

以 AI 人工智慧進行個性翻譯的可行性： 以小說為文本的日中翻譯為例

蔡佩青

台灣 淡江大學日本語文學系 副教授

魏世杰

台灣 淡江大學資訊管理學系 副教授

摘要

本研究主旨在測試 AI 是否能藉由只訓練單一譯者的譯文與其原文，便能如同該譯者所譯出的譯文般，擁有該譯者獨特個性的筆調。本研究使用筆者所譯之日文小說為測試材料，進行了兩次不同階段的實驗。第一階段實驗使用 PyTorch Chatbot 訓練中日文平行語料並以原文的日文做 AI 翻譯測試。其中在語料製作上，有鑒於日文與中文在造句結構上的差異而對日文語料做了許多加工與修整，但似乎反而妨礙了 AI 的機制與性能。加上語料的份量不足，在本階段實驗中並沒有得到預期的結果。但是掌握了類神經機器翻譯的特性，也有些許測試結果已顯示「個性翻譯」是可行的。第二階段實驗使用 PyTorch Chatbot 和 Tensor2Tensor 兩種機器翻譯模組來訓練更多的平行語料，並加入 BLEU 值以評價翻譯測試結果的優劣。從第二階段的實驗結果可知，使用 Tensor2Tensor 時，在短文測試能得到 BLEU 值為 1 的結果，也就是說 AI 譯出了與目標句相同的譯文。本研究得到兩個結論，一是實驗結果顯示 AI 可能仿效單一譯者的譯文，譯出與該譯者相近筆調的文章，二是確認了 Tensor2Tensor 在性能上比 PyTorch Chatbot 更適用於本研究。

關鍵詞： AI 人工智慧，類神經網路，機器翻譯，深度學習

受理日期：2020 年 3 月 10 日

通過日期：2020 年 5 月 1 日

Possibility of stylish translation using artificial intelligence: the case of novel translation from Japanese to Chinese

Tsai, Pei-Ching & Wei, Shih-Chieh

Associate Professors, Department of Japanese &
Information Management, Tamkang University

Abstract

The purpose of this study is to test whether modern artificial intelligence (AI) technologies can learn the unique style of a translator. That is, by training with the source text and its target translation from a single translator, can the system learn to translate like the specific translator? As the dataset, a Japanese novel is used, along with its Chinese translation by one of the authors. Two stages of experiments are conducted. First, a model from the PyTorch Chatbot package is trained on the Japanese-Chinese parallel dataset and tested for translation. During the data preparation, a lot of preprocessing is made to accommodate the difference in sentence structures between Japanese and Chinese. Due to the accommodation and lack of a sufficient dataset, expected results are not obtained but characteristics of neural machine translation (NMT) are learned. It is confirmed that stylish translation is possible by NMT. Second, the models from the PyTorch Chatbot and the Tensor2Tensor packages are trained on a more abundant parallel dataset and evaluated with the BLEU metric. The result shows that Tensor2Tensor performs better. For Tensor2Tensor on some short text tests, a BLEU value of 1 is even obtained which means the exact reproduction of the desired target translation. This study has two conclusions. It is confirmed that the AI technology in use has the potential to imitate a single translator with translations similar in style. It is also confirmed that Tensor2Tensor is more suitable than PyTorch Chatbot in performance for this study.

Keywords : Artificial Intelligence, Neural Network, Neural Machine Translation, Deep Learning

AI 人工知能による個性的翻訳の可能性

—小説の日中翻訳を例に—

蔡佩青

台湾 淡江大学日本語文学科 准教授

魏世杰

台湾 淡江大学情報経営学科 准教授

要旨

本研究の目的は、特定の翻訳者の手による訳文とその原作を AI に学習させることで、AI は当該翻訳者と同じようなスタイルの訳文を生成できるかどうかを明らかにすることである。機械翻訳の実験は筆者が訳した日本語の小説を材料にして 2 段階に分けて行った。第 1 段階では PyTorch Chatbot を以てテキストデータの機械学習を行ってから、データの日本語部分を使って翻訳テストを試みた。データセットを作成する際、日本語と中国語の文構造の相違を考慮して日本語のテキストに多くの修正を入れたが、却ってニューラル機械翻訳モデルの性能発揮を妨げたようである。学習データ量不足の問題も加わって、この実験では良い結果は得られなかった。が、PyTorch Chatbot の特性を把握でき、僅かながらも「スタイルリッシュトランスレーション個性的翻訳」の可能性が確認できた。第 2 段階では、PyTorch Chatbot と Tensor2Tensor を使用し、より語彙量の多いデータセットで機械学習を行い、そして翻訳テストの結果を BLEU 値で測定した。実験は、Tensor 2 Tensor を使用した短文の翻訳では、BLEU 値が 1、即ち参照訳と同様な訳文を得ることに成功した。以上により、PyTorch Chatbot に比べ、Tensor2Tensor はより本研究に適している翻訳モデルだと考え、AI が個性のある訳文を生成する可能性は否定できないと結論付ける。

キーワード：AI 人工知能、ニューラルネットワーク、ニューラル機械翻訳、ディープラーニング

AI 人工知能による個性的翻訳の可能性

—小説の日中翻訳を例に—

蔡佩青

淡江大学日本語文学科准教授

魏世杰

淡江大学情報経営学科准教授

1.はじめに

2016年、Google LCC は深層学習 (Deep Learning) に基づくニューラル機械翻訳 (Neural Machine Translation, 以下 NMT) を導入して¹から、Google 翻訳の品質が飛躍的に向上した。それは、これまでの、構文解析や辞書データの蓄積のもとで文章をパーツごとに訳していく「直訳的」翻訳から、コンテキストを把握することで多くの訳語候補を挙げながら訳していく「意識的」手法に変わったからである。Google 翻訳には世界人口の約 3 分の 1 が話す 8 言語²のデータが仕込まれており、スマートフォン一つで、話せない言語が使われている国への個人旅行も別段怖いものでもなくなろう。

一方、同じ翻訳というジャンルを持つ出版業界では、調査によると、2007 年から 2011 年までの 5 年間、翻訳出版を中心とした台湾出版社の中で図書の発行点数上位 5 社は、合計 1 萬点以上の翻訳図書を本屋に並べたといい³、個人で最も多くの外国語図書を中国語に訳した翻訳者は、年間平均 47 冊もの作品を出していたと分かった⁴。

¹ 「Google 翻訳が進化しました。」(2016年11月16日付)『Google Japan Blog』
<https://japan.googleblog.com/2016/11/google.html> (2019/02/12 閲覧)。

