

# Chunk-based Decoder for Neural Machine Translation

Shonosuke Ishiwatari<sup>1</sup>, Jingtao Yao<sup>2</sup>, Shujie Liu<sup>3</sup>, Mu Li<sup>3</sup>, Ming Zhou<sup>3</sup>, Naoki Yoshinaga<sup>4</sup>, Masaru Kitsuregawa<sup>4,5</sup>, Weijia Jia<sup>2</sup>

 <sup>1</sup>The University of Tokyo

 <sup>2</sup>Shanghai Jiao Tong University

 <sup>3</sup>Microsoft Research Asia

 <sup>4</sup>Institute of Industrial Science, the University of Tokyo

 <sup>5</sup>National Institute of Informatics

## Overview

### Translation between distant language pairs are difficult

1. Some languages use many words to represent one thing while others use less words
2. Some languages are free word-order while others are not

### Idea: Using a chunk rather than a word as basic translation unit

- ☺ Sequence of a sentence becomes much shorter
- ☺ Fixed word order and free chunk order can be modeled independently

### Our decoder for NMT


decodes sentences in a “chunk-by-chunk” manner to overcome the differences of length and word-order


☺ **SOTA performance in distant languages (En -> Ja)**

## Difficulties of translation of En -> Ja

### Japanese sentence has

- longer sequence (En: 25 vs. Ja: 30 [words/sentence])
- free chunk order (e.g., 「だれかが / 犬に」 = 「犬に / だれかが」)

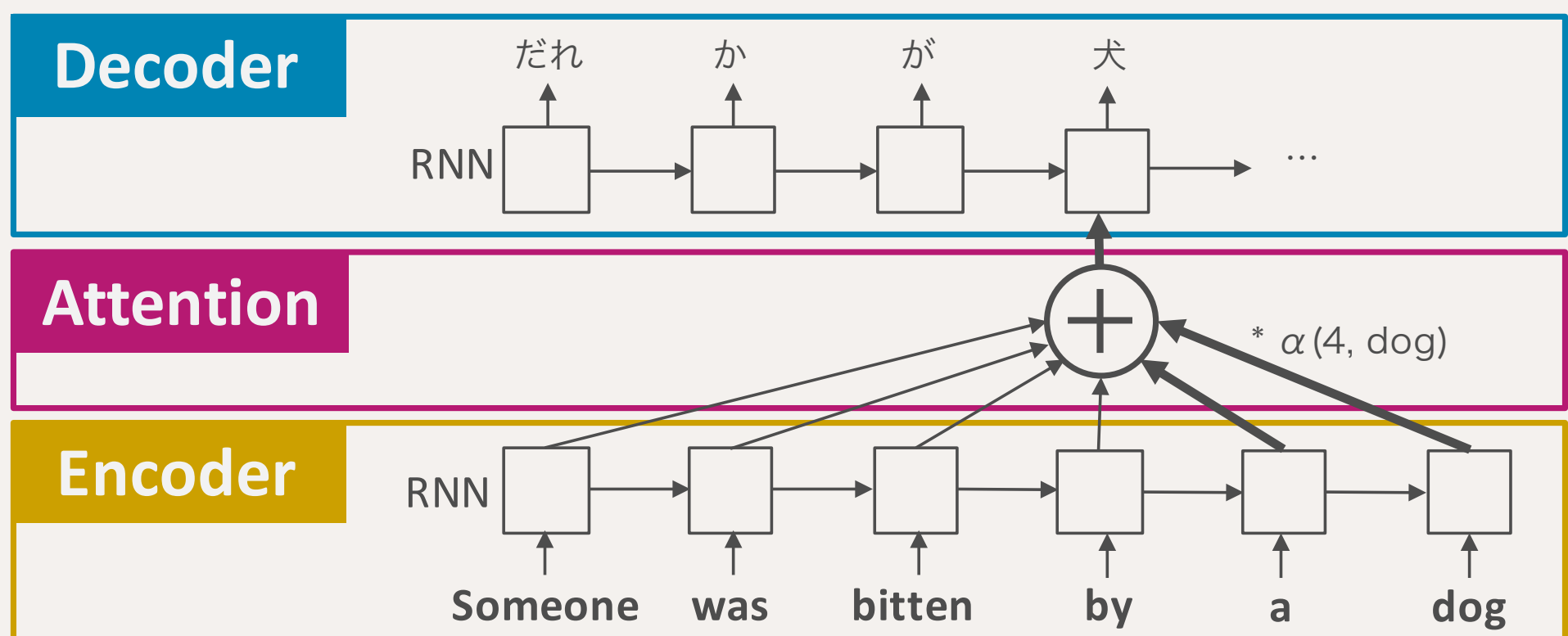
 I heard that someone was bitten by a dog, weren't you injured?

