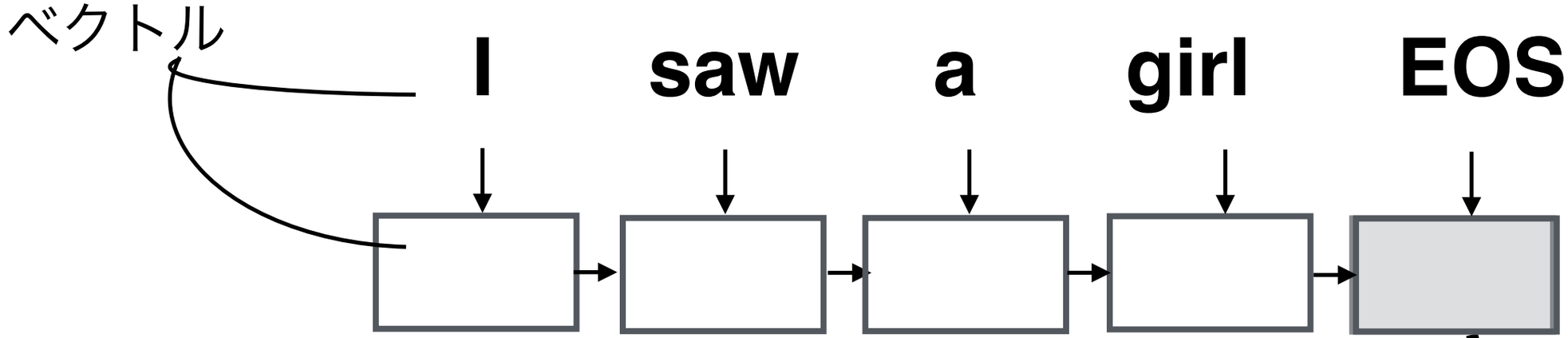
- ニューラル機械翻訳 (NMT) モデルが流行
  - 1つのニューラルネットワークで記述
  - シンプルな仕組み、高性能 (固有の課題あり)
  - Google翻訳 (Wu et al., 2016) などにNMTが導入



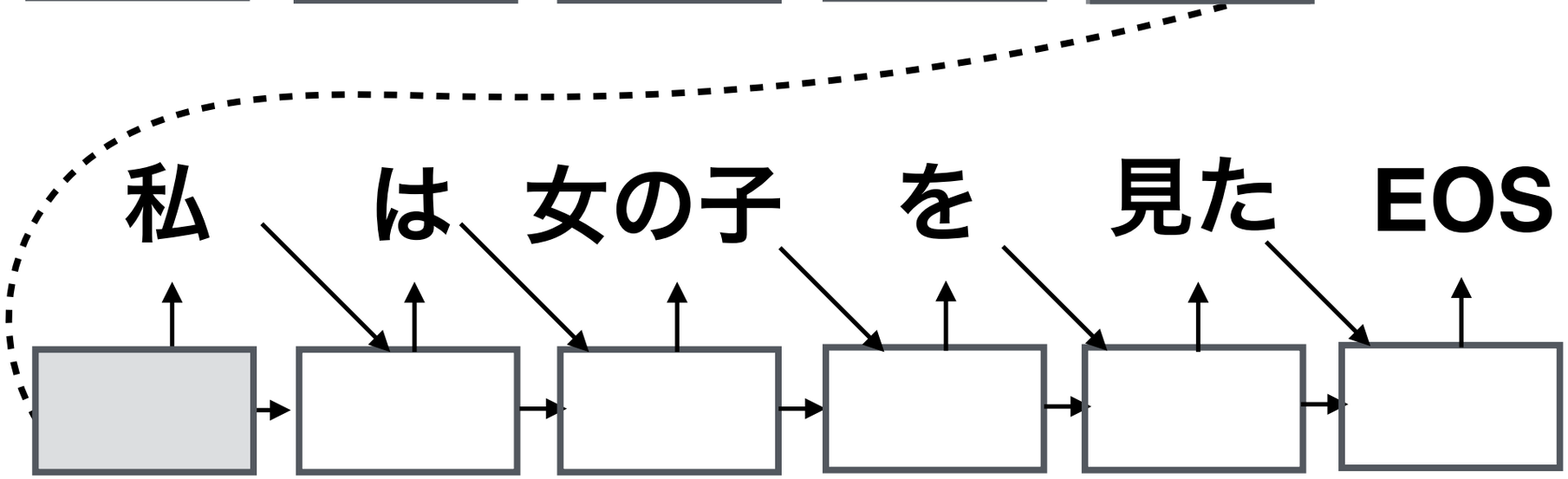
(引用: http://homepages.inf.ed.ac.uk/rsennric/amta2016.pdf)

エンコーダ・デコーダモデル

- ２つのリカレントニューラルネットワークから構成
  - 隠れ層: $\boldsymbol{h}_t = f(\boldsymbol{h}_{t-1}, \boldsymbol{v}_t)$　　(fは非線形関数)



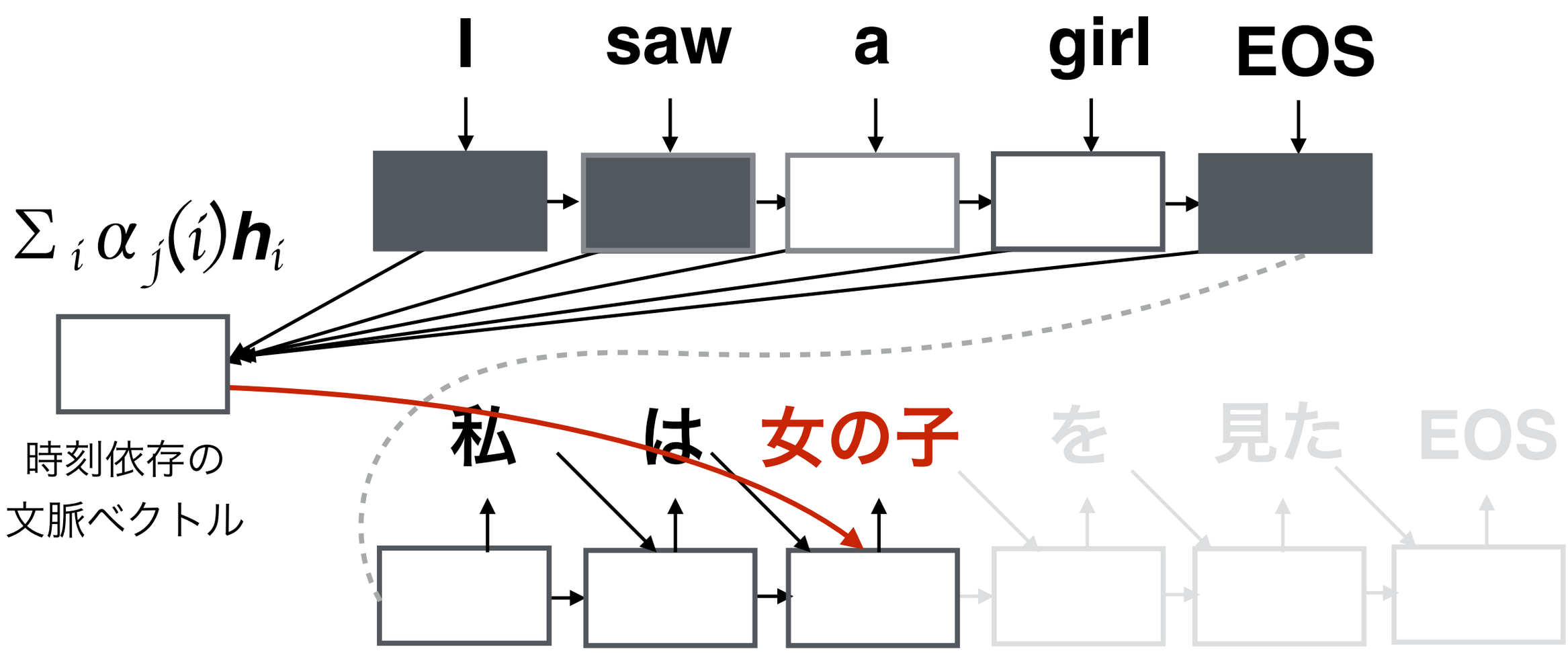
長文出力が苦手という欠点



(Cho et al., 2014a; Sutskever et al., 2014)