² 英語、フランス語、ドイツ語、スペイン語、ポルトガル語、中国語、日本語、韓国語、トルコ語の計 8 言語。

³ 陳子璋 (2012) 《臺灣翻譯產業調查研究 研究報告》国家教育院 (計畫編號 NAER-100-12-F-2-01-00-2-02) 「表 6.3 翻譯出版品 2007-2011 年度出版總件數前 10 大出版社」によると、1 位～5 位にランクインした出版社及び出版した翻訳作品数は次のとおりである。東立：3425 点、青文：1462 点、新視界文化：1434 点、台湾國際角川：1426 点、先端出版：1350 点。

⁴ 同注 3。「表 6.2 翻譯出版品 2007-2011 年度出版總件數前 20 大譯者」によると、ベスト 3 は全て翻訳専門会社名義で出版されたもので、個人名義の翻訳者の最

言ってみれば、機械翻訳即ち AI 翻訳が世界に浸透しながらも、人間の手による翻訳の需要は依然として高いままである。その理由として、AI が作成する訳文はまだ自然な文章とは言えないことや、人間による翻訳には無意識にでも表れてしまう「個性」というものがあることが考えられる。仮に AI が人間に取って代わることは時間の問題だとしても、はたして AI は人間のように個性的な訳文を作成できるか。

本研究は、この素朴な疑問から始まったものであるが、ますます進歩していく AI 翻訳の様々なアプローチへの挑戦でもある。具体的には、一人の翻訳者が作成した訳文とその原作を AI に学習させた後、AI は当該翻訳者と同じような文章スタイルで訳文を作られるかどうかを実験することである。本稿では、ディープラーニングの学習データとなる、日中対訳のデータセットを作成する際の検討事項を示しながら、NMT モデルである PyTorch Chatbot と Tensor2Tensor による AI 翻訳の実験結果を報告する。なお、タイトルに示した「スタイリッシュトランスレーション個性的翻訳」とは、AI は特定の翻訳者の「ライティングスタイル文 体」に似せて訳文を生成することを意味している。

2. 実験概要

実験は 2 段階を経て行った。第 1 段階では、PyTorch Chatbot を用いて、テキストデータの加工やトレーニングのパラメーターなど機械学習に影響を及ぼすだろうと思われる条件を調整しながら、日本語を繁体字中国語に翻訳するテストを行った。日中対訳のデータセットを用いる深層学習の効果を試すために、第 1 段階の実験はテキストデータの整形に重点を置いた。

第 2 段階では、第 1 段階の実験結果を踏まえながらデータセットの作成基準を再調整し、また PyTorch Chatbot の比較対象として

高位は 4 位にランクインした方郁仁氏である。主な翻訳作品は日本の漫画であるが、2007 年から 2011 年までの 5 年間計 236 点の作品が出された。

Tensor2Tensor を加えて実験を行った。そして自動評価尺度 BLEU で AI 翻訳の結果と参照訳との類似度を測定した。

2.1 テキストの選定

深層学習による機械学習の言語研究の多くは、コーパスが単一言語、もしくはフリーの SNS (Social Networking Service) アプリケーションやウェブサイトを通じて集めたビッグデータを用いての実験である。それに対して、本研究は、訳文には翻訳者の個性が表れることを前提にしているため、テキストに次の文学作品を選んだ。

- ・日本語テキスト：北村薫 (1999) 『スキップ』新潮文庫
- ・中国語テキスト：蔡佩青訳 (2008) 『SKIP--快轉』繆思文化

『スキップ』は北村薫が著した《時と人》三部作の一作である。当初の研究計画は三作の対訳データ作成して実験を行う予定だったが、Google Colaboratory を利用する機械学習は一日の稼働時間に制限があることに加え、データセットの作成には多くの時間を要するので、テキストの量を減らし、NMT モデルの有用性を確認することを優先にした。また、中国語テキストに関しては、データセットを作成する際に気づいた誤訳や不自然な訳を私意によって修正したことを断っておく⁵。

2.2 ニューラル機械翻訳モデルについて

第 1 段階の実験で使用する PyTorch Chatbot (以下 PTC) は、注意機構 (Attention Mechanism) が加わる sequence to sequence (以下 seq2seq) を、深層学習フレームワーク PyTorch で実装するものである⁶。再帰型ニューラルネットワーク (Recurrent Neural Network, RNN)

⁵ そうすることで、テキストを忠実に引用していないのではないかという疑念を抱かれるかもしれないが、訳文の修正は誤訳の訂正にとどまっており、また引用テキストと修正の訳文はいずれも同一翻訳者の手によるものであるため、本稿の趣旨なる「個性的翻訳」から外れていないと考える。文末に付録として掲げたデータセットの例を参照されたい。

⁶ STPI (國家實驗研究院科技政策研究與資訊中心) が主催した 2017Formosa Grand Challenge-Talk to AI (2017 台灣科技大擂台【與 AI 對話】熱身賽) で公開されたオープンソースで、パッケージは <https://github.com/ywk991112/pytorch-chatbot> によった。

で構築したオートエンコーダー (Auto Encoder) 及びオートデコーダー (Auto Decoder) を通して、ソースセンテンスとターゲットセンテンスとの対応関係を学習させる。その仕組みのイメージは図 1 のようである。

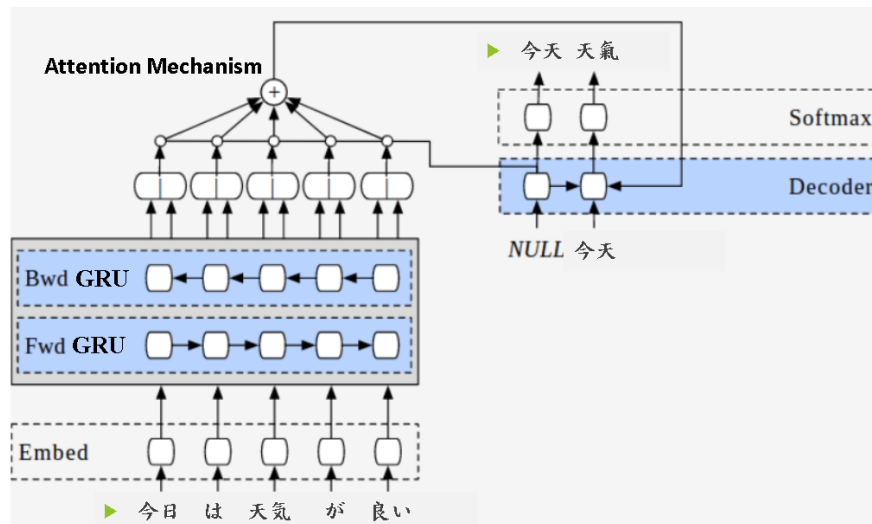


図 1 : PTC に実装した RNN の仕組み (adapted from Thanaki, 2017⁷)

