



# 深層学習による機械翻訳の発展と 同時通訳への挑戦

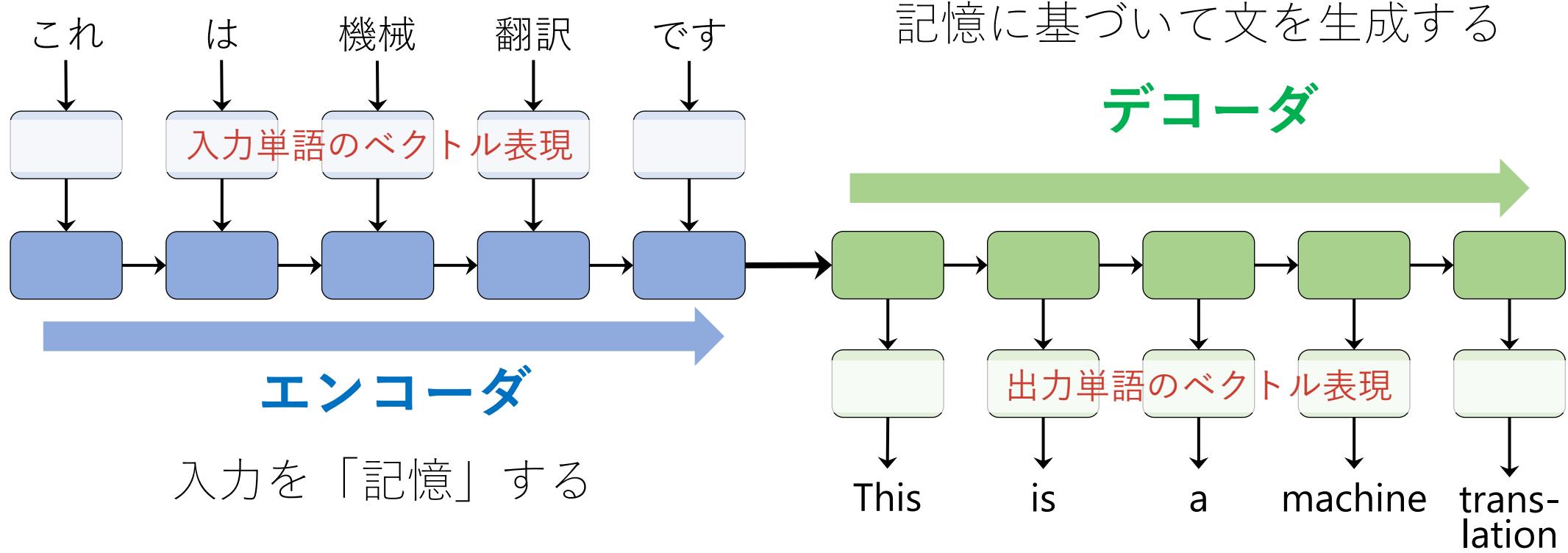
須藤 克仁

奈良先端科学技術大学院大学

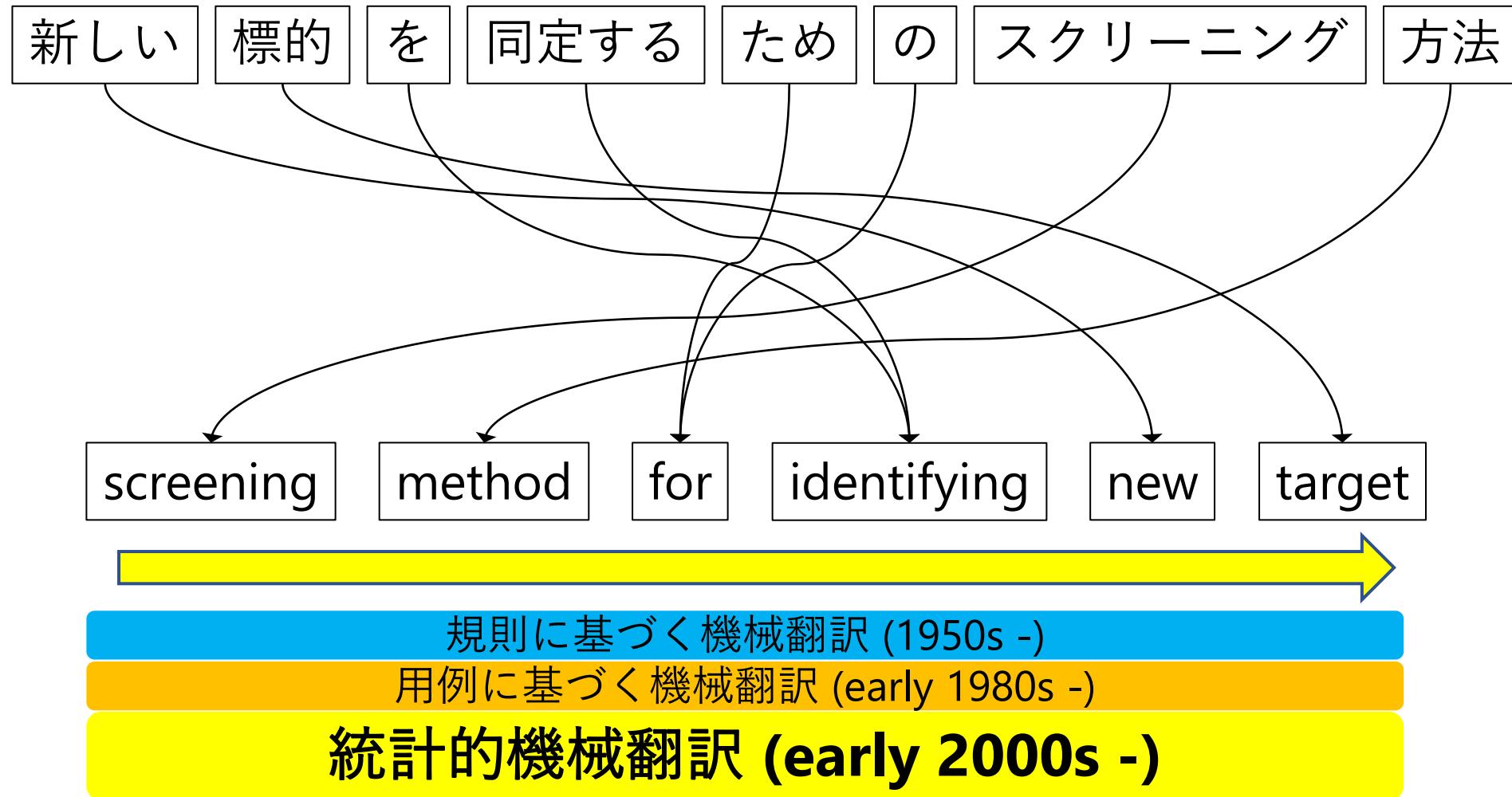
# ニューラル機械翻訳

# ニューラル機械翻訳

- ・系列変換 (seq2seq) 問題として定式化



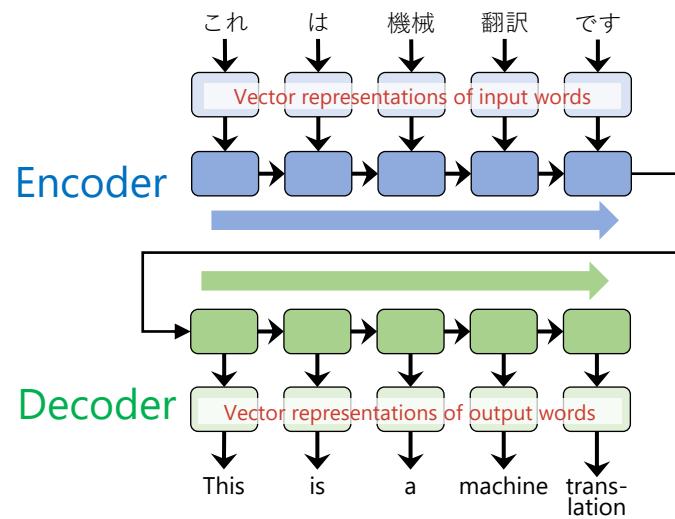
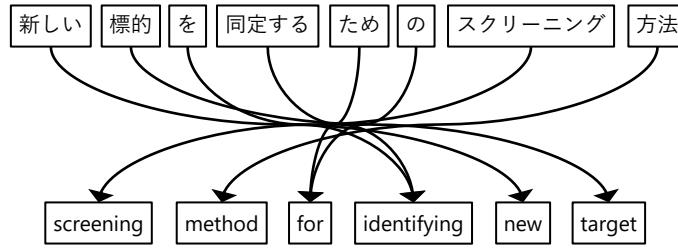
