

# ニューラル機械翻訳

## — 句構造へのアテンションモデル —

東京大学

鶴岡研究室 江里口瑛子

# 自己紹介

- 江里口 瑛子 (えりぐち あきこ)
  - 学士・修士 (お茶の水女子大学)
  - 現在: 東京大学 鶴岡研究室 博士後期課程2年
  - <http://www.logos.t.u-tokyo.ac.jp/~eriguchi/>
- 研究テーマ
  - グラフに基づく半教師あり学習による文書分類
  - 潜在的意味解析手法に基づく多言語解析
  - **ニューラル機械翻訳**

# 本日の発表

- 既存の研究紹介
  - ニューラル機械翻訳
  - アテンションに基づくニューラル機械翻訳
  - 文字列に基づくニューラル機械翻訳
- 我々の取り組み
  - 句構造へのアテンションに基づくニューラル機械翻訳

*“Tree-to-Sequence Attentional Neural Machine Translation”*

Akiko Eriguchi, Kazuma Hashimoto and Yoshimasa Tsuruoka

(<http://arxiv.org/pdf/1603.06075.pdf> → ACL2016)

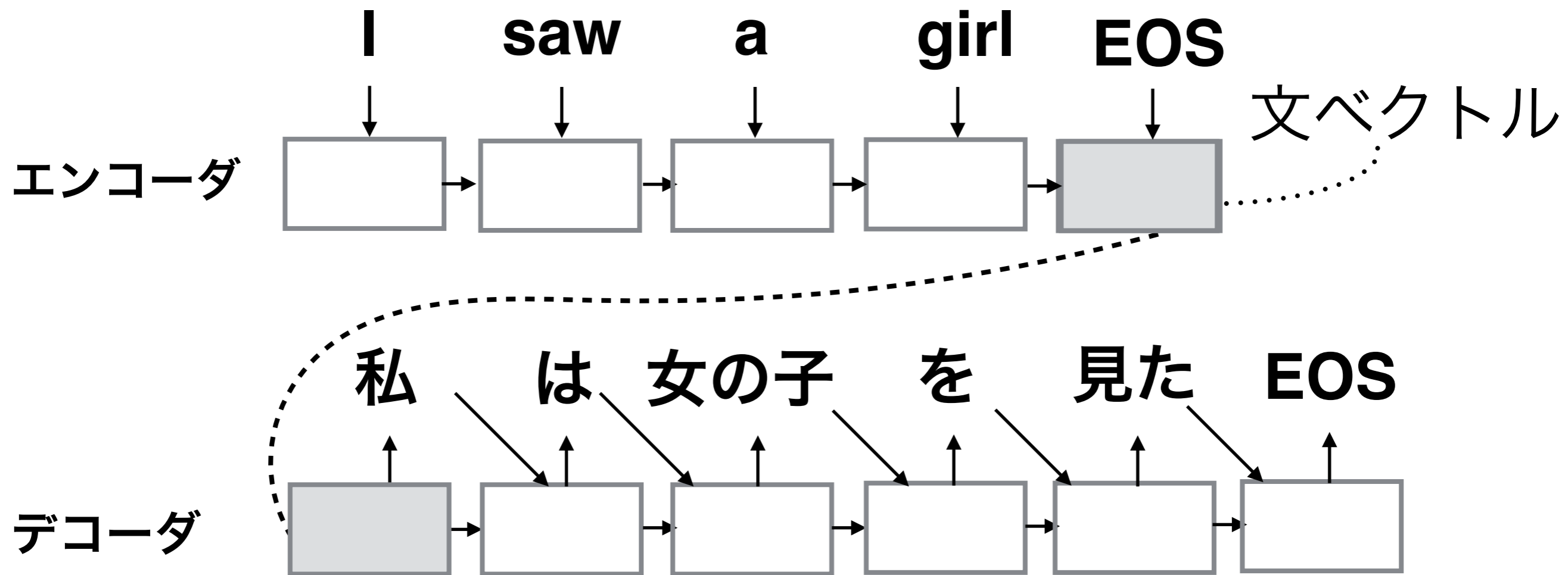
# ニューラル機械翻訳以前の統計的機械翻訳

- 対訳コーパス( $\mathbf{x}, \mathbf{y}$ )から翻訳ルールを学習
  - Noisy-Channel Model (Shannon, 1948)
  - $\mathbf{y} = \operatorname{argmax}_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}) = \operatorname{argmax}_{\mathbf{y}} p(\mathbf{x}|\mathbf{y})p(\mathbf{y})$



