

## 日本語話し言葉書き言葉変換による 大学講義の日英翻訳の精度向上

中尾 亮太<sup>†</sup>・Chenhui Chu<sup>†</sup>・黒橋 禎夫<sup>†</sup>

話し言葉の機械翻訳では、話し言葉に特有の現象が翻訳精度に悪影響を及ぼすことが知られている。本研究では大学講義翻訳システムにおける日英翻訳の前処理として、日本語の話し言葉から書き言葉への自動変換を行うことにより翻訳精度を向上させる。まず大学講義の書き起こしとそれを書き言葉に変換したものを、対応する英文の3つ組からなるコーパスを構築した。次にそれを用いて話し言葉書き言葉変換モデルと日英翻訳モデルを学習させた。その結果、話し言葉書き言葉変換が日英翻訳の精度を向上させることを示した。また、話し言葉に特有の現象の分類に基づき、どのような現象が翻訳精度にどの程度影響するのかを定量化した。

キーワード：機械翻訳, 話し言葉, 書き言葉, 講義翻訳

## Spoken-Written Japanese Conversion for Japanese-English University-Lecture Translation

RYOTA NAKAO<sup>†</sup>, CHENHUI CHU<sup>†</sup> and SADAO KUROHASHI<sup>†</sup>

In machine translation of spoken language, it is known that phenomena specific to spoken language have a negative impact on translation accuracy. Therefore, in this study, as a preprocessing step for Japanese-English translation in our university lecture translation system, we improve the translation accuracy by automatically converting spoken-style Japanese texts to written-style. First, we create a corpus consisting of Japanese transcriptions of university lectures, their conversions into written language, and the corresponding English texts. Next, we train spoken-written conversion models and Japanese-English translation models using the corpus. As a result, we show that spoken-written Japanese conversion improves the accuracy of Japanese-English translation. In addition, we quantify which phenomena affect translation accuracy and to what extent.

**Key Words:** *Machine Translation, Spoken Language, Written Language, Lecture Translation*

---

<sup>†</sup> 京都大学, Kyoto University

## 1 はじめに

グローバル人材育成のため、政府は外国人留学生の受け入れを推進している<sup>1</sup>。また、京都大学でも教養・共通科目を英語で提供したり、日本語・日本文化の教育を組み込んだプログラムを用意するなどの施策を行っている<sup>2</sup>。しかし、主に講義で用いられる言語は日本語であるため、日本人と外国人留学生との間には学習機会の差が未だに存在し、日本語を勉強中の学生をより一層サポートする必要がある。

この解決策として、専門的な内容を通訳できる通訳者を各講義に配置することは、コスト面から現実的でない。一方で、近年、深層学習などによって音声認識や機械翻訳の技術は劇的に進展しており、このような自動処理を活用していくことが今後の方向性であると考えられる。

音声の機械翻訳は、音声から音声の場合 (Jia et al. 2019) と音声からテキストの場合があるが、本研究では後者を対象とする。音声からテキストへの翻訳では、一つのモデルで音声から翻訳済みのテキストまで End to End で変換する手法 (Liu et al. 2019) と、書き起こしと翻訳を別々のモデルで行う手法 (Müller et al. 2016; 須藤 他 2019) がある。前者はエラーの累積が少なく、処理時間も短くできることが多いが、ある言語の音声と別言語のテキストの対応をとったコーパスは少ない (Liu et al. 2019)。また分野適応のための新規コーパスの構築も難しい。このため、翻訳精度を高くすることが困難である。一方、後者の手法は、音声認識器からのエラーの累積があるものの、講義のような比較的落ち着いた話し言葉の音声認識は近年かなり高精度になってきており、単言語内での音声とその書き起こしのコーパスやテキスト形式の翻訳コーパスは利用可能なものが多く、新たなコーパスの構築も難しくない。さらに、日本語の書き起こしと英訳を字幕として同時に提供することで、講義内容の理解の補助と同時に日本語の習得の手助けもできる。

本研究は、以上のメリット・デメリットと、講義音声を他言語字幕に翻訳する既存のシステム (Müller et al. 2016; 須藤 他 2019) の構成を踏まえ、日本語音声を生音声認識器によって書き起こし、次いで機械翻訳により日英翻訳を行う設定で研究を行う。機械翻訳モデルは Transformer (Vaswani et al. 2017) の登場以来翻訳精度が大きく向上しているが、このモデルの性能を十分に発揮するためには大量のテキストデータを用意する必要がある。広く利用可能な日本語テキストの多くは書き言葉を用いたものであり、話し言葉のコーパスは少ないため、機械翻訳モデルは日本語書き言葉から英語への翻訳を学習することになる。一方、講義では話し言葉が用いられ、それを書き起こしたテキストが機械翻訳モデルへの入力となる。このような書き言葉と話し言葉という訓練データと実際のデータ間のギャップは翻訳精度に悪影響を及ぼすことが知られている (Wang et al. 2010; Salesky et al. 2018; 坂本, 田中 2015)。

<sup>1</sup> [https://www8.cao.go.jp/youth/whitepaper/r01honpen/s6\\_1.html](https://www8.cao.go.jp/youth/whitepaper/r01honpen/s6_1.html)

<sup>2</sup> <https://www.kyoto-u.ac.jp/ja/international/students1/study1/undergraduate>