# 以前の機械翻訳のアプローチ



# 従来型機械翻訳とニューラル機械翻訳

## 従来型

- 要素単位の翻訳を組み合わせ
- 全入力単語を被覆するよう翻訳



## Neural

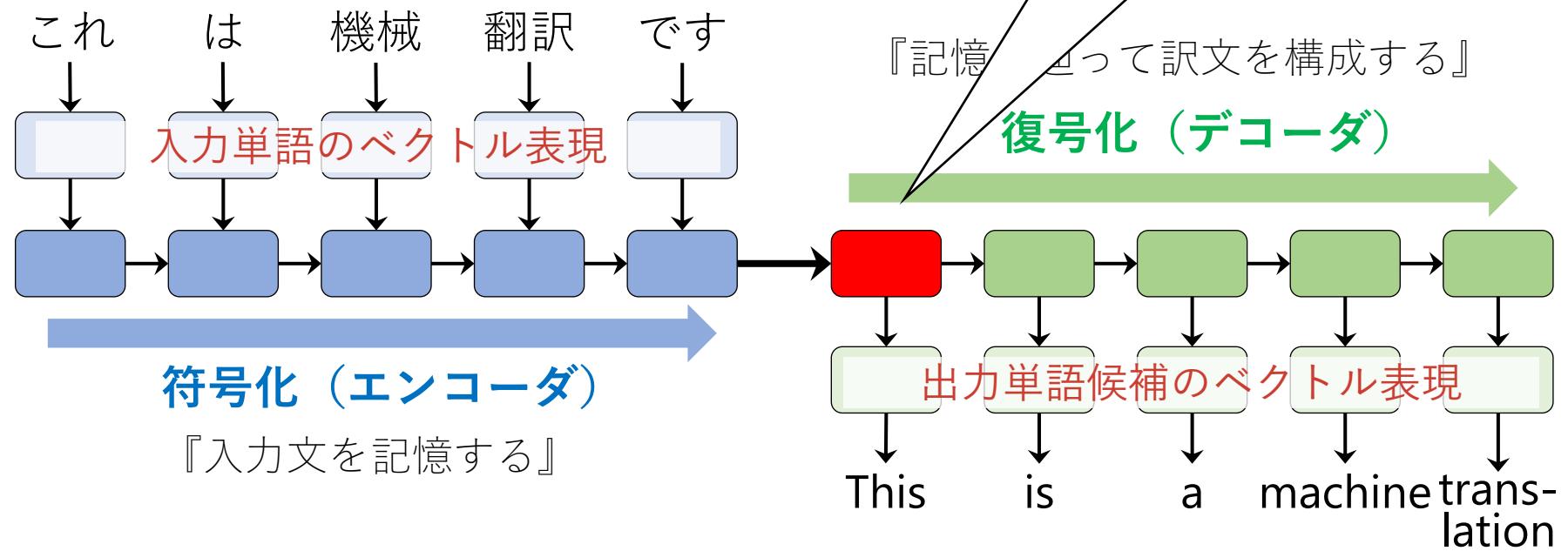
- 入力の全被覆は保証されない
- 訳抜けや過剰な訳語生成が問題

# ニューラル機械翻訳の利点

- 流暢な訳文が得られる
  - 文法的に問題のある入力であっても...
- 実装が簡単
  - Python数百行程度 (+ 深層学習フレームワーク)
- 処理の柔軟性
  - 多言語機械翻訳
  - マルチモーダル機械翻訳（原文 + 画像 → 訳文）