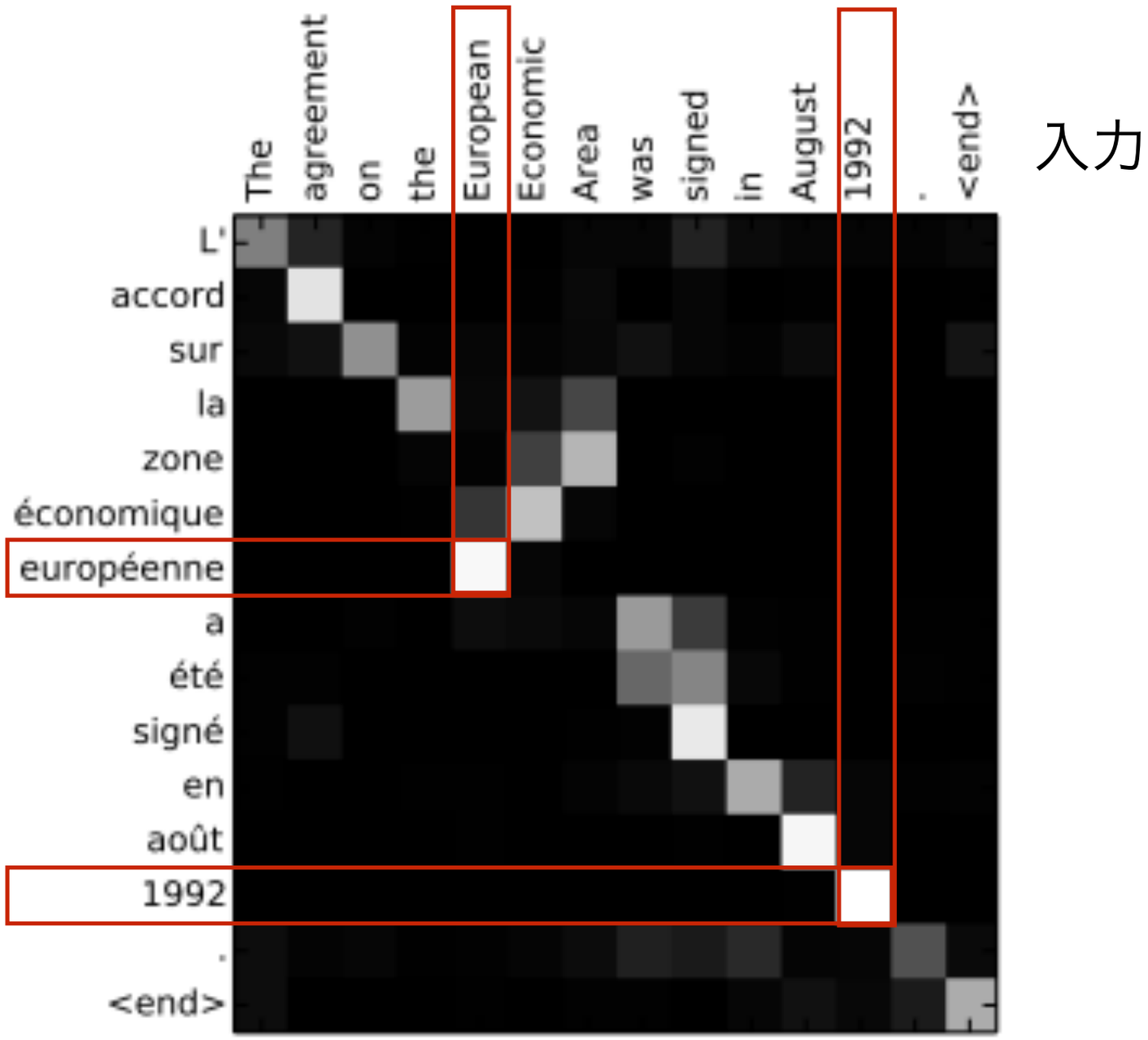
アテンションに基づくニューラル機械翻訳

- 出力時に原言語情報へのアクセスを許すことで改善
  - 関連度合い α で重み付けされた文脈ベクトルを追加



(Bahdanau et al., 2015; Luong et al., 2015b)

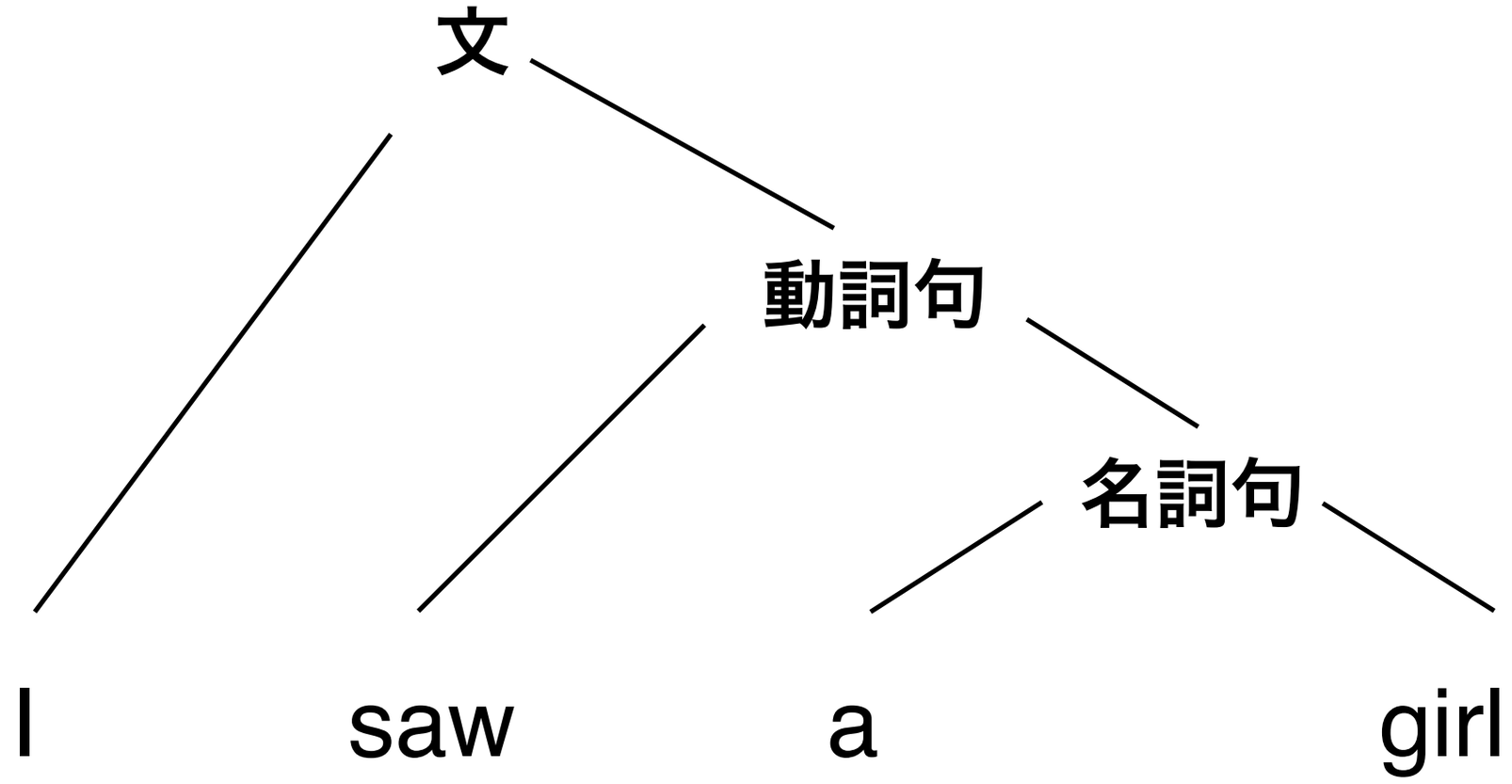
関連度合い α の可視化例



(Bahdanau et al., 2015)