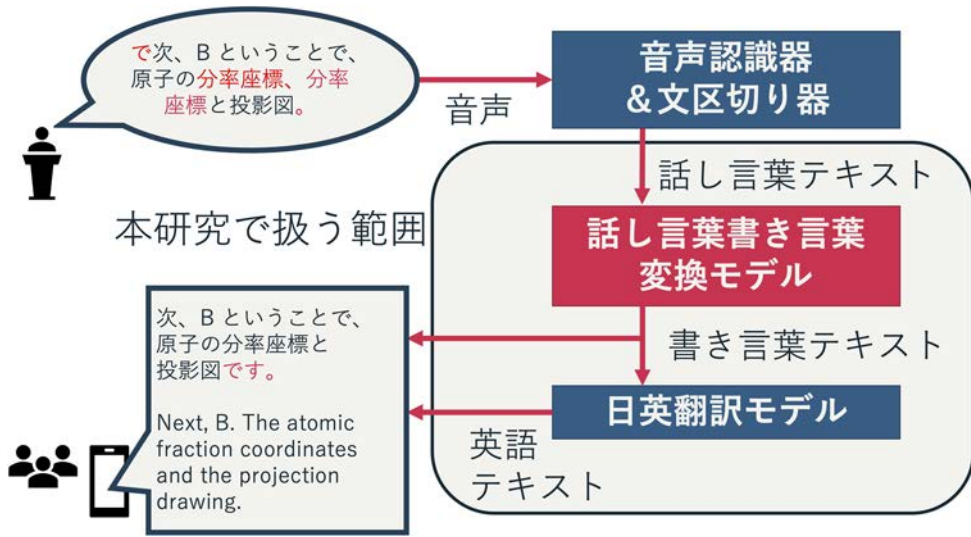


図 1 大学講義の日英翻訳システムの概要

本研究では日本語の話し言葉と書き言葉の違いに着目し、翻訳精度の向上と、標準語に近い整った日本語書き起こしを提供することによる日本語学習の促進を目的として、話し言葉から書き言葉への自動変換を行う (図 1)。タスクとしての話し言葉書き言葉変換には、フィラーや文末の「ですね」の除去等、パターンマッチや簡単な言語モデルである程度解決可能な問題だけでなく、言い直しや言い換えといった冗長な表現の削除や省略された格助詞の補完など文脈に依存した難しい問題も含まれる。本研究ではこれらの問題を単一のモデルで解決することに取り組む。本システムの推論時には、文の区切りとして音声認識器の出力に含まれる無音区間が文末かどうかを判定するニューラルネットワークモデルによる出力を用いる (例:「でアレニウス式は温度依存性を表すやつ [SEP] です [SEP] はい [SEP] そして最後に [SEP] 反応器の [SEP]」の場合は、下線部分の無音区間 ([SEP]) が文末と判定されるような学習を行ったモデルを用いる)。

この話し言葉書き言葉変換を行うモデルを訓練するため、また、その変換による日英翻訳への影響を調査するため、新たなコーパスを構築した。このコーパスは大学講義の書き起こしとそれを書き言葉に変換したもの、対応する英文の 3 つ組からなる。なおコーパス構築の際には、話し言葉特有の現象の分類 (島津 他 2014; 国立国語研究所 2006) を参考にし、どのような現象を変換すべきか取り決めた。

こうして構築したコーパスを用いて話し言葉書き言葉変換モデルと日英翻訳モデルを学習させ、話し言葉書き言葉変換が日英翻訳の精度を向上させることを実験的に示した。また、話し言葉書き言葉変換の複数手法の定量的評価を行った。加えて、話し言葉に特有の現象の分類に基づき、どのような現象がどの程度翻訳精度に影響するのかを定量化した。

## 2 関連研究

本節では、まず話し言葉と書き言葉がどのように違うかを明確化し、話し言葉のコーパスについて述べる。次に話し言葉を書き言葉に変換するに当たり、どのような手法が存在するかを挙げ、最後にそのような変換による翻訳精度への影響についての関連研究を紹介する。

### 2.1 話し言葉と書き言葉の違い

話し言葉に特有の現象を、島津 他 (2014) は 4 つに大別している。本研究においてコーパスを構築するに際し、作業者に話し言葉と書き言葉の違いを理解してもらうために、これを例文とともに提示した。

- 語彙的な現象：音韻的な縮約 / 特有の終助詞 / 口語的な助詞 / 敬語 / 感動詞
- 省略：格助詞の省略・無助詞 / 判定詞の省略
- 冗長な表現：言い直し / 繰り返し / 言い換え / 言い淀み
- 文の概念がないことに関連する現象：名詞句・助詞句発話 / 格助詞での中止 / ねじれ / 倒置

ここで挙げたもののうち、どの現象を本研究の対象とするかは次節以降で議論する。なお、島津らはその他の現象として「数え上げ」と「綴りの説明」も挙げているが、これらについてはあまり現れないとしている。そのため、本研究でもこれらは扱わない。また、文の概念がないことに関連する現象として「先取り」が挙げられているが、これは主に会話で発生する現象であり、講義ではあまり発生しないので省いた。

話し言葉の性質について言及した他の研究の一つである国立国語研究所 (2006) は、作業者によって判断が揺らがないように厳密に定義がなされている。例えば島津 他 (2014) では「音韻的な縮約」と一纏めにしているものを「動詞ラ行音にかかわる撥音化」「テ形の複合動詞にかかわる縮約」などに細かく分類した上で、口語表現であると認める場合を列挙する形で示している。これは本研究では作業者にとって過剰に詳細であるため、本研究のコーパス構築作業にあたっては島津 他 (2014) を参考にした。

日本語の話し言葉のコーパスとして代表的な日本語話し言葉コーパス (CSJ)(前川 2004) は、学会講演、および模擬講演を中心とした音声データを元に、様々な観点からアノテーションを付与し XML 文書の形で利用可能にしたコーパスである。CSJ での話し言葉特有の現象に対するアノテーションとしては、島津 他 (2014) の分類に含まれる音の転訛、フィラー、言い直し・言い淀み等による語断片の他、母音や子音の引き伸ばし、話者の笑いや咳、語の読みに関する知識レベルの言い間違いなども含まれる。

## 2.2 話し言葉の自動整形

尾嶋 他 (2008) は話し言葉に対する整形処理における一次整形として削除, 挿入, 置換の3つの操作を挙げており, CSJ に対してこれらの一次整形を行う箇所を開始タグと終了タグを付与したデータを作成した. また, そのうちの削除する箇所について自動で検出する手法を提案している.

他の日本語の話し言葉を自動で修正し読みやすくする事例として国会会議録を音声データから自動で書き起こすシステム (秋田 他 2010) がある. この研究では書き起こしの忠実性や発話者の意図の尊重のため, 言い直しや繰り返しはそのままテキストに残しており, 冗長な部分を修正するとしても文末の「ですね」の除去などに限られる. しかし, 2.3 項で示す関連研究においては音声認識に続く機械翻訳の段階での話し言葉特有の現象による翻訳精度の低下が示されている. そのため, 本研究では翻訳タスクでの影響を重要視し, 2.1 項に示した話し言葉と書き言葉の違いを可能な限り網羅するような修正を施す.

話し言葉書き言葉変換は Sequence to Sequence のタスクとして解釈することが可能である. そして, 単言語での Sequence to Sequence タスクを解く手法として, 事前訓練済みの言語モデルを fine tuning する手法が近年大きな注目を集めている. 例えば T5 (Raffel et al. 2020) は入出力をタスク固有のプレフィックスとともに Text to Text の形式に変換することで文書分類や自然言語推論など様々なタスクを単一のモデルの転移学習で解くことができると示した. このようなモデルは幅広いタスクに適用可能である一方, 事前訓練に莫大な資金と時間が必要となるため, 英語以外での訓練済みモデルの一般公開は少ない. 日本語で利用可能な事前訓練済みモデルとしては日本語 Wikipedia で事前訓練された BART (田中, 黒橋 2020) が挙げられる.