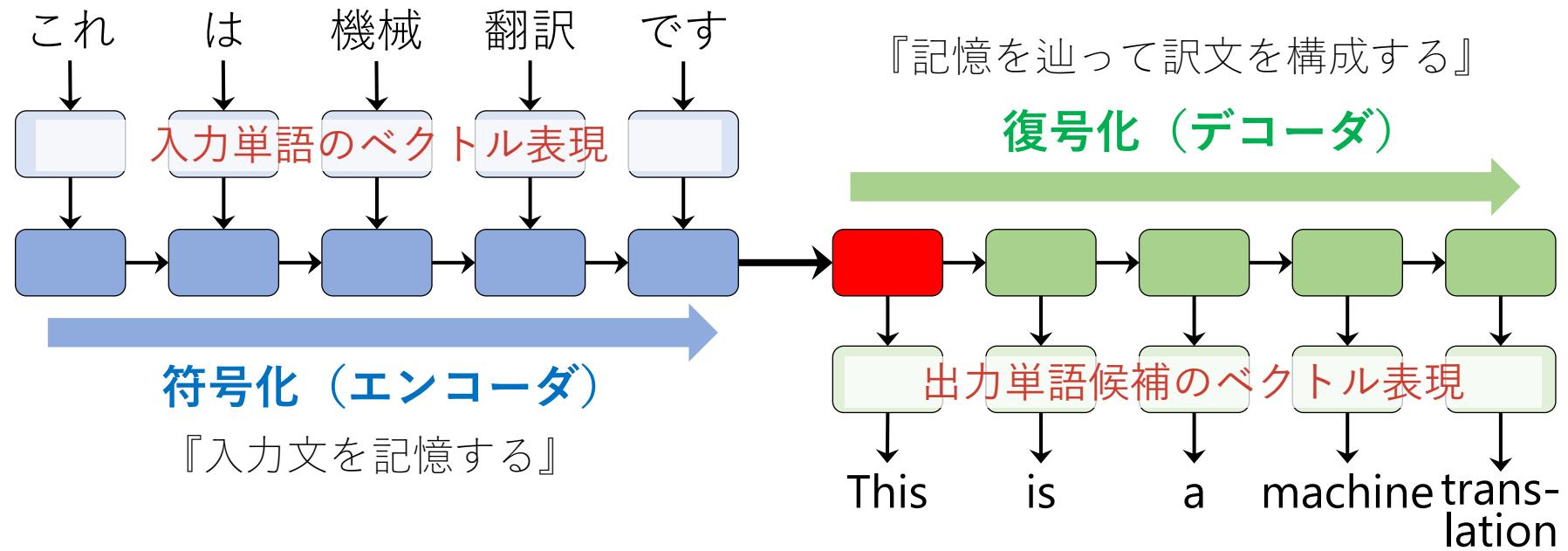
# 単純なニューラル機械翻訳の問題

- ・入力文をLSTMで記憶
- ・LSTMで出力文を生成



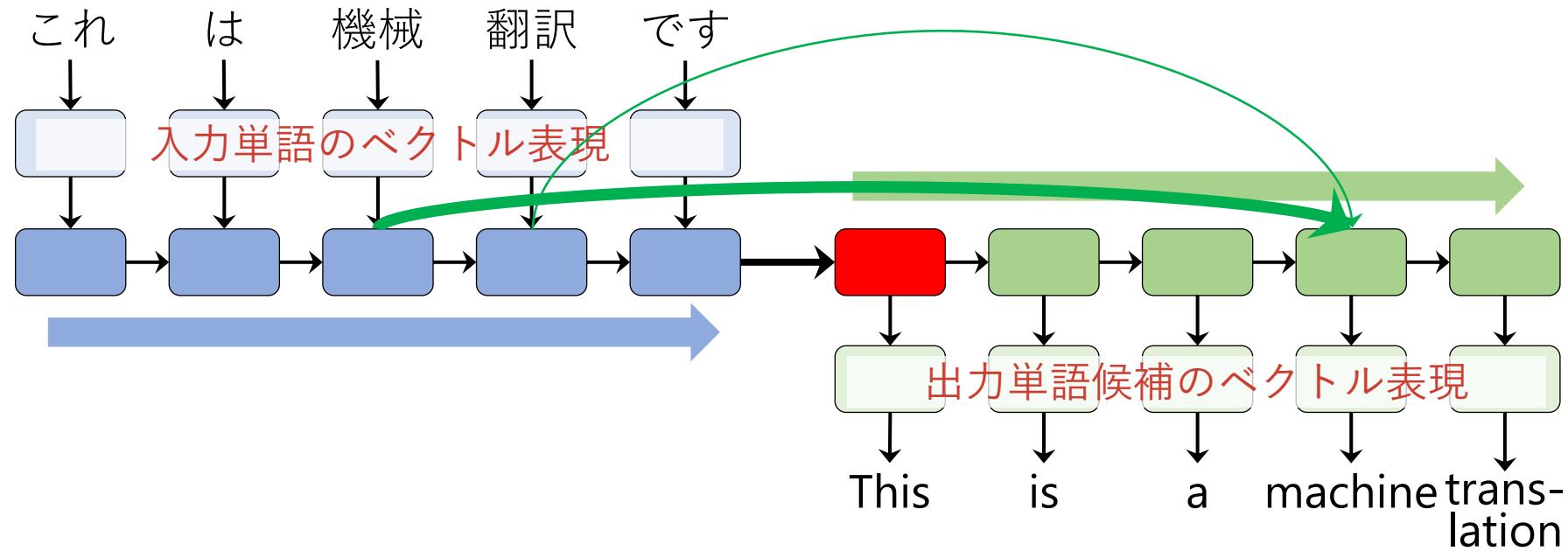
# 単純なニューラル機械翻訳の問題

- 一つのベクトルが全てを担う
  - 読み書きした単語の記憶・忘却
  - 端的に文が長いと荷が重い



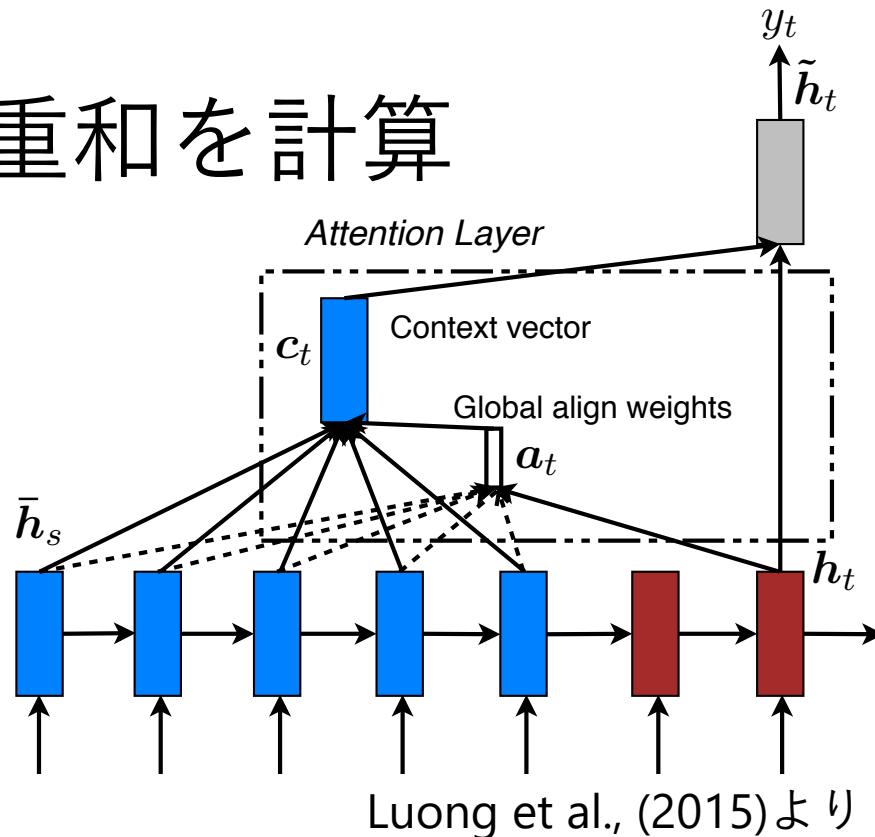
# 注視(attention)機構の導入

- 記憶を分散させ局的に利用
  - 記憶の途中過程を保持
  - 注視による記憶の加重和で訳語選択