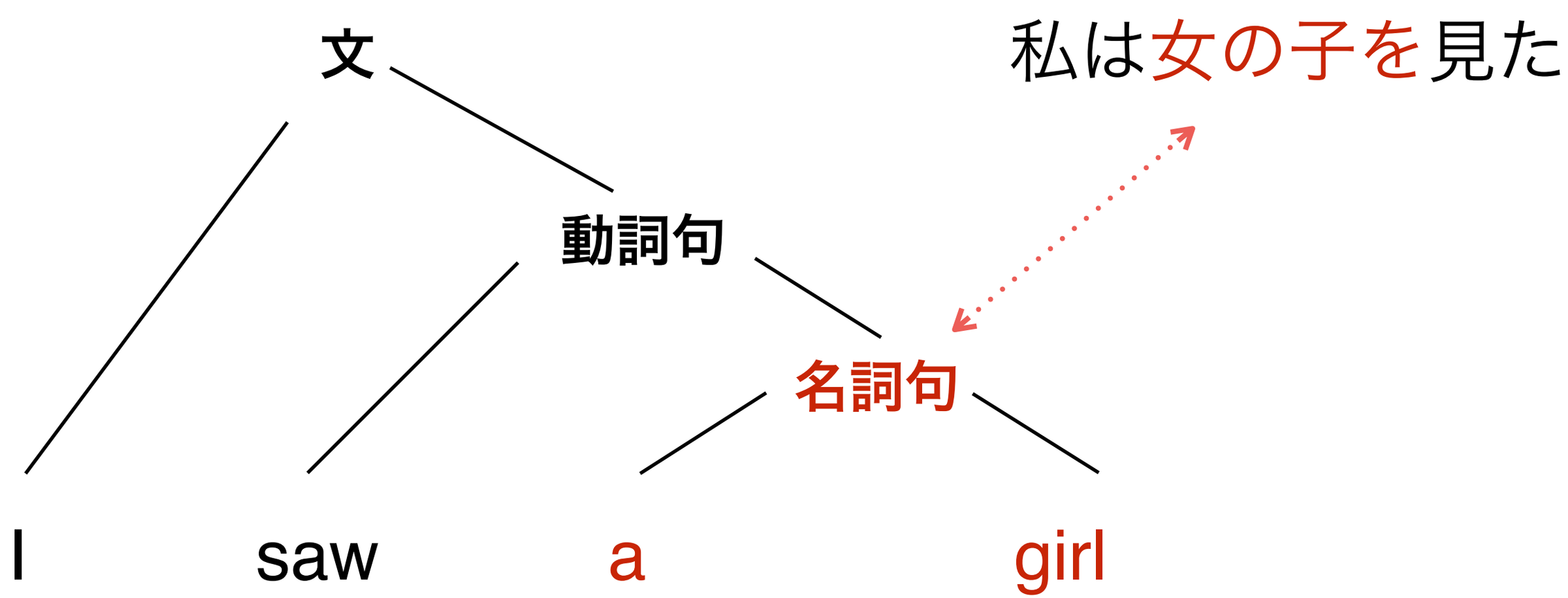
研究目的

- 既存のNMTモデルは、入出力は系列データを想定
- 遠縁の言語間翻訳では構文情報が有用 (Liu et al., 2006)
- 構文情報をNMTモデルへ導入し、翻訳性能を改善