# ニューラル機械翻訳

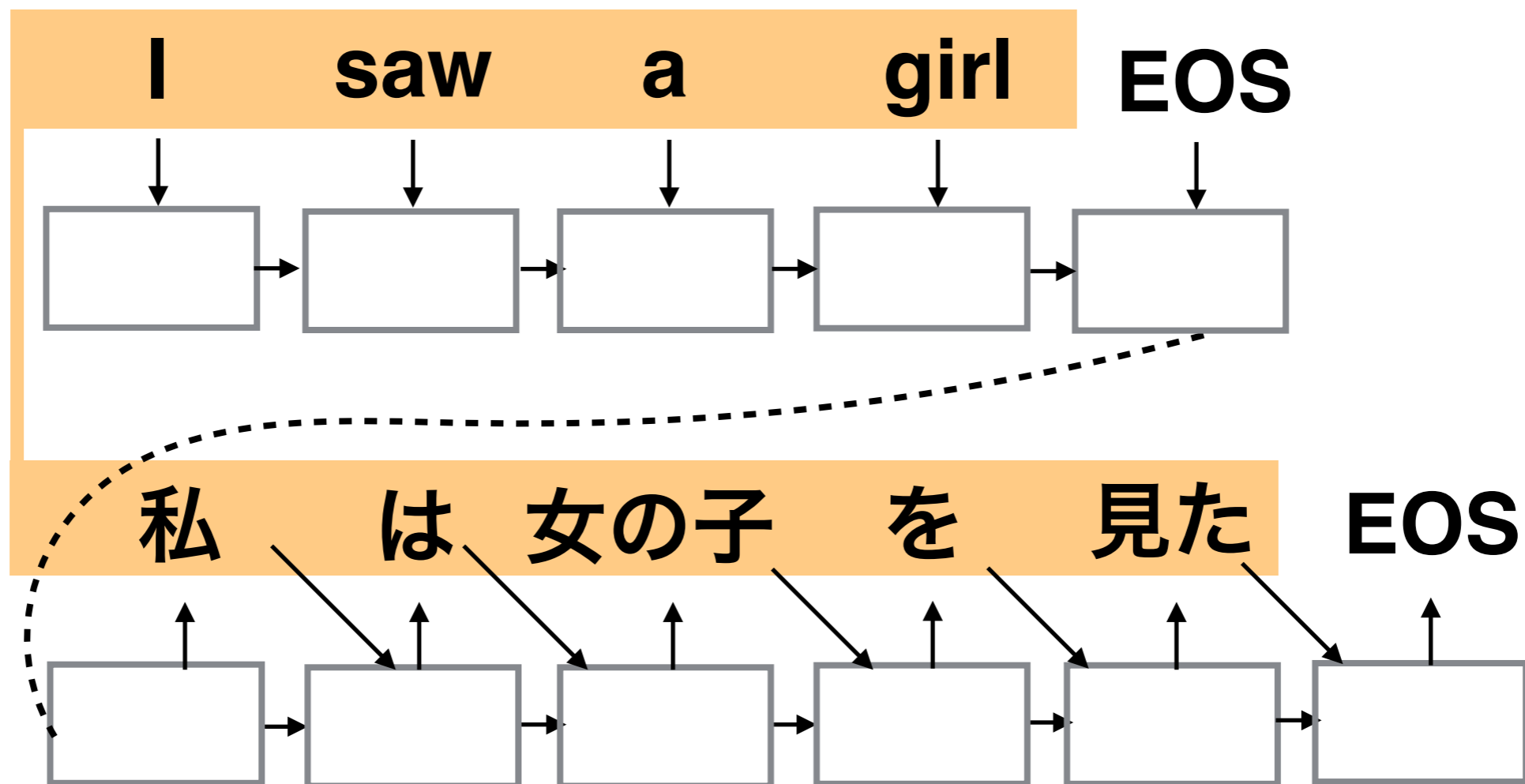
- 「2言語間の系列変換」として直接モデル化
- 2つのリカレントニューラルネットワークを利用



Encoder-Decoderモデル (Cho et al., 2014; Sutskever et al., 2014)

