

ニューラル機械翻訳に対する注意言語モデル

黒澤 道希^{1,a)} 小町 守^{1,b)}

概要: 機械翻訳はニューラルネットワークの活用により、流暢性の高い出力を得られることが報告された。近年ではさらなる流暢性の向上が行われており、その一つに言語モデルを用いた研究がある。

言語モデルを用いる先行研究においては、翻訳機構と言語モデル機構の2つを用意し双方の情報をを用いており、予測を同尺度もしくは動的に重み付けして出力単語を予測するものや言語モデルの予測に翻訳機構の情報を与えるものがある。しかしながら、機械翻訳においては流暢性の向上だけではなく、妥当性を担保することも求められる。つまり翻訳機構の情報を活用し言語モデル機構の情報を補助的に用いるべきであるが、先行研究においては双方の情報を両方用いて出力単語を予測している。そこで本研究では翻訳機構(TM)と言語モデル機構(LM)の2つを用意するが、翻訳機構の予測の補助的要素として言語モデル機構の予測に対してアテンションを取ることで、妥当性を保持した上で出力を流暢にするモデルを提案する。英日翻訳において言語モデル機構を用いた翻訳の先行研究と比較した結果、BLEUスコアとRIBESスコアが向上することを示した。この結果より、英日ニューラル機械翻訳に対しても言語モデル機構を追加することが有用であり、その中でもアテンションが効果的であることを示した。また、言語モデル機構の予測に対するアテンションと実際の出力を分析することにより、言語モデルが文法的性質を用いて予測を補助するために有用な情報である可能性が高いことを合わせて報告する。

1. はじめに

機械翻訳においてニューラル機械翻訳が導入されて以来、統計的機械翻訳と比較して流暢な出力が得られるようになってきている。[15]しかしながらニューラル機械翻訳を流暢にする上では大規模な対訳コーパスが必要であるが、大規模対訳コーパスは容易には準備できない。そこで、大規模対訳コーパスを利用せずに流暢にする研究が行われている。

それらの研究の1つとして単言語コーパスを用いる研究が行われている。単言語コーパスは比較的容易に収集することができ、目的言語において流暢な出力が可能になる場合や、低リソースな言語対においても翻訳が可能になるなど、出力の向上に寄与できる可能性は非常に大きい。単言語コーパスは様々な角度から利用方法が検討されており、翻訳機構の事前学習として利用する手法[10]や逆翻訳を行うことにより擬似対訳コーパスを作成する手法[12]、単言語コーパスで単語分散表現の初期化を行う手法[9]などがある。

本研究では言語モデル機構を新たに作成し、その学習に単言語コーパスを用いる方法[2]、[14]に着目する。近年の

ニューラル機械翻訳においては流暢な文を出力しやすくなっているが、翻訳を行う際に原言語文の情報のみを扱っているため、出力言語の言語的情報が反映されづらいことも分かっている。[11]

ニューラル機械翻訳に対して言語モデル機構を用いた研究においては、従来の翻訳機構の他に単言語コーパスで生成した言語モデル機構を用意し、双方を重み付けして翻訳を行っている。Shallow Fusionモデル[2]では、翻訳機構の出力と言語モデル機構の出力を人手で決めた割合で重み付けしている。Cold Fusionモデル[13]ではゲート関数を作成して言語モデルの重みを動的に決め、言語モデル機構の予測時点で翻訳機構の情報を考慮させている。また、Simple Fusionモデル[14]では双方を同尺度として扱っており重要度に大小は存在しない。

そこで本研究では出力言語側の流暢性を向上させる目的で、翻訳機構を基本とし言語モデル機構を注視しながら予測する**注意言語モデル**を提案する。先行研究では、翻訳と言語モデルを同等または動的に重み付けを行い翻訳性能を向上させた。しかし、仮に言語モデルの信頼度が高くそれを用いた結果流暢な文が出力された場合でも、原言語文に入っていない情報であった場合、翻訳としては不適切である。つまり、言語モデル機構は翻訳機構の予測を向上させるために補助的に用いられるべきだと考える。このモデル

