

自己紹介

- 修士: 東大情報理工
- 博士: 京大情報学
- 現在: 科学技術振興機構(JST) 研究員
– 日中・中日機械翻訳実用化プロジェクト
(2013-2017年度)





自己紹介



- 修士: 東大情報理工
- 博士: 京大情報学
- 現在: 科学技術振興機構(JST) 研究員
– 日中・中日機械翻訳実用化プロジェクト
(2013-2017年度)
- NLP若手の会2017年委員長 <http://yans.anlp.jp>
- NLP2018大会プログラム委員
- 2017年AMCダイヤモンド会員
– 今年は今の所96,406pp



目次

- NMTの基礎
 - SMTとNMTとの比較
 - Encoder-decoderモデル
 - Attention-based NMT
 - NMTの特徴と短所
- NMTの課題
 - 扱える語彙数が少ない
 - 訳抜けと重複
 - 何を学習しているのかわからない
- NMTのその他の話題
 - 多言語翻訳
 - Low/Zero-resource翻訳
 - モデルの軽量化
 - 特殊タグの埋め込み
 - RNN以外のNMTモデル
- まとめ

ニューラル機械翻訳(NMT)の プロダクト化

2014 Sequence to Sequence Learning with Neural Networks
[Sutskever et al., NIPS2014]

2015/05 Baidu 中⇔英

2016/08 SYSTRAN 12言語 (日、韓、英、仏など)

2016/10 WIPO 中、日、韓→英 (特許文)

2016/11 Microsoft 10言語

(英、日、中、仏、独、西、葡、露、伊、刺)

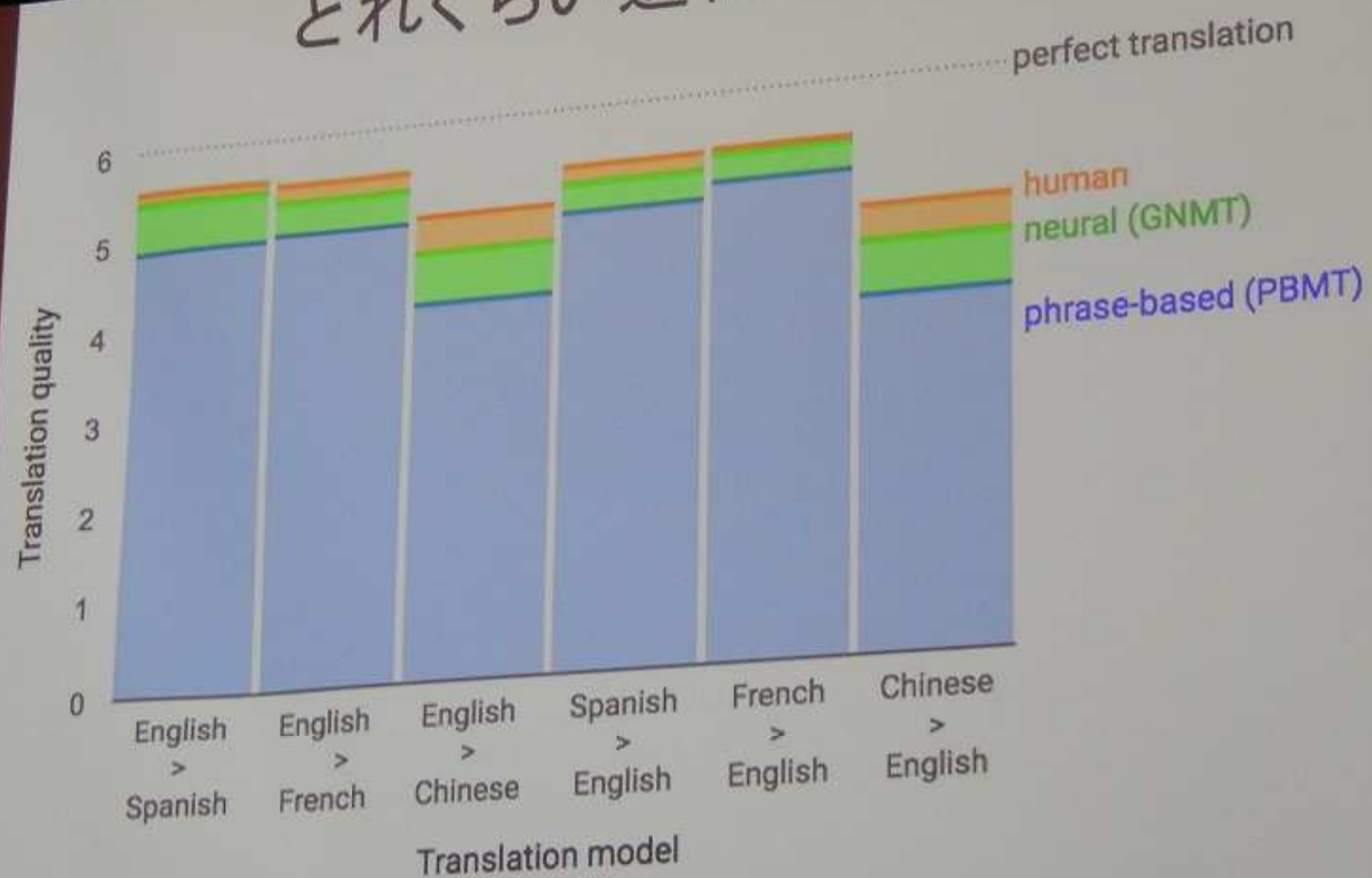
2016/11 Google 8言語

(英、日、中、韓、仏、独、西、葡、土)

