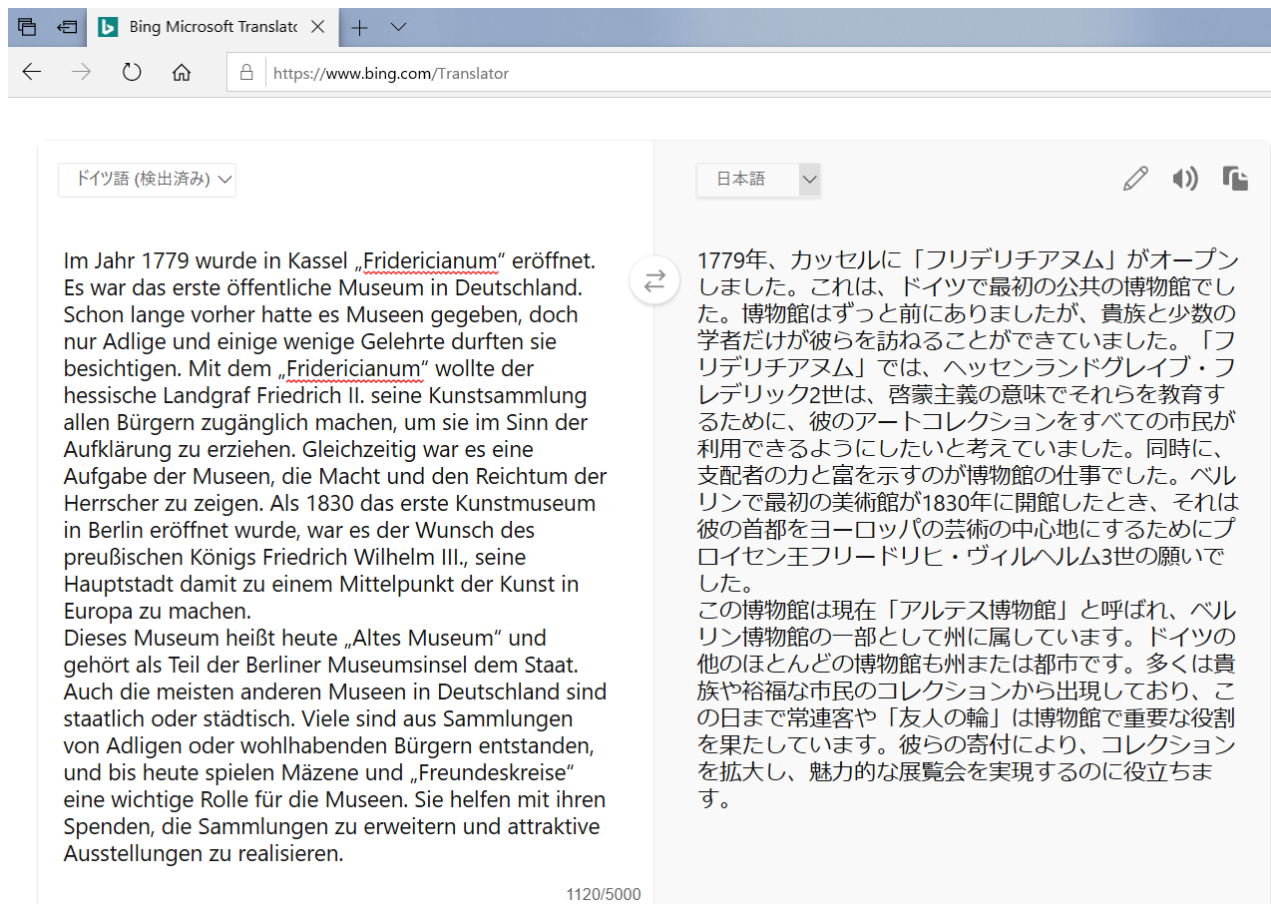


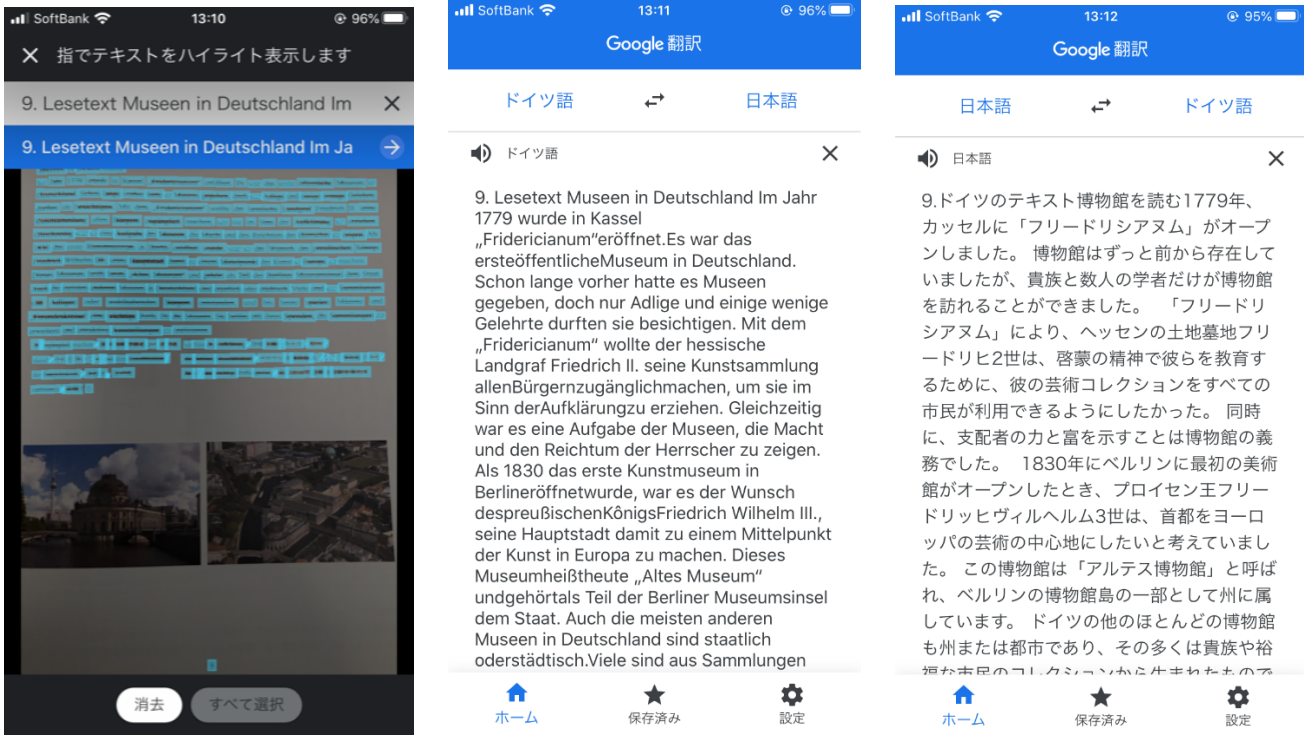
AI 翻訳と外国語教育 資料集

0. シンポジウム開催の契機

【資料1】PC版のGoogle翻訳、Bing (Microsoft) 翻訳の画面



【資料2】スマホで Google 翻訳（スマホカメラでスキャン→ドイツ語文字表示→ワンクリックで日本語訳）



【資料3】ドイツ語作文課題

神様が私たちの祈りをきき入れてくださいますように。（*der Gott, das Gebet, erhören*、接続法第I式を用いて）

答え① *Gott erhöre unser Gebet.*

答え②（Google 翻訳） *Möge Gott unser Gebet erhören!*

1. 機械翻訳の歴史・種類と特徴について

【資料4】機械翻訳（*machine translation, maschinelle Übersetzung*）／自動翻訳の歴史と種類

1933年、ロシア人技術者 Peter Troyanskii (1894-1950) がソ連科学アカデミーで機械翻訳について発表し、特許を出願

1954年、Georgetown 大学や IBM などの研究者グループが機械翻訳システムを発表（実用化へ）

主な翻訳システム

●ルールベース方式

ルールベース機械翻訳 RBMT (Rule Base Machine Translation) 1970年代～

人間が予め作った文法ルールと辞書情報を照らし合わせながら訳文を生成する。

開発者に高度な言語知識が求められ、対象言語ごとに専門家が必要。翻訳精度向上のためには、莫大な人手と作業時間がかかる。

訳抜けや重複訳は起こりにくいが、訳文はぎこちない。

●コーパスベース方式

蓄積された対訳データ（原文と翻訳）を分析・学習した結果からモデルを作成し、そのモデルに基づいて翻訳を行う。人間は規則も辞書も設定しない。機械自身が学習していく。

○統計的機械翻訳 SMT (Statistical Machine Translation) 1990年代～

学習した統計モデルを使って訳文を生成。フレーズベース方式では、ある単語の前後にある単語も含めて句として翻訳し、組み合わせの確率が高いものが訳文として生成される。大量のコーパスをモデルとして用意して学習させる必要があるが、コーパスと学習プログラムがあれば、学習自体はコンピュータが行う。