第 2 段階の実験では、PTC の比較対象として Tensor 2 Tensor⁸ (以下 T2T) を加えた。T2T の入力層 (Input Layer) には、PTC にない、単語の位置情報を与える位置エンコーディング (Positional Encoding) が仕組みられている。また、隠れ層 (Hidden Layer) において、PTC ではゲート付き回帰型ユニット (Gated recurrent unit, GRU) が使われているのに対して、Transformer Block が使用されている。そのうえ、マルチヘッドの注意機構 (Multi-Head Attention) が用いられることで、次にアウトプットする単語を予測する際に、ソースセンテンスから関連性の高いと思われる単語を多く見つけ出すことが可能になる。そのメカニズムは図 2 のようである。

⁷ Jalaj Thanaki (2017) Python Natural Language Processing: Advanced machine learning and deep learning techniques for natural language processing, Packt Publishing.

⁸ Google Brain とユーザーコミュニティの研究者によって共同開発された、深層学習モデルとデータセットのライブラリである。オープンソースは <https://github.com/tensorflow/tensor2tensor> によった。

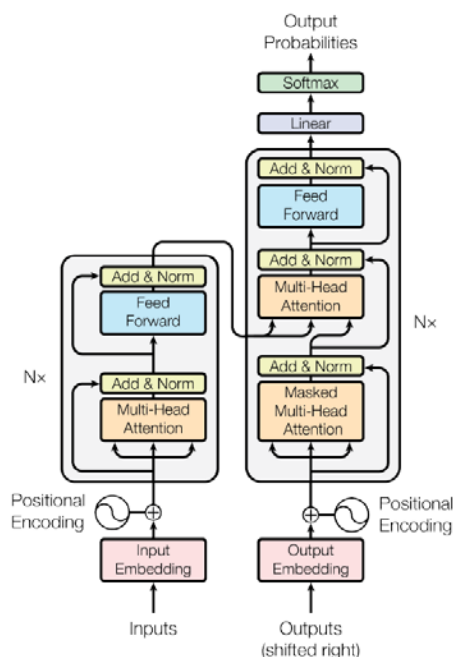


図 2: T2T に実装した Transformer の仕組み (Vaswani et al., 2018⁹)

3.第 1 段階の実験

第 1 段階の実験は 3 回にわたって行った。1 回目は、句読点を含む文単位で作成した日本語と中国語のテキストデータ¹⁰を、それぞれ形態素解析エンジン MeCab¹¹と jieba¹²にかけて分かち書きにしたうえで、PTC に読み込んで機械学習を行った。その後、学習済みのデータセットの日本語部分を使って機械翻訳のテストを試みた。しかし、結果は正確な訳文を生成することはできなかった。その理由の一つとして、文を分かち書きにした後のトークン数 (Number of Tokens) の受け入れをコントロールするパラメーター (初期設定は

⁹ Ashish Vaswani et al. (2018) Tensor2Tensor for Neural Machine Translation, arXiv:1803.07416, p2.

¹⁰ 日本語はテキスト北村薫 (1999) 『スキップ』 pp.145-174 から、中国語はテキスト蔡佩青訳 (2008) 『SKIP--快轉』 pp.112-135 からデータを抽出した。

¹¹ MeCab (和布蕪) は京都大学情報学研究科が日本電信電話株式会社コミュニケーション科学基礎研究所との共同研究ユニットプロジェクトを通じて開発した形態素解析エンジンである。詳細は公式 HP, <http://taku910.github.io/mecab/> を参照されたい。

¹² jieba-zh_TW (結巴斷詞台灣繁體版) は、簡体字中国語自然言語処理パッケージ jieba をベースに設計された繁体字中国語テキストを形態素解析するツールである。パッケージは https://github.com/ldkrsi/jieba-zh_TW によった。

20) が存在しており，テキストデータの一部は学習対象から外されてしまったことが挙げられる。しかし，機械学習の所要時間の増加を避けたいため，パラメーターの数値を上げるのではなく，長い文を短くする方法を取ってみた。

3.1 テキストデータの修正

2 回目のテストでは，文を 20 以下のトークンに分割するが，対訳データを増やすために，もともとトークン数が 20 を超えていない文も，さらに短く分割することがある。その際，文の構造を崩さないように，原文における単語レベルの会話文を除き，単語のみの対訳は作らなかった。それで 818 セット（日本語と中国語の対訳を 1 セットとする）の対訳データを作成した。但し，両言語の語順や言語習慣の違いによって，対訳に齟齬をきたす場合は適宜中国語訳を改めた。

また，jieba が認識できない日本人の氏名について，例えば「真理子」を「真理」と「子」に分割した場合，辞書ファイルに修正を入れた。最後に，文の意味を捉えるのに関わる疑問符を除いて，句読点や鉤括弧・リーダーなどの符号も削除した。

以上のようにテキストデータを整形して，再度翻訳テストを試みた。今回は機械学習に使ったテキストデータの文だけではなく，学習したことのある単語で任意に作った文（以下テスト文）も試してみたが，やはり正確な訳文を生成することに失敗した。それは，PTC にとって，判別可能な言葉が異なるソースセンテンスに散在し，テスト文における言葉と言葉の接続関係を学習できていないためではないかと推測する。

そこで，PTC の，テキストデータに対する判別精度を如何に高めるかを考えるとき，日本語の言語の特質に目を向けた。MeCab で日本語テキストを形態素解析すると，助詞・助動詞・補助動詞など，言葉と言葉の関係を示したり言葉に意味を付加したりするための付属語まで，文が細かく分割されるので，それらの文中における意味を

示すことが困難になる。例えば、例 1J に示したとおりに、「知っていた」は「知っ」「て」「い」「た」のように切られている。「知っ」あるいは「て」だけではその文中における意味が理解しがたく、例 1C の中国語訳から該当の訳語を取り出すのに難しい。