→ 現在は英語と全ての言語との間でNMT

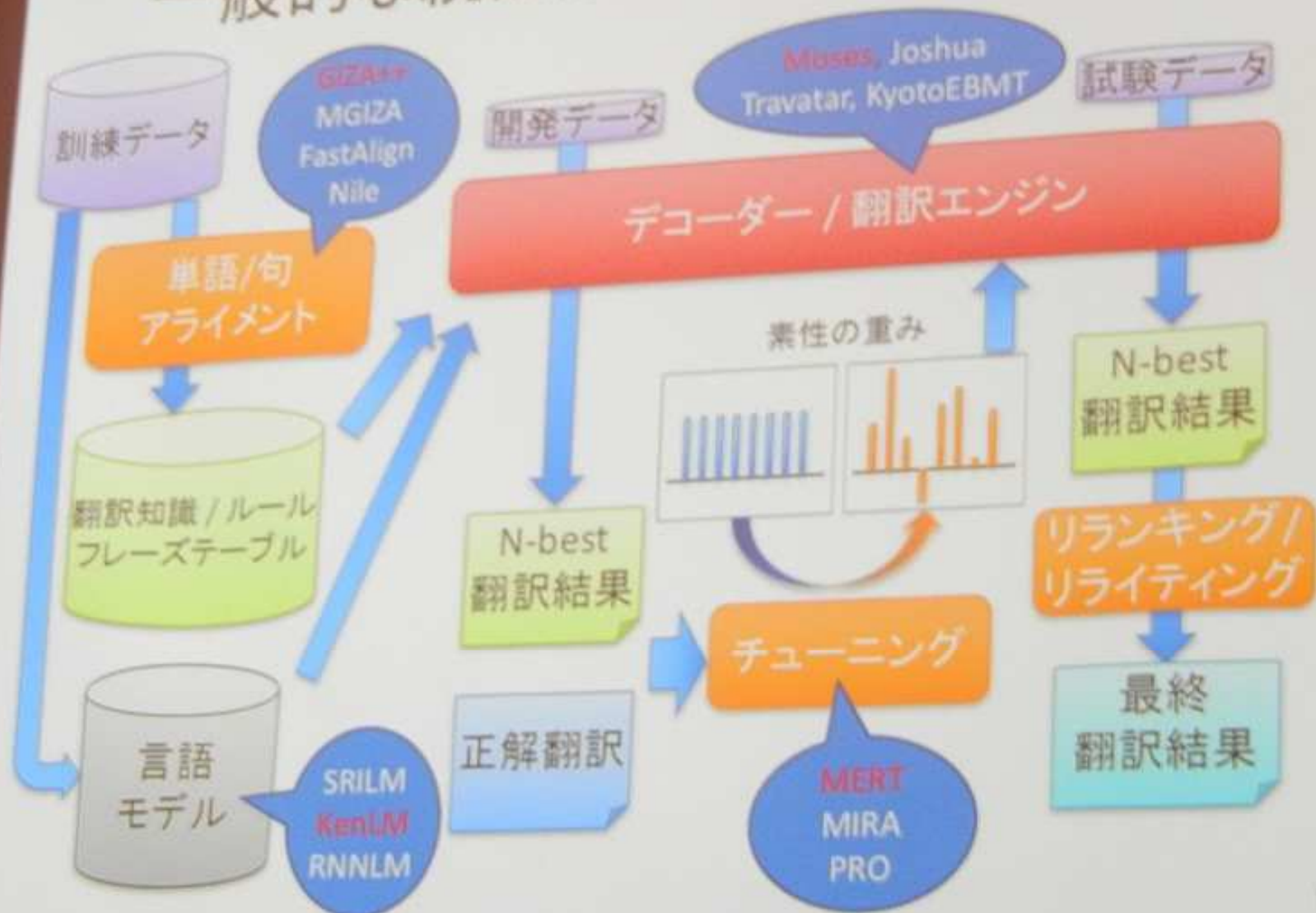
英語以外の言語間はおそらく英語を介してのブリッジング

どれくらい進化したの？

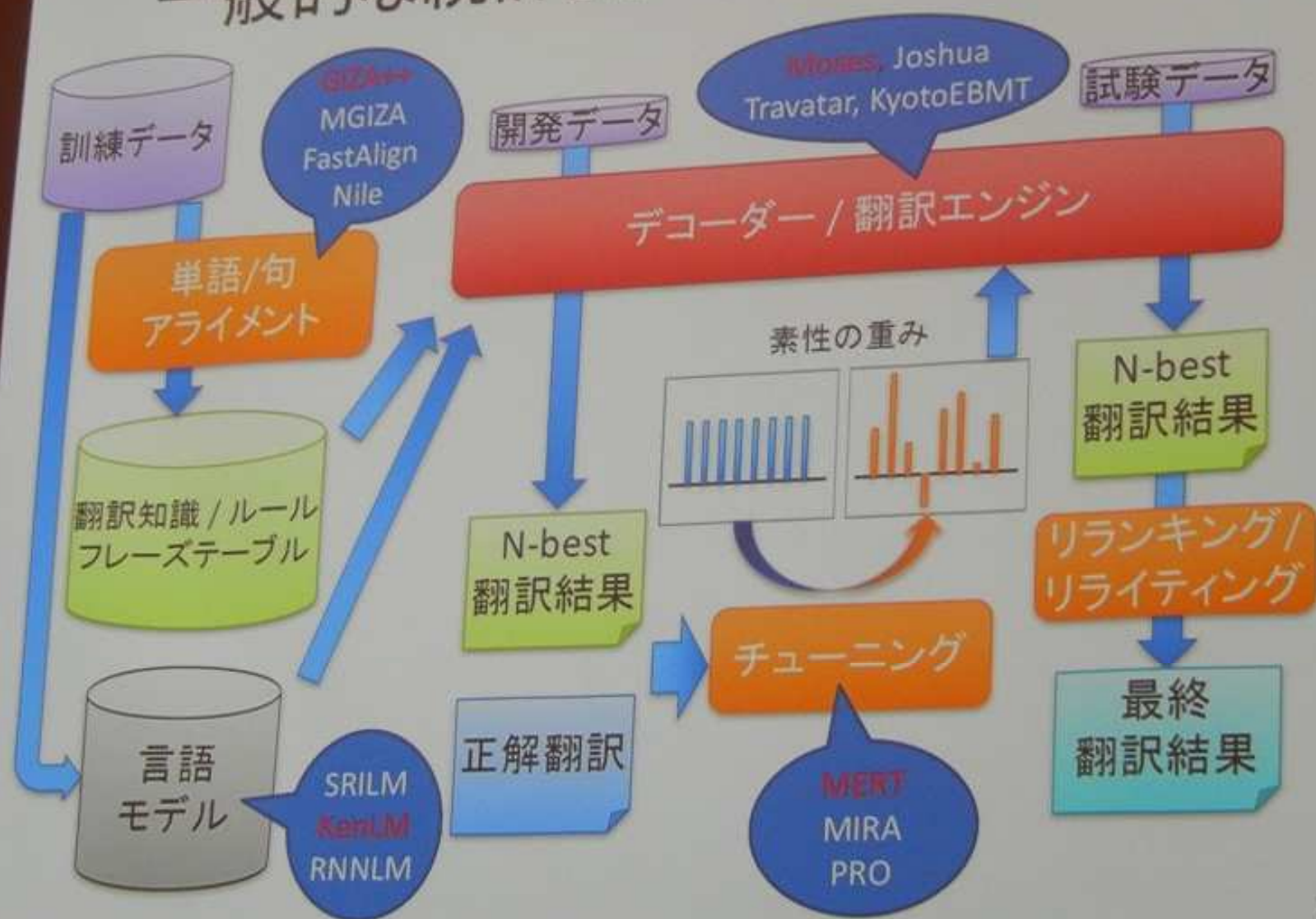


<https://research.googleblog.com/2016/09/a-neural-network-for-machine.html>⁶

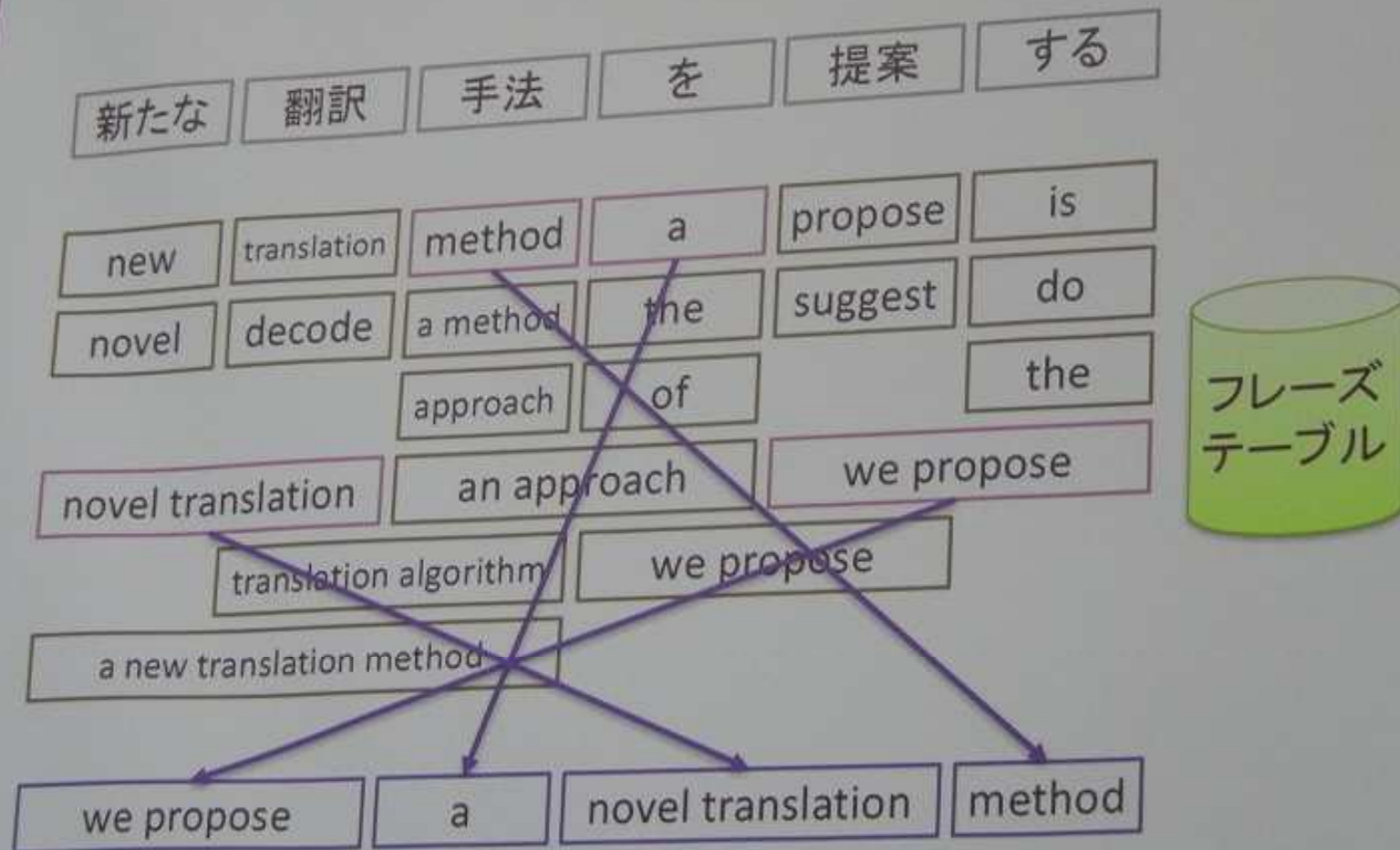
一般的な統計翻訳(SMT)の枠組み



一般的な統計翻訳(SMT)の枠組み