# 注視の仕組み

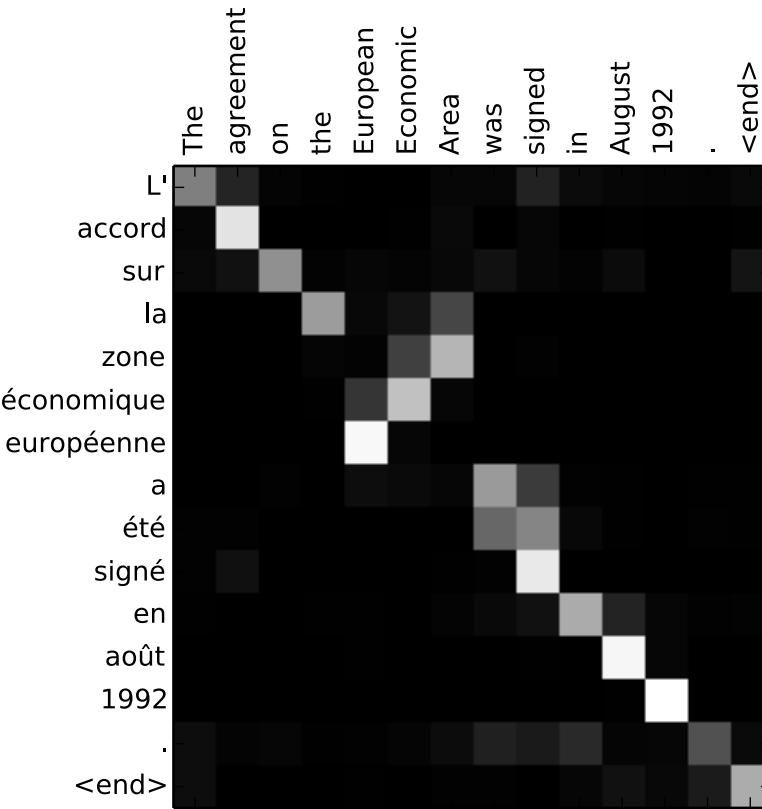
- ・デコーダ状態ベクトル( $h_t$ )に基づき
  - ・注視重み( $a_t$ )を計算
  - ・記憶ベクトル( $\bar{h}_s$ )の加重和を計算文脈ベクトル( $c_t$ )
- ・ $h_t$ と  $c_t$ を用いて訳語を選択する



Luong et al., (2015)より

# 注視と単語対応

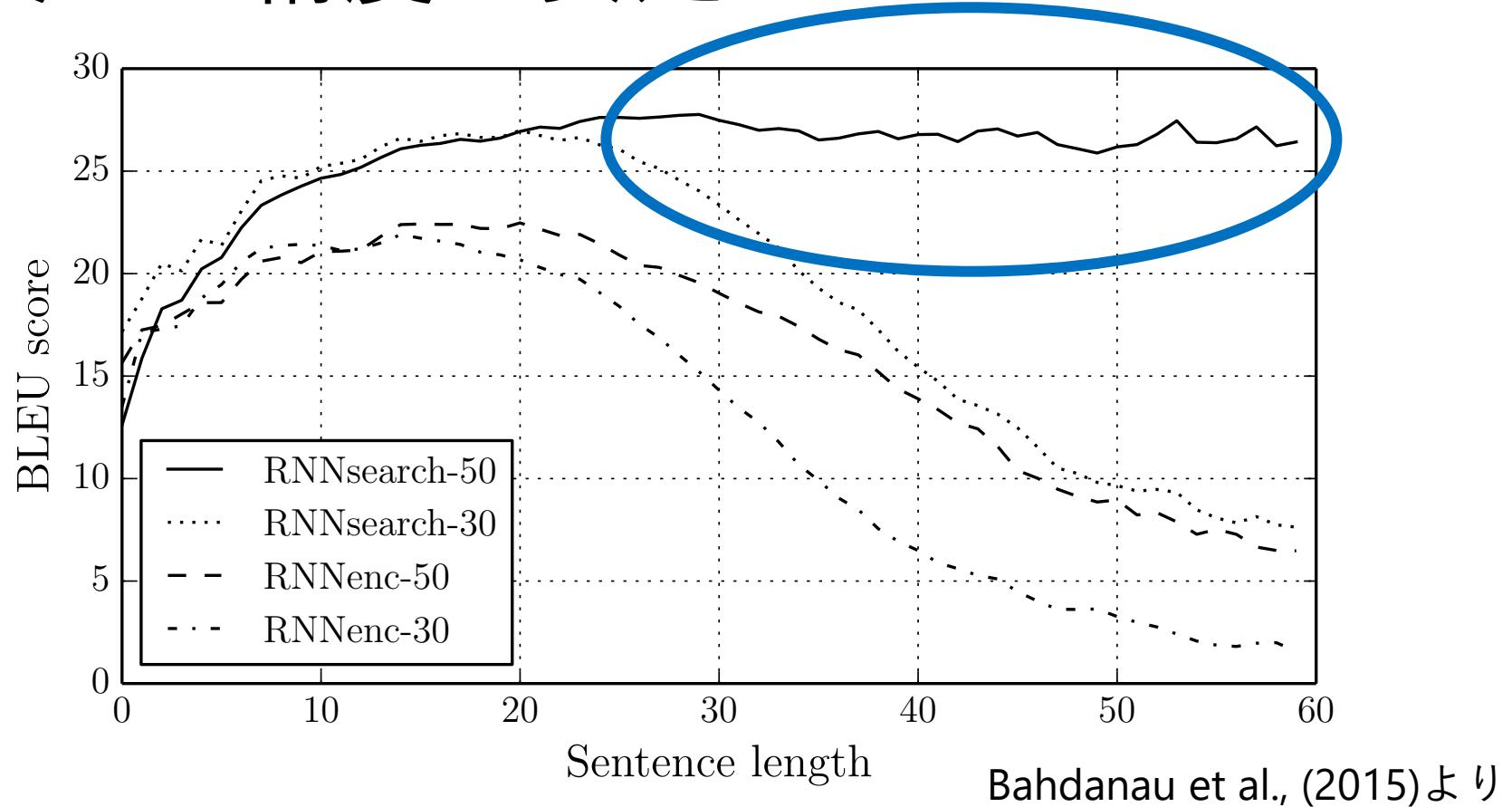
- 注視により単語対応が得られる？
- 得られることが多いが  
対応は保証されない
- ニューラル翻訳は  
訳せるように学習  
できれば何でもよい