話し言葉は, フィラーや特有の助詞が用いられるものの, 一般的には文法的に間違っているわけではない. しかし同じ言語内で文の意味を維持したまま, トークンを挿入したり削除したりといった文の操作を行うという観点では, 時制や冠詞などの文法上の間違いを訂正するタスクである文法誤り訂正と類似している. LaserTagger (Malmi et al. 2019) は, 文法誤り訂正を含む Sequence to Sequence のタスクを Sequence Labeling として解くことができる手法で, 小さいデータセットでも良い性能を出せることと推論が高速であることが特徴である.

## 2.3 話し言葉の整形による翻訳精度向上

Wang et al. (2010) は, 翻訳の前処理としてフィラーや冗長な表現を除去することで翻訳精度が向上することを示している. また, Salesky et al. (2018) は Transformer のような深層学習に基づくモデルでも訓練に利用可能な流暢なコーパスと実際に入力される非流暢なテキストの不一致が翻訳精度の低下を引き起こすと主張している. しかし日本語では, それらだけでなく「ですんで」「こっから」のような音韻的な縮約による違いも多く見られる. 本研究では非流暢性だけでなく語彙的な現象による翻訳精度への影響も調査する. また, 整形されたテキストを字幕

として提供することは日本語初学者が勉強するに当たり、辞書を引きやすくなる、標準的な日本語を覚えられるなどのメリットがある。

坂本, 田中 (2015) は話し言葉の日英翻訳における前処理として話し言葉を簡潔な表現に変換する手法を提案しているが、前処理、日英翻訳ともにルールベースであり、End to End な手法と比べて構築・管理の手間がかかる。また、2.1 項で述べた冗長な表現を代表として、話し言葉においては単純なパターンマッチや形態素解析で検出するのが難しい現象が存在するため、坂本らの研究ではこれらの変換は取り扱っていない。加えて、翻訳の手法は Transformer アーキテクチャの登場により飛躍的に精度を向上させており、再度話し言葉の変換による翻訳精度向上の有効性を検証する必要がある。

### 3 コーパスの構築

#### 3.1 文区切り器の訓練

講義翻訳システムで用いる文区切り器は独自のコーパスで訓練されたものであるため、ここでそのコーパスについて述べる。文区切り器は事前訓練と fine tuning で別々のコーパスを使用している。

事前訓練に使用したのは日本語話し言葉コーパス (CSJ) をベースとしたコーパスで、CSJ の noncore に含まれる ClauseBoundaryLabel が文末とアノテーションされた箇所を区切るべき位置としている。区切り位置を推定する訓練を行うため、5 から 10 の文をひとまとめにし、正解の文末位置トークンを文と文の間に挿入する。そして、各文を分かち書きして形態素の間にダミーの文末位置トークンをランダムに挿入する。このようにして作成されたのが文区切り器の事前訓練コーパスである。

fine tuning には京都大学の講義を AMI の音声認識器 AmiVoice エンジン<sup>3</sup>で書き起こしたものをベースとしたコーパスを用いる。このコーパスには音声認識器が出力する無音区間トークンが発話内に含まれており、この無音区間の中からアノテーション作業を行った者が文末としてふさわしいと考えた箇所を文末の正解データとした。文区切り器は入力文字列中の無音区間が文末かどうかの判別を学習する。

文区切り器は、これらのコーパスを用いて訓練された BERT<sub>base</sub> モデルであり、入力トークン列に含まれる文末候補位置（音声認識器の出力する無音区間トークン）の中から文末位置を推測する Sequence Labeling モデルとなっている。fine tuning 用コーパスのテストデータに対する f 値は 82.2% となっている。

<sup>3</sup> <https://www.advanced-media.co.jp/amivoice>

### 3.2 アノテーション

ここで示すデータは京都大学で2019年度に行われた5つの講義の書き起こしをもとに構築した。表1に講義名と文の数を示す。以後、この書き起こしのデータをJ<sub>話</sub>と表記する。

J<sub>話</sub>をベースに2.1項で述べた話し言葉と書き言葉の違いを具体例とともに作業者に提示して変換を依頼した。ただし、それらの話し言葉と書き言葉の違いのうち、「名詞句・助詞句発言」「ねじれ」はコンテキストへの依存性の高さや専門的な内容に言及している場合の判断の難しさから、作業者には変換を指示しなかった。また、講義で使われる敬語の多くは丁寧語であり、書き言葉でもよく見られる表現であるため、「敬語」に関する違いも変換対象としなかった。また、話し言葉から書き言葉への変換は一意的なものではないため、本データセットでは多様性を重視し変換ルールの統一は行わず、各作業者に判断を委ねることとした。作業は日本語を母語としテープ起こし作業を業務としている者8名で、担当する量がほぼ等分となるように行った。

実際に変換作業を行った所、2.1項で述べた話し言葉と書き言葉の違いに含まれないが修正を施すべき点がいくつか見つかった。以下に具体例を示す。

- 文頭の「で」の除去  
「それで」と同じ働きをする接続助詞の「で」は、話者の癖により本来必要としない箇所でも出現することが多いことがわかった。
- 句読点位置の修正  
話し言葉では文という概念がないため、それを書き起こす際の句読点位置についても曖昧性が生じる。そのため、より適切な位置に句読点をずらしたり、過剰な句読点を削除する修正を行った。この修正は曖昧性があるが、各作業者の判断に委ねた。  
例文：「菱面对称というのは、括弧ですね、括弧で菱面对称であるんですが」→「菱面对称というものの括弧です。括弧で菱面对称であるのですが」
- 文法上の間違いの訂正  
接続詞ではないのに文頭で使われている「なので」、明らかな格助詞のミスなど、文法的に不適切な箇所を訂正した。

講義名	文数
有機化学基礎および演習	2,815
物理化学基礎及び演習	4,520
電気回路基礎論	4,159
化学プロセス工学基礎	2,049
基礎無機化学	4,041
合計	17,584

表 1 書き起こしの講義名と文数

分割	J <sub>話</sub> -J <sub>書</sub>	J <sub>話</sub> -J <sub>書</sub> -E
訓練	11,072	3,353
開発	1,384	275
テスト	1,384	367
合計	13,840	3,995