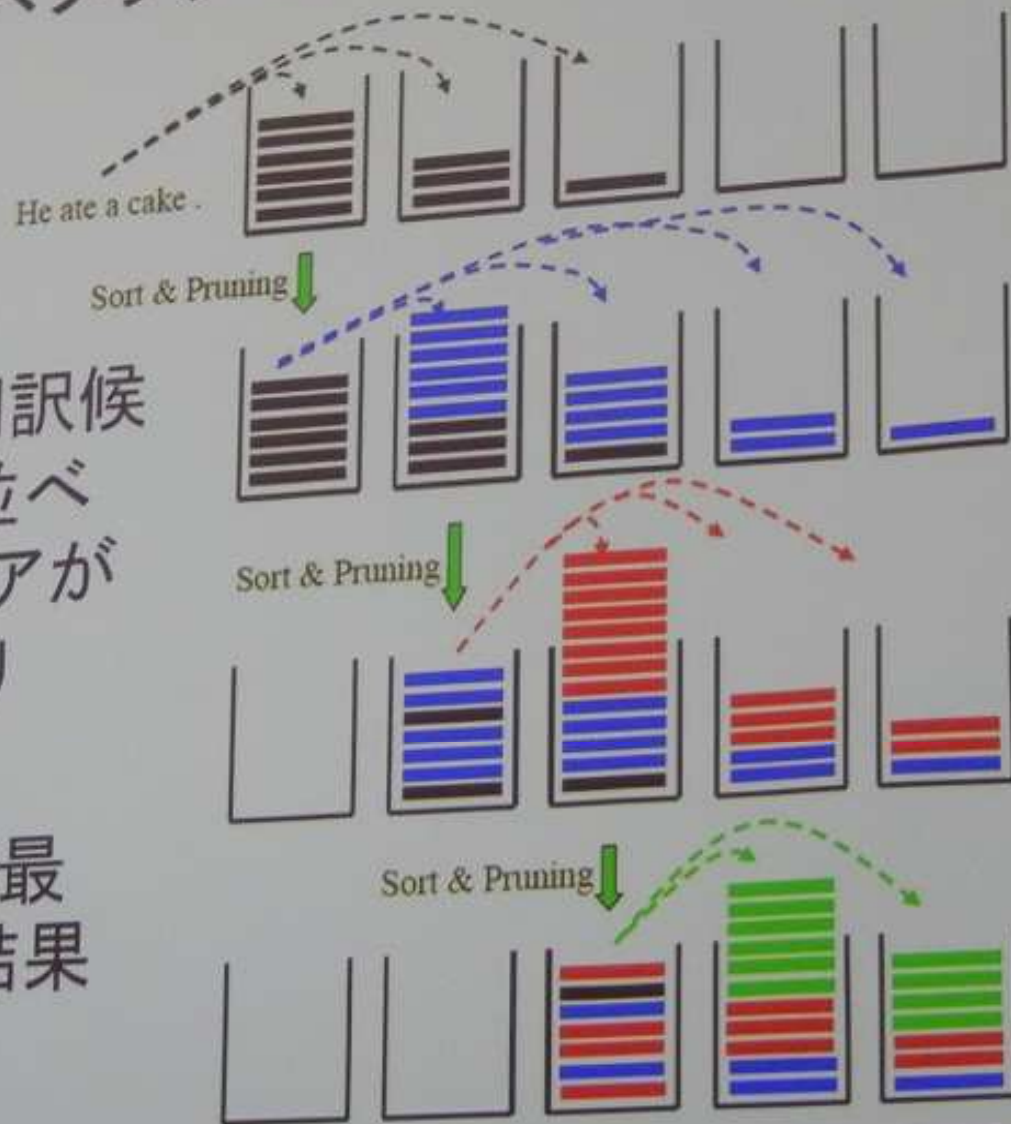


フレーズベースSMTのデコーディング



マルチスタックビームサーチ

- 翻訳された入力
単語数ごとに
スタックを用意
- 各スタック内で翻訳候
補をスコア順に並べ
替え (sort)、スコアが
低いものを枝刈り
(pruning)
- 最後のスタックの最
高スコアが翻訳結果



一般的な統計翻訳(SMT)の枠組み

- トレーニング、チューニング、テストの大きく3ステップに分かれる
- トレーニングに使うモデルと翻訳(デコーディング)に使うモデルは完全独立なことが多い
 - トレーニング = IBM model、デコーディング = log linear