Bahdanau et al., (2015) より

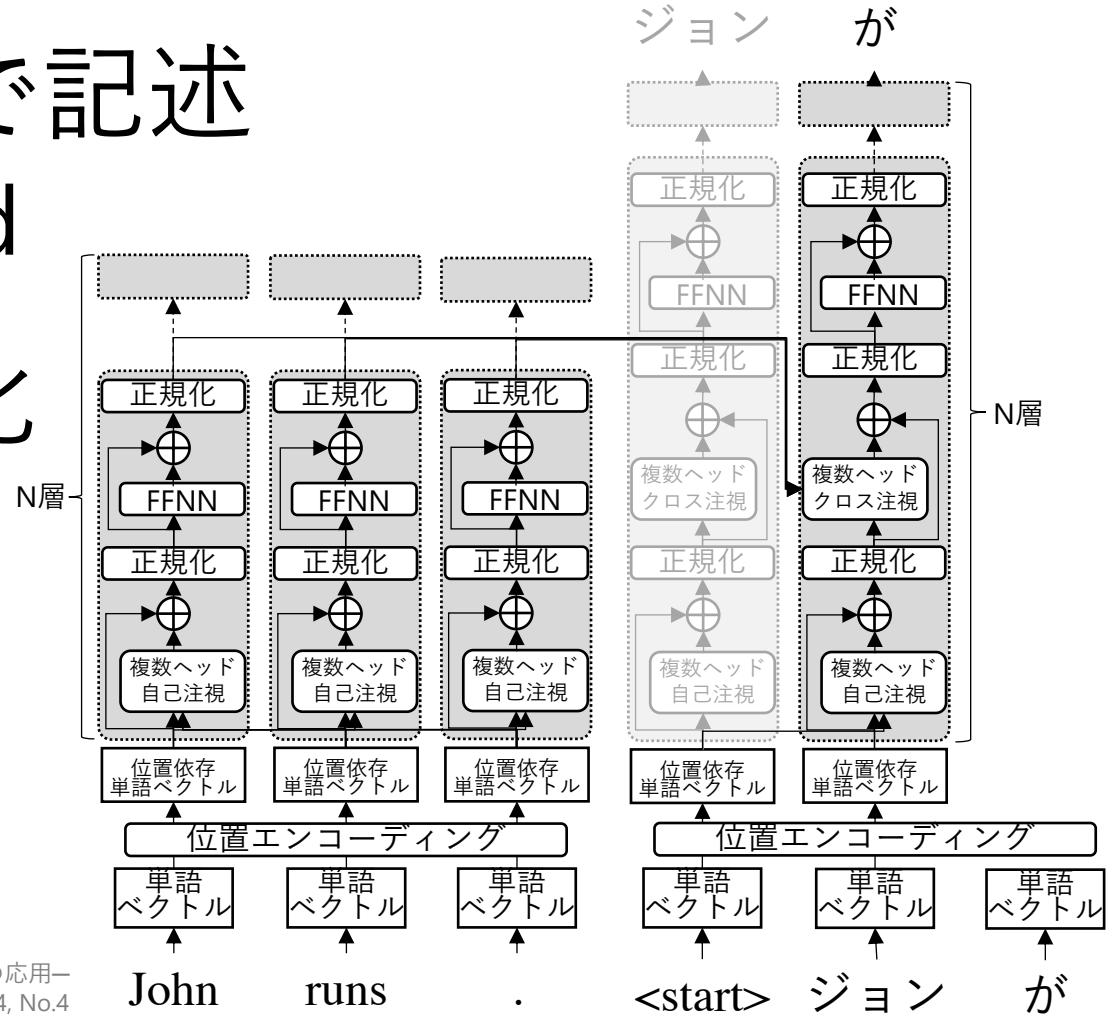
# 効果：長文における精度改善

- 長い文でも精度が安定



# Transformerの構成

- 要素間関係を注視のみで記述
  - Attention is All You Need
- 複数回注視を行い抽象化



須藤, ニューラル機械翻訳の進展—系列変換モデルの進化とその応用—  
人工知能学会誌 Vo.34, No.4

# Transformerの特徴

- エンコーダを並列化
  - 位置情報を周期関数で埋め込む (positional encoding)
- 注視の工夫
  - 自己注視 (self attention)
    - エンコーダが入力、デコーダが（直前までの）出力を注視可能
  - 複数ヘッド (multi-head) 注視
    - 異なるタイプの情報を独立に注視可能

# Transformerの応用

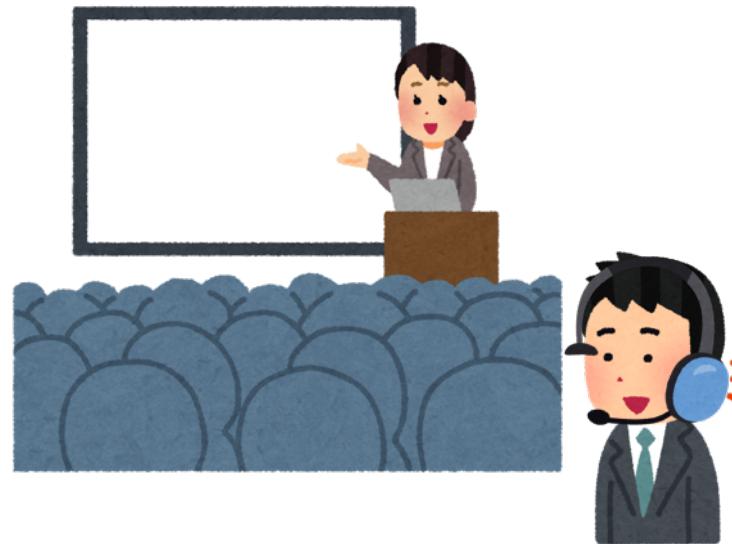
- NMTの新しい de facto standard
  - 対訳データが大量にあれば特に強力
- BERT: Transformerを用いた文符号化
  - 自然言語処理の事前知識として利用
  - 質問応答等の様々な応用で精度向上
  - 様々な言語で学習済みモデル公開