表 2 構築したデータセットの文数

- 意味的に冗長な箇所の削除

以下の例文のように、発話の意味的な内容において重要でない部分を削除し、簡潔かつ整った文体にした。

例文：容易に作ることができるという形です。

- 話し言葉的な語の変換

音韻的な縮約があるものや話し言葉に特有の助詞に限らず、話し言葉で多く用いられると考えられる語を同じ意味で書き言葉で主に用いられているものに置換した。

例：さっき（先ほど）、なんか（など）

変換後のデータを以後 J<sub>書</sub> と表記する。J<sub>話</sub>-J<sub>書</sub> の訓練・開発・テストセットの文数を表 2 に示す。なお本システムは、毎年度同じ講師がほぼ同じ内容の講義を行い、そのたびにデータが蓄積され、改めてモデルが訓練されるという想定のもとで開発されている。そのため、未知の語彙や熟語が出現する頻度は低いと仮定し、開発セットとテストセットは全ての文からランダムに選択した。

本コーパスに含まれるデータの具体例を以下に《変換前/変換後》の表記で示す。

- 《でえーとまあ/》《こっから/ここから》RCO 《/,》ここ《ですね/です》《。/,》《えーカルボニル、/》ここがカルボニル基《なんだけど/なのですけれど》、《ここがカルボニル基なんですけど、/》
- 《あ/》ページが《変わっちゃうのか、/変わってしまいます。》《いいですか、理想気体のほうは/理想気体のほうはいいですか》。
- 今日の分は《やっというて/行っておいて》いただいた方がいいと思います。

J<sub>書</sub> の一部の翻訳を翻訳業者に依頼し、日英対訳コーパスを構築した。この英訳を以後 E と表記する。これにより、話し言葉と書き言葉と英訳の対応をとった 3 つ組のコーパス J<sub>話</sub>-J<sub>書</sub>-E データセットを構築した。統合されたデータの文数は表 2 に示す。文数とあるが、実際には句読点位置の修正の結果や翻訳の都合で、2 つ以上の文が 1 つの文と対応づいている場合も存在する。

## 4 話し言葉書き言葉変換

### 4.1 手法

話し言葉書き言葉変換を行う手法として、LaserTagger と BART を比較する。J<sub>話</sub>-J<sub>書</sub> データに対してこの 2 モデルを fine tuning し、評価には SARI (Xu et al. 2016) を用いる。実装は LaserTagger 著者らと同じものを用いる。なお、この実装では、出力と正解の完全一致（修正箇所が 0 個の場合）に対して高いスコアを与えるため、スコアの計算時に  $0/0 = 1$  として計算している。また、ランダムなテストデータ 100 件に対する推論にかかる平均時間も計測する。



#### 4.1.1 LaserTagger

LaserTagger は、ある種の Sequence to Sequence のタスクを Sequence Labeling として解く手法である。入力の各トークンを維持するか削除するかをラベル付けし、必要ならトークン列を挿入する。エンコーダ部分は、マスキングされたトークンを復元する Masked LM と 2 文間の関係が連続する文かそれ以外かを判断する Next Sentence Prediction の 2 タスクで事前訓練された Transformer Encoder である BERT を用いる。デコーダ部分は単一の全結合層を用いる  $\text{LASERTAGGER}_{\text{FF}}$  (以後  $\text{LT}_{\text{FF}}$  とする) と一層の Transformer Decoder を用いる  $\text{LASERTAGGER}_{\text{AR}}$  (以後  $\text{LT}_{\text{AR}}$  とする) のバリエーションがある。

話し言葉書き言葉変換のタスクは、2.1 項で示したように、多くは語彙の変換や削除で対応できるため、LaserTagger が適用できる。ただし、倒置の修正は多くのトークンにまたがる変換であるため LaserTagger では解くことが難しいというデメリットがある。一方 Sequence to Sequence の手法はそのような制限はないため解ける可能性がある。

Sequence to Sequence で問題を解く際、Transformer を始めとする多くの主流なモデルは一定数の語彙以外のトークンを表す特殊トークン (<UNK> など) で表記される) を出力してしまう可能性がある。このような不要なトークンが翻訳モデルの入力に混じると、翻訳精度の低下につながる恐れがある。また、Sequence to Sequence のモデルには同じトークン列を繰り返し出力してしまう問題も存在する。Sequence Labeling の場合は、トークンの維持と削除、そして予め決められた出力フレーズ語彙からのトークン列の挿入しか操作のパターンがないため、それらの可能性をなくすことができる。より品質の保証されたサービスを提供するという観点において、機械翻訳の前処理として不要なノイズを加えてしまう可能性をコントロールできることも LaserTagger のメリットの一つである。

LaserTagger に必要となる事前訓練された BERT モデルは、日本語 Wikipedia で訓練された  $\text{BERT}_{\text{large}}$  (柴田 他 2019) を用いる。また、LaserTagger の Decoder は全結合層の場合と Transformer Decoder の場合の両方を検討する。J<sub>話</sub>-J<sub>書</sub> データセットにおける fine tuning は、Tesla V100 GPU を 1 枚使用してバッチサイズ 8 で 20 エポック訓練し、各エポック終了時にモデルのパラメータを保存した。開発セットに対する exact スコアが最も高いパラメータを最終的なスコアの算出に用いる。

LaserTagger は倒置などの複数のトークンにまたがる修正や出力フレーズ語彙に含まれないトークン列の挿入を必要とする変換を行うことができないため、訓練セットからはそれらのデータは削除される。出力フレーズ語彙の数を Malmi et al. (2019) にならって 500 としたとき、訓練セット 11,072 件のうち、Sequence Labeling 形式に変換できたものは 8,832 件であった。倒置が発生している文や頻度の低い語彙が挿入されている文に限って訓練上重要な変換が含まれるとは考えにくいいため、この 2,340 件の削除の影響は大きくないと想定される。

### 4.1.2 BART

比較対象として, Sequence to Sequence のモデルである BART を用いる. 田中, 黒橋 (2020) が日本語 Wikipedia で行った事前訓練には次の 2 種類が併用されている.

- Text Infilling : 長さ 0 以上のトークンのスパンが単一の特殊トークンで置き換わっており, これを復元する
- Sentence Permutation : 複数文がランダムな順に並べ替えられており, これを復元する  
モデルサイズは BART<sub>large</sub> を用いる. この事前訓練は Tesla V100 GPU を 4 枚使って約 1 ヶ月行われた. なお語彙数は 32,000, バッチサイズは 512 である. fine tuning 時には通常の Transformer アーキテクチャとして Sequence to Sequence のタスクを学習する. J<sub>話</sub>-J<sub>書</sub> データセットに対する fine tuning は, Tesla V100 GPU を 1 枚使用して 40 エポック訓練し, 各エポック終了時にモデルのパラメータを保存した. 開発セットに対する loss が最も低いパラメータを最終的なスコアの算出に用いる. なお, 出力に <UNK> が含まれている場合, 入力に 1 対 1 で対応が取れる <UNK> が存在する場合はもともとのトークンを復元する処理を加えている.