- チューニングは一般にBLEUに対して行われる
- デコーディングは入力文を小さな部分ごとに目的言語に「置き換える」ことで行い、全ての部分が置き換えられたら翻訳終了

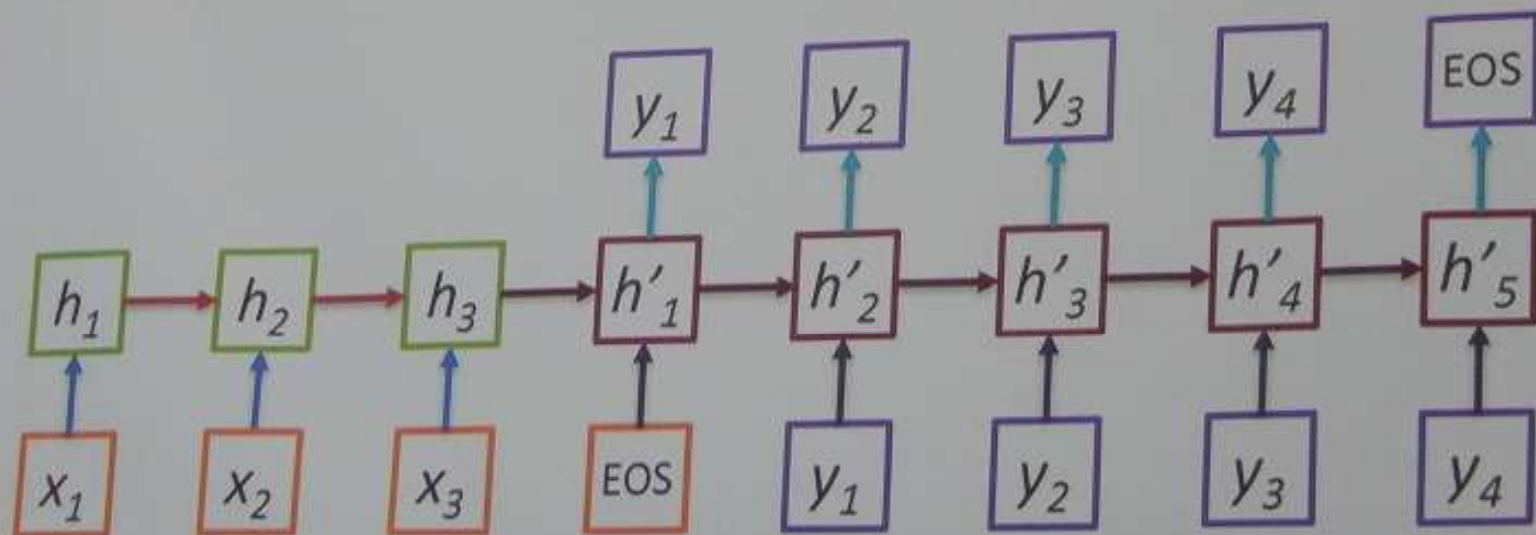
一般的なニューラル翻訳(NMT)の枠組み



Sequence to Sequence Learning

[Sutskever et al., 2014]

- 通常のRNNは入力と出力の長さが同じでなければならないし、順序も一致していなければならない
- 機械翻訳では当然入力と出力の長さが違うし、語順も異なる
- 入力側と出力側で別々のRNNを使おう！