¹ 首都大学東京大学院 システムデザイン研究科

^{a)} kurosawa-michiki@ed.tmu.ac.jp

^{b)} komachi@tmu.ac.jp

では、翻訳機構の情報を基にした attention と言語モデル機構の単語予測確率を合わせており、言語モデル機構と翻訳機構の双方が独立に予測しながらも attention によって言語モデル機構の予測に翻訳機構の情報を考慮させることが可能である。これによって翻訳機構の情報を適切に活用しながら流暢性の高い翻訳を実現できると考える。

これらのモデルについて実験した結果、注意言語モデルにおいて翻訳の流暢さが向上した、また、attention の重みをもとに、言語モデル機構の予測の傾向と翻訳への影響について分析を行った。

本研究の主な貢献は以下の通りである。

- ニューラル機械翻訳に対して言語モデル機構を導入する新たな注意言語モデルを提案した。
- 英日翻訳において言語モデル機構を活用することにより流暢かつ妥当な出力が可能になることを示した。
- 言語モデルによって翻訳が改善される理由について分析した。

2. 先行研究

2.1 Shallow Fusion

Gulcehre ら [2] は、翻訳モデルと言語モデルの予測を元に翻訳を行う *Shallow Fusion* (1) を提案した。このモデルでは単言語コーパスを活用し言語モデルを事前学習し、翻訳モデルに目的言語の知識を導入することで翻訳を改善する。

Shallow Fusion では、予測単語 \hat{y} は、

$$\hat{y} = \operatorname{argmax}_y \log P_{\text{TM}}(\mathbf{y}|\mathbf{x}) + \lambda \log P_{\text{LM}}(\mathbf{y}) \quad (1)$$

によって決定される。ここで、 \mathbf{x} は原言語の入力である。また、 $P_{\text{TM}}(\mathbf{y}|\mathbf{x})$ は翻訳モデルの単語予測確率であり、 $P_{\text{LM}}(\mathbf{y})$ は言語モデルの単語予測確率である。

このときの λ は言語モデルを考慮する割合を決めるハイパーパラメータとして定義されており、手動で決めている。

2.2 Cold Fusion

Gulcehre ら [2] は、翻訳機構と言語モデル機構を同時に学習する *Deep Fusion* を同時に提案した。それを改善したモデルとして Sriram ら [13] は、言語モデル機構の予測のために翻訳機構の情報を渡す *Cold Fusion* を提案した。

このモデルでは、事前学習された言語モデル機構に対し、翻訳機構と言語モデル機構の両方を考慮して重みを動的に決めるゲート関数を導入しており、言語モデル機構は翻訳機構の情報も利用しながら予測される。翻訳機構と言語モデル機構を合わせて1つの機構として学習することで精度と流暢性を改善した。Cold Fusion では予測単語 \hat{y} は、

$$h_{\text{LM}} = W_{\text{LM}} S_{\text{LM}}(\mathbf{y}) \quad (2)$$

$$g = W_{\text{gate}}[S_{\text{TM}}(\mathbf{y}|\mathbf{x}); h_{\text{LM}}] \quad (3)$$

$$h' = [S_{\text{TM}}(\mathbf{y}|\mathbf{x}); g \cdot h_{\text{LM}}] \quad (4)$$

$$S_{\text{cold}} = W_{\text{output}} h' \quad (5)$$

$$\hat{y} = \operatorname{argmax}_y \operatorname{softmax}(S_{\text{cold}}) \quad (6)$$

によって決定される。ここで、 $S_{\text{TM}}(\mathbf{y}|\mathbf{x})$ と $S_{\text{LM}}(\mathbf{y})$ はそれぞれ翻訳機構と言語モデル機構の予測値（正規化前）である。また、 g とは言語モデル機構の予測を考慮する割合を決める関数であり、 h_{LM} にかけることで用いる。 W_{LM} は隠れ層の次元数から単語次元数に変換する重みで、 W_{gate} は隠れ層の次元数の2倍から隠れ層の次元数に変換する重み、 W_{output} は隠れ層の次元数の2倍から単語次元数に変換する重みである。なお、翻訳機構の予測である S_{TM} とはベクトルを結合させる。

