

修士学位論文公聴会

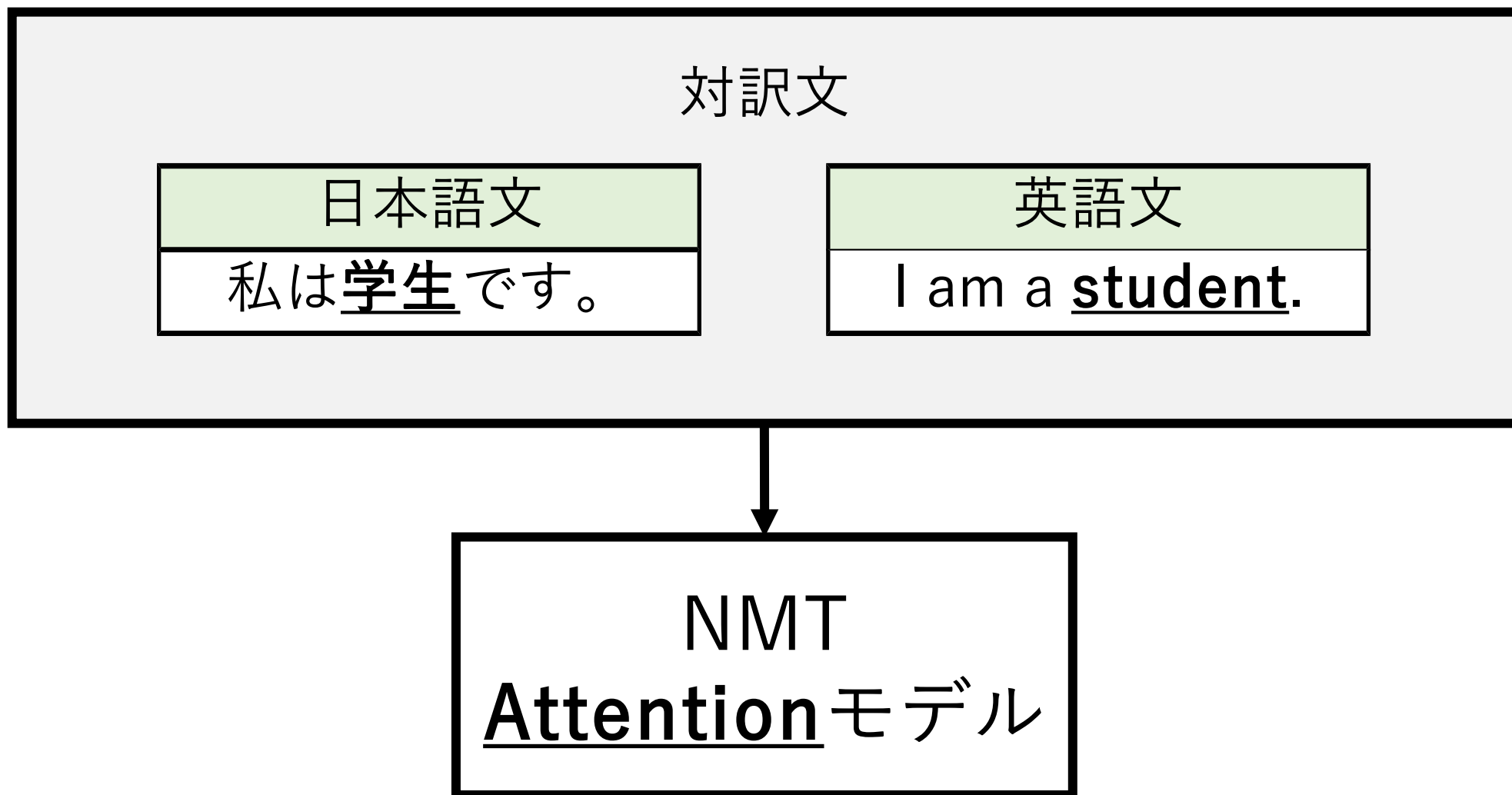
日英ニューラル機械翻訳における  
対訳句を利用した精度向上

自然言語処理研究室

M18J4007Z 今仁優希

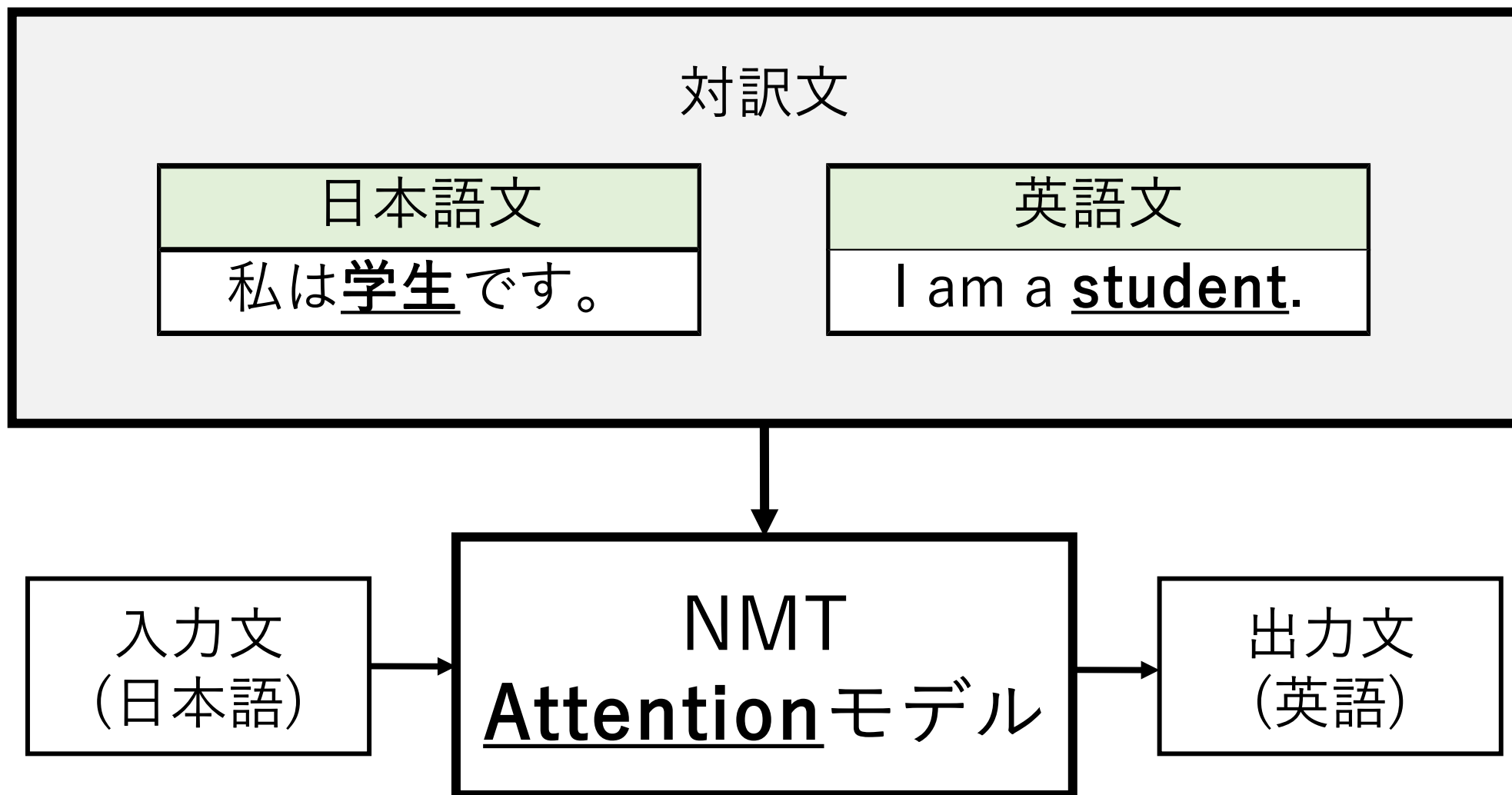
令和2年2月12日

# ニューラル機械翻訳(NMT)<sup>[1]</sup>



[1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In Proceedings of ICLR, 2015.

# ニューラル機械翻訳(NMT)<sup>[1]</sup>

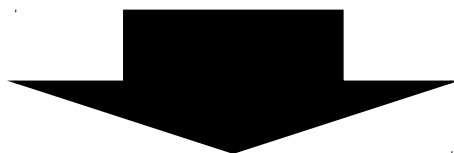


[1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In Proceedings of ICLR, 2015.

# NMTの問題点

---

語句の対応(Attention)の学習に誤り



誤った翻訳語句の生成



翻訳精度低下

# 先行研究

---

- 統計翻訳の問題：  
語句の対応（翻訳モデル）に誤り

学習データに対訳句を追加<sup>[2]</sup>  
語句の対応を強調



統計翻訳の精度向上

[2] 池淵堅斗. 対訳文から抽出した対訳句を用いた統計翻訳. 2015 年度修士論文, 2015.