Sequence to Sequence Learningの欠点

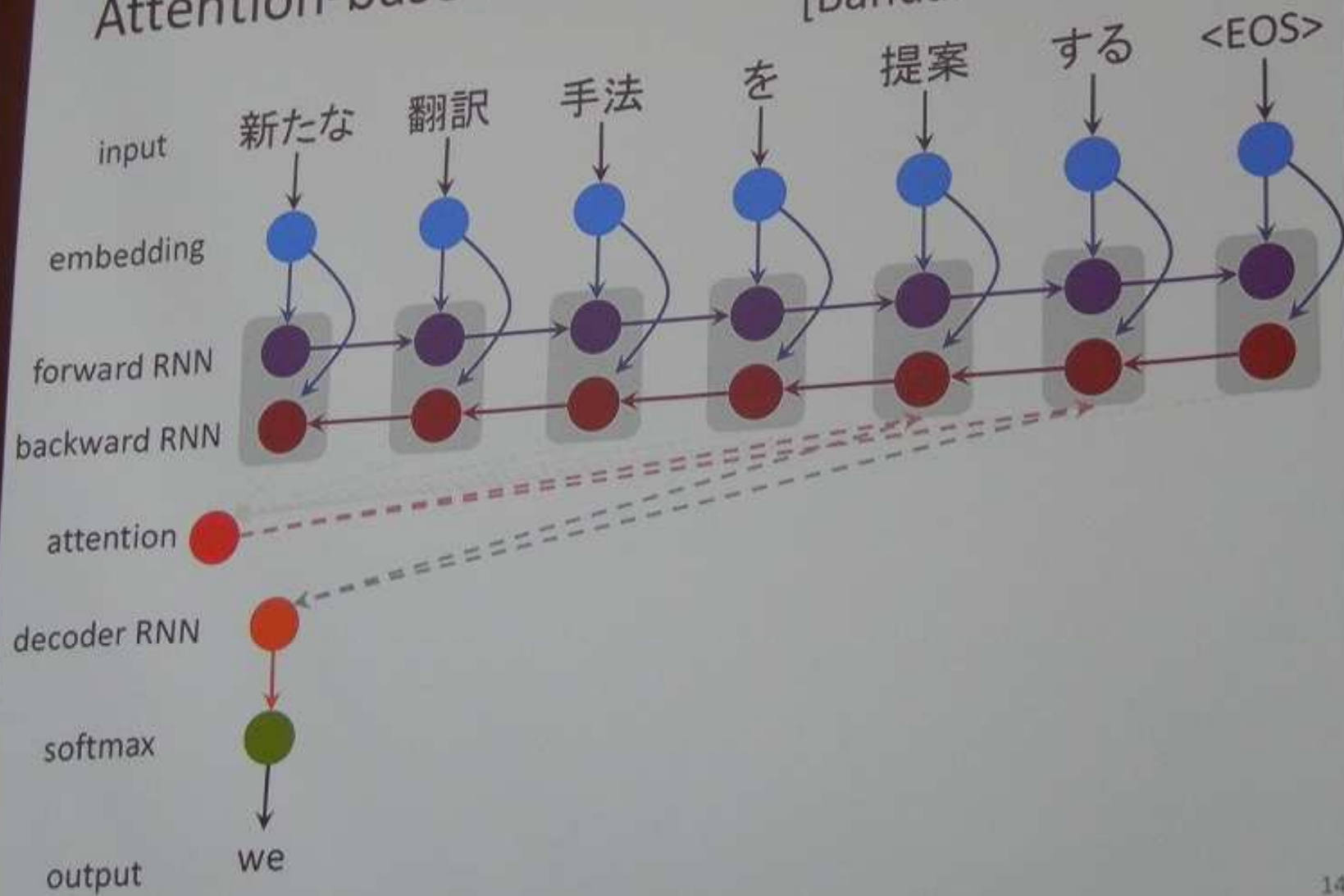
- 入力文の情報を1つの固定長のベクトルに押し込めている
 - 長い文だと翻訳精度が極端に落ちる
 - 入力文を逆順に入れると翻訳精度が上がるという黒魔術が発見される

隠れ層の状態を全て保存しておいて
翻訳時に適宜参照(=attention)しながら
翻訳しよう！

ついでにRNNを両方向走らせよう！
(=bidirectional RNN)

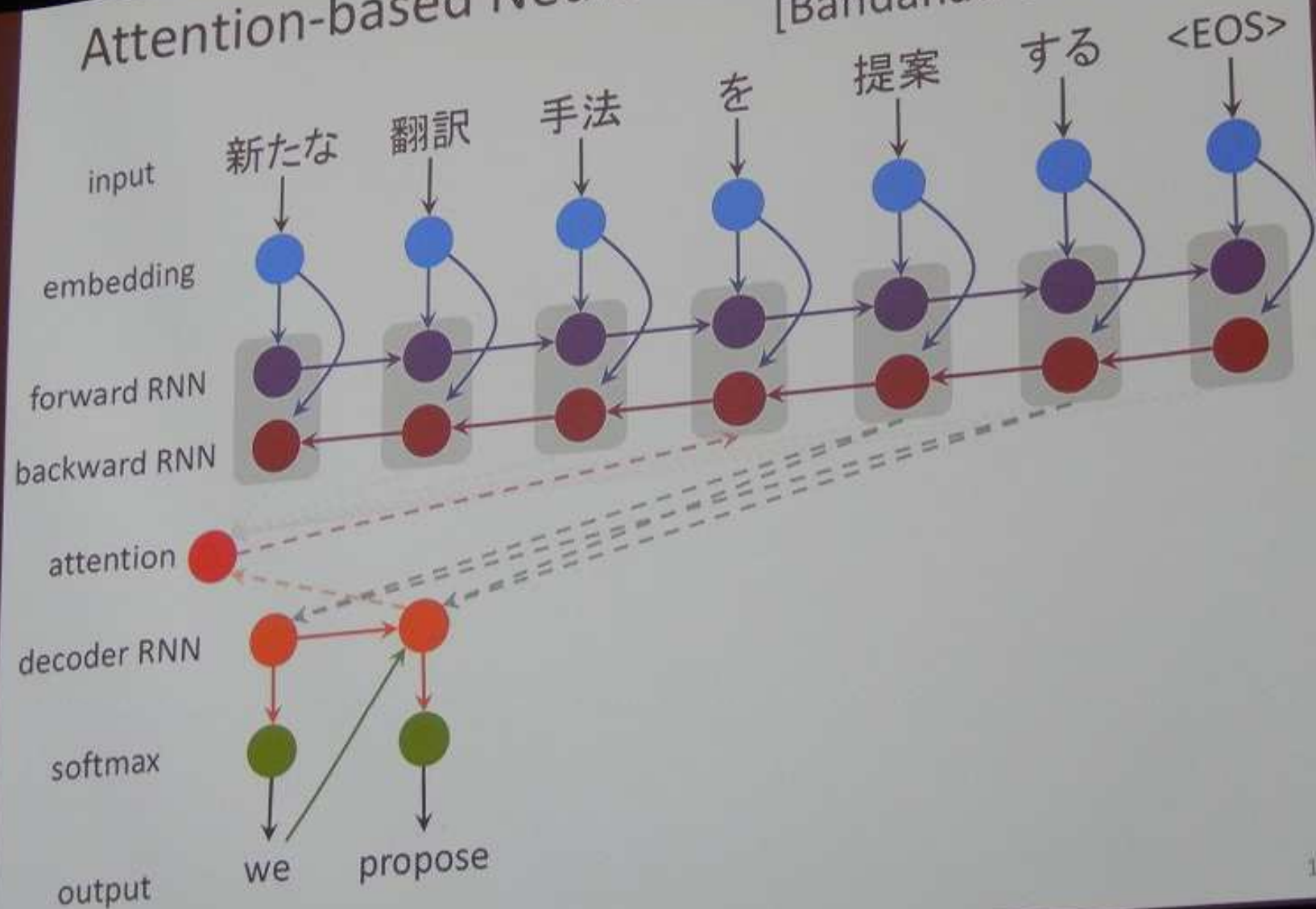
Attention-based Neural Machine Translation

[Bahdanau et al., 2014]