2.3 Simple Fusion

Stahlberg ら [14] は、Cold Fusion をシンプルにした *Simple Fusion* を提案した。このモデルでは Cold Fusion とは異なり言語モデル機構の予測に翻訳機構の情報は用いられない。

彼らは2つの類似した手法 POSTNORM (7) と PRENORM (8) を提案した。

POSTNORM と PRENORM における予測単語 \hat{y} は、

$$\hat{y} = \operatorname{argmax}_y \operatorname{softmax}(\operatorname{softmax}(S_{\text{TM}}(\mathbf{y}|\mathbf{x})) \cdot P_{\text{LM}}(\mathbf{y})) \quad (7)$$

$$\hat{y} = \operatorname{argmax}_y \operatorname{softmax}(S_{\text{TM}}(\mathbf{y}|\mathbf{x}) + \log P_{\text{LM}}(\mathbf{y})) \quad (8)$$

によって決定される。ここで、 $S_{\text{TM}}(\mathbf{y}|\mathbf{x})$ は翻訳モデルの予測値（正規化前）であり、 $P_{\text{LM}}(\mathbf{y})$ は言語モデルの予測確率である。

POSTNORM では、言語モデル機構の出力する確率と翻訳機構の出力する確率を掛け合わせており、言語モデル機構と翻訳機構をそれぞれ同尺度の単語確率として扱うことにより出力を改善した。

PRENORM では、言語モデルの対数確率と翻訳モデルの予測値（正規化前）を足し合わせており、言語モデル機構と翻訳機構をそれぞれ別尺度の予測値として扱うことにより出力を改善した。

これらの Simple Fusion モデルは言語モデルを用いた他の手法と比較して単純であり、かつ BLEU スコアが向上した。

3. 注意言語モデル

我々は新たに注意言語モデルを提案する。Shallow Fusion と Simple Fusion では、言語モデル機構の情報を等価もしくは固定した重みを決定させて考慮していた。しかしながら翻訳では入力言語の情報が保持されていることが前提であるため、言語モデル機構の予測をそのまま利用する

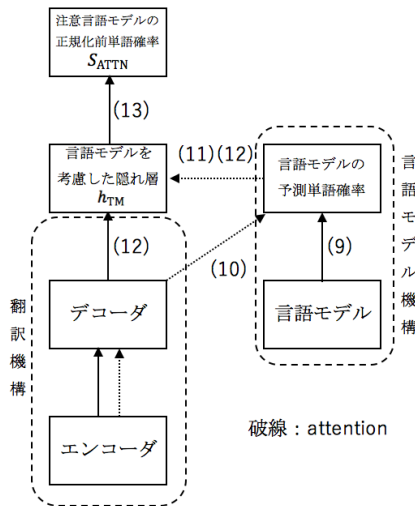


図 1 注意言語モデルの概略図

のではなく、翻訳機構の情報を考慮して言語モデル機構を活用することが好ましいと考える。また、Cold Fusion では翻訳機構の情報を言語モデル機構に対して予測前に渡すため、言語モデル機構独自の予測はされていない。

そこで我々が提案したモデルでは、翻訳機構とは独立に行われる言語モデル機構の予測を、翻訳機構との関連性 (attention) を活用したうえで利用することで言語モデル機構を予測の補助的情報として用いる。今回のモデルではデコードと同時に言語モデルの予測も行われる。すなわち、言語モデル機構の出力は現在予測している語までの情報しか得ることができない。そこで我々はアテンションの対象として言語モデルが出力した単語予測確率を用いた単語アテンションを用いることとする。

まず、アテンションの対象となる言語モデルの単語予測確率 $P_{LM}(y)$ は、

$$P_{LM}(\mathbf{y}; y = \text{word}) = \text{softmax}(S_{LM}(\mathbf{y})) \quad (9)$$