例 1J

わたし__が__こちら__の__世界__で__、__どう__い__う__姓__を__持__
っ__て__い__る__か__は__知__っ__て__い__た__。

例 1C

我__知__道__我__在__這__個__世__界__裡__，__我__有__著__怎__麼__樣__
的__姓__氏__。

* 下線は分かち書きを示すものである。以下同じ（例 2～例 4）。

そこで、3 回目のテストで日本語のテキストデータにさらなる修正を加えた。まず、文中における言葉の意味や働きを明確にするために、用言とそれに接続している助動詞や接続助詞を一つのまとまりにする（例 2J）。次に、短文をより短いフレーズに分割する場合、フレーズとして不要な助詞を削除した（例 3J）。また、PTC の学習量を上げるために単語の対訳も作成した。最後に、形態素解析では日本語と中国語の判別基準が異なる複合語や定型表現に関しては、適宜修正を入れた（例 4J，例 4C）。

例 2J

修正前：持__っ__て__い__る__か__は__知__っ__て__い__た

修正後：持__っ__て__い__る__か__は__知__っ__て__い__た

例 3J

修正前：どう__い__う__姓__を__持__っ__て__い__る__か__は__知__っ__て__
__い__た

修正後：どう__い__う__姓__／持__っ__て__い__る__／知__っ__て__い__た

例 4J

修正前：カミナリ__族__／お__医__者__さん__／紛__れ__も__ない

修正後：カミナリ族／お医者さん／紛れもない

例 4C

修正前：閃電__族／分__毫__不差__地

修正後：閃電族／分毫不差地

3.2 実験結果

3.1 で 1 回目と 2 回目のテスト結果に触れながらテキストデータ修正の基準と経緯について述べた。3 回目のテストは、延べ 1966 セットの対訳データを PTC に学習（トレーニング回数は 4800 回）させた後、2 回目と同様なテスト文で翻訳テストを行ったが、PTC はまたしても正確な訳文を作ることはできなかった。しかし、2 回目と 3 回目のテスト結果を並べて見ると、相異なる訳文が生成されていることが分かる（図 3, 図 4）。また、面白いことに、(A') と (B') は文末の「から」と「ある」から訳し始めるという訳し方をしているのである。これは人間が日本語を中国語に訳す時の一般的な思考パターンに似ているように思われる。

(A)	(B)
source: 'そこで、言葉に詰まったのは、何と呼んでいいか分からなかったから。' filter: 'そこで、言葉に詰まったのは、呼んでいいか分からなかったから。' '我知道我擁有怎樣的姓氏' '我曾經很感動' '我知道我擁有怎樣的畫面' '我知道我擁有怎樣的話' '我知道我擁有怎樣的少'	source: '「見たことある」' filter: '見たことある' '也有義務架設有線電話' '也難為情得叫人生氣' '也難為情得叫人' '也難為情得叫' '也難為情得叫人感到'

図 3：2 回目の翻訳テストの例

(A')	(B')
source: 'そこで言葉に詰まったのは何と呼んでいいか分からなかったから' filter: 'そこで言葉に詰まったのは何と呼んでいいか分からなかったから' '因為不能在教職員室裡' '因為不能在教職員室裡商量晚餐要吃什麼吧？' '因為不能在教職員室裡商量晚餐要吃什麼吧' '因為不能在教職員室裡商量晚餐要吃什麼' '如果一次有太多事進到腦袋裡來'	source: '見たことある' filter: '見たことある' '有' '有種所在之處' '有疑問' '有種所在之' '公告'

図 4：3 回目の翻訳テストの例

* source は翻訳テスト用のソースセンテンス、filter は学習しなかった言葉が source 文から排除された後の文である。また、中国語の訳文は順位付けされている。

seq2seq に基づく機械翻訳は、ある言葉に接続できると判断される候補の言葉をいくつか挙げながら訳していくという仕組みである。絵に描くと樹形図になるが、ここで注目したいのは、PTC は文を「作る」のではなく、既存の訳語から言葉を「選ぶ」ことである。ところが、それは無造作な選び方でもないようだ。図 3 の (A) はその顕著な例である。候補に挙げられた訳文を見ると、「怎麼樣の」の後ろにそれぞれ「姓氏／畫面／話／小」が接続されており、異なる訳語は選ばれているが、単語間の接続としては成立可能であることが理解される。つまり、AI 翻訳は文法的概念に基づく「作文」ではなく、学習した言葉から「選択」によって文を生成しているのである。

ということは、PTC に決まっている表現による中国語のテキストデータのみ与えれば、生成する訳文の表現も制限されるはずである。そうすると、本研究の目的である「個性的翻訳」の実現も不可能ではなかろう。ただ、原段階のようなやり方では、テキストデータの整形に莫大な時間を要するため、機械学習に適していると言えるほどのデータ量の作成は極めて困難だという課題が残ってしまう。

4. 第 2 段階の実験

第 1 段階の実験では、日本語と中国語の文構造の相違に注目してテキストデータの整形に重点を置いた。ところが、機械翻訳を実用化するのに対訳データの作成に手間がかかると非現実的であり、一方で機械翻訳モデルの特性をより活かすべきだと考え、テキストデータの人工修正を最小限にし、データ量を増やすことを最優先課題にした。また、文における単語間の接続も翻訳精度を左右していると確認したので、第 2 段階の実験ではテキストデータを加工する時に文を短く分割しすぎないように心がけた。

4.1 データセットの作成

テキストは第 1 段階の実験と同じ小説『スキップ』を使用するが、分量は原作のページ数で言うと前段階の 12 倍弱の 345 ページとな

る¹³。日本語・中国語の順に 1 行ずつ並べて対訳データを作っていくが、句読点や記号が必ずしも対応しているわけではないので、原則として記号と文末に来る句読点を削除する。但し、文意が不明瞭にならないように、文中の読点や三点リーダーや感嘆符などは原文のままにする。また、日本語の会話文によく見られる疑問を示す終助詞「か」の省略については、原文に疑問符「？」がある場合は残すこととする。「か」も「？」も省略されている場合は、疑問文であることが分かるように、中国語訳では疑問符を捨てずに取っておく。そして、1 文を 50 字以内（読点・記号を含む）に切ってデータセットを作成した。両言語あわせて 8356 文となり、約 21 万字ある。

データセットの作成例を次に掲げる。

(a) 会話文を示す括弧を取り除く例

日：池ちゃん、脈絡ないよお
中：小池，妳的話毫無脈絡可循欸

(b) 「か／？」はないが中国語訳では疑問符を補う例

日：何が、あぶない
中：什麼東西很危險？

(c) 感嘆符を保留する例

日：演劇部に入れ！
中：來參加話劇社！

(d) 長文を短文に区切る例

日：貼られた写真の、小さな長方形の窓から、こちらを見つめていたのは
中：從貼著照片的小長方形窗戶那頭，凝視著我的是
日：昨日、美也子さんに視線を投げていた、あの男の子だった
中：昨天，對美也子投以視線的，那個男孩子