フレーズテーブルや言語モデルといった独立した翻訳知識を用意したり、翻訳時に考慮する素性を設計したりする必要がある。

機械的な訳文で流暢さが足りないが、数字は絶対間違えない。語順が異なる言語対では句の並べ替えを行い、そこで失敗が起こることがあった。

○ニューラル機械翻訳 NMT (Neural Machine Translation)

人間の脳に見られる神経の特性を模倣したニューラルネットワークにディープラーニング（深層学習）を適用。

ニューラルネットワークは 1943 年に考案され 1950 年代に試作されたが、計算速度の遅いコンピュータでは難しく、2010 年代になって再び注目を集めた。

リカレントニューラルネットワークを用いた翻訳方式は 1997 年に発表。SMT と同等以上の性能を達成し、機械翻訳の研究の中心となったのは 2015 年頃から。2016 年 11 月には Google が NMT のサービスを開始。

文全体の情報を文脈として扱うことで、語順や構造が異なる言語間でも高い翻訳精度に達している。人間が翻訳した文章をもとに学習し、訳文の特徴も捉えるので、生成される訳文も人間が書いたものに近い自然な文章になる。入力文をベクトル表現に変換するエンコーダー（encoder）、出力すべき単語を決定する時にどこに注目すべきかをコントロールするアテンション機構（attention mechanism）、出力文を生成するデコーダー（decoder）の 3 つのパーツで構成される。ひとつのニューラルネットワークを用意するだけで訓練も翻訳もすべて同じ枠組みで行い、自動的に学習。入力から出力までが単一のモデルで完結するような枠組みをエンドツーエンド（end-to-end）という。

◇訳文の評価 人手評価の例 ATR…流暢さ（fluency）と適切さ（adequacy）の 2 軸で 5 段階評価
自動評価の例 BLEU…機械訳と参照訳の n-gram の一致度を計算

【資料 5】ニューラル機械翻訳について

- ・ NMT は適切性のうち訳漏れの数は最大だが、誤訳は SMT より少なく、流暢さは全項目で上回る。
- ・ NMT は従来方式に比較して多くの学習時間や計算資源（数百万文～）が必要。
- ・ NMT は、原文と訳文の単語対応がとりにくい。一方、単語をベクトルで表現するために、同じ意味だが別の単語であっても、同じような翻訳を出力できる。
- ・ 翻訳文の生成においては、SMT が入力文の部分的な置き換えの組み合わせで実現され、入力文をすべて置き換えることで翻訳が終了したが、NMT では、翻訳の終了をコントロールすることができない。このため訳抜け（under translation、約 18%）や重複訳（over translation、約 4%）が起こる。
- ・ NMT の翻訳は、訳文だけ見ると日本語の文として成立していることが多い。原文と照らし合わせて初めて翻訳の誤りが含まれていることが判断できるが、訳文だけを見ても気づきにくい。
- ・ 言語ペアや分野、あるいは文章の構造によっては、SMT や RBMT の方が品質がよいこともあるため、ハイブリッド型エンジンも開発されている。
- ・ 機械翻訳では、人間と違って意味は考えていないため、複数の意味を持つ語は正しく訳しにくい。
例) My siste is older than I. → 「妹は私より年上です。」
- ・ 単語の訳も正確ではないことがある。
例) Landeskunde ([英] regional studies) → 「研究」 (Bing 翻訳)、「地域顧客」 (みらい翻訳)

【資料 6】総務省や NICT の事業

総務省「グローバルコミュニケーション計画」 https://www.soumu.go.jp/main_content/000285578.pdf

NICT (情報通信研究機構) 「翻訳バンク」 <https://h-bank.nict.go.jp/>

「みんなの自動翻訳@TexTra」 15 言語 <https://mt-auto-minhon-mlt.ucrj.jgn-x.jp/>

「多言語音声翻訳アプリ VoiceTra」 31 言語 <http://voicetra.nict.go.jp/>

2. 機械翻訳の教育への活用

- ・後編集＝ポストエディット（post-editing）：誤訳を修正し、適切かつ流暢な表現に改める作業。
- ・前編集＝プレエディット（pre-editing）：コンピュータが翻訳しやすいように、原文の曖昧さを除去し、省略を補う作業。