# ニューラル機械翻訳の学習・性能

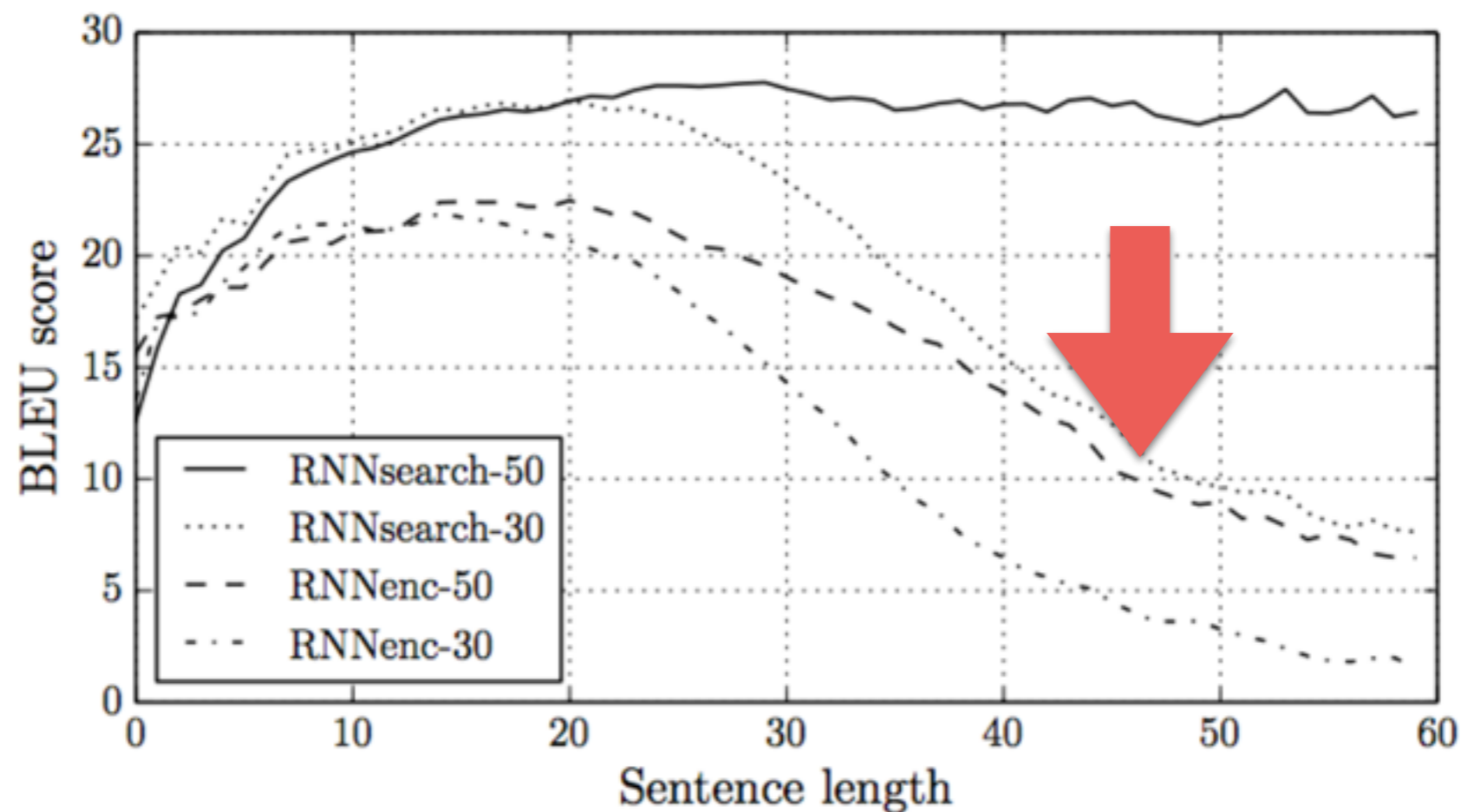
- 学習は、対訳コーパスから  $p_{\theta}(\mathbf{y}|\mathbf{x})$  を最大化
- 英仏・英独で最高精度 (Luong et al., 2015a; Luong et al., 2015b)



Encoder-Decoderモデル (Cho et al., 2014; Sutskever et al., 2014)

# Encoder-Decoderにおける問題点

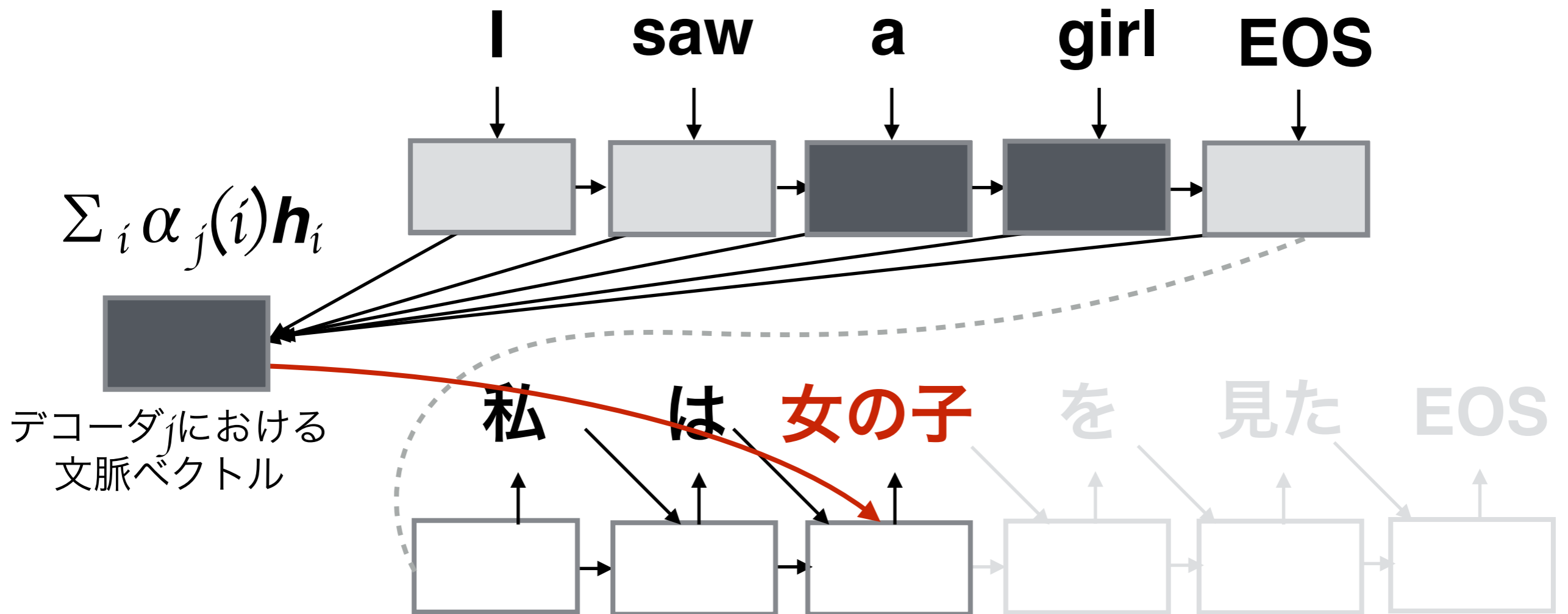
- 文が長くなるにつれ、翻訳の性能が低下
- エンコーダとデコーダの個々の繋がりが弱い