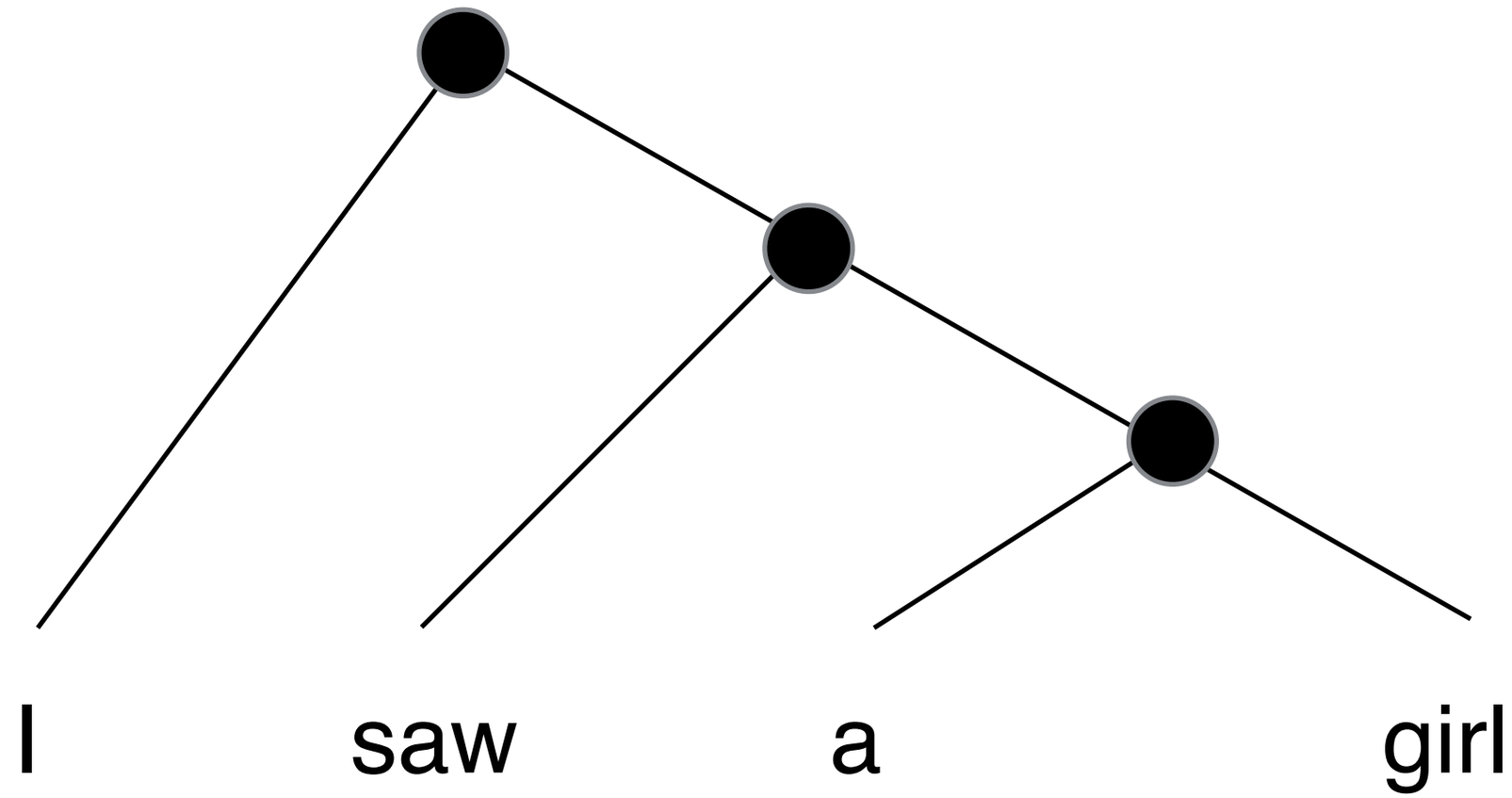
提案手法

- 翻訳元言語における文の構文構造情報をエンコード
- 単語・句へのアテンション機構により、
  単語-単語ならびに**句-単語**の関連度合いを学習



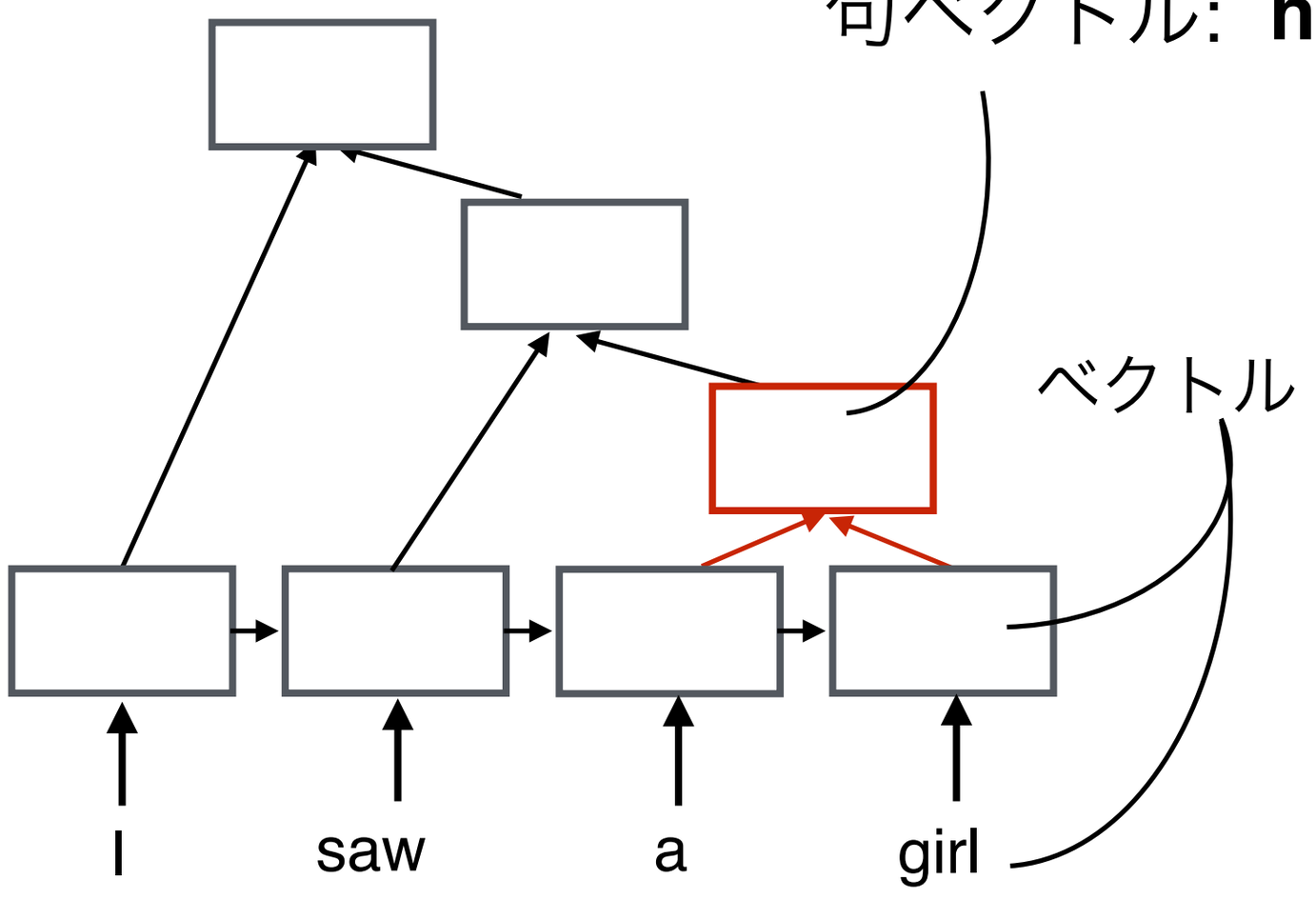
構文情報の抽出

- 翻訳元言語の句構造情報に着目
  - 句構造解析器により2分木を取得
  - 構造情報のみを利用するため、句ラベルは利用しない



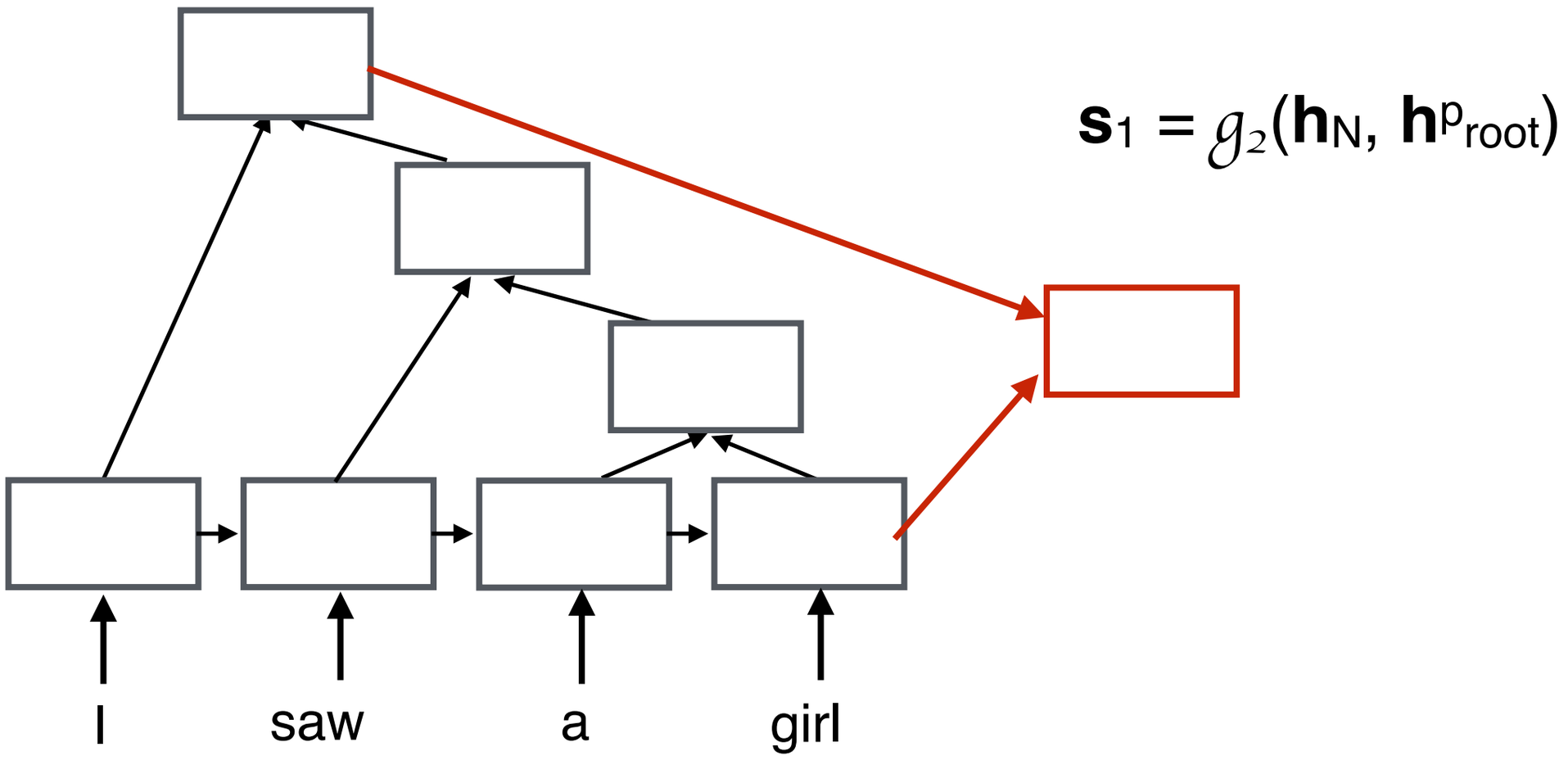
木構造に基づくエンコーダ

- 得られた構造に従って、句ベクトルを計算
  - 系列エンコーダ上で、ボトムアップに



デコーダの初期化

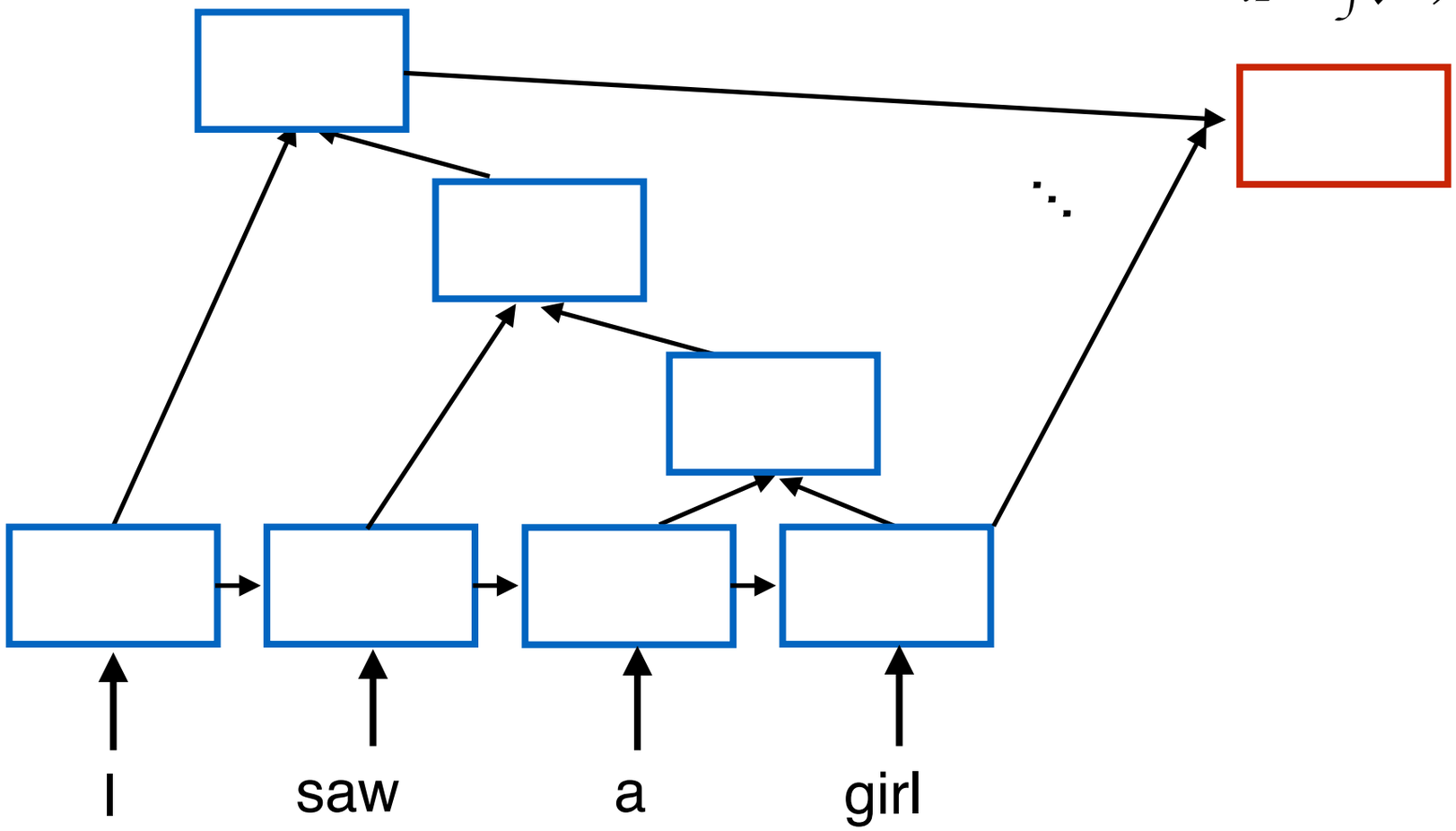
・ ２つの文ベクトルからデコーダの初期隠れ層を算出



木構造エンコーダへのアテンション

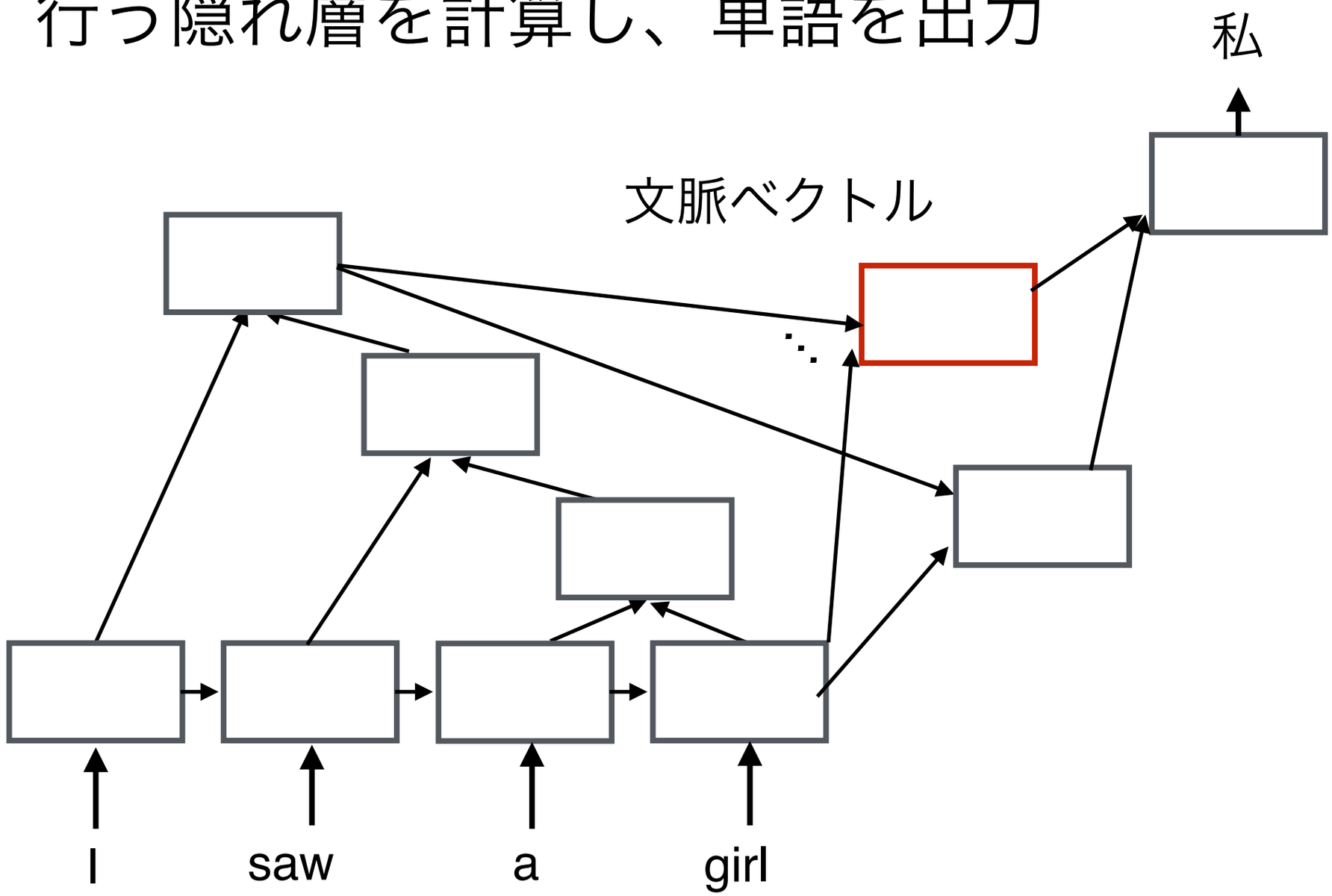
- アテンション計算の対象に句の隠れ層を含める
  - *N*個の単語があるとき、2分木内の節数は *(N-1)*個

文脈ベクトル: $\Sigma_{i1} \beta_j(i1) \boldsymbol{h}_{i1} + \Sigma_{i2} \beta_j(i2) \boldsymbol{h}^p_{i2}$



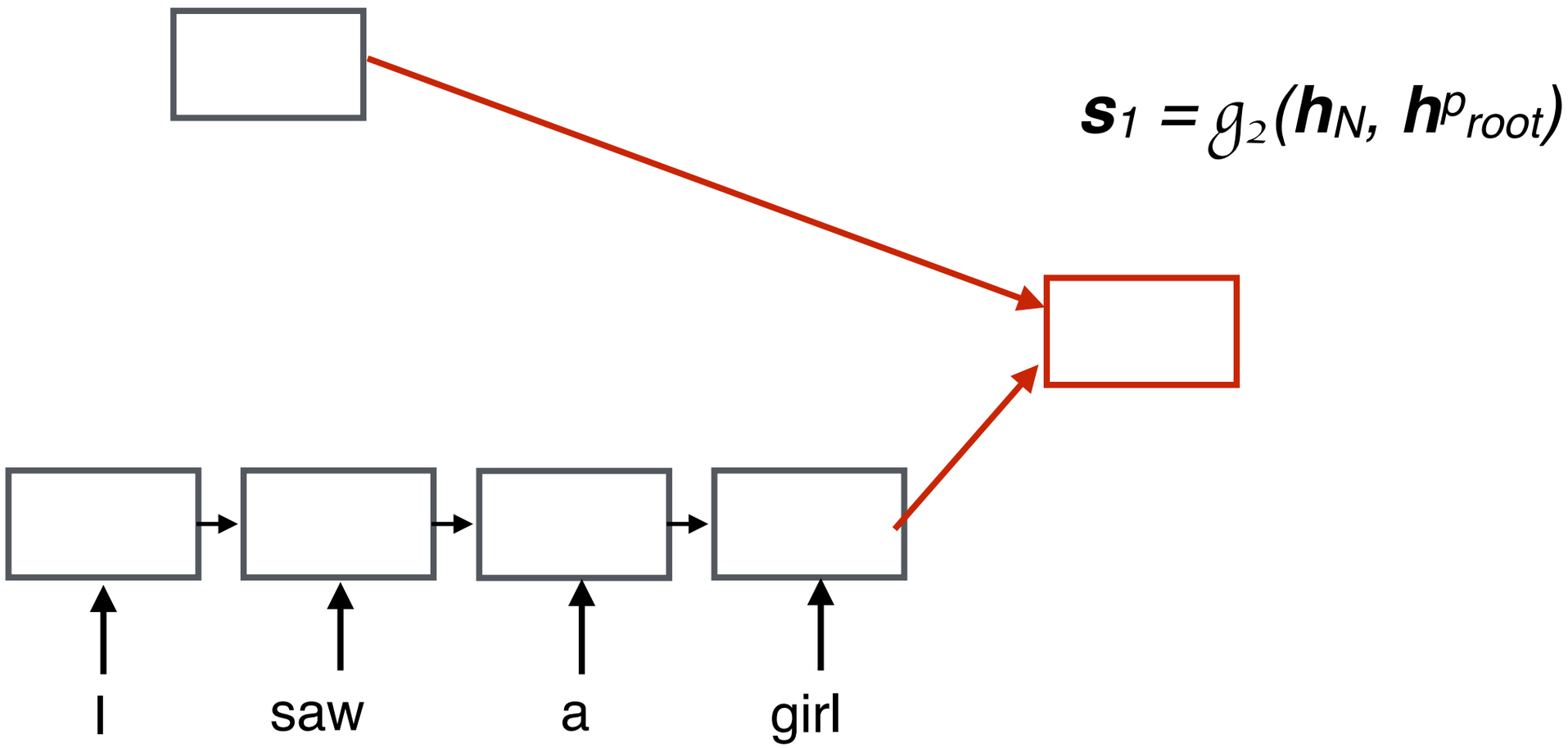
デコーダ

・ デコーダの隠れ層と文脈ベクトルから単語予測を行う隠れ層を計算し、単語を出力



構文解析が失敗した場合

- 既存のアテンションに基づくニューラル機械翻訳モデルに帰着
  - $\boldsymbol{h}^{p}_{root}$ をゼロベクトルとする