## 4.2 結果と考察

結果を表 3 に示す. LaserTagger が BART<sub>large</sub> と並ぶ精度を持っていることがわかる. ここから, 話し言葉と書き言葉でほとんどのトークンが同一であり, 基本的にはトークンを維持または削除するだけでよく, 挿入する場合もそのパターンは 500 個あれば語彙数 32,000 の Sequence to Sequence モデルに匹敵することが可能であるとわかる. また, LT<sub>AR</sub> はトークンの追加でわずかに LT<sub>FF</sub> に勝るものの, ほぼ同等の性能を持つことがわかった. なお, BART<sub>large</sub> の出力において <UNK> の復元に失敗した例はテストセット中には含まれなかった. 処理時間については, 機械翻訳モデルの推論時間が約 1 秒であるため, リアルタイムでの使用においては BART は推論時間が長すぎて実用性に欠けると言わざるを得ない. よって, 続く実験では LT<sub>FF</sub> を主な実験対象とする. LT<sub>AR</sub> は LT<sub>FF</sub> に比べ処理時間が長いものの翻訳処理よりは短いため実用性があると考え, わずかながら性能で勝るため比較対象として実験対象とする.

変換の具体例を表 4 に示す. 《なってる/なっている》と《ですよね/です》の修正は全てのモデルが成功しているが, 冗長な表現である「電荷, 正電荷を」の修正については LT<sub>AR</sub> は失敗している. なお, LaserTagger に対しては 5 節でさらなる具体例を示し翻訳に対する影響の考察を行う.

手法	維持	追加	削除	SARI	処理時間
LT <sub>FF</sub>	91.9	65.4	<b>87.7</b>	81.7	<b>38 ms</b>
LT <sub>AR</sub>	91.9	65.6	<b>87.7</b>	81.7	406 ms
BART <sub>large</sub>	<b>92.0</b>	<b>68.3</b>	87.0	<b>82.5</b>	2,511 ms

表 3 話し言葉書き言葉変換の評価

手法	出力
J <sub>話</sub>	+1 にな <u>って</u> る <u>という</u> か、 <u>これが電荷、正電荷を</u> <u>帯びて</u> る <u>という</u> ことは <u>考えにくい</u> です <u>よね</u> 。
J <sub>書</sub>	+1 にな <u>って</u> いる <u>という</u> か、 <u>これが正電荷を</u> <u>帯びて</u> いる <u>という</u> ことは <u>考えにくい</u> です。
LASERTAGGER <sub>FF</sub>	+1 にな <u>って</u> いる <u>という</u> か、 <u>これが、正電荷を</u> <u>帯びて</u> いる <u>という</u> ことは <u>考えにくい</u> です。
LASERTAGGER <sub>AR</sub>	+1 にな <u>って</u> いる <u>という</u> か、 <u>これが電荷、正電荷</u> <u>を帯びて</u> いる <u>という</u> ことは <u>考えにくい</u> です。
BART <sub>large</sub>	+1 にな <u>って</u> いる <u>という</u> か、 <u>これが正電荷を</u> <u>帯びて</u> いる <u>という</u> ことは <u>考えにくい</u> です。

表 4 話し言葉書き言葉変換の具体例

## 5 話し言葉書き言葉変換の日英翻訳への影響

### 5.1 手法 (LaserTagger & Transformer)

日本語話し言葉書き言葉変換モデルには、4.2 項で述べたように LT<sub>FF</sub> を主として用い、比較用として LT<sub>AR</sub> を用いる。日英翻訳モデルには Transformer の標準実装である tensor2tensor (Vaswani et al. 2018) に実装された Transformer<sub>big</sub> を、科学技術論文の概要のアラインメントから構築された約 1380 万文の日英対訳コーパスで事前訓練 (Tesla V100 GPU 4 枚で 11 万 step) したものをを用いる。

この実験では、英文の対応付けの際に文の区切りが調整されたことにより、J<sub>話</sub>-J<sub>書</sub>-E のテストデータが J<sub>話</sub>-J<sub>書</sub> の訓練データに含まれていないことを保証することが困難であったため、表 2 の J<sub>話</sub>-J<sub>書</sub> のデータ全体は使わず、J<sub>話</sub>-J<sub>書</sub>-E の対応がついている部分のみを用いる。LaserTagger は J<sub>話</sub>-J<sub>書</sub> のペアで fine tuning (4.1.1 目と同様の条件) を行い、Transformer は J<sub>書</sub>-E のペアで fine tuning (Tesla V100 GPU 4 枚で開発データにおける BLEU スコアが上昇しなくなるまで 200 step 単位で訓練) を行う。これによって話し言葉を自動で書き言葉に変換した後に日英翻訳を行うことができる。比較対象として、J<sub>話</sub>-E で fine tuning を行った Transformer に J<sub>話</sub> のテストデータを入力する場合と、J<sub>書</sub>-E で fine tuning を行った Transformer に J<sub>書</sub> のテストデータを入力する場合を考える。前者は話し言葉書き言葉変換による精度の変化を定量化するのに使い、後者は話し言葉書き言葉変換による精度向上の上限の指標とする。なお、最終的な評価指標には SacreBLEU (Post 2018) による BLEU スコアを使用する。

## 5.2 結果と考察

以後, \*Eはそれぞれ対応する日英翻訳モデルからの出力を表すこととする. 結果を表5に示す. 話し言葉書き言葉変換により BLEU スコアが向上しており,  $LT_{AR}$  は  $J_{書-E}$  と同等,  $LT_{FF}$  もほぼ同じスコアを示した.

表6に話し言葉書き言葉変換により翻訳精度が改善している例を示す. 下線で示した部分が冗長な箇所であるが,  $J_{話-E}$  は英訳が不適切になっている. 一方, この冗長な箇所が修正されている場合の翻訳はより適切なものとなっている.  $LT_{FF}$  は不要なトークン (二重下線部分) を挿入してしまっているが, 英訳には影響は見られない.

## 5.3 話し言葉特有の現象の種類ごとの翻訳精度への影響

$J_{話-J_{書-E}}$  の全テストデータに対して, 2.1項で述べた4種類の現象と, 3節で述べたその他の現象のそれぞれが現れているかどうかをラベル付けした. ただし,  $J_{書}$  にて修正されていないもの (名詞句発言やねじれなど) はカウントしない.