と表される。そして、言語モデルに対するアテンションを考慮した翻訳モデルの隠れ層 h_{TM} は、

$$\alpha_{\text{word}} = \frac{\exp(e_{\text{word}}^T S_{TM}(\mathbf{y}|\mathbf{x}))}{\sum_{\text{word} \in V} \exp(e_{\text{word}}^T S_{TM}(\mathbf{y}|\mathbf{x}))} \quad (10)$$

$$c_{\text{word}} = \alpha_{\text{word}} e_{\text{word}} \quad (11)$$

$$c = \sum_{\text{word}} c_{\text{word}} \cdot P_{LM}(\mathbf{y}; y = \text{word}) \quad (12)$$

$$h_{TM} = [S_{TM}(\mathbf{y}|\mathbf{x}); c] \quad (13)$$

$$S_{ATTN} = W h_{TM} \quad (14)$$

と表される。ここで、 e_{word} は単語のエンベディングを表し、 c_{word} は各単語の一般的な単語アテンションを表している。 c は言語モデルを考慮したアテンションをとった隠れ層を表している。また、 W は隠れ層の次元数の2倍から単語次元数に変換する重みであり、 $P_{LM}(\mathbf{y}; y = \text{word})$ は単語 word に対する言語モデル ($P_{LM}(\mathbf{y})$) の予測確率である。

表 1 実験に用いたコーパスの詳細。

用途	コーパス	最大文長
学習用	827,188	60
開発用	1,790	
評価用	1,812	
言語モデル用 (日本語文のみ)	1,909,981	60

表 2 実験の詳細設定。

項目	設定
事前学習エポック	15 epoch
翻訳機構の学習最大エポック	30 epoch
最適化手法	AdaGrad
初期学習率	0.01
埋め込み層の次元数	512
隠れ層の次元数	512
学習バッチサイズ	128
語彙サイズ	30,000

また、 α_{word} はその単語の重要度を表す重みである。この時、式 (12) において $P_{LM}(\mathbf{y}; y = \text{word})$ をかけることにより言語モデルを考慮した単語アテンションとなっている。

最終的な予測単語 \hat{y} は、

$$\hat{y} = \underset{y}{\text{argmax softmax}}(S_{ATTN}) \quad (15)$$

で求められる。

注意言語モデルの概略図は図 1 に示す。このモデルでは、翻訳モデルの情報と言語モデルの情報を用いて求められるアテンションを考慮することで、言語モデルの情報を翻訳モデルに活用している。

なお、学習の順序は Simple Fusion [14] を踏襲し次の順番で行われる。

- (1) 言語モデル機構の学習を行う。
- (2) 言語モデル機構を固定して翻訳機構および言語モデル機構に対するアテンションの学習を行う。

4. 英日翻訳実験

4.1 実験設定

本研究では、比較手法として従来の NMT と Simple Fusion モデル (POSTNORM, PRENORM) を、提案手法として注意言語モデルを用意し、英日翻訳を行い BLEU [8] と RIBES [3] で評価した。

実験には翻訳用と言語モデル用の 2 つのコーパスを用いる。今回は同一ドメインの言語モデルでの学習を行うため、Asian Scientific Paper Excerpt Corpus (ASPEC) [7] の学習データを対訳コーパスと単言語コーパスの 2 つに分割する。対訳コーパス (翻訳用) には学習用データ約 300 万文のうち、文アライメントの確信度が高い 100 万文を用い、最大トークン数が 60 トークン以下のものに限定する。

表 3 英日翻訳実験結果.

モデル	BLEU スコア	RIBES
ベースライン	29.62	79.68
Simple Fusion (PRENORM)	30.30	80.08
Simple Fusion (POSTNORM)	29.70	79.74
注意言語モデル	30.89	80.90

単言語コーパス（言語モデル用）には対訳コーパスに使用していない文を用い、同様に最大トークン数が 60 トークン以下のものに限定する。また、コーパスの前処理として、日本語側に関しては MeCab (IPADic) を用いて形態素解析を行い、英語側に関しては Moses^{*1} の Tokenizer および Truecaser を用いて処理を行った。開発用と評価用にも ASPEC のデータを用い、それぞれの詳細を表 1 に示す。なお、今回は翻訳に重点を置くため、単語分割、語彙制約は翻訳用コーパスのみで設定しており、例えば単言語コーパスにのみ存在する語は高頻度であったとしても未知語として処理される。