実験設定

- 英日の科学論文データ ASPEC (130万文)
  - 学習: 134万文, 開発: 1789文, テスト: 1811文
  - 語彙数: (英, 日) = (87K, 65K)
  - 構文解析器: *Enju* (Miyao and Tsujii, 2008)
- モデルパラメータ
  - 隠れ層: {512, 768, 1024}次元, 単語ベクトル: 512次元
  - 128ミニバッチ学習
  - パラメータ更新: SGD
  - *BlackOut* (負例サンプリングによるSoftmax近似) (Ji et al, 2016)
  - 出力: ビーム探索 + 長さ制約 (ビーム幅: 20)

実験結果

- BLEU (Papieni et al., 2002) で同等,
  RIBESスコア (Isozaki et al., 2010) で最高性能を達成

	BLEU ↑	RIBES ↑
tree-to-string (Baseline)	29.80	69.19
tree-to-string (Neubig and Duh., 2014)	33.44	75.80
+ NMTモデルによるリランキング (Neubig et al., 2015) ※	**38.17**	81.38
既存のNMTモデル (512次元) (Luong et al., 2015b)	34.64	81.60
提案手法 (512次元)	35.05	81.67
アンサンブル (3モデル)	**38.00**	**83.27**

※ WAT'15の最高性能システム

翻訳例

[翻訳元]

SiO2 films showed excellent performance even at 430°C or less, and the memory effect of Si dot MOS capacitor was confirmed.