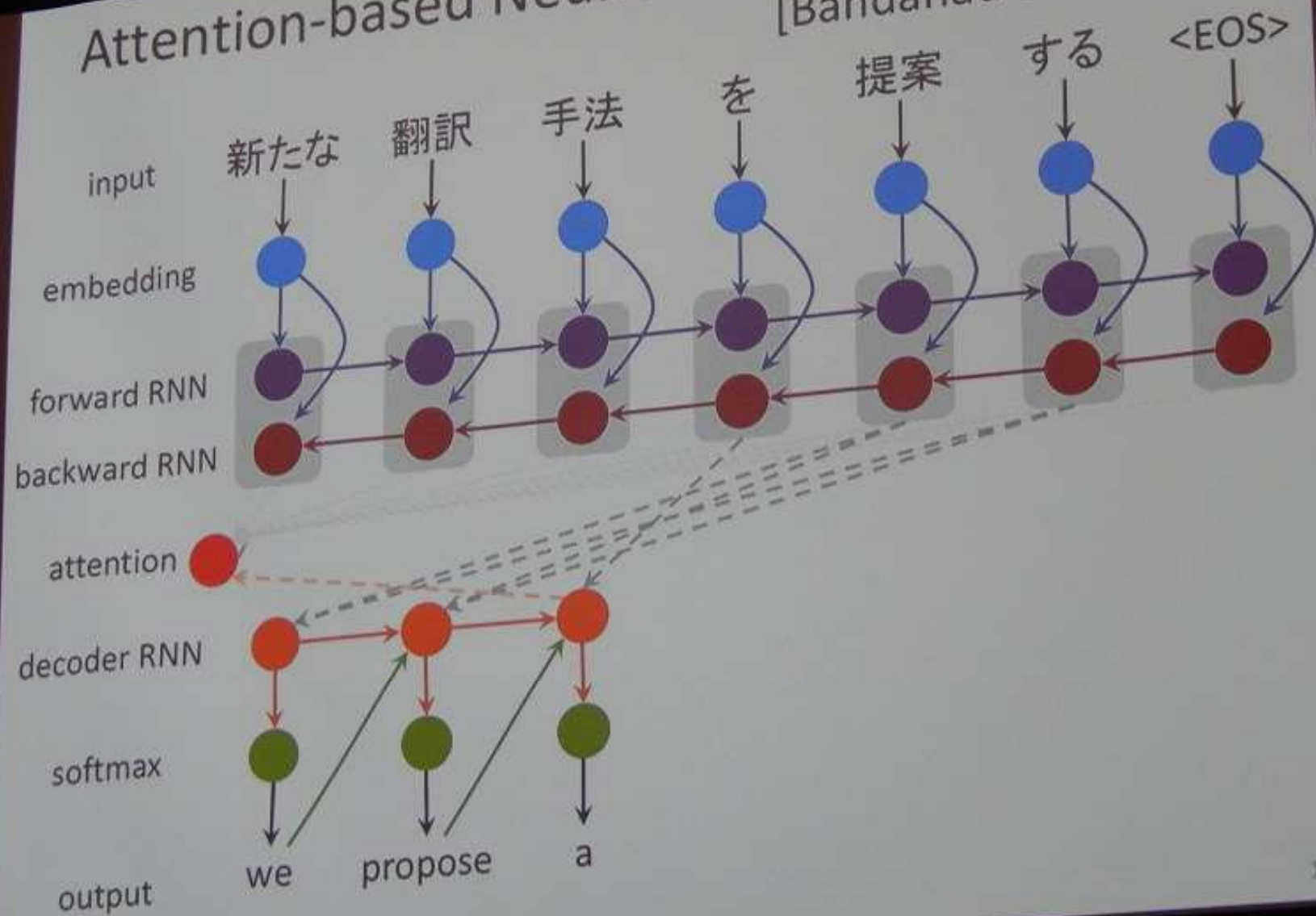
Attention-based Neural Machine Translation

[Bahdanau et al., 2014]



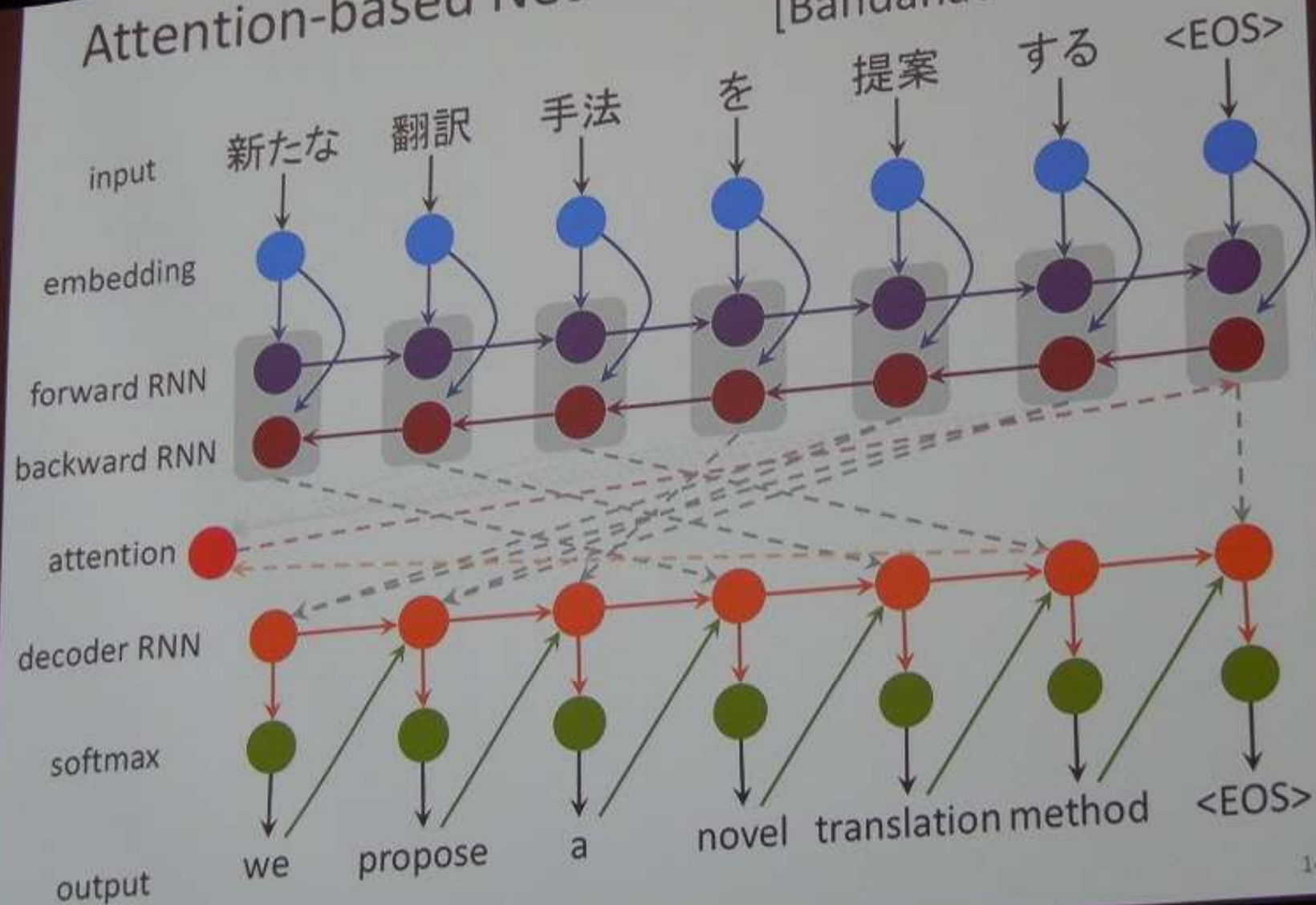
Attention-based Neural Machine Translation