# 研究の目的

---

- NMTの問題：語句の対応(Attention)に誤り

NMTの学習データに対訳句を追加

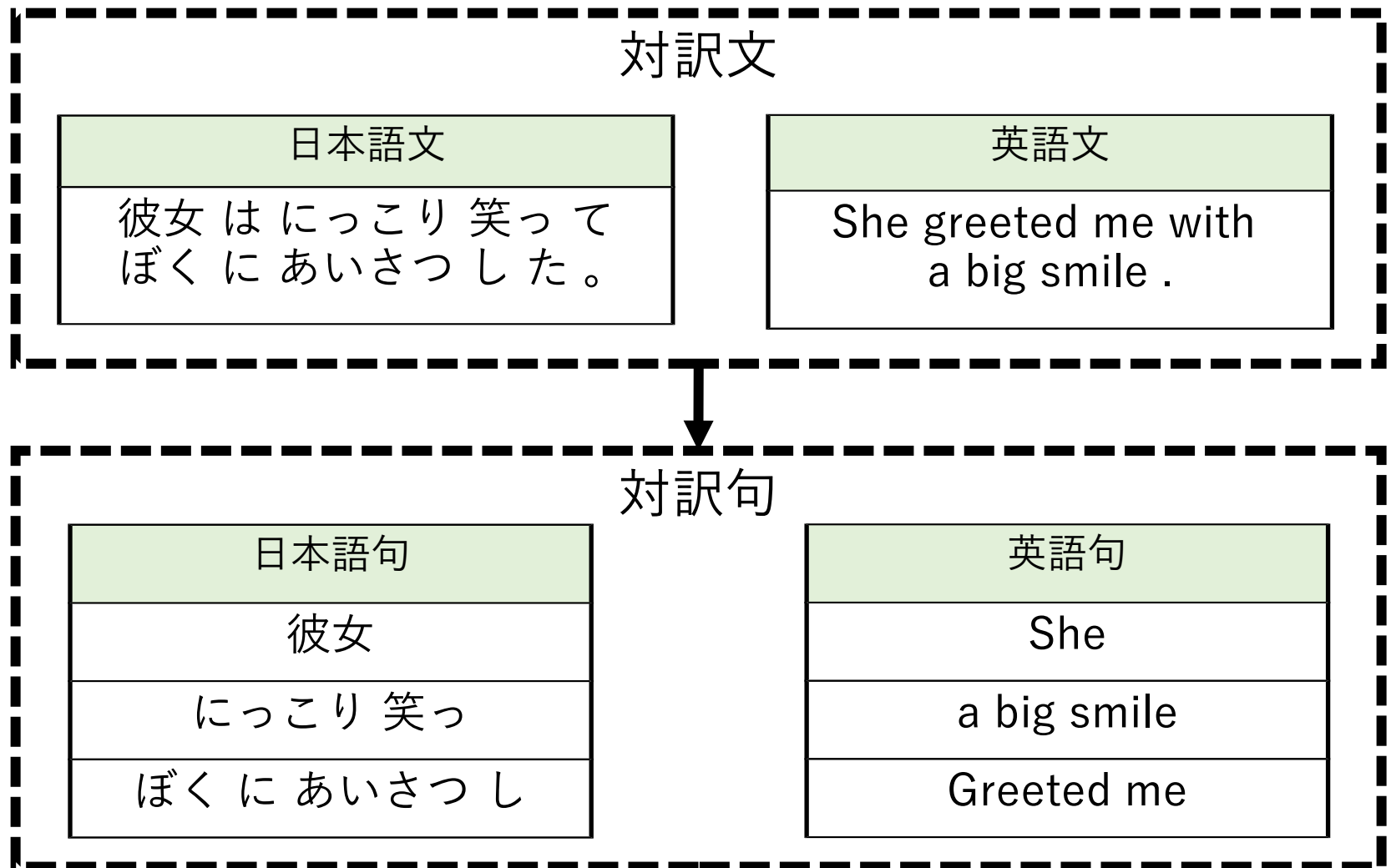


学習時の語句の対応を強調  
Attentionを強化

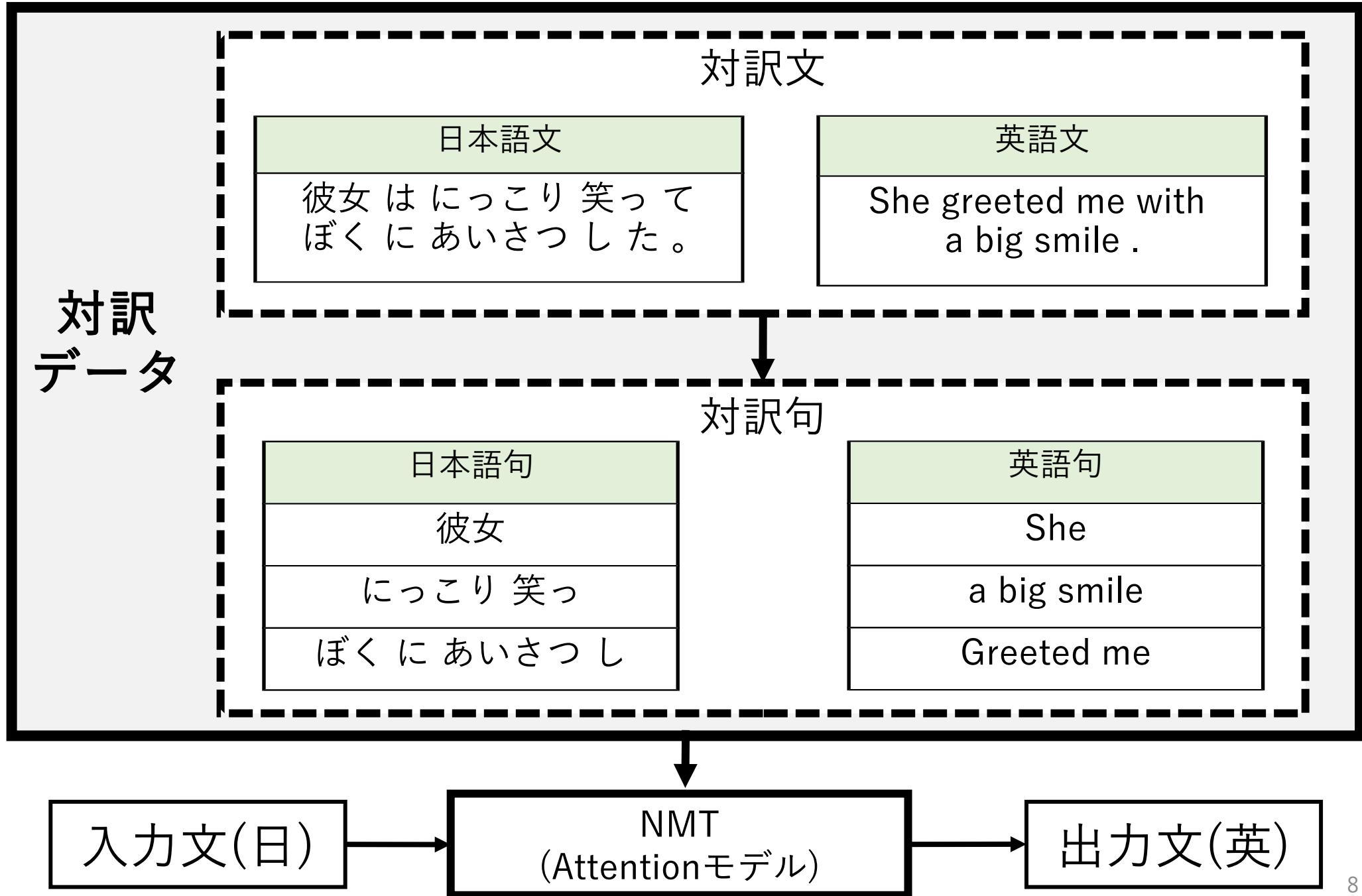


**NMTの翻訳精度向上**

# 対訳句を利用したニューラル機械翻訳



# 対訳句を利用したニューラル機械翻訳





# 実験環境

---

- 対訳文(重文複文10万文)

彼女はにっこり笑ってぼくにあいさつした。
She greeted me with a big smile .