(Bahdanau et al., 2015)

# アテンションに基づくニューラル機械翻訳

- デコーダとの関連度に応じてエンコーダ情報を追加
- 関連度  $\alpha$  も同時学習 (単語アライメントのような働き)

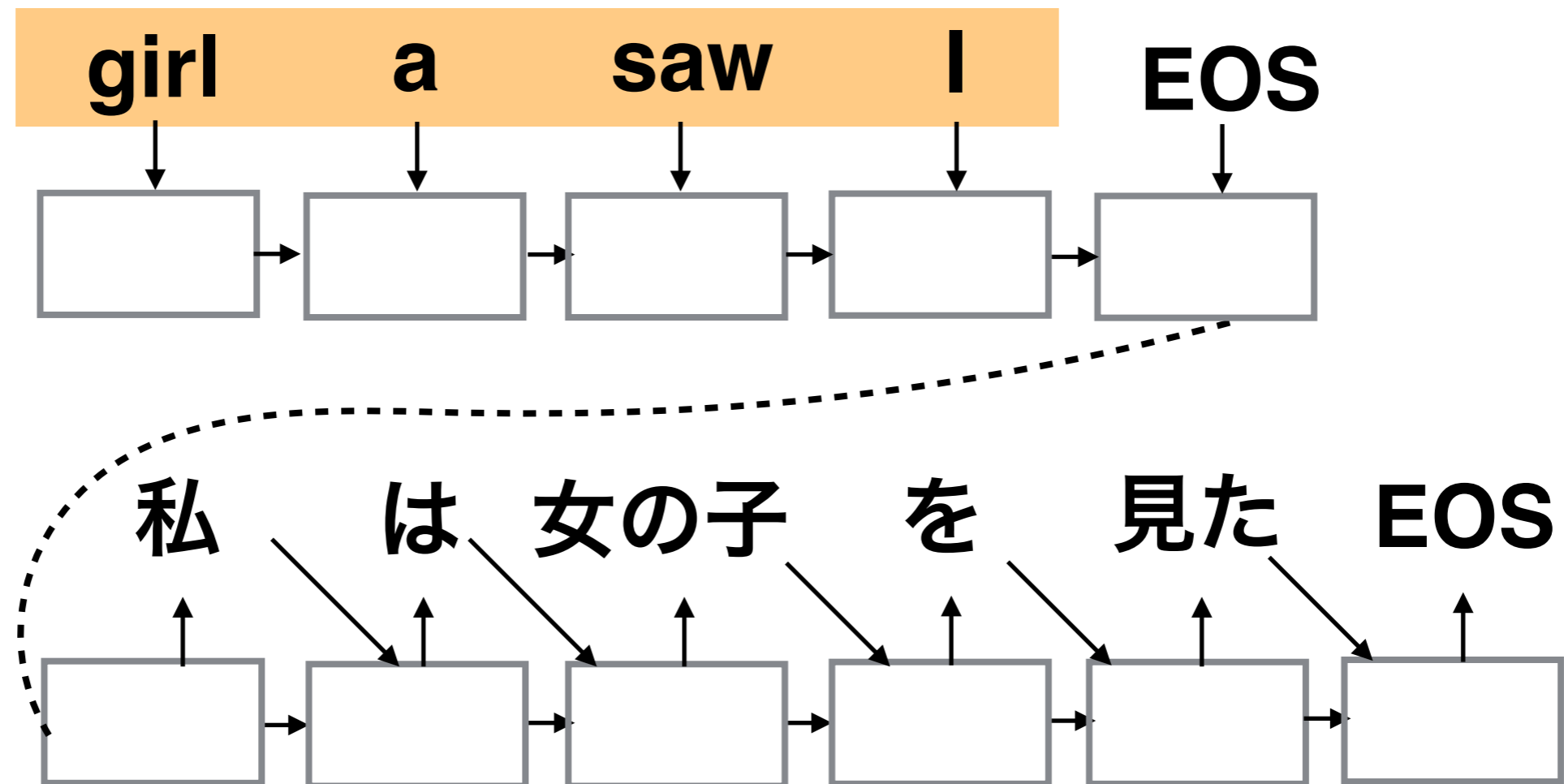


アテンションに基づくニューラル機械翻訳モデル (Bahdanau et al., 2015; Luong et al., 2015b)