# 自動同時通訳の実現に向けて

※ 本研究の一部はJSPS科研費JP17H06101の助成を受けたものです

# 逐次通訳と同時通訳

- 逐次通訳
  - 発話ごとに通訳
- 同時通訳
  - 発話と **同時並行で** 通訳



# 同時通訳における遅延

(1) The relief workers (2) say (3) they don't have (4) enough food, water, shelter, and medical supplied (5) to deal with (6) the gigantic wave of refugees (7) **who are ransacking the countryside** (8) **in search of the basics** (9) to stay alive.

出典: 水野 的 (Akira Mizuno), 『同時通訳の理論: 認知的制約と訳出方略』

# 同時通訳における遅延

(1) The relief workers (2) **say** (3) they don't have (4) enough food, water, shelter, and medical supplied (5) to deal with (6) the gigantic wave of refugees (7) who are ransacking the countryside (8) **in search of the basics** (9) **to stay alive.**

(1) 救援担当者は (9) **生きるための** (8) **食料を** 求めて (7) **村を荒らし回っている** (6) **大量の** 難民たちの世話をするための (4) **十分な食料** や水, 宿泊施設, 医薬品が (3) **無いと** (2) **言っています**

# 順送りの訳

(1) The relief workers (2) **say** (3) they don't have (4) enough food, water, shelter, and medical supplied (5) to deal with (6) the gigantic wave of refugees (7) who are ransacking the countryside (8) in search of the basics (9) **to stay alive.**

(1) 救援担当者たちの (2) **話では** (4) 食料, 水, 宿泊施設, 医薬品が (3) **足りず** (6) 大量の難民たちの (5) 世話ができない**とのことです。**  
(7) **難民たちは**今村々を荒らし回って, (9) 生きるための (8) 食料を求めているのです。

# 遅延削減に向けた戦略

1. 部分的な訳出が可能になった時点で訳出を開始する
2. 訳出の順序を入力に近いものに変更する
3. 訳出を簡明 (concise) にする  
*(現時点では未だ「翻訳」)*
4. 入力を予測す  
*(現時点では未だ「翻訳」)*

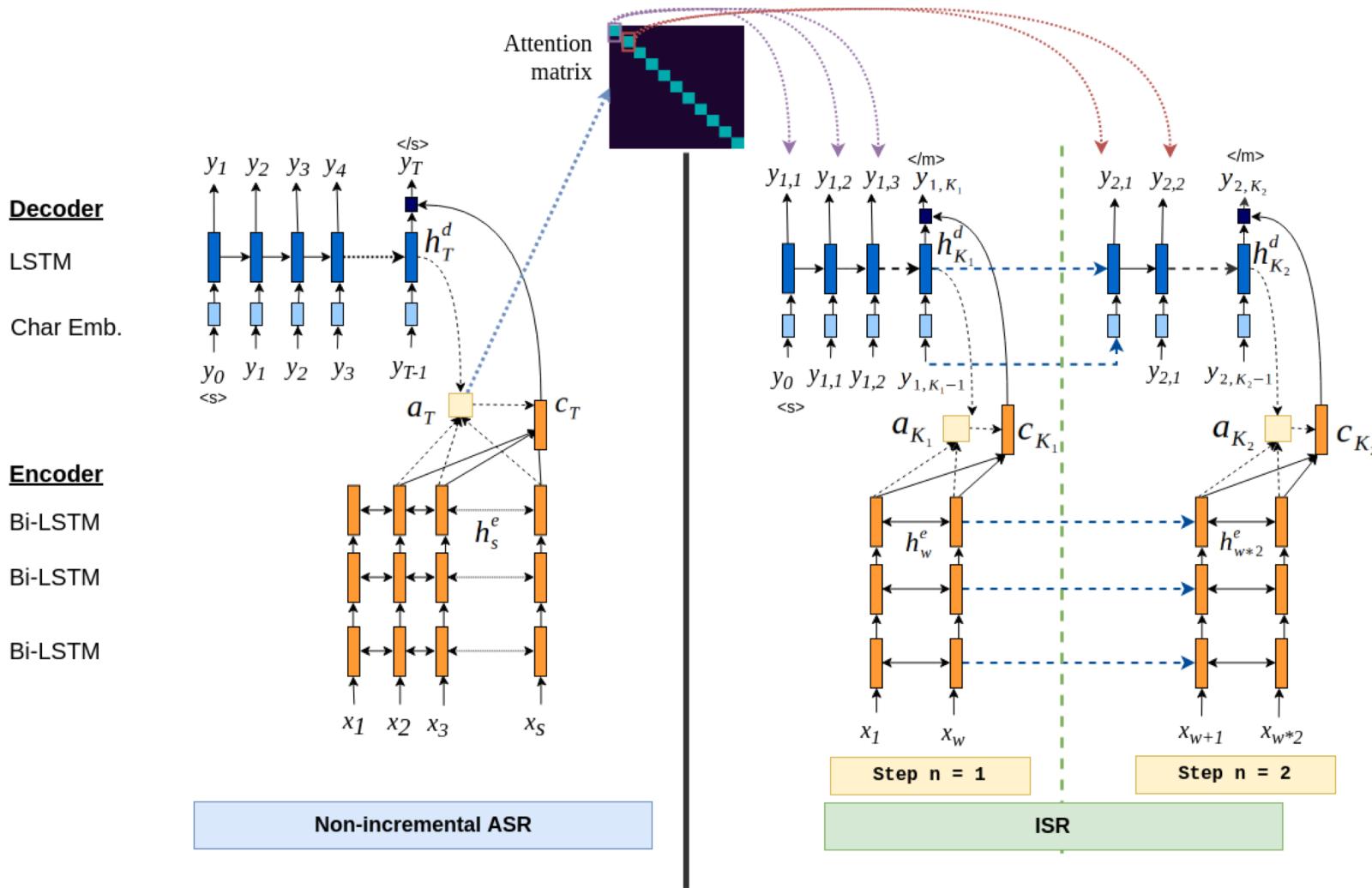
# 漸進的音声認識

Novitasari, Tjandra, Sakti, and Nakamura: Sequence-to-Sequence Learning via Attention Transfer for Incremental Speech Recognition, Proc. Interspeech 2019, pp. 3835-3839 (2019)