[正解翻訳]

SiO2 膜 は, 430 °C 以下 でも 優れた 性能 を 示 し , Si ドット MOS コンデンサ の メモリ 効果 を 確認 し た 。

[提案手法による翻訳]

SiO2 膜 は 430 °C 以下 でも 優れた 性能 を 示 し , Si ドット MOS コンデンサ の メモリ 効果 を 確認 し た 。

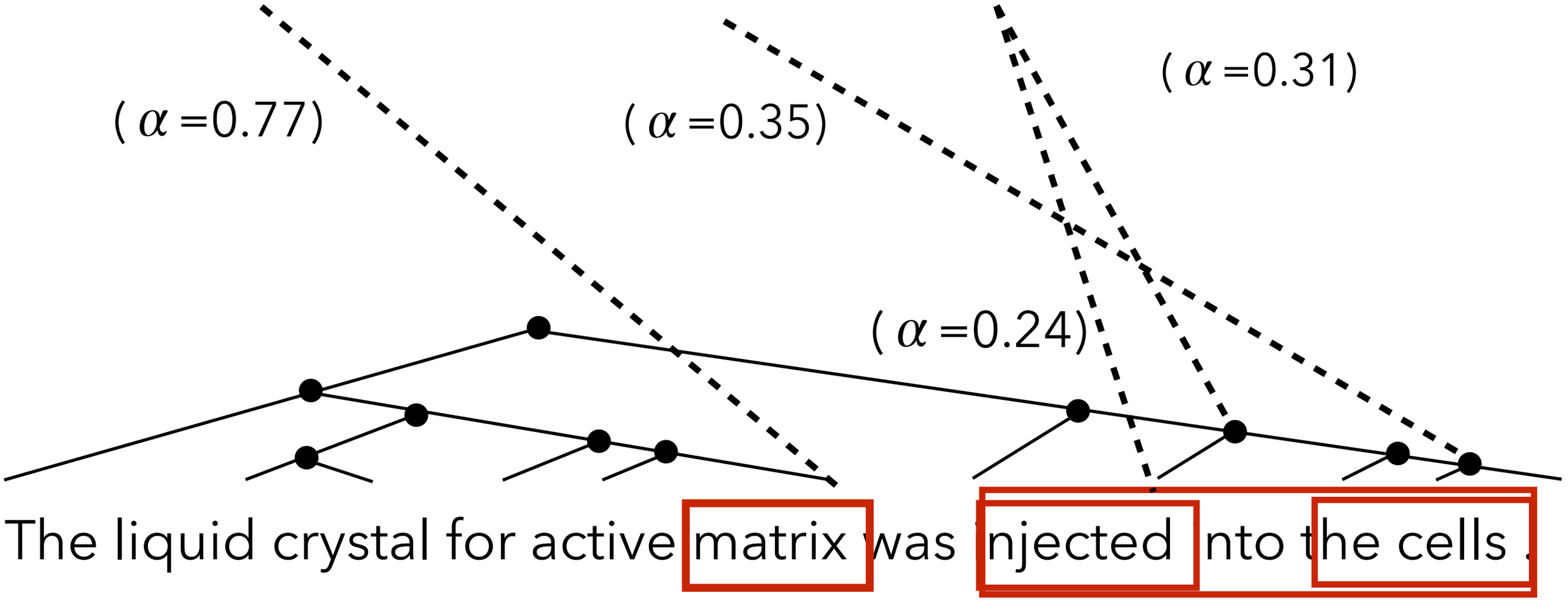
翻訳例とアテンションの学習の様子

[正解翻訳]

セル に は アクティブ マトリックス 用 液晶 を 注入 し た 。

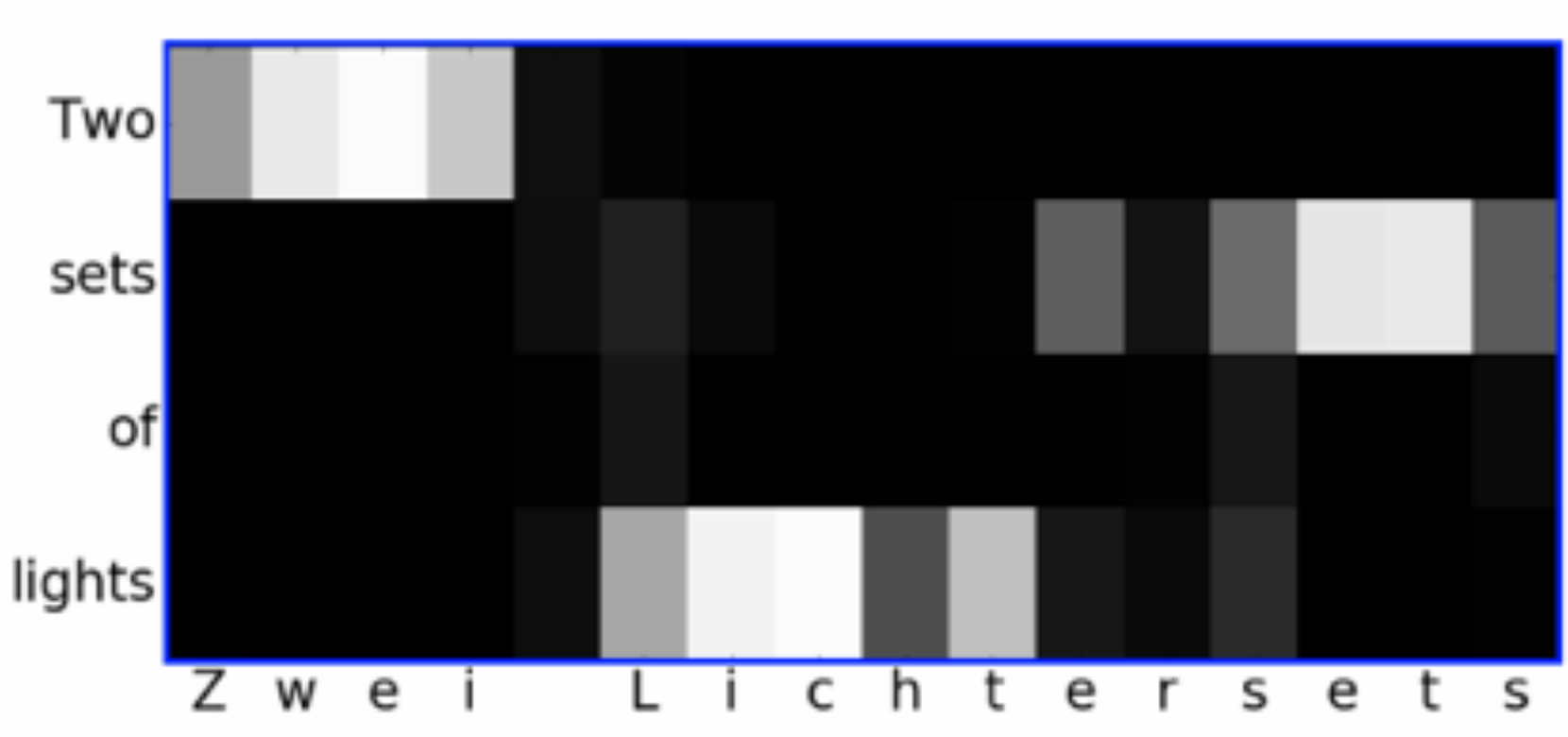
[提案手法による翻訳]

活性 マトリックス の 液晶 を セル 内 に 注入 し た 。



文字ベース翻訳

- 文は単語の系列でもあるが、文字の系列でもある
  - 日本語テキストなどの場合、単語分割器の精度が影響
  - 活用が多い言語や、容易に合成語をつくれる場合、データ数の増加に従い、語彙は爆発的に増加