[Bahdanau et al., 2014]



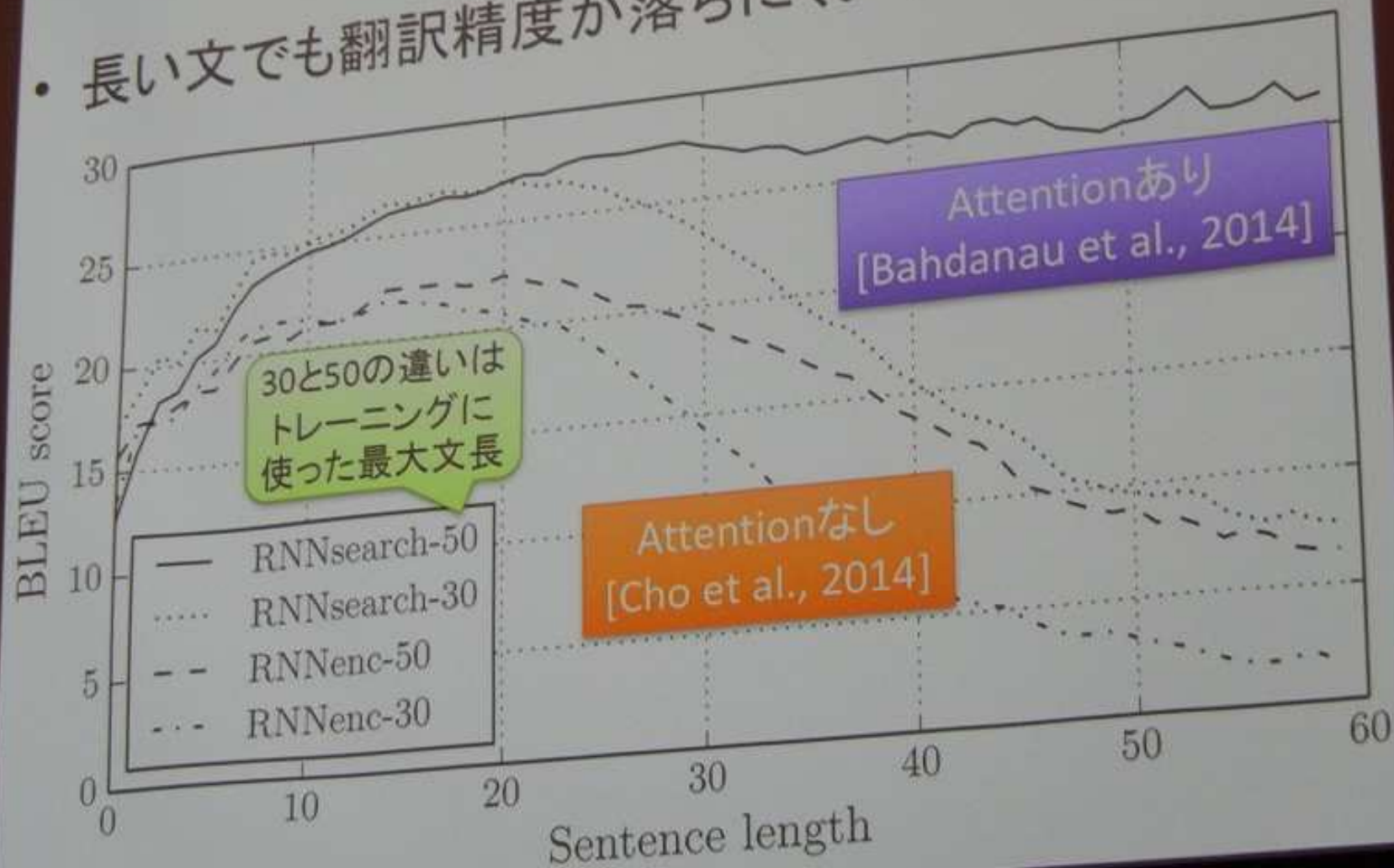
Attention-based Neural Machine Translation

[Bahdanau et al., 2014]