# 音声認識と漸進的処理

- 音声認識における文脈情報
  - タスク 자체はmonotoneで並べ替えはない
  - 言語モデル観点での文脈は必要
- 漸進的処理を実現する方式
  - HMM: 長距離/後方文脈を考慮できない → 自明
  - End-to-End: 入力全体を注視機構で参照可能
    - 処理単位を分割して参照範囲を限定
    - どう注視範囲を限定させるか？

# 提案手法: 注視転移による漸進的音声認識



- 教師モデルは全体を参照可能
- 生徒モデルは局所情報のみ利用
  - 教師モデルの注視を最大限真似る学習

# 漸進的機械翻訳

帖佐, 須藤, 中村: 英日同時翻訳のためのConnectionist Temporal Classificationを用いたニューラル機械翻訳, 情報処理学会研究報告 2019-NL-241 (2019)

# 遅延と流暢さのトレードオフ

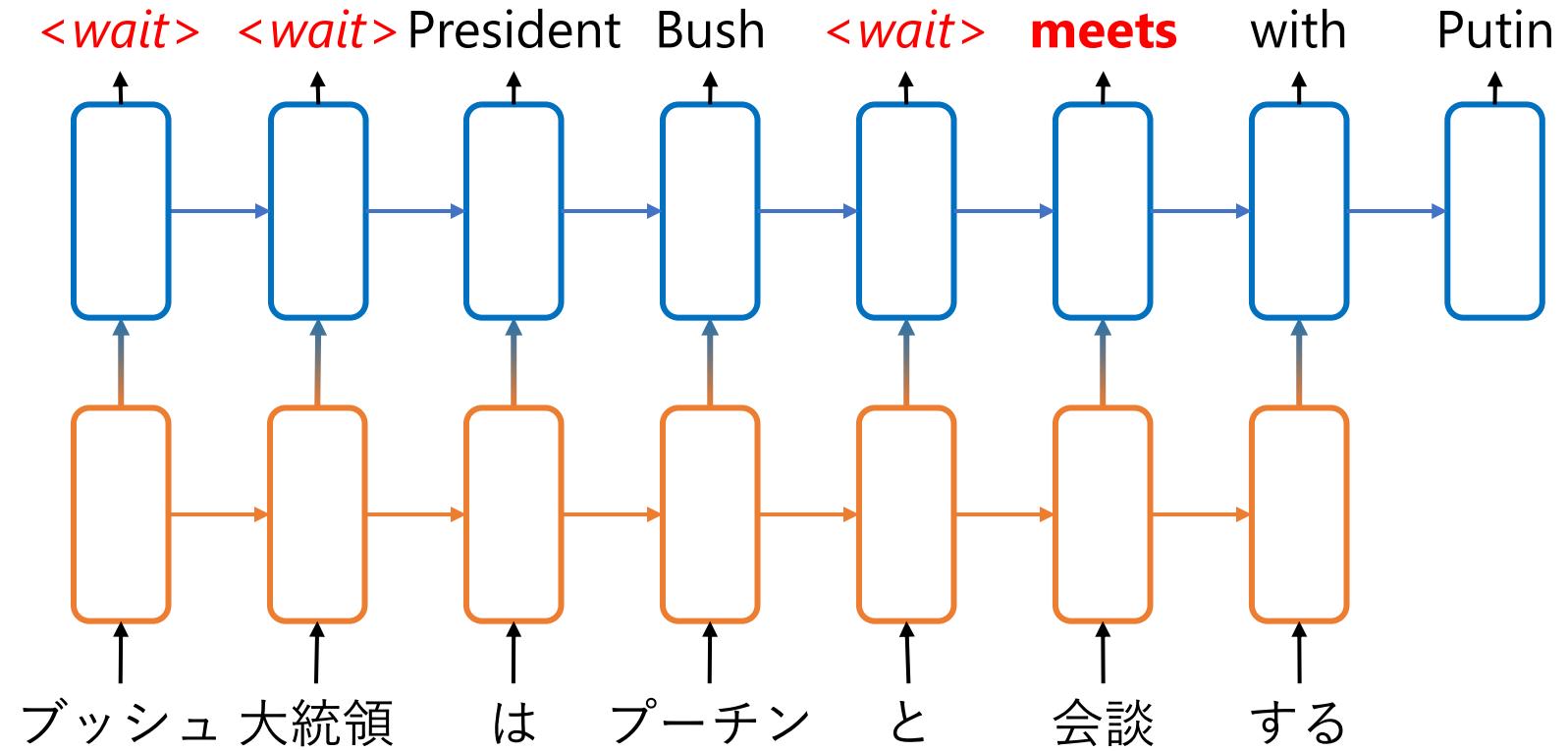
- ・遅延を短くすれば流暢さは一般に低下

I'm going to take you  
on a journey  
in the next 18 minutes.