 だれ かが / 犬 に / 噛ま れ た そう だ けれど、 / 君 は / 怪我 し なかっ た ？

## Word-based Encoder-Decoder

### Enc-Dec with Attention [Bahdanau+ 15]

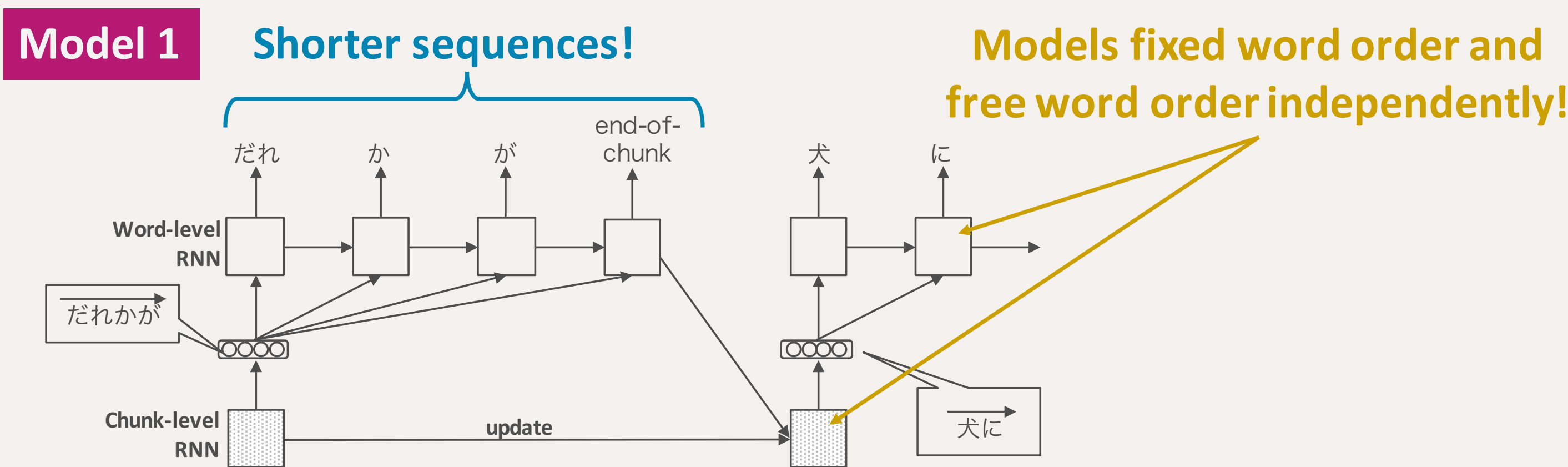
- encodes / decodes “word-by-word”