¹³ 日本語はテキスト北村薫（1999）『スキップ』pp.5-350 から、中国語はテキスト蔡佩青訳（2008）『SKIP--快轉』pp.5-269 からデータを抽出した。内容としてはそれぞれのプロローグから第 7 章までである。

4.2 機械学習の結果

実験は PTC と T2T を使用したが，それぞれは字句解析プログラムの SentencePiece と SubwordTextEncoder で文の分かち書き作業を行った¹⁴。機械学習のトレーニング回数と翻訳テストの BLEU スコアは次の表 1 に示すとおりである。

モデル	PTC		T2T
トレーニング回数	num_hidden_layers: 1	6000	250,000 (num_hidden_layers: 6)
	num_hidden_layers: 2	6000	
	num_hidden_layers: 3	6000	
BLEU	num_hidden_layers: 1	0.855	0.911
	num_hidden_layers: 2	0.85	
	num_hidden_layers: 3	0.84	

表 1 : PTC と T2T による翻訳テストの結果

PTCの場合，レイヤー数（num_hidden_layers）¹⁵1～3でそれぞれ6千回のトレーニングを行った。その後，ソースセンテンス即ちテキストデータの日本語の部分をテスト文にしてPTCに翻訳させた結果，BLEUはいずれも0.85前後にとどまった。それに対して，T2Tによるテストでは，25万回のトレーニングを行った後，PTCと同様な翻訳テストをして得たBLEUは0.91までに上った。

BLEU は 0～1 の実数で表われ，数値が 1 に近づけば近づくほど良好な訳文であると判断される。数値が 1 になることは，機械翻訳による訳文は参照訳と全く同じであることを意味している。実験に使用した PTC と T2T はモデルが異なるために単純比較はできないが，BLEU を見る限り，PTC に比べ T2T はより本研究に適していると言えよう。

ところが，翻訳テストでは，トレーニングに使用したデータセッ

¹⁴ subword 数を PTC は 31，T2T は max_length の 256 に設定した。

¹⁵ seq2seq の基本構造について，Denny Britz , Anna Goldie, Minh-Thang Luong, Quoc V. Le (2017) Massive Exploration of Neural Machine Translation Architectures, Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp.1142-1451 を参照されたい。

トと同様な日本語のテキストデータを用いたのにも関わらず、BLEU は 1 に達していない。その理由として挙げられるのは、図 5 と図 6 に示すように、句読点の全角と半角の違いや、一つのソースセンテンスには二つ以上のターゲットセンテンスが対照されていることである。それ以外に、まるで幼児が初めて言葉を口にしたような、トンチンカンな訳文も見られる。しかし、それはむしろ、AI は与えられた正解をそのまま提出しているのではなく、自ら訳文を書こうとしていることを物語っているのではないか。

```
BLEU:0.858。
S:関係ないけど白い巨塔も終わるんだよ。
T:雖然不相關，不過想想白色巨塔也快演完了。
A:雖然不相關，不過想想白色巨塔也快演完了。
。
BLEU:0.622。
S:母の具合でも悪くなったのか、と心配になる。
T:真擔心是不是母親身體惡化了。
A:有擔心是不是母親惡化了。
。
BLEU:0.000。
S:しずくが垂れるので、廊下でスカート脱いだ。
T:因為滴著水，所以我在走廊上把裙子脫下。
A:因為我 因為屋子站在那裡，走廊上了裙子。
。
BLEU:0.000。
S:なあに。
T:什麼啦？
A:幹嘛？
```

図 5 : PTC (BLEU < 1 の例)

```
BLEU:0.000。
S:明日、フォークダンス、出来るかなあ。
T:明天，能跳土風舞嗎？
A:明天，公布在我就會燈前的事物上清晨。
。
BLEU:0.370。
S:自分が主任だったら、どうするんです。
T:如果我是主任，怎麼辦？
A:如果自己已经是主任，該怎麼辦？
。
BLEU:0.000。
S:あのね。
T:那個啊。
A:喂！
。
BLEU:0.000。
S:そうか。
T:原來如此。
A:對噢。
```

図 6 : T2T (BLEU < 1 の例)

* S はソースセンテンス，T はターゲットセンテンスの参照訳，A は翻訳テストの結果を指している。

4.3 翻訳テスト

また、機械学習を行っていないテキストデータによる翻訳テストも行った。用意したテスト文は次の(1)～(3)から抽出したものである。参照訳はいずれも筆者の手によった。

- (1) 実験用学習データと同じ小説¹⁶
- (2) 実験用学習データと同一作者による同シリーズの小説¹⁷
- (3) 実験用学習データと同一作者によるその他の小説¹⁸

各テキストから一定量の文を抽出して翻訳テストを行ったが、紙幅の都合上、その一部のみ次の (t1) ～ (t3) で示す¹⁹。

¹⁶ 北村薫 (1999) 『スキップ』新潮文庫。

¹⁷ 北村薫 (2003) 『リセット』新潮文庫。

¹⁸ 北村薫 (1999) 『六の宮の姫君』創元社推理文庫。

¹⁹ PTC による翻訳テストは num_hidden_layer: 3 のトレーニング結果を用いて行った。

- (t1) S 途中から、わたしは速足になり、そして走った
T 中途，我加快腳步，然後用跑的

PTC

topn	response	bleu
1	途中,我又停步他走來說,他	0
2	途中,我又夠雙腳步	0
3	途中,我又停下腳步,被住了	0
4	途中,我又停步他走來說	0
5	途中,我又快步住了	0
6	途中,我又停下腳步	0
7	途中,我又快步住他,走了一步	0
8	途中,我又夠雙腳步過來	0
9	途中,我又快步住他	0
10	途中,我停下說步	0

T2T

topn	response	bleu
1	從途中,我停下腳步	0
2	途中,因為我停下腳步	0
3	以途中,我停下腳步	0
4	但因為途中,我停下腳步	0
5	途中,因為我停下腳步,用跑的	0
6	過了中途中,我立足在跑步地跑的	0
7	我停下腳步,在途中,然後跑的	0
8	過了中途中,我立足在跑步地跑步地跑步	0
9	途中,因為我停下腳步,跑步地跑步	0
10	途中,因為我停下腳步,做跑步地	0