# 逆順入力による改善

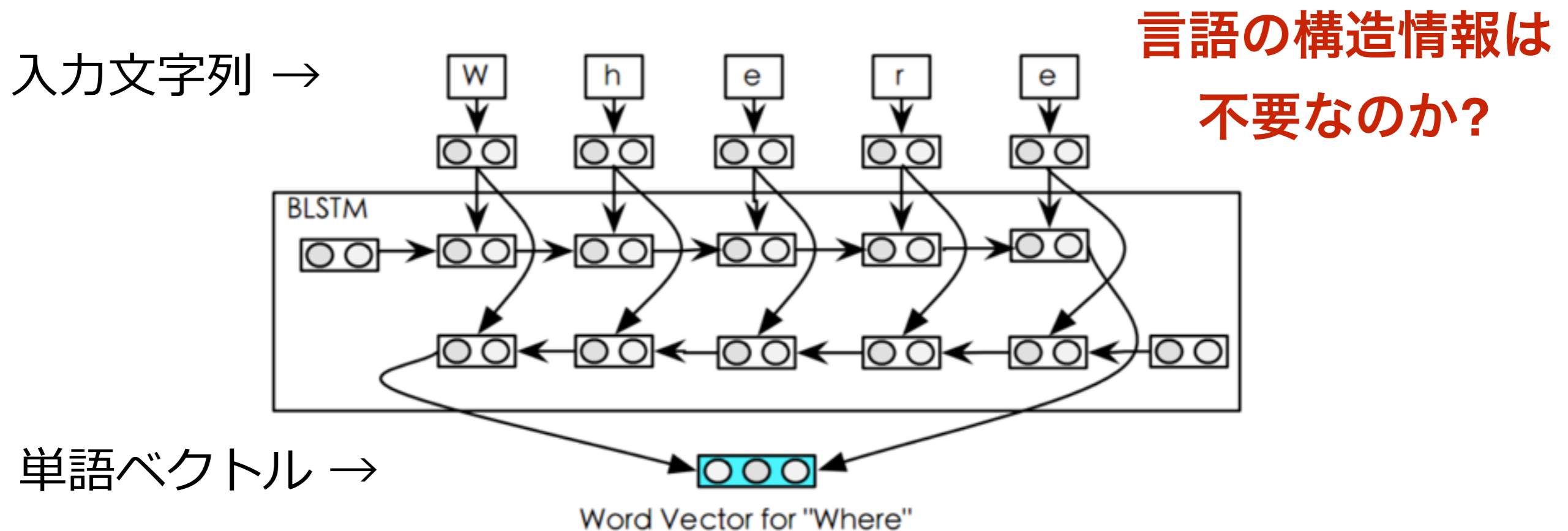
- 訳出単語間の平均距離が短くなる (Sutskever et al., 2014)
- 語順の似通った言語対 (英仏・英独) で有効  
(Sutskever et al., 2014; Luong et al., 2015b)



Encoder-Decoderモデル (Cho et al., 2014; Sutskever et al., 2014)

# 最近のニューラル翻訳の動向

- 文字列に基づくモデル (Ling et al., 2016; Chung et al., 2016)
  - 入出力は**文字 (アルファベット) 系列**
  - 未知語の訳語推定に利用 (Luong et al., 2016)



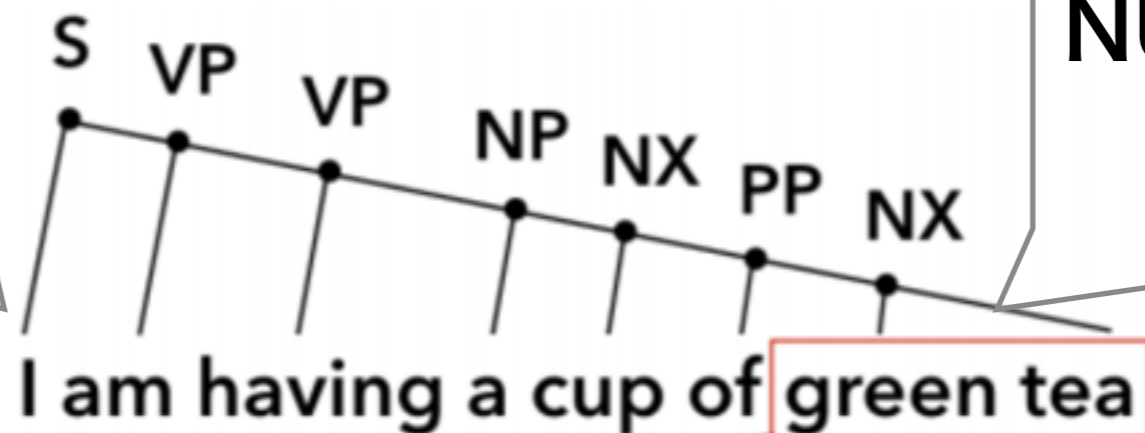
(Costa-jussà and Fonollosa, 2015)