## Proposal: Chunk-based Decoder for Neural Machine Translation

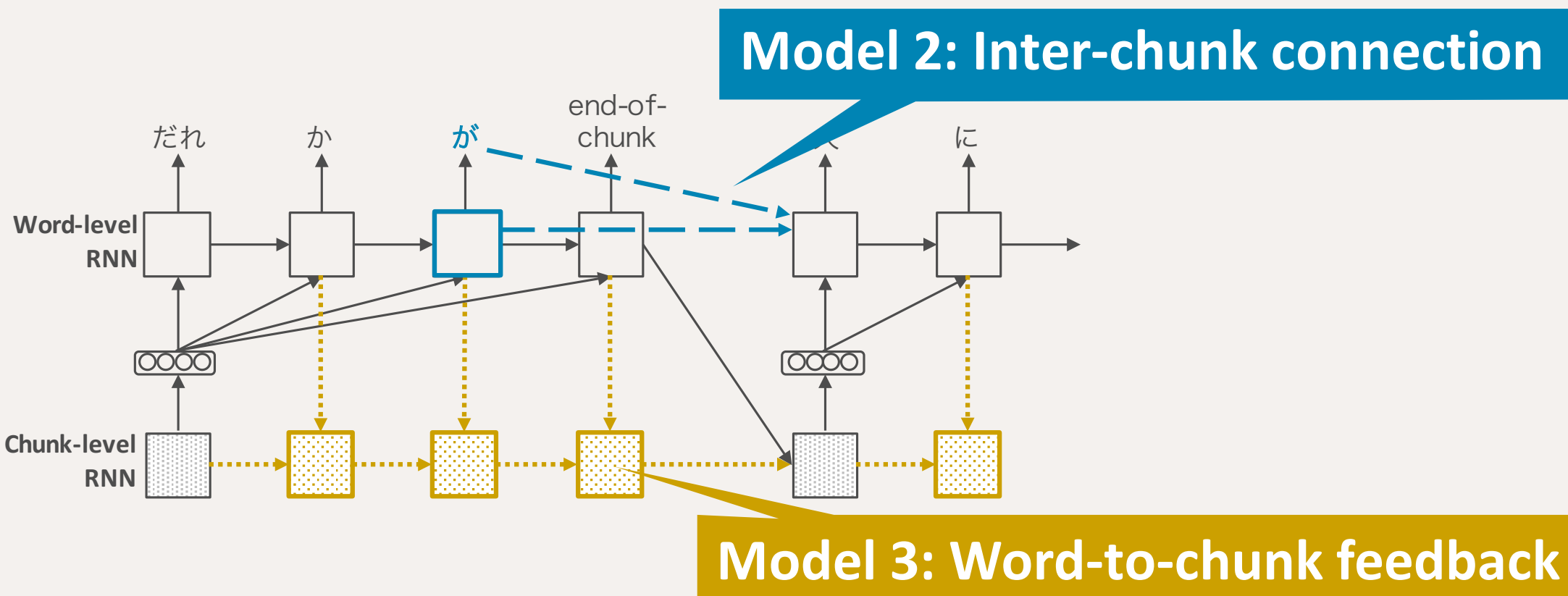
### Two step decoding

- First a chunk, then words inside the chunk



### Two additional connections

1. to capture the interaction between chunks
2. to memorize previous outputs well



## Experiments

### Data

- ASPEC [Nakazawa+ 16], 1.6M En/Ja pairs

### Preprocessing

- Bunsetsu chunking with J.DepP [Yoshinaga & Kitsuregawa 09]

### Baseline systems

1. Word-based encoder-decoder [Bahdanau+ 15]
2. Tree-based encoder [Eriguchi+ 16] (SOTA)

### Results

#### Quality of translation

Model		BLEU	RIBES
Word-based encoder + Chunk-based decoder [Proposed]	Model 3	<b>37.26</b>	<b>82.23</b>
	Model 2	35.81	81.29
	Model 1	34.70	81.01
Tree-based encoder [Eriguchi+ 16] + Word-based decoder		(34.91)	(81.66)
Word-based encoder-decoder [Bahdanau+ 15]		36.33 (34.64)	81.22 (81.60)

#### Quality of generated chunks

Decoder type	BLEU	RIBES
Chunk-based (Model 3)	<b>8.69</b>	<b>52.82</b>
Chunk-based (Model 2)	7.78	51.48
Chunk-based (Model 1)	7.59	50.47
Word-based	7.56	50.73

※ Scores inside () are not our implementations

### Translation examples

Source	the atmospheric glow discharge is a homogeneous electric discharge obtained by applying alternating voltage after introducing atmospheric He gas in a typical dielectric barrier discharge reactor .	because user operation is important for the idea support in material development, an interface for a substance operation at atomic level was developed .
Reference	大気圧グロー放電は典型的な誘電体バリア放電リアクタに大気圧ヘリウムを導入して交流電圧を印加すると得られる均一な放電である。	材料開発における発想支援のためには、ユーザの側の操作が重要であるため、原子レベルでの物質操作のためのインターフェイスを開発した。
Word-based [Bahdanau+ 15]	Wrong chunk location 典型的な誘電体バリア放電反応器において大気中のHeガスを導入した後、交流を印加することにより、大気グロー放電は均一な放電である。	支援 (support) is missing 材料開発の考え方にはユーザ操作が重要であり、原子レベルでの物質操作のためのインターフェイスが開発されている。
Chunk-based [proposed]	大気中グロー放電は、 / 典型的な / 誘電体バリア放電反応器に / 大気 He ガスを / 導入した / 後、 / 交流を / 印加して / 得られる / 均一な / 放電である。	材料開発に / おける / アイデア支援には / ユーザ操作が / 重要であるので、 / 原子レベルでの / 物質操作の / ための / インタフェイスを / 開発した。