- (t2) S 本棚を見ているうちに、その前にあるごみ箱が目に入った
T 看著書架時，書架前的垃圾桶進入眼簾

PTC

topn	response	bleu
1	我對著書架在書之間,朝面的盒子裡	0
2	我對著書架在書之間,朝面的盒子	0
3	我並在書架看	0
4	我對著書架在書之間	0
5	我在書架時,朝在我面前的盒子裡	0
6	我在書架時,朝向那前	0
7	我在書架時,朝在我面前	0
8	我在書架時,朝在我面前的盒子	0
9	我在書架時,朝在我面前的一位盒子	0
10	我在書架時,朝向那前的盒子裡	0

T2T

topn	response	bleu
1	放眼前隨即出現了,有幾個書架	0
2	放眼前隨處理,而且我眼前	0
3	放眼前隨處理,可以在書桌前	0
4	找書架,有幾個箱子在裡我眼前	0
5	找書架,我很認真地進了那本	0
6	放眼前隨即出現了,在書架前	0
7	找書架,有幾個箱子在裡面前	0
8	找書架,我很認真地進了那本子	0
9	找書架,有幾個箱子在裡不看到印的	0
10	放眼前隨處理,有幾個箱子在我眼前	0

- (t3) S 家に帰ってから、自転車に乗って、またやって来た
T 回家之後，再騎腳踏車過來

PTC

topn	response	bleu
1	回到家	0
2	回到家,過了腳踏車就放了一會兒,又是過來	0
3	回到家,過了腳踏車就放了一會兒	0
4	回到家,過了腳踏車騎,被腳踏車過來,又上了共過來	0.17
5	回到家,過了腳踏車就放了一會兒,又是騎過來	0
6	回到家,過了腳踏車騎,被腳踏車過來,又上了共	0.18
7	回到家,母親站著腳踏車去,過來騎過來	0
8	回到家,過了腳踏車就放了一會兒,又是坐見	0
9	回到家,過了腳踏車就放了	0
10	回到家,過了腳踏車騎,是騎腳踏車	0.18

T2T

topn	response	bleu
1	但因為回到家,所以變暗	0
2	但因為回到家,所以放了車站腳踏車	0
3	但因為回到家,所以變得難車來了	0
4	但因為回到家,所以放了車子,下腳踏車	0
5	但因為回到家,所以繞了車來了	0
6	但因為回到家,所以搭車去	0
7	但因為回到家,所以搭車站了天回家	0
8	但因為回到家,所以排了車子,下腳踏車	0
9	但因為回到家,所以繞了車來車回來	0
10	但因為回到家,所以問了車站,下腳踏車	0

*Sはソースセンテンスのテスト文，Tはターゲットセンテンスの参照訳，topnは訳文候補の順位，responseは翻訳テスト後の訳文を指している。

テスト文は構文が複雑すぎず固有名詞など特定の語彙が含まれていないのを意図的に選択したとはいえ、ランダムに抽出した文であ

る。また、今回の翻訳テストに限って見ると、PTCとT2Tはテスト文のすべての単語を認識していることから、データセットの語彙量が不足していないと言える。ところが、結果は上記 (t1) ~ (t3) の bleu 欄に示しているように、殆どの訳文はゼロもしくはゼロに近い数値しか得られなかった。response 欄の中国語を見ても、翻訳はできていないことが分かる。しかし、topn1~10 に挙げられた訳文候補を逐次見ていくと、機械が深層学習を通して覚えた言葉を一つずつ翻訳（選択）しながら文を作ろう（単語を並べよう）としていることが明らかである。そこで、より短い文で試してみた。結果は (t1') と (t2') のとおりに、PTCもT2Tもターゲットセンテンスの参照訳と全く同様な、もしくはかなり近い訳文を見事に作ってくれた。

(t1') S わたしは、じっと池ちゃんを見た

T 我直盯著小池

PTC

topn	response	bleu
1	我,他往前看著	0
2	我,他往前	0
3	我,他往前看著我	0
4	我,他往前看著小池	0
5	我,一直往美也子的	0
6	我,一直往美也子	0
7	我,一瞬間,我低了	0
8	我,直往一看	0
9	我,一瞬間,我低了頭	0
10	我,一瞬間,我低了一頭	0

T2T

topn	response	bleu
1	我凝視著我	0
2	我盯著那邊	0
3	我盯著小池	0.58
4	我盯著那頭	0
5	我直盯著小池	1.00
6	我凝視著小池	0
7	我曠地看著小池	0
8	我盯著小池看著	0.43
9	我盯著小池望向	0.43
10	我凝視著那陣子	0

(t2') S でも、今の私は一人です

T 但是，現在的我是獨自一人

PTC

topn	response	bleu
1	但是,現在的我,一個人是一個人的	0.20
2	但是,現在的我,一個人是一個人的一個人	0.17
3	但是,現在,我剛一個人是一個人的	0
4	但是,現在,現在的我是一個人	0.39
5	但是,現在,我剛一個人是一個人的一個人	0
6	但是,現在的我負責一個人	0.27
7	但是,現在,我剛一個人	0
8	但是,現在的我是一個人的一個人	0.30
9	但是,現在的我的	0.26
10	但現在的我是一個人	0.36

T2T

topn	response	bleu
1	但是, 現在的,我, 是一個人	0.56
2	但是, 現在的,我, 是要一個人	0.51
3	但是, 現在的,我, 是一個人	0.56
4	但是, 現在的,我, 是只我一人	0.47
5	但是, 現在的,我, 是意	0.54
6	但是, 現在的,我, 是別人	0.55
7	但是, 現在的,我, 是就一個人	0.51
8	但是, 現在的,我, 是有一個人	0.51
9	但是, 現在的,我, 是有一個人	0.51
10	但是, 現在的,我, 是只我一人	0.53

5. 考察

第1段階の実験において、翻訳精度が低かった原因は日本語と中国語の構文上の相違にあると考えたため、日本語のテキストデータ

にかなりの人工修正を入れた。一方、第2段階の実験では、文を適当な長さに分割したことにとどまって、テキストデータの加工は殆どしなかった。結果は、第1段階の実験よりも、より良いと思われる訳文の生成ができたことから、翻訳精度を左右するのは文の長さや形ではないと考えられる。ただ、機械学習はテキストデータを字句解析した後に実行するので、字句解析プログラムによって文の分割の仕方も異なる。実験の第1段階では MeCab と jieba、第2段階では SentencePiece と SubwordTextEncoder を使用した。MeCab と jieba はそれぞれ日本語と中国語のためにデザインされ、対応する辞書に従って文を解析する形態素解析器である。それに対して、多言語対応の SentencePiece と SubwordTextEncoder は、高い頻度で現れる単語やフレーズを一つの単位として認識し、出現頻度の低い字句をよく出現する単位まで細かく分割するという特徴がある。実験時の条件設定は異なるものの、今回使用したデータセットのような、一冊の小説から抽出したテキストデータには、相似する表現や関連性の高い言い回しが多く含まれているので、単一言語専用の字句解析プログラムよりも SentencePiece や SubwordTextEncoder のほうが適していると考えられる。