# 我々の取り組み

- 語順の異なる言語間翻訳に構文情報は有用  
(Yamada and Knight, 2001; Neubig, 2013)
- 構文情報を利用したニューラル翻訳モデルの提案
  - 翻訳元言語の句構造情報に着目

異なる構文構造

英語: SVO



NULLアライメント

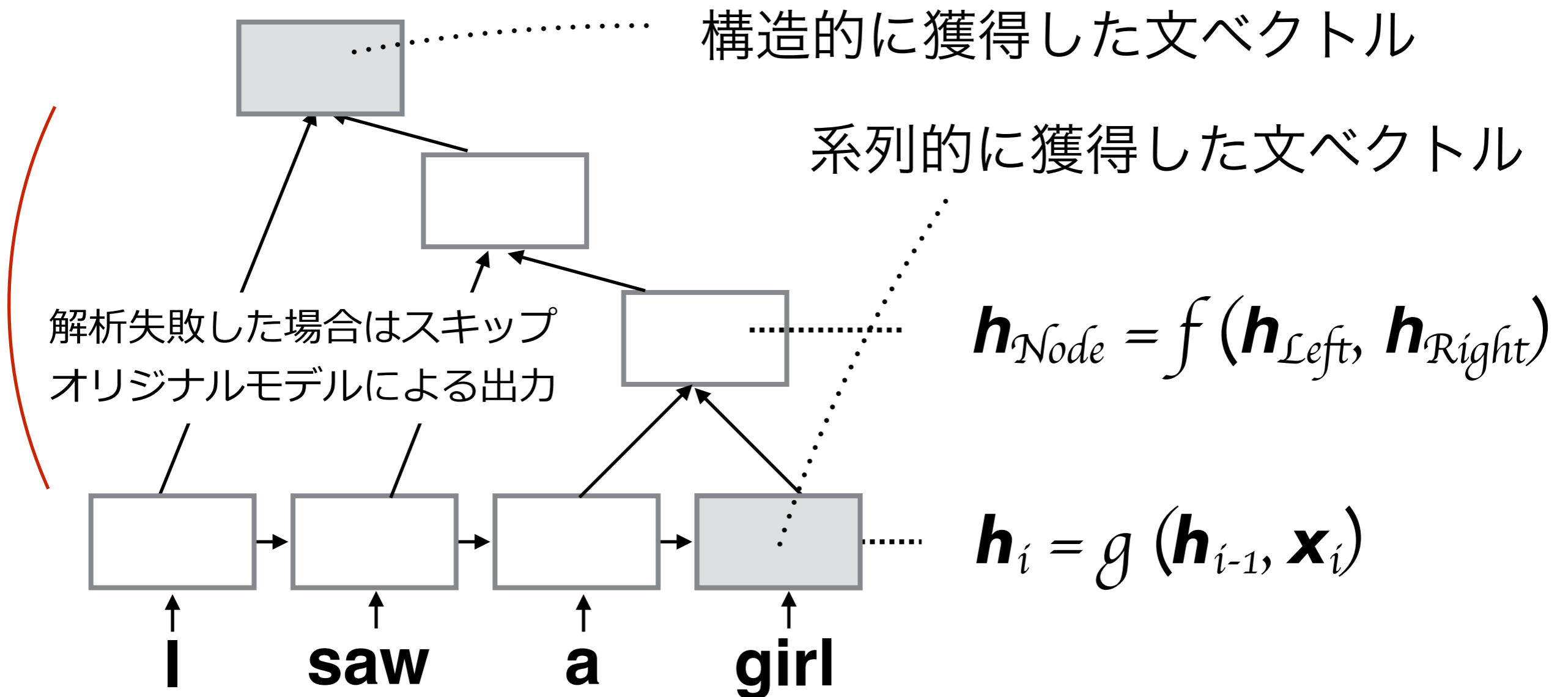
“a cup of”

私は緑茶を飲んでいる  
.....  
(S) (O) (V)