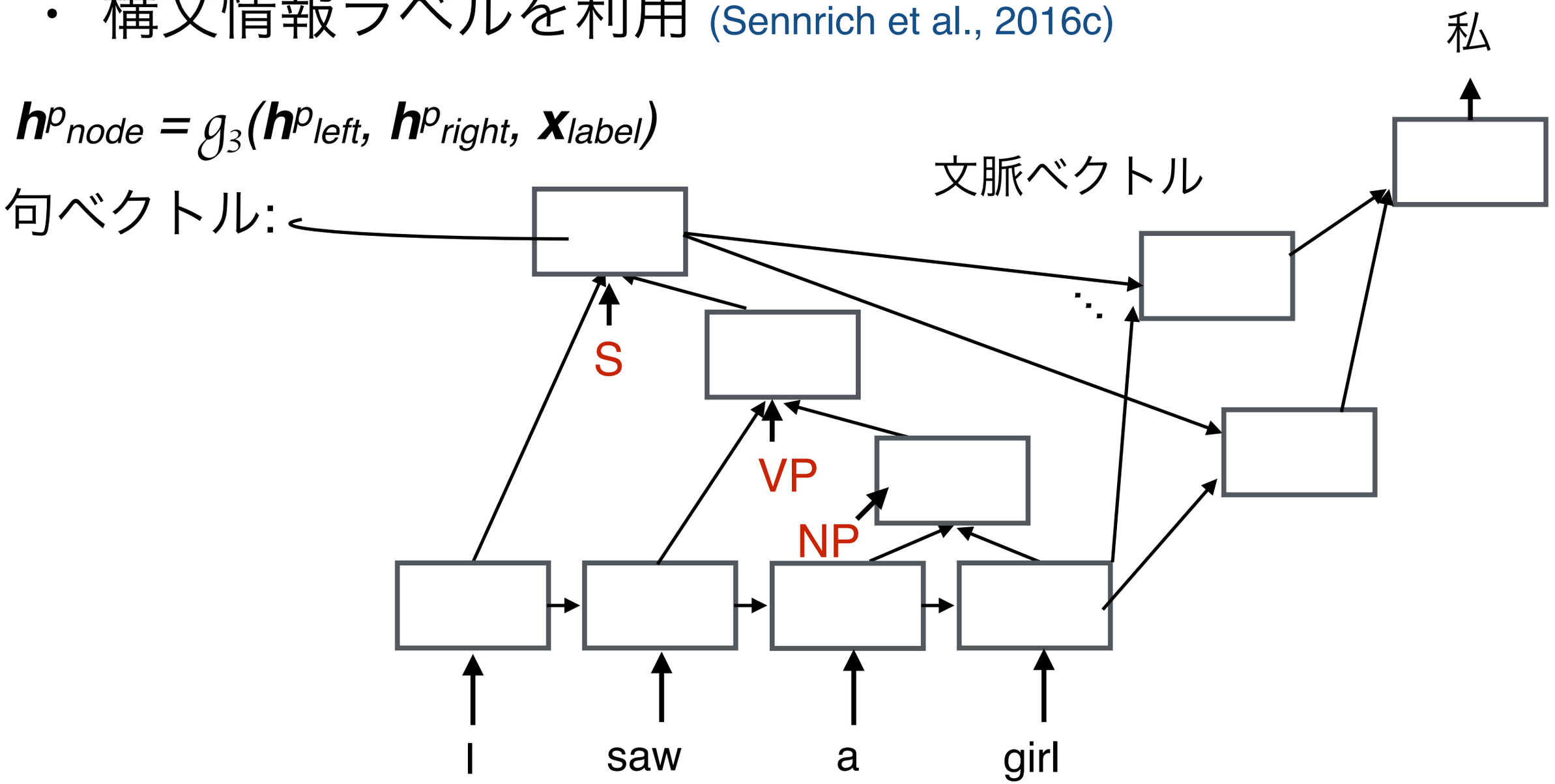
(アルファベット)
1文字ずつの翻訳

(Chung et al., 2016)

提案手法に文字ベースデコーダを適用

- 1文字ずつデコード
- 構文情報ラベルを利用 (Sennrich et al., 2016c)

$$\boldsymbol{h}^{p}_{node} = g_3(\boldsymbol{h}^{p}_{left}, \boldsymbol{h}^{p}_{right}, \boldsymbol{x}_{label})$$



実験設定

- 同ASPECコーパスを使用
- モデルパラメータ
  - 隠れ層: 512次元, 単語ベクトル: 256次元
  - 128ミニバッチ学習
  - パラメータ更新: SGD
  - Softmax
  - 出力: ビーム探索 + 長さ制約 (ビーム幅: 20)

	英語	日本語
単語ベースデコード	87,796	65,680
文字ベースデコード	87,796	**3,004**

実験結果

- 単語ベースデコーダにはおよばないものの、
  文字ベースデコーダでも翻訳は可能であることを確認

	BLEU ↑	RIBES ↑
提案手法 (単語ベースデコーダ)	**35.05**	**81.67**
提案手法 (文字ベースデコーダ)	34.47	80.73
提案手法 (文字ベースデコーダ) + 句ラベル	34.36	81.12
既存のNMTモデル (Luong et al., 2015b)	34.64	81.60
tree-to-string (Baseline)	29.80	69.19

翻訳例1

[翻訳元]

The electric power generation was the 380 micro watt .

[正解翻訳]

発電量は３８０マイクロワットであった。

[単語ベースデコード]

発電は３８０UNKWであった。

[文字ベースデコード]

発電は３８０マイクロワットであった。

(α = 0.78)

翻訳例2

[翻訳元]

This paper describes development outline of low-loss forsterite porcelain .

[正解翻訳]

低損失フォルステライト磁器の開発概要などを述べた。

[単語ベースデコーダ]

ここでは，UNKUNKの開発概要を述べた。

[文字ベースデコーダ]

低損失フォルステライト磁器の開発概要を述べた。

文字とsub-wordと単語の関係

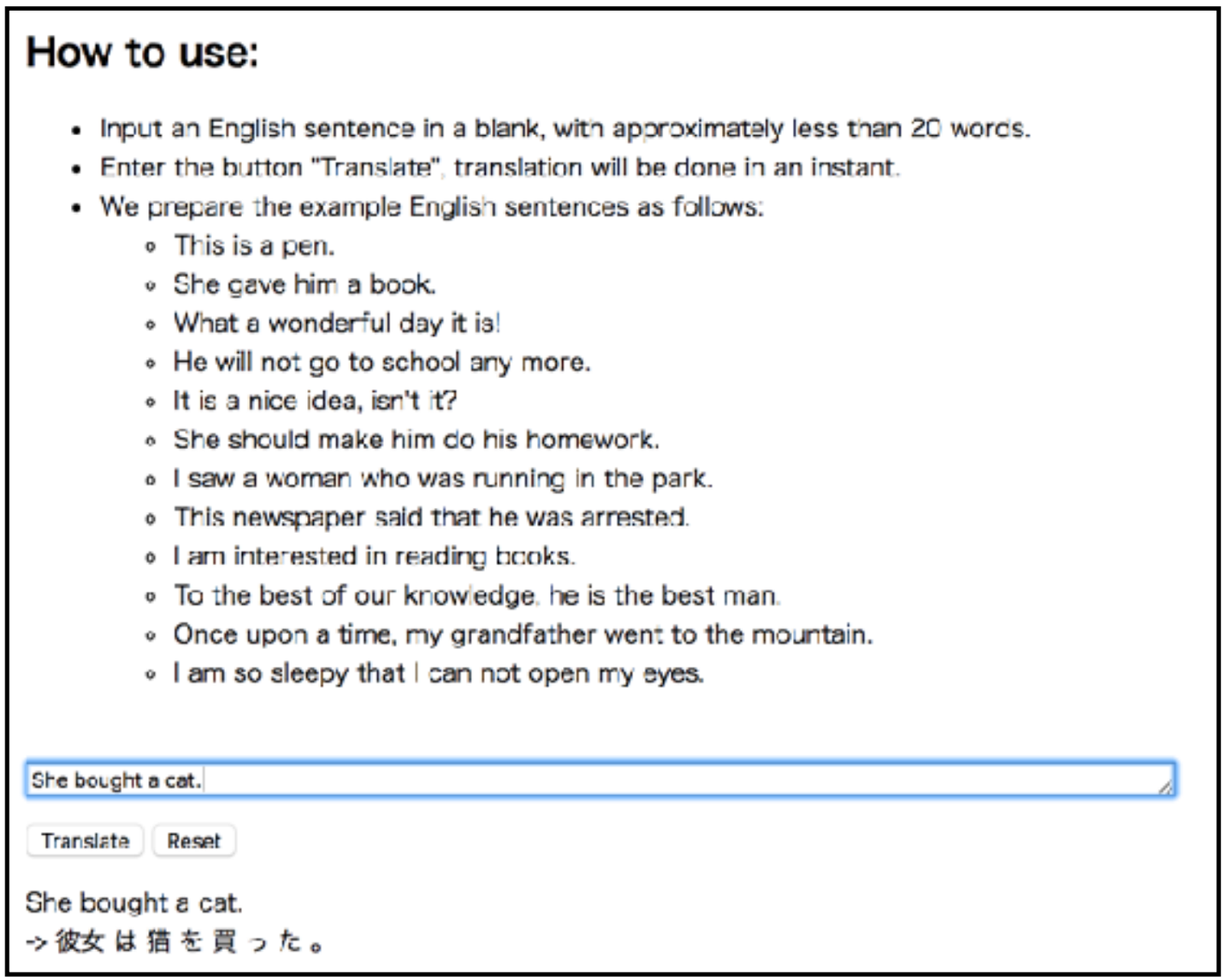
- 語彙数 (扱う言語に依存)
  - 文字 > sub-word (Sennrich et al., 2016a) >> 単語
- 系列の長さ
  - 単語 > sub-word > 文字
- 語彙数と系列の長さは学習時間と性能のトレードオフ (Wu et al., 2016)
  - 単語出力予測 (Softmax) の学習は計算コストが高く、語彙数に依存
  - RNNにおける処理は系列の長さに依存
  - NMTモデル短い出力を好み (Cho et al., 2014b)、BLEUが下がる
  - 記述したニューラルネットワークのモデル構造にも依存