手法	BLEU
$J_{話-E}$	20.7
$J_{書-E}$	<b>21.4</b>
$LT_{FF-E}$	21.3
$LT_{AR-E}$	<b>21.4</b>

表5 話し言葉書き言葉変換による翻訳精度への影響

データ	具体例
$J_{話}$	な必ず物質が出てきたらですね、あの空間群とかあの規定する、対称性を規定する記号みたいなが出てきます。
$J_{書}$	必ず物質が出てきたら、空間群とか対称性を規定する記号のようなものが出てきます。
$LT_{FF}$	必ず物質が出てきたら、空間群とか、対称性を規定する記号を <u>み</u> たいのが出てきます。
$LT_{AR}$	必ず物質が出てきたら、空間群とか対称性を規定する記号 <u>み</u> たいのが出てきます。
$J_{話-E}$	If matter comes out, symbols define symmetry according to the space group.
$J_{書-E}$	If a substance appears, you will have a symbol that defines the space group and symmetry.
$LT_{FF-E}$	If matter always appears, you will find symbols that define the space group or symmetry.
$LT_{AR-E}$	If matter always appears, there are symbols that define the space group or symmetry.
E	Whenever a material appears, something like a symbol that defines a space group or symmetry appears.

表6 話し言葉書き言葉変換による日英翻訳の改善の具体例

ラベル付けした結果を表7に示す。また、その他の現象の修正 (e) の内訳を表8に示す。この付与したラベルに基づき、テストセットを分割してBLEUスコアを算出し、話し言葉特有の現象の種類ごとの翻訳精度への影響を調査した。

表9に省略の修正 (b) を含む場合とそれ以外とで分割した際のBLEUスコアの違いを示す。(b) を含む場合は含まない場合に比べてBLEUスコアで6.0~6.7ポイントの差がある。これは他の修正に関する場合と比べても大きい差であり、助詞や判定詞の省略と翻訳精度の低下に大きな相関があることを示す。しかし、BLEUスコアの差が最も大きいのはJ書-Eの場合であり、省略を修正することでは翻訳精度はそれほど改善しない(J話-Eと比較し0.4ポイント差)。これは省略が発生するような発話はそもそも翻訳が難しいことを示していると考えられる。例えば文としての構造が整っておらず、英訳時に語順を入れ替えるなどの調整を入れる余地がある場合や、格助詞や判定詞だけでなく主語や目的語まで省略されている(これらはコーパス構築時に修正要件に含めていない) ためにそれらを無理やり英訳した結果が間違っている場合などが考えられる。

表10に冗長な表現の修正 (c) を含む場合とそれ以外とで分割した際のBLEUスコアの違いを示す。まず、省略の修正 (b) に関するBLEUスコアの違いと異なり、修正を含むか含まないかの分割によるBLEUスコアの差が0.4~2.1と小さい。これは冗長な表現はどのような状況でも発生し、そのため翻訳の難しさの分散がこの修正を含む場合と含まない場合とで大きく異なることに起因すると考えられる。次に、(c) を含む場合はLT<sub>FF</sub>-EとLT<sub>AR</sub>-Eのスコアが

ラベル	件数
(a) 語彙的な現象の修正 を含む	294
(b) 省略の修正 を含む	58
(c) 冗長な表現の修正 を含む	82
(d) 文の概念がないことに関連する現象の修正 を含む	14
(e) その他の現象の修正 を含む	236
(z) 修正箇所なし	31

表7 テストデータ 367 件の分類

修正内容	件数
句読点位置の修正	148
文頭の「で」の除去	99
話し言葉的な語の変換	68
意味的に冗長な箇所の削除	23
文法上の間違いの訂正	8

表8 その他の現象の修正 (e) の内訳 (重複を含む)

J<sub>話</sub>-E と J<sub>書</sub>-E のほぼ中間に位置していることがわかる。これは冗長な表現の修正に関しては LaserTagger にまだ改善の余地があり、いくつかの場合で修正に失敗していることが原因であると考えられる。

表 11 に文の概念がないことに関連する現象の修正 (d) を含む場合とそれ以外とで分割した際の BLEU スコアの違いを示す。なお、(d) のラベルが付与されたデータが 14 件と少ないため、ここで述べることは一般性に欠ける可能性がある。(d) を含むか含まないかの分割に関して、J<sub>話</sub>-E の場合に最も大きい 6.3 ポイントの BLEU スコアの差があり、文としての体をなしていないテキストの機械翻訳が難しい事がわかる。また、J<sub>書</sub>-E の場合は 2.4 ポイントに縮まっているため、この現象を修正することで翻訳精度を大きく向上させることができると考えられる。ただし、文の概念がないことに関連する現象の発生頻度が低いいため、翻訳精度全体としての影響は小さい。文の概念がないことに関連する現象の修正 (d) を含む場合について、冗長な表現の修正 (c) の場合と同じく、LT<sub>FF</sub>-E と LT<sub>AR</sub>-E は J<sub>話</sub>-E と J<sub>書</sub>-E のほぼ中間の BLEU スコアとなっている。すなわち修正に改善の余地があると考えられる。しかし、LaserTagger は原理的に倒置を修正することがほぼ不可能であるため、これを改善するためには訓練セットの増量などではなく手法自体の改良が必要である。

語彙的な現象の修正 (a) とその他の現象の修正 (e) は不要なトークン列の削除や同じ意味のトークン列との置換など操作として類似している部分が多く、またその共起する割合も高い ((a) : 202/294 = 68.7%, (e) : 202/236 = 85.6%) ため、まとめて扱う。付与したラベルに (a), (e) のどちらか、または両方のみを含む場合を (x) とする。これとの対比として、省略の修正

手法	(b) を含む (58 件)	(b) を含まない (309 件)
J <sub>話</sub> -E	15.7	21.9
J <sub>書</sub> -E	16.1	22.8
LT <sub>FF</sub> -E	16.1	22.7
LT <sub>AR</sub> -E	16.6	22.6

表 9 省略の修正 (b) を含むかどうかによる BLEU スコアの違い

手法	(c) を含む (82 件)	(c) を含まない (285 件)
J <sub>話</sub> -E	19.1	21.2
J <sub>書</sub> -E	21.1	21.5
LT <sub>FF</sub> -E	20.3	21.7
LT <sub>AR</sub> -E	20.6	21.5

表 10 冗長な表現の修正 (c) を含むかどうかによる BLEU スコアの違い

手法	(d) を含む (14 件)	(d) を含まない (353 件)
J <sub>話</sub> -E	14.6	20.9
J <sub>書</sub> -E	19.1	21.5
LT <sub>FF</sub> -E	17.1	21.5
LT <sub>AR</sub> -E	18.6	21.5