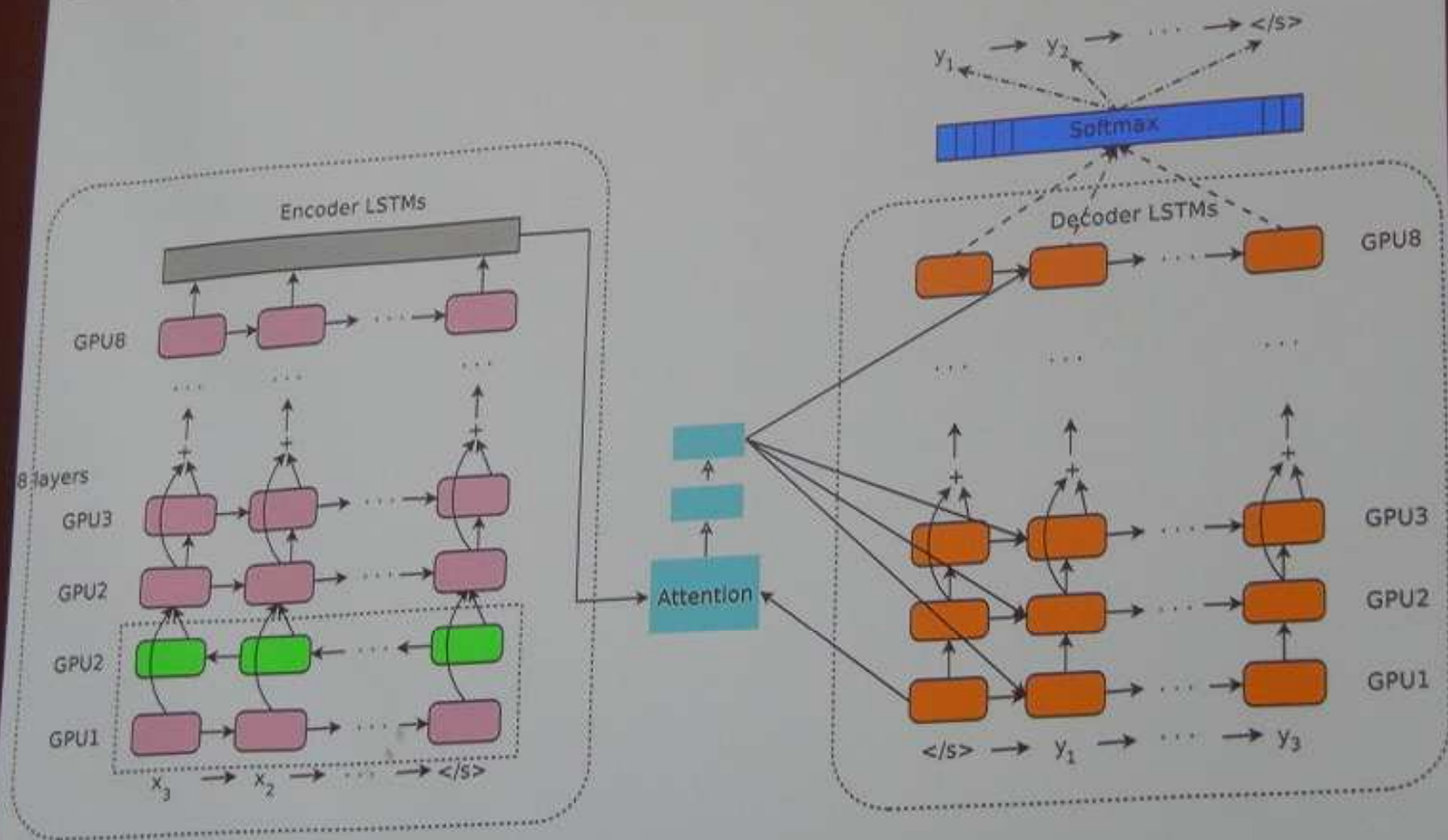
Attentionの導入による効果

- 長い文でも翻訳精度が落ちにくい！



Google's Neural Machine Translation (GNMT)

[Wu et al., 2016]



ビームサーチ

- SMTと同様、翻訳の途中状態を複数保持しておく
 - 途中状態のスコアを定義し、スコア上位を残す
 - スコアの計算例
 - 単語の生成確率の積を出力単語数で正規化
 - 過去のattentionの状態をスコアに反映 [Wu et al., 2016]
- SMTとの違い = いつ終わるかわからない
 - <EOS>が出るまでサーチが続く
- 並列計算がしにくく、GPUとの相性が良くない
- softmaxが必要になるため、計算が重たい
- ビームサーチの有無で精度が大きく異なる

アンサンブル

- 複数のNMTを平行に実行し、単語を一つ出力する際に各NMTからの予測を統合して決定
 - NMTは同じモデルでも、全く異なるモデルでもよい
 - 1つのモデルの訓練途中の複数のパラメータを使ってもよい (loss最小、BLEU最大、最終状態など)が、独立に訓練した複数のNMTを使う方が若干精度が良い

入力: what is the next word? 出力: 次の単語 ...

(は, が, を, に, ...)

NMT1 = (0.3, 0.1, 0.2, 0.1, ...)

NMT2 = (0.2, 0.1, 0.6, 0.1, ...)

NMT3 = (0.4, 0.2, 0.1, 0.2, ...)

NMT4 = (0.2, 0.2, 0.3, 0.1, ...)

相加平均 $\frac{1}{M} \sum_{m=1}^M p_m$

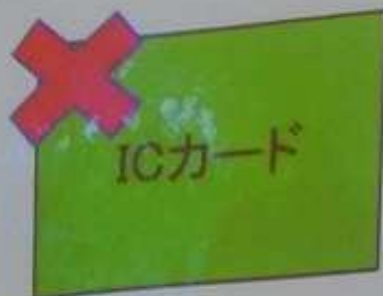
(0.28, 0.15, **0.30**, 0.13, ...)

相乗平均 $\left(\prod_{m=1}^M p_m \right)^{1/M}$

(**0.26**, 0.14, 0.24, 0.12, ...)

NMTの特徴

- フレーズテーブルは存在しない
 - 単語アライメント不要、同じモデルで訓練もテストも行う
- SMTのように入力文を「置き換える」ことで翻訳するのではなく、入力文も見ながら、言語モデルのように翻訳文を作り出す
 - <EOS>が(どこかで)出力されたら完了
 - **入力文を過不足なくカバーして翻訳することができない**
- 入口(入力文)と出口(翻訳文)以外は全て数値計算(行列の積など)だけで動いている(end-to-end)
- embeddingのおかげで柔軟な翻訳が可能



大変恐れ入りますが、富良野駅ではICカードが
ご利用になれません。

Thank you very much can not use an IC card
by furano station very much.

JR北海道 富良野駅

Please

You can not a IC card (Kitaca
Suica PASMO ICOCA) by furano
station.

Also IC card charge also
unavailable.

Very I'm sorry but tickets
counter in cash and credit card
is available in a ticket to please.

Or cash only available ticket
purchase please.

今RTしたやつが、最近のGoogle翻訳でどうなるか見てみたら、Googleの文間を読み取る優しさ能力がスゴイ。

逆に、Google翻訳で正しそうな日本語になったからと言って、元の英文が正しいと信じてはいけないということだと再認識した。

英語 ▾



日本語 ▾



You can not a IC card by Furano station.

Also IC card chage also unavailable.

Very I'm sorry but tickets counter in cash and credit card is available in a ticket to please.

Or cash only available ticket purchase please.

富良野駅ではICカードが使用できません。
また、ICカードのチャージも利用できません。

非常に申し訳ありませんが、チケットは現金でカウンターで、クレジットカードはチケットでご利用いただけます。

またはご利用可能なチケットのみご購入ください。

<https://twitter.com/yasutoc/status/834651903984152576> 21

NMTは本質的には言語モデルと同じ

- エンコードされた入力 x と過去の出力 $y_1 \dots y_{t-1}$ から次の出力単語 y_t を予測

$$p(y_t | y_1, \dots, y_{t-1}, x)$$

- RNNなので $t = 0$ から K (<EOS>出力)までかけると

$$\prod_{t=1}^K p(y_t | y_1, \dots, y_{t-1}, x) = p(y | x)$$

SMTで求めたい
ものと同じ！

	言語モデル	NMT
入力(x)	考慮しない	考慮する
過去の出力	数単語前まで考慮	全て考慮
単語の表現	離散的 (単語そのものを利用)	連続的 (実数値ベクトル)

NMTの短所

- BLEUに対して最適化されていない
 - 通常のトレーニングは1単語出力するごとに、正解の単語と比較(cross-entropy)
 - BLEUなど任意の評価指標を損失関数として使う方法(minimum risk training)もある [Shen et al., 2016]
 - BLEUは向上するが人手評価は変わらない？ [Wu et al., 2016]
- embeddingのせいで全然違う訳が出ることがある
 - I come from Tunisia → ノルウェーの出身です [Arthur et al., 2016]
- attentionの結果はそこまで信用できない