そして、第2段階の実験において、T2T は、PTC より高い翻訳精度を出した。学習データの量が決して十分とは言えない厳しい条件制限の中で、PTC の 6 千回より遥かに超えた 25 万回のトレーニング回数は翻訳精度に一定の影響を与えたと推測される。

最後に、BLEU による評価について触れておく。T2T の翻訳テストにおいて、学習しなかったテキストデータを使ってテストして得た BLEU の平均値は 0.2140 である。同じく日中対訳の機械翻訳に関する実験には、データセット・字句解析・パラメーターなどの条件は異なるが、これに近い BLEU 値を得ている報告がある²⁰。した

²⁰ Longtu Zhang, Mamoru Komachi (2019) Chinese-Japanese Unsupervised Neural Machine Translation Using Sub-character Level Information. arXiv:1903.00149。日本語と簡体字中国語との双方向翻訳の実験ノートだが、日本語から中国語への

がって、本研究における T2T の実験結果は客観的な指標となりうると言える。ただ、「BLEU の特徴として、局所的に流暢な文、参照文に表現法やスタイルが一致する文などに高い評価値を与えることが挙げられ」るが、意味的妥当性との相関が低く、語順の誤りに対してそれほど敏感ではないことが指摘されている²¹。Neubig (2013) によれば、文レベルの機械翻訳において人間による評価が自動評価尺度よりの確だという²²。確かに、たとえ評価基準としているのは人間の手による訳文であっても、BLEU は人間が意図している評価の幅をどれほど付度できるか、という疑問は大いに残る。例えば(t2')では、PTC が訳した「但是, 現代的我是一個人」よりも、T2T が訳した「但是, 現代的我, 是一個人」のほうが高い BLEU 値を得ている²³が、大方の中国語話者の考えでは読点の一つ多いと思うのではなからうか。

6. おわりに

AI が小説を書く日も東大に入る日²⁴もまだまだ遠いようである。それは、AI がゼロから物語を構想して作文することができなく、大学入試センター試験の国語問題がうまく解けなかったからだ。しかし、翻訳は AI にとってはどちらかという得意な科目のようだ。同じコーパスを利用した自然言語処理でも、ある程度の量の多言語コーパスを AI に暗記させれば、対訳だけではなく、異なる言語の会話も成立する（というような勘違いは我々がしてしまう）。AI のその

翻訳テストでは 19.72~23.4(BLEU 値×100)という実験結果が報告されている。

²¹ Graham Neubig (2013)「文レベルの機械翻訳評価尺度に関する調査」情報処理学会研究報告・自然言語処理研究会報告, pp.2-3。

²² ここで言う自動評価尺度は BLEU, WER, TER, RIBES, METEOR を指している。同注 21。

²³ その原因の一つとして、参照訳の中国語では全角の記号が使われていることが挙げられる。

²⁴ AI に小説家や受験生にさせるプロジェクトは、公立はこだて未来大学の松原仁氏が中心とした「きまぐれ人工知能プロジェクト 作家ですよ」と、国立情報学研究所が中心とした「ロボットは東大に入れるか」プロジェクトが有名である。

特技を利用して、一人の作家の作品と一人の翻訳者による訳文のみ
与え、AIはその翻訳者の文章にそっくりな訳文を生成することを期
待して、本研究を始めた。幾度の試行錯誤の結果、(t1')と(t2')
で示したとおりに、学習したことのないテスト文でも、AIは確かに
特定翻訳者の^{スタイル}文体を模倣して翻訳していた。

これまでの実験及び上述の結果分析により、AIによる「^{スタイリッシュ}個性的な
翻訳」の可能性があると考えたい。無論、真の個性のある翻訳と言
えるまでには、テクニカルな面でもプラクティカルな面でも解決せ
ねばならぬ問題点が山積している。今後、学習データの量を増やす
こともさることながら、異なる機械翻訳モデルによる実験も重ねて
研究を進めていく予定である。

テキスト

北村薫（1999）『スキップ』新潮文庫

北村薫（1999）『六の宮の姫君』創元社推理文庫

北村薫（2003）『リセット』新潮文庫

北村薫著、蔡佩青譯（2008）《SKIP--快轉》繆思文化

参考文献

Graham Neubig（2013）「文レベルの機械翻訳評価尺度に関する調
査」情報処理学会研究報告・自然言語処理研究会報告

国立国語研究所

https://pj.ninjal.ac.jp/corpus_center/bccwj/morphology.html

Google Japan Blog（2016/11/16 付）

<https://japan.googleblog.com/2016/11/google.html>（2019/02/12 閲
覧）

科技大擂台「與 AI 對話」<https://fgc.stpi.narl.org.tw/activity/techai>

陳子瑋（2012）臺灣翻譯產業調查研究報告，國家教育院

Ashish Vaswani, Samy Bengio, Eugene Brevdo, Francois Chollet, Aidan

- N.Gomez, Stephan Gouws, Llion Jones, Łukasz Kaiser, Nal Kalchbrenner, Niki Parmar, Ryan Sepassi, Noam Shazeer, Jakob Uszkoreit (2018) Tensor2Tensor for Neural Machine Translation. arXiv:1803.07416
- Denny Britz, Anna Goldie, Minh-Thang Luong, Quoc V. Le (2017) Massive Exploration of Neural Machine Translation Architectures. EMNLP 2017, pp.1442-1451. arXiv:1703.03906v2
- Jalaj Thanaki (2017) Python Natural Language Processing: Advanced machine learning and deep learning techniques for natural language processing, Packt Publishing
- Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu (2002) BLEU: a Method for Automatic Evaluation of Machine Translation. ACL 2002, pp.311-318
- Minh-Thang Luong, Hieu Pham, Christopher D. Manning (2015) Effective Approaches to Attention-based Neural Machine Translation. EMNLP 2015, pp.1412-1421. arXiv:1508.04025v5
- Longtu Zhang, Mamoru Komachi (2019) Chinese-Japanese Unsupervised Neural Machine Translation Using Sub-character Level Information. 33rd Pacific Asia Conference on Language, Information and Computation, pp.309-315. arXiv:1903.00149