表 11 文の概念がないことに関連する現象の修正 (d) を含むかどうかによる BLEU スコアの違い

(b), 冗長な表現の修正 (c), または文の概念がないことに関連する現象の修正 (d) のいずれかを含む場合を (y), 修正を含まず  $J_{\text{話}}$  と  $J_{\text{書}}$  が完全に同一の場合を (z) とし, これらと比較した結果を表 12 に示す. まず, (x) と (y) を比較すると BLEU スコアに 3.6~4.9 の差があることは, (b) と (d) の修正を含む場合は含まない場合と比べて BLEU スコアが小さいことと関連している. (x), (y) と比べて (z) の BLEU スコアが高いことは, 修正の必要がないほど元から整った文であったことを考えると自然である. (z) に関して, 入力と同じであるにも関わらず  $J_{\text{話-E}}$  のスコアが低くなっているが, これは訓練セット内でのノイズ (話し言葉特有の現象) がモデルの出力に影響を及ぼしている結果と考えられる.

その他の現象の修正 (e) の細分類ごとの BLEU スコアへの影響を表 13 に示す. まず,  $J_{\text{話-E}}$  と  $J_{\text{書-E}}$  を比較すると, 最も差が大きいのは意味的に冗長な箇所削除となっている. これは冗長な表現の箇所を無理に訳出しようとした結果として不要なトークンが英訳に現れるために BLEU スコアが低下していると考えられる. 次に,  $LT_{\text{FF-E}}$  と  $LT_{\text{AR-E}}$  を比較すると, 件数的に多い句読点位置の修正では  $LT_{\text{AR}}$  が優れているものの, 話し言葉的な語の変換や意味的に冗長な箇所削除, 文法上の間違いの訂正では  $LT_{\text{FF}}$  が優れている. これらの現象については平均の BLEU スコアが他の現象と比べ低いいため, 全体的な翻訳品質の底上げという観点から評価すると  $LT_{\text{FF}}$  のほうが  $LT_{\text{AR}}$  よりも優れていると考えられる.

このラベル付けしたデータに対する  $LT_{\text{FF}}$ ,  $LT_{\text{AR}}$  の話し言葉書き言葉変換の精度を表 14 か

手法	(x) : (a), (e) のどちらかまたは両方のみを含む (213 件)	(y) : (a) と (e) 以外の修正を含む (123 件)	(z) : 修正なし (31 件)
$J_{\text{話-E}}$	22.5	17.7	24.3
$J_{\text{書-E}}$	22.8	19.2	24.8
$LT_{\text{FF-E}}$	23.2	18.3	24.8
$LT_{\text{AR-E}}$	23.1	18.7	24.8

表 12 語彙的な現象の修正 (a), その他の現象の修正 (e) を含むかどうかによる BLEU スコアの違い

手法	全体 (236 件)	句読点位置の修正を含む (148 件)	文頭の「で」の除去を含む (99 件)	話し言葉的な語の変換を含む (68 件)	意味的に冗長な箇所削除を含む (23 件)	文法上の間違いの訂正を含む (8 件)
$J_{\text{話-E}}$	19.3	20.1	20.6	17.4	15.0	19.0
$J_{\text{書-E}}$	20.4	21.6	22.1	18.8	18.8	21.5
$LT_{\text{FF-E}}$	20.1	20.9	21.8	18.2	18.9	19.7
$LT_{\text{AR-E}}$	20.1	21.2	21.8	17.2	16.3	18.5

表 13 その他の現象の修正 (e) の細分類ごとの BLEU スコアの違い

手法	(b) を含む (58 件)	(b) を含まない (309 件)
LT <sub>FF</sub>	66.9	83.4
LT <sub>AR</sub>	69.6	83.0

表 14 省略の修正 (b) を含むかどうかによる SARI スコアの違い

手法	(c) を含む (82 件)	(c) を含まない (285 件)
LT <sub>FF</sub>	71.0	83.6
LT <sub>AR</sub>	71.3	83.6

表 15 冗長な表現の修正 (c) を含むかどうかによる SARI スコアの違い

手法	(d) を含む (14 件)	(d) を含まない (353 件)
LT <sub>FF</sub>	70.0	81.2
LT <sub>AR</sub>	73.2	81.2

表 16 文の概念がないことに関連する現象の修正 (d) を含むかどうかによる SARI スコアの違い

手法	(x) : (a), (e) の どちらかまたは両方 のみを含む (213 件)	(y) : (a) と (e) 以外の修正 を含む (123 件)	(z) : 修正なし (31 件)
LT <sub>FF</sub>	84.4	70.8	96.0
LT <sub>AR</sub>	84.0	71.6	96.3

表 17 語彙的な現象の修正 (a), その他の現象の修正 (e) を含むかどうかによる SARI スコアの違い

ら 17 に示す. 語彙的な現象 (a), その他の現象 (e) のようなパターンマッチに近い修正は LT<sub>FF</sub> が優れている一方, それ以外の修正では LT<sub>AR</sub> のほうが精度が高いことがわかる.

## 6 おわりに

本研究では日本語の話し言葉と書き言葉の違いに着目し, 両者の違いを表すデータセットを構築した上で, 話し言葉から書き言葉に自動で変換する手法を用いて日英翻訳の精度が向上することを示した. さらに, どのような違いが翻訳精度に影響するのかを定量化した.

今後の課題として, 本研究では取り扱わなかったが, 音声認識に誤りがある場合はそれがそのまま英訳されてしまう問題がある. 今回のように翻訳の前処理としてそれらを修正したり, 翻訳モデルを誤りに対してロバストにすることで対処するか, 音声から直接翻訳テキストを生成する手法によりこれを解決できる可能性がある.