- 比較実験

- ベースライン

- 対訳文のみを学習データに利用

- + 人手対訳句

- 人手で作成した対訳句を対訳文に追加

- + 自動対訳句

- 自動で作成した対訳句を対訳文に追加

# 対訳句の数と精度\*

## 人手対訳句 (549,694句)

\* 100句評価

○適切 86 句		△部分的に適切 10 句		×不適切 4 句	
条件	the conditions	暇	will be free	山田さん	he
演説し	made a speech	歌詞を書け	Fit the words	自社	their

## 自動対訳句 (26,757句)

○適切 36 句		△部分的に適切 44 句		×不適切 20 句	
食事の用意をし	preparing the meals	はしごを掛け	a ladder	入りて	Ogata
同じ道	same path	フラスコ	flask and heat	使っ	money

# 翻訳実験結果（人手対比較評価\*）

- ベースライン vs + 人手対訳句 \* 100文評価

ベースライン○	+人手対訳句○	差なし	一致
14文	<b>44文</b>	38文	4文

- ベースライン vs + 自動対訳句

ベースライン○	+自動対訳句○	差なし	一致
10文	<b>29文</b>	56文	5文

翻訳精度：提案手法 > ベースライン

# 出力例（提案手法○）

- +人手対訳句 > ベースライン

入力文	彼は自尊心が強すぎてとても <u>引き返せ</u> なかった。
参照訳	He was far too proud to go back .
ベースライン	He was too proud for his pride .
+人手対訳句○	He was too proud to <u>turn back</u> .

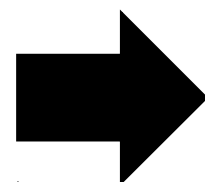
# 考察（精度向上の原因分析）

入力文	彼は自尊心が強すぎてとても <u>引き返せ</u> なかった。
ベースライン	He was too proud for his pride .
+人手対訳句○	He was too proud to <u>turn back</u> .

- 追加した人手対訳句

日本語側	英語側
引き返せ	turn back

正しい対応を強調



翻訳精度向上

# 出力例（ベースライン○）

- ベースライン > +自動対訳句

入力文	両国は <u>兵力を1万に抑える</u> ことで合意した。
参照訳	Both countries agreed to limit troops to 10,000.
ベースライン○	The two countries agreed to <u>restrict troops to 10,000</u> .
+自動対訳句	The two countries agreed to chase <u>to 10,000</u> .

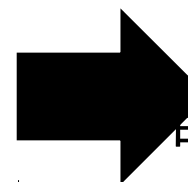
# 考察（精度低下の原因分析）

入力文	両国は <u>兵力を1万に</u> 抑えることで合意した。
ベースライン	The two countries agreed to restrict troops to 10,000.
+自動対訳句	The two countries agreed to chase <u>to 10,000</u> .

- 追加した自動対訳句

日本語側	英語側
兵力を1万	to 10,000

不適切な対応の対訳句



翻訳精度低下

## 考察（人手対訳句と自動対訳句の比較）

---

- 対訳句の数と精度  
人手対訳句 > 自動対訳句
- 翻訳精度の向上  
人手対訳句 > 自動対訳句
- 作成コストの利点  
自動対訳句 > 人手対訳句

より多量かつ高精度の自動対訳句  
➡ 低コスト & 更に精度向上



# まとめ

---

- 目的

  - NMTの語句の対応を強調

  - Attentionの学習を強化

- 提案手法

  - 対訳文より作成した対訳句を  
対訳文に追加

- 結果

  - 翻訳精度:

  - 提案手法 > ベースライン

- 課題

  - より多量かつ高精度の自動対訳句作成

# 精度向上の原因分析

- +人手対訳句 (44文向上)

正しい訳語 対訳句と相関あり	正しい訳語 対訳句と相関なし	その他 (ベースラインの出力文が著しく悪い、など)
28文	9文	7文

- +自動対訳句 (29文向上)

正しい訳語 対訳句と相関あり	正しい訳語 対訳句と相関なし	その他 (ベースラインの出力文が著しく悪い、など)
14文	9文	5文

# 精度低下の原因分析

- +人手対訳句 (14文低下)

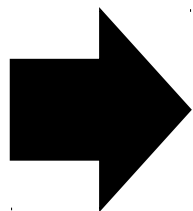
誤った訳語 <u>対訳句と相関あり</u>	誤った訳語 <u>対訳句と相関なし</u>	その他 (+自動対訳句の出力文が 著しく悪い、など)
6文	5文	3文

- +自動対訳句 (10文低下)

誤った訳語 <u>対訳句と相関あり</u>	誤った訳語 <u>対訳句と相関なし</u>	その他 (+自動対訳句の出力文が 著しく悪い、など)
4文	4文	2文

# 精度向上の原因分析（+ 人手対訳句） 対訳句作成に利用した対訳文

日本語文	その国は、 <u>引き返せ</u> ないところまでじきに行ってしまうことであろう。
英語文	he country will soon reach a point from which it will be unable to <u>turn back</u> .



日本語側	英語側
引き返せ	turn back

# 精度向上の例（+自動対訳句）

- +自動対訳句 > ベースライン

入力文	<u>警官に護送されて</u> 来た。
参照訳	He came under police escort .
ベースライン	He has been sent to the police .
+自動対訳句○	He came <u>under police escort</u> .