※句構造解析器Enju (Miyao and Tsujii, 2008) を使用

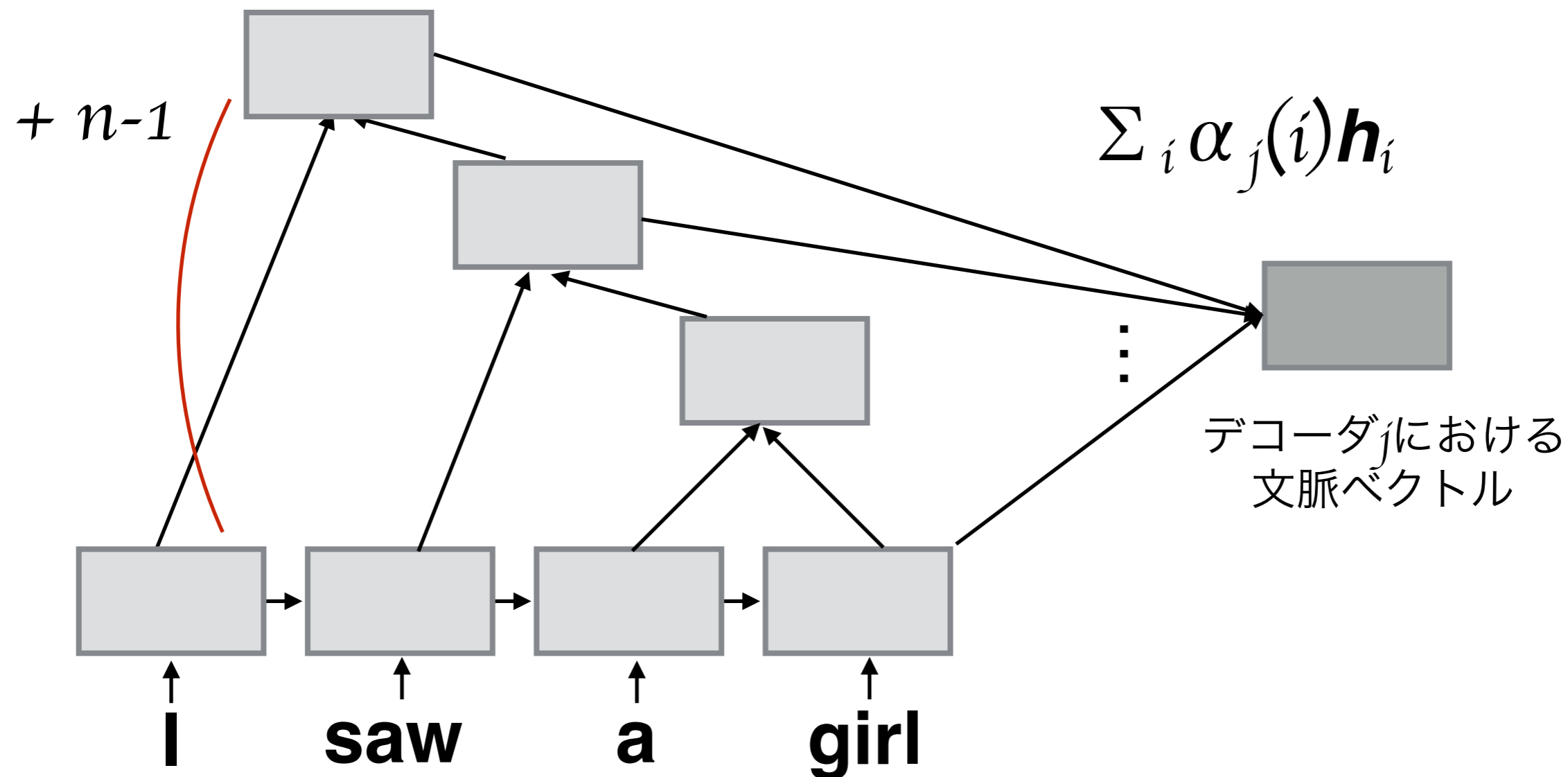
# 句構造のための木エンコーダ

- リカレントネットワークによる単語の系列情報の獲得
- 句の隠れ層は左右の子から再帰的に獲得 (Tai et al., 2015)



# 句構造へのアテンション機構

- 系列/ 句ベクトルから文脈ベクトルを構成
- 単語と句の情報のどちらが重要かはモデルに任せる



# 予備実験 (WAT'15 英日翻訳タスク)

- 科学技術論文の対訳コーパス: 10万文
  - 語彙数: (英, 日) = (約2万, 約2万) + unknown
  - 開発データ: 1790文
- 単語ベクトルと隠れ層の次元: 256
- 確率的勾配降下法により学習

	RIBES	BLEU
<b>提案モデル</b>	<b>72.6</b>	<b>20.5</b>
系列アテンションモデル	71.5	19.4
系列アテンションモデル (逆順入力)	69.5	17.5

# WAT'15 英日翻訳タスク

- 科学技術論文の対訳コーパス: 134万文
  - 語彙数: (英, 日) = (約6万, 約8万) + unknown
  - 開発データ: 1790文
  - テストデータ: 1812文
- 単語ベクトルと隠れ層の次元: 512
- 確率的勾配降下法により学習
- 計算環境: CPU, 16スレッド (約1週間)