付録

引用テキスト及び機械学習用データセットを例として一部掲載する。

(A) 日本語テキスト pp.148-149

「はい」

相手は少年か青年。とにかく、若い男だった。

「先生？」

桜木さんに用のようだ。

「桜木はただいま、出ております」

まるで奥さんのように、そう答えるのも、くすぐったく腹だたしい。

「……」

とまどったような沈黙。

「どちら様でしょうか」

「真理子先生でしょ？」

「は？」

「桜木真理子先生でしょ？」

う、と返事に詰まった。わたしがこちらの世界で、どういう姓を持っているかは知っていた。だが、その《桜木真理子》が、どういう人物なのか、——奥さんであり、母である他にどんな役割を果たしているのかは知らない。聞いていない。

言葉に窮している内についた息が、相手には曖昧ながらも肯定の返事と受け取れたようだ。溺れる人間も、川は遠慮なく流してしまうものだ。もがいている、そのままの形で。

「新田ですよ」

「にった？」

「ほら、また先生のクラスになった新田」

誰かのこもったような笑い声が微かに響いてきた。かけているところには複数の子がいるらしい。何がなんだか分からない。だが、《桜木真理子先生》のために、ぼろを出してはいけないと思う。

「ああ……」

受話器を持ったまま、頷いてしまう。それからやっと、馬鹿だわ、わたし、と思う。ごまかすのなら、真理子はいない、わたしは妹だ、とか何とかいって、とにかく切つてしまえばよかったのだ。

泥沼に足を突っ込んでしまった。

「先生、あのさあ」

「うん」

「俺、——暴走族に入っちゃった」

(B) 中国語テキスト pp.114-116

「喂。」

對方是少年或青年。總之，是年輕男子。

「老師嗎？」

好像是找櫻木先生的。

「櫻木現在不在家。」

簡直就像太太，只是這樣的回答，也難為情得令人生氣。

「……」

似乎是猶豫不決的沉默。

「您是哪一位？」

「是真理子老師吧？」

「啊？」

「是櫻木真理子老師吧？」

唔，我的回答卡住了。我知道在這世界裡，我擁有怎樣的姓氏。然而，那位「櫻木真理子」是個怎麼樣的人？——我不知道

除了太太和母親之外，還扮演著什麼角色？沒問過。

詞窮之際我發出的歎息，雖然曖昧不清，但對方仍認定是肯定的回答吧。溺水的人，河川是毫不客氣地載流而去的。維持著掙扎的形狀順水流而去。

「我是新田啊。」

「新田？」

「就是，又回到老師班上的新田。」

小小聲地響起某人悶笑的聲音。電話那頭似乎有很多人的樣子。完全不知所云。但我想，為了「櫻木真理子老師」，不能露出馬腳。

「啊啊……」

我拿著話筒，點起頭來。然後，終於，發現，我是個笨蛋。要矇混的話，說真理子不在，我是她妹妹之類的，總之切掉就好了嘛。

腳踩進泥沼了。

「老師，是這樣的。」

「嗯。」

「我——加入暴走族了。」

(C) 日中対訳のデータセット

はい

喂

相手は少年か青年

對方是少年或青年

とにかく、若い男だった

總之，是個年輕男子

先生？

老師嗎？

桜木さんに用のようだ

好像是要找櫻木先生

桜木はただいま、出ております

櫻木現在不在家。

まるで奥さんのように、そう答えるのも、くすぐったく腹だたしい

簡直就像太太一樣，只是這樣回答，也難為情得叫人生氣

とまどったような沈黙

一種猶豫不決的沉默

どちら様でしょうか
您是哪一位？

真理子先生でしょ？
是真理子老師吧？

は？
啊？

桜木真理子先生でしょ？
是櫻木真理子老師吧？

う、と返事に詰まった
唔，我的回答卡住了

わたしがこちらの世界で、どういう姓を持っているかは知っていた
我知道我在這個世界裡，我擁有怎麼樣的姓氏

だが、その桜木真理子が、どういう人物なのか
然而，那位櫻木真理子是個怎麼樣的人？

奥さんであり、母である他にどんな役割を果たしているのかは
知らない
我不知道除了太太和母親之外，還扮演著什麼角色？

聞いていない
沒聽說

言葉に窮している内についた息が
詞窮之際我發出的嘆息

相手には曖昧ながらも肯定の返事と受け取れたようだ
雖然曖昧不清，但對方仍認定是肯定的回答吧

溺れる人間も、川は遠慮なく流してしまうものだ
即使是溺水的人，河川依舊會毫不客氣地載流而去

もがいている、そのままの形で
保持那掙扎的樣態

新田ですよ
我是新田啊

にった？
新田？

ほら、また先生のクラスになった新田
就是，又回到老師班上的新田

誰かのこもったような笑い声が微かに響いてきた
小小地響起有人悶笑的聲音

かけているところには複数の子がいるらしい
電話那頭似乎有很多人的樣子

何がなんだか分からない
完全不了狀況

だが、桜木真理子先生のために、ぼろを出してはいけないと思
う
但我想，為了櫻木真理子老師，不能露出馬腳

ああ
啊啊

受話器を持ったまま、頷いてしまう
我拿著電話筒，點起頭來

それからやっと、馬鹿だわ、わたし、と思う
然後，終於，發現，我是個笨蛋

ごまかすのなら、真理子はいない、わたしは妹だ、とか何とか
いって
要蒙混的話，說真理子不在，我是她妹妹之類的

とにかく切ってしまうえばよかったのだ
總之掛掉就好了嘛

泥沼に足を突っ込んでしまった
自己把腳踩進泥沼了

先生、あのさあ
老師，是這樣的

うん
嗯

俺、暴走族に入っちゃった
我，加入暴走族了

付記

本稿は、「AI と日本語教育」国際シンポジウム（2019年3月9日，淡江大学），及び2019年度「台湾日本語教育研究」国際シンポジウム「AI と日本語教育との対話」（2019年11月30日，淡江大学）における口頭発表原稿に加筆・修正を施したものである。席上貴重なご意見を賜った各位に深く感謝する。また，本実験の実施及び分析にあたり，熱心なご協力をいただいた淡江大学資訊管理系（院）蔣宜靜氏にお礼申し上げる。