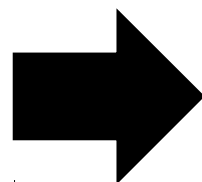
# 精度向上の原因分析（+自動対訳句） 追加した対訳句

入力文	<u>警官に護送されて</u> 来た。
ベースライン	He has been sent to the police .
+自動対訳句○	He came <u>under police escort</u> .

- 追加した自動対訳句

日本語側	英語側
警官に護送されて	under police escort

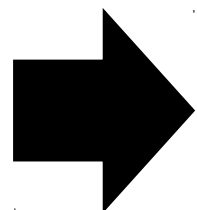
正しい対応を強調



翻訳精度向上

# 精度向上の原因分析（+自動対訳句） 対訳句作成に利用した対訳文

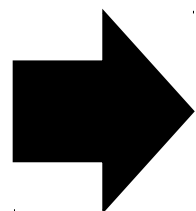
日本語文	犯人は <u>警官に護送されて</u> 東京へ連行された。
英語文	The criminal was brought to Tokyo <u>under police escort</u> .



日本語側	英語側
警官に護送されて	under police escort

# 精度向上の原因分析（+自動対訳句） 対訳句作成に利用した対訳文パターン

日本語側	英語側
犯人は <u>警官に護送されて</u> 東京へ連行された。	The criminal was brought to Tokyo <u>under police escort</u> .
N05 は <u>N04</u> N01 N02 N00 N03 た。	N05 N03 N00 N02 N01 <u>N04</u> .
彼は <u>日本</u> に帰ると決めた。	He ecided to return to <u>Japan</u> .



日本語側	英語側
警官に護送されて	under police escort



## 精度向上の例2（+自動対訳句）

- +自動対訳句 > ベースライン

入力文	彼女は <u>うつろな目</u> で事故現場を見つめていた。
参照文	She stared vacantly at the scene of the accident .
ベースライン	She gazed at the scene of the accident with her eyes .
+自動対訳句○	She gazed at the scene of <u>unseeing eyes</u> .

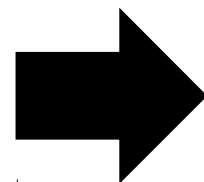
## 精度向上の原因分析 2 (+ 自動対訳句) 追加した対訳句

入力文	彼女はうつろな目で事故現場を見つめていた。
ベースライン	She gazed at the scene of the accident with her eyes .
+自動対訳句○	She gazed at the scene of unseeing eyes .

- 追加した自動対訳句

日本語側	英語側
うつろな目	unseeing eyes

正しい対応を強調



翻訳精度向上

# 精度低下の例（+人手対訳句）

- ベースライン > +人手対訳句

入力文	それを見て私は <u>どきどき</u> した。
参照訳	I throbbed at the sight .
ベースライン○	The sight made me <u>feel nervous</u> .
+人手対訳句	The sight <u>thumped</u> me .

# 精度向上の原因分析（+ 人手対訳句） 追加した対訳句

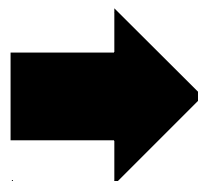
入力文	それを見て私は <u>どきどき</u> した。
ベースライン○	The sight made me <u>feel nervous</u> .
+ 人手対訳句	The sight <u>thumped</u> me .

- 追加した人手対訳句

日本語側	英語側
<u>どきどき</u> し	<u>thumped</u>
胸が <u>どきどき</u> し	My heart <u>thumped</u>
どきどきする	go thump
胸はどきどきし	his heart raced

部分的にのみ

適切な対応を強調



文全体として誤り

# 精度低下の例（+自動対訳句）

- ベースライン > +自動対訳句

入力文	ワインは飲み過ぎると <u>悪酔い</u> することがある。
参照訳	Too much wine can make you sick .
ベースライン○	Too much wine may make you <u>feel blue</u> .
+自動対訳句	Too much wine can <u>quarrel with one another</u> .

# 精度向上の原因分析（+ 人手対訳句）

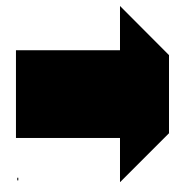
## 追加した対訳句

入力文	ワインは飲み過ぎると <u>悪酔いする</u> ことがある。
ベースライン○	Too much wine may make you <u>feel blue</u> .
+自動対訳句	Too much wine can <u>quarrel with one another</u> .