【資料7】前編集の時に気をつけるべき日本語の特性の例

主語や目的語の省略が多い。

定冠詞・不定冠詞がない。

指示代名詞が多用される。（指示先を特定する必要がある。）

単数・複数が明示されないことがめずらしくない。

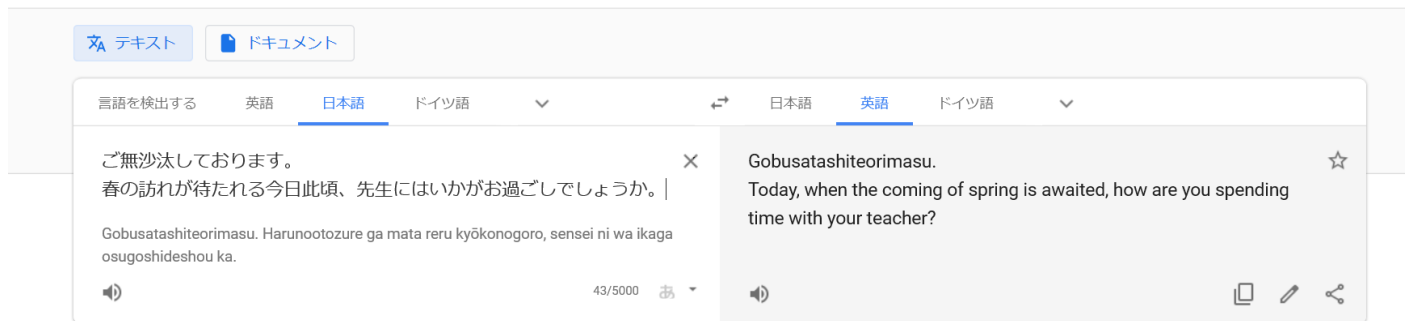
英語にはない助数詞（二台、三着など）が用いられる。

語義の多義性。

〇〇さんの場合、Mr./Ms.が不明。

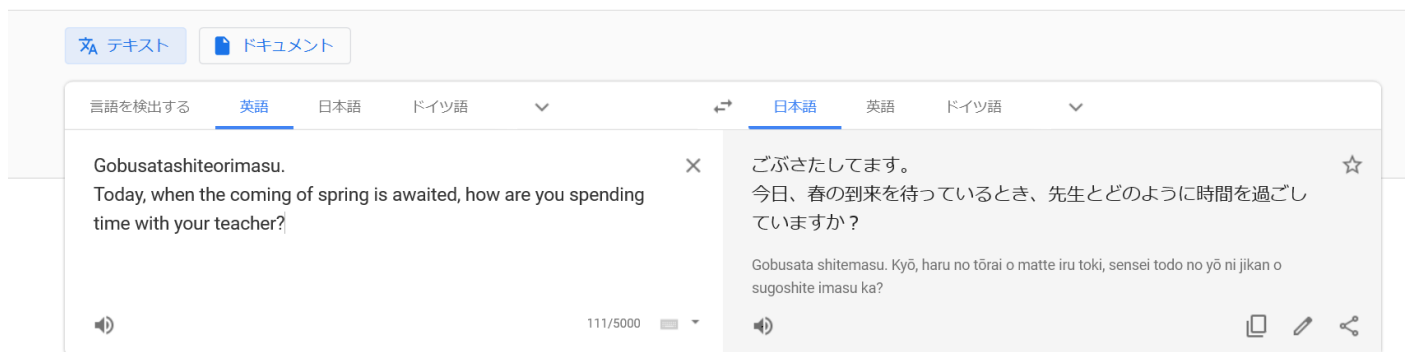
【資料8】日本語から外国語に訳し、それを日本語に戻して翻訳が合っているかチェックする方法

Google 翻訳



The screenshot shows the Google Translate interface. On the left, the source text in Japanese is: "ご無沙汰しております。春の訪れが待たれる今日此頃、先生にはいかがお過ごしでしょうか。" Below it is the English translation: "Gobusata shiteorimasu. Harunootozure ga mata reru kyōkonogoro, sensei ni wa ikaga osugoshideshou ka." The right side shows the English text being translated back into Japanese: "Gobusata shiteorimasu. Today, when the coming of spring is awaited, how are you spending time with your teacher?"

Google 翻訳



The screenshot shows the Google Translate interface. On the left, the source text in English is: "Gobusata shiteorimasu. Today, when the coming of spring is awaited, how are you spending time with your teacher?" Below it is the Japanese translation: "ごぶさたしています。今日、春の到来を待っているとき、先生とどのように時間を過ごしていますか？" The right side shows the Japanese text being translated back into English: "Gobusata shiteorimasu. Kyō, haru no tōrai o matte iru toki, sensei todo no yō ni jikan o sugoshite imasu ka?"