ベースラインとして Bahdanau ら [1] と Luong ら [4] が提案したニューラル機械翻訳モデルを元に独自に実装したもの [6] を使用し、比較手法及び提案手法は前述のベースラインをもとに作成した。なお比較を行うため、モデル以外の設定は全ての実験で統一している。また、本実験では Byte Pair Encoding (BPE) [5] は行っていない。^{*2}

今回の実験における詳細な設定は表 2 に示す。なお、事前学習は言語モデル機構のみを学習しており、ベースラインは言語モデル機構がないため事前学習は行われない。

4.2 結果

実験の結果を表 3 に示す。結果より注意言語モデルはベースラインと比較して BLEU が 1.27, RIBES が 1.32 向上した。また、先行研究である Simple Fusion については、提案されている 2 手法ともにベースラインと比較して BLEU が向上しており、今回の設定においては PRENORM のほうが高かった。なお、注意言語モデルは PRENORM と比較しても BLEU が 0.59, RIBES が 0.82 向上した。

5. 考察

5.1 モデル間の BLEU スコアと RIBES スコアに関する考察

実験結果よりベースラインと比較して言語モデルを用いた各手法の BLEU と RIBES の双方が向上した。このことより言語モデルを出力時に用いることによって妥当性を保ちながら英日翻訳において確認できた。さらに、先行研究と比較して注意言語モデルの BLEU と RIBES が向上していることから、言語モデルはアテンションとして翻訳機構

^{*1} <http://www.statmt.org/moses/>

^{*2} 本手法は BPE を適用することも可能だが、本報告提出時点で実験途中であるため割愛する。

との関連性を用いて補助的に用いることが有用であることが示される。

5.2 出力文に関する定性的な考察

それぞれのモデルの出力文について、例を表 4, 5 に示す。

表 4 では、ベースラインと比較して PRENORM と注意言語モデルの流暢性が向上している。また、注意言語モデルの方が文内での前後関係をより明確に示していることがわかる。一方で、POSTNORM では“線量”を“容量”と翻訳しており、妥当性を欠く結果となってしまっていることが見える。

表 5 の PRENORM を見ると端的かつ流暢な出力のように見える。しかし、Simple Fusion モデルのどちらも原言語文を正しく訳すことはできず、むしろ妥当性がベースラインと比較して同等もしくは悪化している。反面、今回提案した注意言語モデルでは原言語文の内容を妥当性を損なわず参照訳以上に正確に翻訳でき、流暢性も参照訳と同等レベルに翻訳できている。

このことから、言語モデル機構が出力の流暢性の向上性に寄与することが示され、また、注意言語モデルでは妥当性についても保持されていることがわかる。

5.3 注意言語モデルにおける言語モデルの影響

5.3.1 流暢性の向上

注意言語モデルの出力と単語アテンションについて一部を抜粋して表 6 に示す。

上位 5 単語を示した単語アテンションを確認すると、1, 2 トークン目を除いて適切な予測がなされている。例えば、始め鉤括弧 (「) で始まった文に対し、3 トークン目以降は終わり鉤括弧 (」) で閉じようとする傾向が見られる。また、“発電”で終えて鉤括弧を閉じるのが好ましくないと判断され、続く語は“所”であると予測している。これは、アテンションにより翻訳機構との関連性を保持した状態で言語モデルが文法や接続など流暢性を向上させる予測ができていることを示していると考えられる。

重みについて注目すると、一部の単語のみが高い重みを持つ場合もあれば、複数の単語が平均的な重みを持つ場合も存在することがわかる。これは翻訳機構の予測と言語モデル機構の予測がともに高い場合には特定の語の重みが高くなり、どちらか一方でも信頼の高くない予測を出した時、すなわち幅広い翻訳が考えられる時に平均的な重みを持つと考えられるが、この点についてはさらなる分析が必要である。

5.3.2 妥当性の担保

一方で、注意言語モデルによるアテンションが、翻訳の妥当性を向上させうる内容に関する語を導いていることは極めて少ない。仮に、注意言語モデルによるアテンションにおいてある特定の語が他の語と比較して大幅に高い重み

表 4 言語モデル機構による流暢性の向上例.