- 追加した自動対訳句

日本語側	英語側
悪酔い	with someone
悪酔いし	quarrel with someone

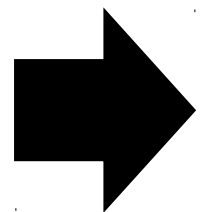
不正確な対応の対訳句



翻訳精度低下

# 精度低下の原因分析（+自動対訳句） 対訳句作成に利用した対訳文パターン

日本語側	英語側
目を <u>開</u> けて眠る。	He sleeps with his eyes <u>open</u> .
N04 N01 <u>N02</u> N00 N03 。	He N03 N00 N01 N04 <u>N02</u> .
彼はすぐ <u>悪酔い</u> して人からむ。	He got extremely drunk and picked a <u>quarrel with someone</u> .

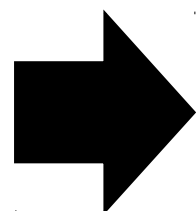


日本語側	英語側
悪酔いし	quarrel with someone

## 誤った自動対訳句を含む対訳文

日本語文	緒方氏の門に <u>入りて</u> 医学を学ぶ。
英語文	He studied medicine under Dr. <u>Ogata</u> .

- 作成された自動対訳句



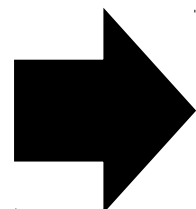
日本語側	英語側
入りて	Ogata



## 誤った自動対訳句を含む対訳文パターン

日本語側	英語側
緒方氏の門に <u>入りて</u> 医学を学ぶ。	He studied medicine under Dr. <u>Ogata</u> .
N04 N01 <u>N02</u> N00 N03 。	He N03 N00 N01 N04 <u>N02</u> .
目を <u>開け</u> て眠る。	He sleeps with his eyes <u>open</u> .

- 作成された自動対訳句



日本語側	英語側
入りて	Ogata

# NMTの低頻度語翻訳の問題

- **低頻度語**：  
対訳文中に出現する頻度の低い語
- 全く異なる訳語を生成

クリントン頻  
度6

入力文	クリントン 政権 はこれを 変えようと 試みた。
ベースライン	The <b>announcer</b> attempted to change this .
+自動対訳句	The <b>Clinton Administration</b> attempted to change this .

- 訳語を過剰に生成

微々たる  
頻度4

入力文	防衛 費 に 比べれば 文教 予算 など <b>微々たる</b> ものだ。
ベースライン	In comparison with defense costs , we are <b>small and small</b> accommodation .
+自動対訳句	Compared with our defense expenses , the budget budget is <b>almost negligible</b> .

# 低頻度語を含む文のみの評価

- 低頻度語を含む文のみ

翻訳手法	BLEU	METEOR	RIBES	WER
ベースライン	0.1434	0.4020	0.7257	0.7254
+人手対訳句	0.1693	0.4400	0.7364	0.6921
+自動対訳句	0.1528	0.4101	0.7268	0.7136

翻訳精度の向上：大

# 低頻度語を含む文のみの評価

- 低頻度語を含む文のみ

翻訳手法	BLEU	METEOR	RIBES	WER
ベースライン	0.1434	0.4020	0.7257	0.7254
+人手対訳句	0.1693	0.4400	0.7364	0.6921
+自動対訳句	0.1528	0.4101	0.7268	0.7136

- 低頻度語を除いた文

翻訳手法	BLEU	METEOR	RIBES	WER
ベースライン	0.1865	0.4603	0.7563	0.6625
+人手対訳句	0.1956	0.4790	0.7626	0.6483
+自動対訳句	0.1931	0.4670	0.7606	0.6514

# 実験結果（自動評価）

## 自動評価(10,000文)

翻訳手法	BLEU	METEOR	RIBES	WER
ベースライン	0.1534	0.4147	0.7312	0.7107
+人手対訳句	0.1701	0.4430	0.7411	0.6873
+自動対訳句	0.1585	0.4194	0.7332	0.7017

翻訳精度：

+ 人手対訳句 > + 自動対訳句 > ベースライン

# 外部辞書データの対訳句を利用

英辞郎<sup>[3]</sup> (**550,000句**)

日本語側	英語側
シャンデリア	chandelier
継承する	inherit
激烈な論争	bitter controversy

+ 英辞郎 vs ベースライン

ベースライン○	+英辞郎○	差なし	一致
14文	<b>34文</b>	49文	3文

[3] 英辞郎 <http://www.alc.co.jp/>.