## 謝 辞

本研究は、科研費#19K20343の助成を受けたものです。また、京都大学における講義翻訳プロジェクトを立ち上げられた大嶋 正裕 工学部長とシステムとコーパス構築にご協力いただいた株式会社アドバンスド・メディアに感謝申し上げます。

## 参考文献

- 秋田祐哉, 三村正人, 河原達也 (2010). 会議録作成支援のための国会審議の音声認識システム. 電子情報通信学会論文誌, **93-D** (9), pp. 1736–1744. [Y. Akita et al. (2010). An Automatic Transcription System for Creation of Meeting Records in the Japanese Congress. *Denshi Joho Tsushin Gakkai Ronbun-shi*, 93-D(9), pp. 1736–1744.].
- Jia, Y., Weiss, R. J., Biadys, F., Macherey, W., Johnson, M., Chen, Z., and Wu, Y. (2019). “Direct Speech-to-Speech Translation with a Sequence-to-Sequence Model.” In *Proceedings of the Interspeech 2019*, pp. 1123–1127.
- 国立国語研究所 (2006). 日本語話し言葉コーパスの構築法. 国立国語研究所. [Kokuritsu Kokugo Kenkyu-jo (2006). *Nihongo Hanashi-kotoba Kopasu No Kochiku-ho*. Kokuritsu Kokugo Kenkyu-jo.].
- Liu, Y., Xiong, H., Zhang, J., He, Z., Wu, H., Wang, H., and Zong, C. (2019). “End-to-End Speech Translation with Knowledge Distillation.” In *Proceedings of the Interspeech 2019*, pp. 1128–1132.
- 前川喜久雄 (2004). 『日本語話し言葉コーパス』の概要. 日本語科学, **15**, pp. 111–133. [K. Maekawa (2004). “Nihongo Hanashi-kotoba Kopasu” No Gaiyo. *Nihongo Kagaku*, 15, pp. 111–133.].
- Malmi, E., Krause, S., Rothe, S., Mirylenka, D., and Severyn, A. (2019). “Encode, Tag, Realize: High-Precision Text Editing.” In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 5054–5065, Hong Kong, China. Association for Computational Linguistics.
- Müller, M., Nguyen, T. S., Niehues, J., Cho, E., Krüger, B., Ha, T.-L., Kilgour, K., Sperber, M., Mediani, M., Stüker, S., and Waibel, A. (2016). “Lecture Translator - Speech Translation Framework for Simultaneous Lecture Translation.” In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pp. 82–86, San Diego, California. Association for Computational Linguistics.

- 尾嶋憲治, 河原達也, 秋田祐哉, 内元清貴 (2008). 話し言葉の整形作業における削除箇所  
の自動同定. 情報処理学会研究報告. NL, 自然言語処理研究会報告, **185**, pp. 85–91. [K. Osaki et al. (2008). Automatic Detection of Portions to be Deleted during Cleaning Transcripts. *Joho Shori Gakkai Kenkyu Hokoku. NL, Shizengengo Shori Kenkyukai Hokoku*, 185, pp. 85–91].
- Post, M. (2018). “A Call for Clarity in Reporting BLEU Scores.” In *Proceedings of the 3rd Conference on Machine Translation: Research Papers*, pp. 186–191, Belgium, Brussels. Association for Computational Linguistics.
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. (2020). “Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer.” *Journal of Machine Learning Research*, **21** (140), pp. 1–67.
- 坂本明子, 田中浩之 (2015). 話し言葉機械翻訳のための日本語前編集. 言語処理学会第 21 回  
年次大会発表論文集, pp. 4–11. [A. Sakamoto and H. Tanaka (2015). Hanashi-kotoba Kikai  
Honyaku No Tame No Nihongo Mae-henshu. Proceedings of the 21st Annual Meeting for the  
Association for Natural Language Processing, pp. 4–11.].
- Salesky, E., Burger, S., Niehues, J., and Waibel, A. (2018). “Towards Fluent Translations From  
Disfluent Speech.” In *Proceedings of the 2018 IEEE Spoken Language Technology Workshop  
(SLT)*, pp. 921–926.
- 柴田知秀, 河原大輔, 黒橋禎夫 (2019). BERT による日本語構文解析の精度向上. 言語処理学会  
第 25 回年次大会発表論文集, pp. 205–208. [T. Shibata et al. (2019). BERT Ni Yoru Nihongo  
Kobun Kaiseki No Seido Kojo. Proceedings of the 25st Annual Meeting for the Association  
for Natural Language Processing, pp. 205–208.].
- 島津明, 中野幹生, 堂坂浩二, 川森雅仁 (2014). 話し言葉対話の計算モデル (1 版) . コロナ社.  
[A. Shimatsu et al. (2014). Hanashi-kotoba Taiwa No Keisan Moderu. Korona-sha.].
- 須藤克仁, 林輝昭, 西村優汰, 中村哲 (2019). 授業アーカイブの翻訳字幕自動作成システムの試  
作. 情報処理学会研究報告. NL, 自然言語処理研究会報告, **2019** (15), pp. 1–4. [K. Suto et  
al. (2019). Jugyo Akaibu No Honyaku Jimaku Jido Seisei Sisutemu No Shisaku. *Joho Shori  
Gakkai Kenkyu Hokoku. NL, Shizengengo Shori Kenkyukai Hokoku*, 15, pp. 1–4.].
- 田中佑, 黒橋禎夫. BART 日本語 Pretrained モデル. <https://nlp.ist.i.kyoto-u.ac.jp/?BART>  
日本語 Pretrained モデル. [Y. Tanaka and S. Kurohashi (2020). BART Nihongo Pretrained  
Moderu.].
- Vaswani, A., Bengio, S., Brevdo, E., Chollet, F., Gomez, A., Gouws, S., Jones, L., Kaiser,  
L., Kalchbrenner, N., Parmar, N., Sepassi, R., Shazeer, N., and Uszkoreit, J. (2018). “Ten-  
sor2Tensor for Neural Machine Translation.” In *Proceedings of the 13th Conference of the As-  
sociation for Machine Translation in the Americas (Volume 1: Research Track)*, pp. 193–199,

Boston, MA. Association for Machine Translation in the Americas.

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). “Attention Is All You Need.” In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (Eds.), *Proceedings of the Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc.
- Wang, W., Tur, G., Zheng, J., and Ayan, N. F. (2010). “Automatic Disfluency Removal for Improving Spoken Language Translation.” In *Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 5214–5217.
- Xu, W., Napoles, C., Pavlick, E., Chen, Q., and Callison-Burch, C. (2016). “Optimizing Statistical Machine Translation for Text Simplification.” *Transactions of the Association for Computational Linguistics*, 4, pp. 401–415.

## 略歴

**中尾 亮太**：2018年京都大学工学部電気電子工学科卒業。現在、京都大学大学院情報学研究科修士課程に在籍。言語処理学会学生会員。

**Chenhui Chu**：2015年に京都大学大学院情報学研究科博士課程修了。博士（情報学）。2020年より同研究科特定准教授。機械翻訳、自然言語処理の研究に従事。言語処理学会、情報処理学会、人工知能学会、ACL、各会員。

**黒橋 禎夫**：1994年京都大学大学院工学研究科電気工学第二専攻博士課程修了。博士（工学）。2006年より京都大学大学院情報学研究科教授。自然言語処理、知識情報処理の研究に従事。2014年より日本学術会議連携会員。

(2021年3月1日 受付)

(2021年5月31日 再受付)

(2021年7月10日 採録)