実装

- GitHubにて公開
  - *https://github.com/tempra28/tree2seq*
  - N3LPライブラリを利用
- N3LPライブラリ (https://github.com/hassyGo/N3LP)
  - 鶴岡研究室で作成 (作成者: 橋本和真)
  - C++言語, 行列演算ライブラリEigenを利用
  - マルチコアCPU上で動作
  - 詳細は「鶴岡研におけるニューラルネット+NLP」まで *(http://www.logos.t.u-tokyo.ac.jp/~hassy/publications/talk/pfi_dl2016/slides.pdf)*

オンラインデモの公開

- 英日データ (約15万文) で学習



(http://www.logos.t.u-tokyo.ac.jp/~eriguchi/demo/tree2seq/index.php)

NMT研究の方向性は多種多様

- 未知語処理 (Luong et al., 2015a)
- マルチタスク (Luong et al., 2016c)
- ドメイン適用 (Hashimoto et al., 2016)
- 複数多言語翻訳 (Johnson et al., 2016)
- 言語毎の課題に着目
  - 語彙爆発 (Sennrich et al., 2016a)

    敬語の訳出 (Sennrich et al., 2016b)
- 文構造情報の潜在化 (Kim et al., 2017; Hashimoto and Tsuruoka, 2017)
- 目的言語側での構文構造利用 (Eriguchi et al., 2017)

おまけ: WAT'16 英日タスク

- 単体NMTシステムモデルで最高性能を達成
- 計算時間: 5日
  - (Cromieres et al., 2016)は2週間以上

	BLEU	RIBES
Hashimoto and Tsuruoka (2017)	**39.19**	82.66
Cromieres et al. (2016)	38.20	82.39
Neubig et al. (2015)	38.17	81.38
本発表	38.00	83.27

研究をしていて感じること

- 研究用資源 (対訳コーパス)
  - 大量の (良質な) 対訳コーパスは中々手に入らない
    - デモ用の学習に特許データは不適当
  - 英日データは複数の理由で海外研究者に勧めにくい
    - WMT'17に中英タスクが追加
- 計算機環境の差
- 人手評価の必要性
  - BLEUなどの自動評価指標では、類義語翻訳 = 翻訳誤り
  - 人手評価による分析は不可欠

まとめ

- ニューラル機械翻訳モデルの紹介
- 構文構造を導入した新たなモデルの提案
  - 翻訳元言語における句構造情報を利用したニューラル機械翻訳モデル (Eriguchi et al., 2016a)
  - 文字ベースデコーダを適用 (Eriguchi et al., 2016b)
  - 英日翻訳において提案手法による性能改善を確認

引用文献リスト

(Bahdanau et al., 2015) Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. In Proceedings of the 3rd International Conference on Learning Representations.

(Chung et al., 2016) Junyoung Chung, Kyunghyun Cho, and Yoshua Bengio. 2016. A character-level decoder without explicit segmentation for neural machine translation. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, pages 1693–1703.

(Cho et al., 2014a) Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, pages 1724–1734.

(Cho et al., 2014b) KyungHyun Cho, Bart van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014a. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. In Proceedings of Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8).

(Eriguchi et al., 2016a) Akiko Eriguchi, Kazuma Hashimoto, and Yoshimasa Tsuruoka. 2016. Tree-to-sequence attentional neural machine translation. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, pages 823–833.

(Eriguchi et al., 2016b) Akiko Eriguchi, Kazuma Hashimoto, and Yoshimasa Tsuruoka. 2016. Character-based Decoding in Tree-to-Sequence Attention-based Neural Machine Translation. In Proceedings of the 3rd Workshop on Asian Translation (WAT2016). The COLING 2016 Organizing Committee, pages 175–183.

(Eriguchi et al., 2017) Akiko Eriguchi, Yoshimasa Tsuruoka, and Kyunghyun Cho. 2017. Learning to Parse and Translate Improves Neural Machine Translation. arXiv preprint arXiv:1702.03525.

引用文献リスト

(Hashimoto et al., 2016) Kazuma Hashimoto, Akiko Eriguchi, and Yoshimasa Tsuruoka. 2016. Domain Adaptation and Attention-Based Unknown Word Replacement in Chinese-to-Japanese Neural Machine Translation. In Proceedings of the 3rd Workshop on Asian Translation (WAT2016). The COLING 2016 Organizing Committee, pages 75–83.

(Hashimoto et al., 2017) Kazuma Hashimoto and Yoshimasa Tsuruoka. 2016. Neural Machine Translation with Source-Side Latent Graph Parsing. 2017. arXiv preprint arXiv:1702.02265.

(Isozaki et al., 2010) Hideki Isozaki, Tsutomu Hirao, Kevin Duh, Katsuhito Sudoh, and Hajime Tsukada. 2010. Automatic Evaluation of Translation Quality for Distant Language Pairs. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pages 944–952.

(Ji et al., 2016) Shihao Ji, S. V. N. Vishwanathan, Nadathur Satish, Michael J. Anderson, and Pradeep Dubey. 2016. BlackOut: Speeding up Recurrent Neural Network Language Models With Very Large Vocabularies. In Proceedings of the 4th International Conference on Learning Representations.

(Johnson et al., 2016) Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, Jeffrey Dean. 2016. Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. arXiv preprint arXiv:1611.04558.

(Kim et al., 2017) Yoon Kim, Carl Denton, Luong Hoang, Alexander M. Rush. 2017. Neural Machine Translation with Source-Side Latent Graph Parsing. arXiv preprint arXiv:1702.00887.

(Liu et al., 2006) Yang Liu, Qun Liu, and Shouxun Lin. 2006. Tree-to-string alignment template for statistical machine translation. In Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, pages 609–616.

引用文献リスト

(Luong et al., 2015a) Thang Luong, Ilya Sutskever, Quoc Le, Oriol Vinyals, and Wojciech Zaremba. 2015b. Addressing the Rare Word Problem in Neural Machine Translation. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, pages 11–19.

(Luong et al., 2015b) Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective Approaches to Attention-based Neural Machine Translation. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 1412–1421.

(Luong et al., 2015c) Minh-Thang Luong, Quoc V. Le, Ilya Sutskever, Oriol Vinyals, Lukasz Kaiser. 2016. Multi-task Sequence to Sequence Learning. In Proceedings of the 4th International Conference on Learning Representations.

(Miyao and Tsujii, 2008) Yusuke Miyao and Jun'ichi Tsujii. 2008. Feature Forest Models for Probabilistic HPSG Parsing. Computational Linguistics, 34(1):35–80.

(Neubig et al., 2014) Graham Neubig and Kevin Duh. 2014. On the elements of an accurate tree-to-string machine translation system. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pages 143–149.

(Neubig et al., 2015) Graham Neubig, Makoto Morishita, and Satoshi Nakamura. 2015. Neural Reranking Improves Subjective Quality of Machine Translation: NAIST at WAT2015. In Proceedings of the 2nd Workshop on Asian Translation (WAT2015), pages 35–41.

(Sutskever et al., 2014) Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to Sequence Learning with Neural Networks. In Advances in Neural Information Processing Systems 27, pages 3104–3112.

引用文献リスト

(Sennrich et al., 2016a) Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural machine translation of rare words with subword units. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, pages 1715–1725.

(Sennrich et al., 2016b) Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Controlling Politeness in Neural Machine Translation via Side Constraints. In Proceedings of Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, pages 35–40.

(Sennrich et al., 2016c) Rico Sennrich and Barry Haddow. 2016. Linguistic input features improve neural machine translation. In Proceedings of the First Conference on Machine Translation, pages 83–91.

(Papineni et al., 2002) Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: A Method for Automatic Evaluation of Machine Translation. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pages 311–318.

(Wu et al., 2016) Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Lukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144.