# 人手対訳句の作成方法

- 対訳文

彼の お母さんがああ若いとは思わなかった。

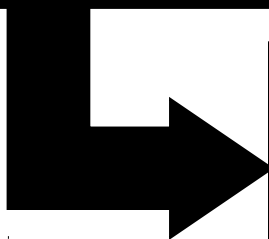
I never expected his mother to be so young .

- 人手対訳文パターン

NP Nがああ AJとは Vなかった。

I never V NP N to be so AJ .

対訳文と  
パターンを照合



日本語側	英語側
<u>彼の</u>	<u>his</u>
お母さん	mother
若い	expected

# 自動対訳句の作成方法

- 対訳文

彼女 は 3 か国語を話します。

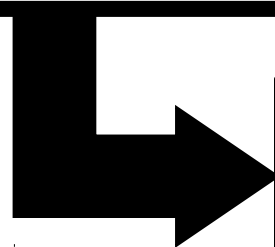
She speaks three languages .

- 自動対訳文パターン

N1 は N2 を話します。

N1 speaks N2 .

対訳文と  
パターンを照合



日本語側	英語側
<u>彼女</u>	<u>She</u>
3 か国語	three languages



# 実験環境

## 対訳文(重文複文10万文)

彼女はにっこり笑ってぼくにあいさつした。

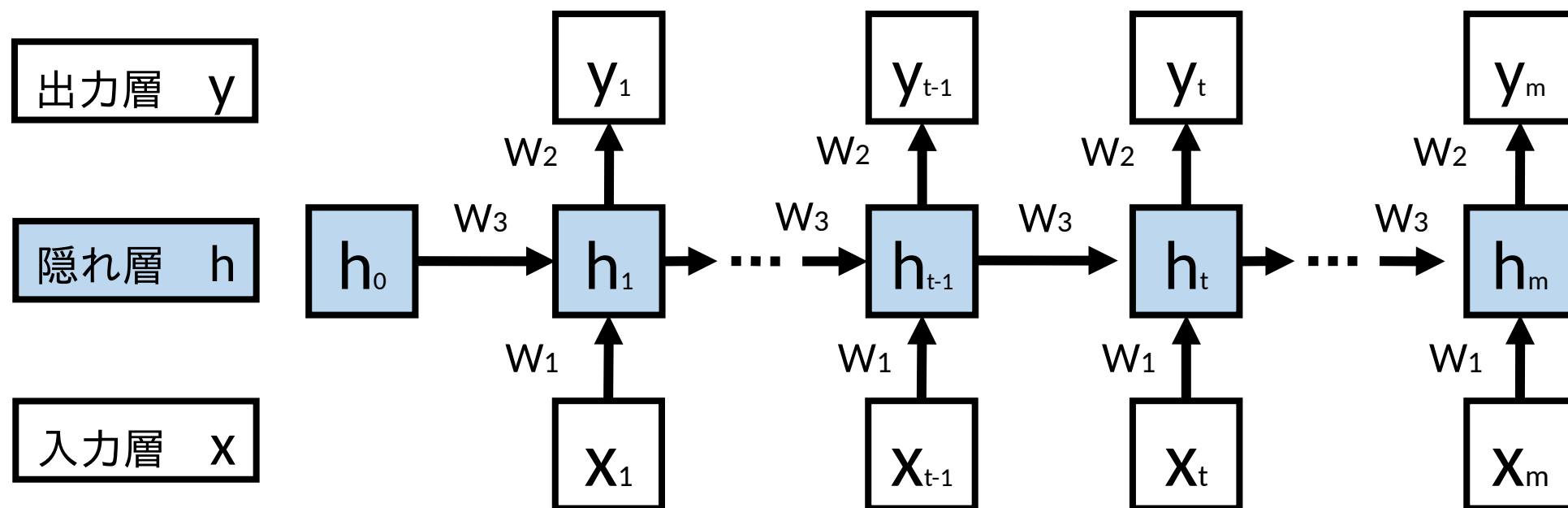
She greeted me with a big smile .

## 対訳句

- 人手対訳句  
(549,694対)
- 自動対訳句  
( 26,757対)

彼女	She
にっこり笑っ	a big smile
ぼくにあいさつし	Greeted me

# リカレントニューラルネットワーク(RNN)



$$h_t = f(w_1 x_t + w_3 h_{t-1})$$
$$y_t = g(w_2 h_t)$$

$(f, g: \text{活性化関数})$   
 $w: \text{重み}$

# 注意機構(Attention)モデル