皆様をお連れします  
旅へ  
今から18分後です  
今から18分で  
皆様を旅へお連れします

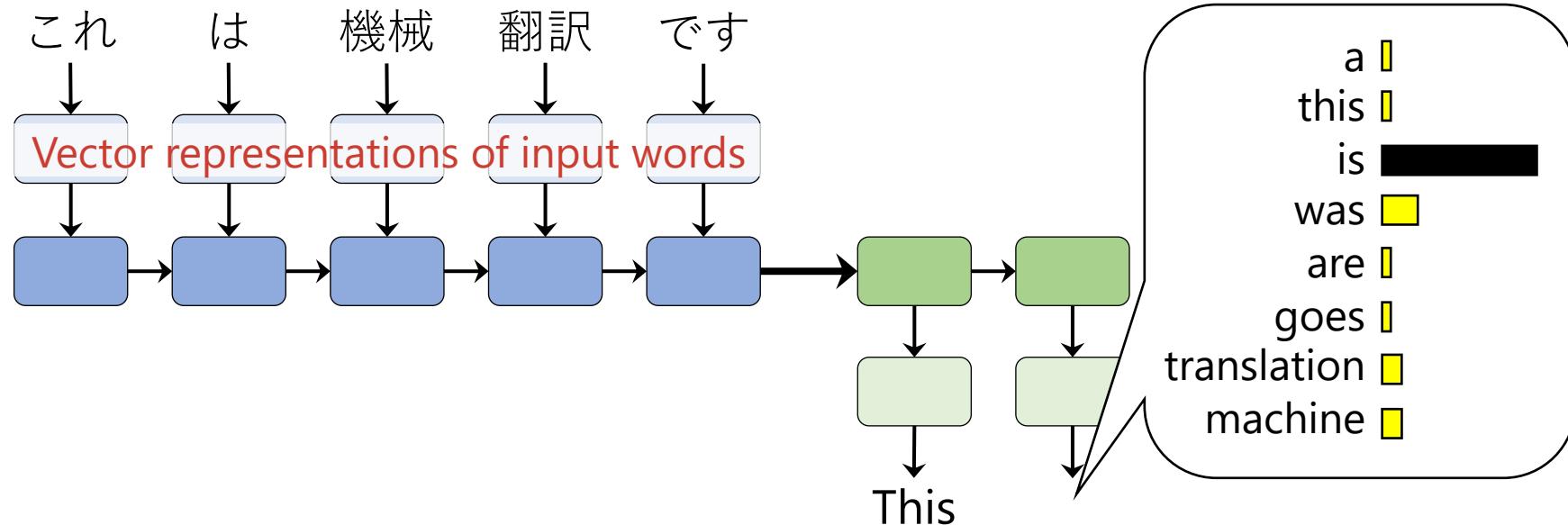
# 提案手法: デコーダにおける適応的遅延

- 特殊記号 *<wait>* を定義し遅延を制御



# 目的関数

- 二種類の損失を最小化
  - 遅延
  - 単語予測誤り



# <wait>挿入位置の問題

- <wait>の位置は自由度が高い
  - President Bush meets <wait> with
  - President Bush <wait> meets with
  - President <wait> Bush meets with
  - ...
- *Connectionist Temporal Classification (CTC)*に基づく最適化（動的計画法）  
[A. Graves+ ICML 2006]

# 漸進的テキスト音声合成

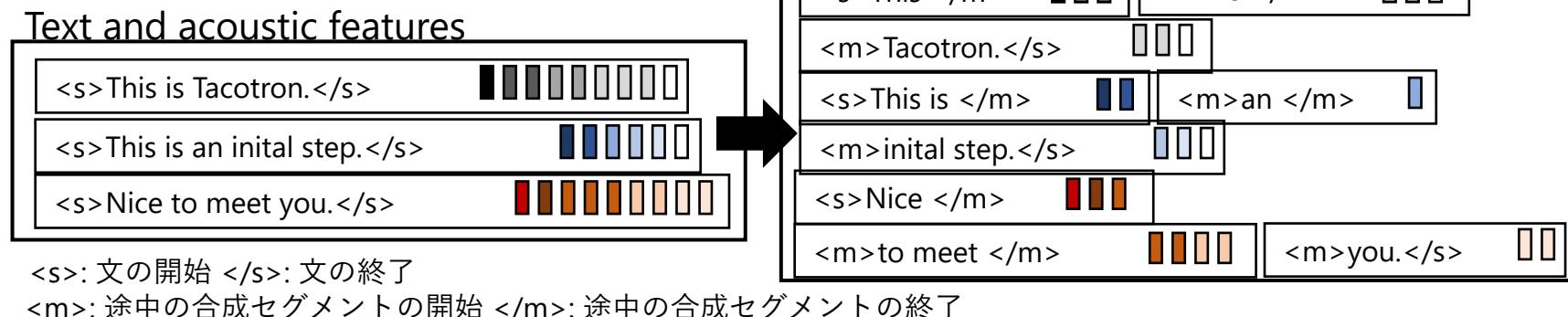
Yanagita, Sakti, and Nakamura: Neural iTTS: Toward Synthesizing Speech in Real-time with End-to-End Neural Text-to-Speech Framework, Proc. of the 10th ISCA Speech Synthesis Workshop, pp. 183-188 (2019)

# テキスト音声合成と漸進的処理

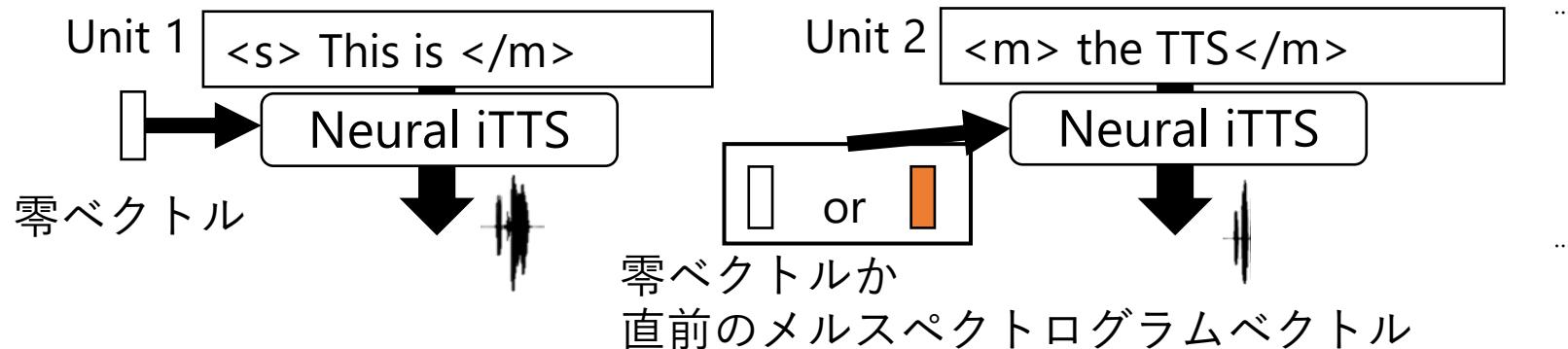
- ・アクセント/イントネーション決定に文脈が必要
  - ・英語: 2単語程度? 日本語: 2アクセント句程度?  
(提案手法における実験結果より)
- ・漸進的合成でも音声を滑らかに繋ぐ
  - ・分割して出力しても繋ぎ目の連續性が欲しい

# 提案手法: 合成過程の分割と直前結果参照

- 学習データも分割



- 合成時は直前の結果を初期値として再利用



# 音声から音声への 同時通訳システム

# デモ: 同時音声翻訳

# おわりに

# おわりに

- 深層学習による機械翻訳の進展
  - 急速に実用化、実装技術も含め日進月歩
  - 研究レベルでは対訳文によらない「教師なし機械翻訳」も
- 自動同時通訳への挑戦
  - より挑戦的な課題へ
- 表層的でない「深い」翻訳が将来的な課題