モデル	文
原言語文	responding to these changes DERS can compute new dose rate .
参照訳	DERS はこれらの変化に対応して新たな線量率を計算できる。
ベースライン	これらの変化に対応する応答は、新しい線量率を計算できる。
Simple Fusion (PRENORM)	これらの変化に対応すると、新しい線量率を計算できる。
Simple Fusion (POSTNORM)	これらの変化に対応する応答は新しい用量率を計算できる。
注意言語モデル	これらの変化に対応することにより、新しい線量率を計算できる。

表 5 Simple Fusion における妥当性の低下例.

モデル	文
原言語文	the magnetic field is given in the direction of a right angle or a parallel (reverse to the flow) to the tube axis .
参照訳	磁場は管軸に直角か平行逆方向に加えた。
ベースライン	磁場は、直角または平行(流れ)の方向に与えられた。
Simple Fusion (PRENORM)	磁場は、左右角または平行(流れに逆)になる。
Simple Fusion (POSTNORM)	磁場は、管軸に平行(逆に直角)または平行(流れに逆)方向に与えられた。
注意言語モデル	磁場は直角または平行(流れへの逆)方向に与えられる。

を示したとしても、その語が出力されるとは限らず、翻訳機構の予測が出力に用いられる可能性も高い。実際に表6の例を見ても、文頭を始めとして言語モデルの出力を考慮していない部分が多く見られる。

そもそも言語モデルは言語として確からしい出力をするものであり、1章でも述べた通り、原言語文の情報を用いた出力をしないことは明白である。そのことから鑑みても言語モデル機構は独立に予測しながらも、その情報を鵜呑みにせず翻訳機構との関連性 (attention) を考慮しているため、言語モデルを活用した翻訳ができており適切である。その反面、言語モデル機構が示す文法的情報は有用であり、翻訳に対して流暢な出力になりやすいものを情報として与えることにより流暢性の高い出力ができていますと考えられる。

以上のことから、本手法における言語モデルの役割として、出力文の流暢性を向上させるために翻訳機構に対して情報を与えていることとなり、例えば原言語文から得られる文意だけでは翻訳がしづらい細かな表記方法などの確度を上げるなどの、一種の正則化的役割が強いことを推測できる。

6. おわりに

今回は言語モデル機構を考慮した翻訳モデルについてアテンションを用いる手法を提案した。結果から、ニューラル機械翻訳に対して言語モデルを用いることは有用であることを再確認し、言語モデル機構と翻訳機構を同等に扱うのではなく、言語モデルに対してアテンションを用いることで妥当性をあまり下げずに流暢性の向上ができ、BLEUとRIBESをさらに向上することができた。

一方で、今回の実験はコーパスや単語分割の条件を限定

した状況で行ったものである。今回の設定では翻訳コーパス(翻訳機構)を基準に言語モデルを生成した。この状況では語彙の作成が翻訳コーパスにのみ依存しており、言語モデル機構も同等の設定で作られていない限り利用することはできない。すなわち、翻訳の設定に合わせて言語モデルを都度作り直す必要がある。しかしながら実際には翻訳モデルに依存する言語モデルを利用しなければならない状況は好ましくない。そのため、語彙に影響を受けないモデルについて検討すべきである。

参考文献

- [1] Bahdanau, D., Cho, K. and Bengio, Y.: Neural Machine Translation by Jointly Learning to Align and Translate, *Proc. of ICLR* (2015).
- [2] Gulcehre, C., Firat, O., Xu, K., Cho, K., Barrault, L., Lin, H.-C., Bougares, F., Schwenk, H. and Bengio, Y.: On Using Monolingual Corpora in Neural Machine Translation, arXiv (2015).
- [3] Isozaki, H., Hirao, T., Duh, K., Sudoh, K. and Tsukada, H.: Automatic Evaluation of Translation Quality for Distant Language Pairs, *Proc. of EMNLP*, Cambridge, MA, pp. 944–952 (online), available from <https://www.aclweb.org/anthology/D10-1092> (2010).
- [4] Luong, T., Pham, H. and Manning, C. D.: Effective Approaches to Attention-based Neural Machine Translation, *Proc. of EMNLP*, pp. 1412–1421 (2015).
- [5] Luong, T., Sutskever, I., Le, Q., Vinyals, O. and Zaremba, W.: Addressing the Rare Word Problem in Neural Machine Translation, *Proc. of ACL*, pp. 11–19 (2015).
- [6] Matsumura, Y. and Komachi, M.: Tokyo Metropolitan University Neural Machine Translation System for WAT 2017, *Proc. of WAT*, pp. 160–166 (online), available from <https://www.aclweb.org/anthology/W17-5716> (2017).
- [7] Nakazawa, T., Yaguchi, M., Uchimoto, K., Utiyama, M., Sumita, E., Kurohashi, S. and Isahara, H.: ASPEC:

表 6 注意言語モデルの出力と言語モデルの例 (抜粋).

モデル		出力														
原言語文		details of dose rate of " Fugen Power Plant " can be calculated by using <unk> software .														
参照訳		DERS ソフトウェアを用いて「ふげん発電所」の線量率を詳細に計算できる。														
注意言語モデル		「ふげん発電所」の線量率の詳細を、<unk> ソフトウェアを用いて計算できる。														
注意言語モデル (抜粋)		「	ふ	げん	発電	所	」	の								
単語 attention (上位 5 単語)	とその重み	本 9.9e-1	この 5.5e-1	」 9.9e-1	」 1.0	所 9.9e-1	」 1.0	」 1.0	について 7.7e-1	標記 8.7e-5	その 3.5e-1	ね 3.2e-6	号 2.7e-8	機 1.3e-4	発電 3.2e-12	の 1.7e-1
		この 4.2e-5	日本 7.0e-2	げん 2.0e-9	げん 1.4e-11	」 1.2e-6	の 1.7e-18	」 1.2e-6	における 4.5e-2	また 8.5e-6	1 2.7e-2	出 1.1e-10	<unk> 1.1e-12	設備 7.7e-11	<unk> 7.6e-19	で 6.4e-3
		これら 1.5e-6	高 4.7e-3	り 3.6e-11	・ 1.8e-14	装置 2.6e-12	用 6.3e-19	用 6.3e-19	と 3.2e-3							

Asian Scientific Paper Excerpt Corpus, *Proc. of LREC*, pp. 2204–2208 (2016).

- [8] Papineni, K., Roukos, S., Ward, T. and Zhu, W.-J.: BLEU: A Method for Automatic Evaluation of Machine Translation, *Proc. of ACL*, pp. 311–318 (2002).
- [9] Qi, Y., Sachan, D., Felix, M., Padmanabhan, S. and Neubig, G.: When and Why Are Pre-Trained Word Embeddings Useful for Neural Machine Translation?, *Proc. of NAACL*, pp. 529–535 (online), DOI: 10.18653/v1/N18-2084 (2018).
- [10] Ramachandran, P., Liu, P. and Le, Q.: Unsupervised Pretraining for Sequence to Sequence Learning, *Proc. of EMNLP*, pp. 383–391 (online), DOI: 10.18653/v1/D17-1039 (2017).
- [11] Sennrich, R. and Haddow, B.: Linguistic Input Features Improve Neural Machine Translation, *Proc. of WMT*, pp. 83–91 (online), DOI: 10.18653/v1/W16-2209 (2016).
- [12] Sennrich, R., Haddow, B. and Birch, A.: Improving Neural Machine Translation Models with Monolingual Data, *Proc. of ACL*, pp. 86–96 (online), DOI: 10.18653/v1/P16-1009 (2016).
- [13] Sriram, A., Jun, H., Satheesh, S. and Coates, A.: Cold Fusion: Training Seq2Seq Models Together with Language Models, arXiv (2017).
- [14] Stahlberg, F., Cross, J. and Stoyanov, V.: Simple Fusion: Return of the Language Model, *Proc. of WMT*, pp. 204–211 (online), available from <http://aclweb.org/anthology/W18-6321> (2018).
- [15] Tu, Z., Lu, Z., Liu, Y., Liu, X. and Li, H.: Modeling Coverage for Neural Machine Translation, *Proc. of ACL*, pp. 76–85 (online), DOI: 10.18653/v1/P16-1008 (2016).