# WAT'15 英日翻訳タスク

- ニューラル機械翻訳モデルとの比較
  - RIBESスコアは、英日翻訳間で人手判断との相関が高い  
(Isozaki et al., 2010)

	RIBES	BLEU
<b>提案モデル</b>	<b>81.42</b>	34.65
系列アテンションモデル (GPU) (Zhu, 2015)	79.70	32.19
+ 未知語処理, 事前並び替え処理, SMTモデル, アンサンブル	80.27	36.21
系列アテンションモデル (GPU) + 英語のスペル訂正, 文字列デコード (Lee et al, 2015)	81.15	35.75



# WAT'15 英日翻訳タスク

- 統計的機械翻訳モデルとの比較
  - RIBESスコアにおいて最高精度と同等

	RIBES	BLEU
<b>提案モデル</b>	<b>81.42</b>	34.65
ベースライン: 句ベース翻訳モデル	69.19	29.80
ベースライン: 階層句ベース翻訳モデル	74.70	32.56
ベースライン: tree-to-string翻訳モデル	75.80	33.44
tree-to-string翻訳モデル (Neubig and Duh, 2014)	79.65	36.58
+ 系列へのアテンションモデルによる リランキング (Neubig et al., 2015)	81.38	38.17

# 翻訳例1

[Reference]

セルにはアクティブマトリックス用液晶を注入した。

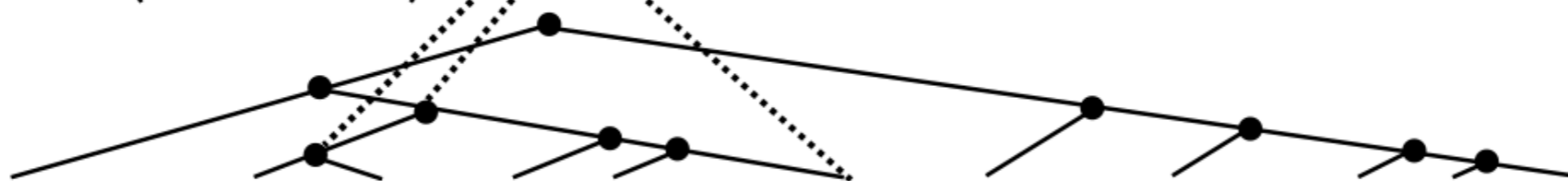
[Translation]

(a) 活性マトリックスの液晶をセル内に注入した。EOS

( $\alpha=0.77$ )

( $\alpha=0.41$ )

( $\alpha=0.07$ )



The liquid crystal for active matrix was injected into the cells. EOS

[Translation]

(b) 活性マトリックスの液晶をセル内に注入した。EOS

( $\alpha=0.31$ )

( $\alpha=0.24$ )

( $\alpha=0.35$ )



The liquid crystal for active matrix was injected into the cells. EOS

# 自動評価指標におけるとりこぼし

- 自動評価指標は参照訳との $n$ -gramマッチ
  - 単語の表層が異なる場合、スコアに反映されない
- 提案モデルにおける同義語の翻訳例：
  - **アクティブ** — **活性**
  - バージョン — 版
  - 女 — 女性
  - 2次に比例 — 二乗に比例

# 翻訳例2

The characteristics of r5 version of this software, instruction manual, and design document were summarized.

## Reference:

このソフトウェアのR5バージョンの特徴、利用マニュアルと設計文書をまとめた。

## Translation:

このソフトウェアのR5版の特性、命令マニュアル、設計文書についてまとめた。

# 翻訳例3

A case (85-year-old female) with dissecting aneurysm of the aorta of the \*UNK\* A type with a unique form (de \*UNK\* ii type) is experienced.

## Reference:

特異な形態を示したStanfordA型(De Bakey II型)の解離性大動脈りゅうの1例(女85歳)を経験した。

## Translation:

A型(de \*UNK\* II型)の大動脈の解離性動脈りゅうの1例(85歳女性)を経験した。

# 翻訳例4

A photocurrent signal is produced using light of 420nm, which is directly proportional to the square of radiation intensity .

## Reference:

420nm の光により, 光強度の2次に比例した光電流信号が発生する。

## Translation:

光電流信号は420nmの光を用いて生成され, それは放射強度の二乗に比例する。

# まとめ

- 既存のニューラル機械翻訳モデルの紹介
  - 最近のトレンドは**単語列から文字列へ**
- 句構造へのアテンションに基づくニューラル機械翻訳モデルを提案
  - 翻訳元言語の**構文構造**を陽に利用
  - アテンションにより**入力句/単語との関係**を柔軟に学習
  - WAT'15 英日翻訳タスク: 最高精度モデルと同等の RIBESスコアを達成

コード: <https://github.com/tempra28/tree2seq>