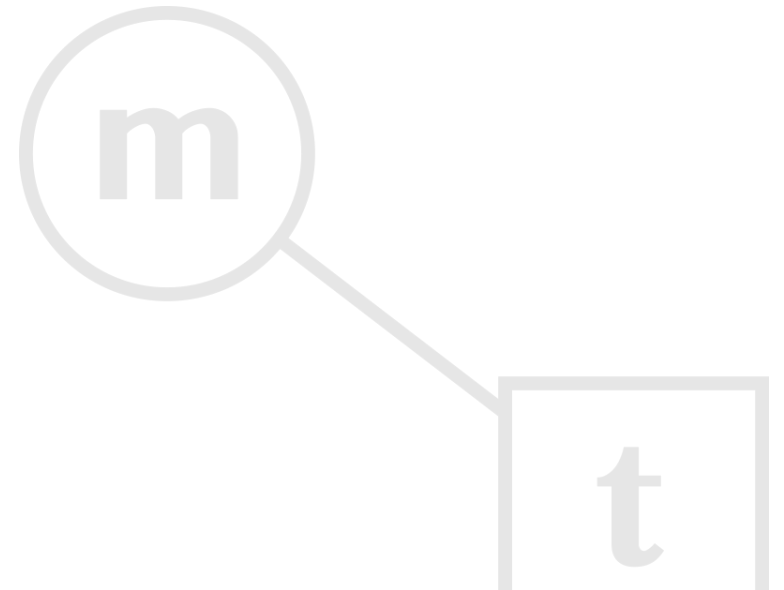




翻訳Tips

株式会社みらい翻訳
Mirai Translate, Inc.

〒150-0002 東京都渋谷区渋谷 2-22-3 渋谷東口ビル2F
TEL: 03-6434-1340 FAX: 03-6434-1590
<https://miraitranslate.com>



0. [はじめに](#)

1. [ニューラル機械翻訳の性質](#)

2. [翻訳しやすい原文をつくるために](#)

3. [テキスト翻訳とファイル翻訳](#)

4. [用語集](#)

1. [用語集とは](#)

2. [用語集 登録のヒント（基礎編）](#)

3. [用語集 登録のヒント（実践編）](#)

*) 本資料の内容は、今後の機械翻訳エンジンのアップデートによって変わる可能性があります。

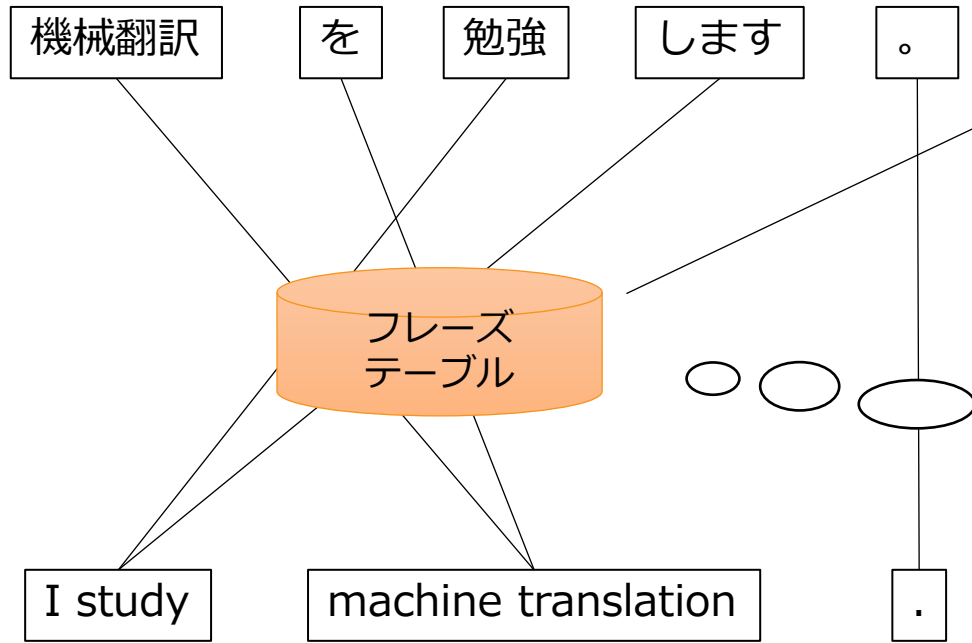
参考：[翻訳までの流れ](#)

0. はじめに (今までの機械翻訳の概要)

2010年代前半までは「機械翻訳」といえばフレーズベース統計的機械翻訳 (Phrase-Based Statistical Machine Translation, PBSMT) と呼ばれる、単語列単位の組み替えで翻訳文生成する技術が主流でしたが、単語列単位で切り貼りして文を構成するため、全体で見ると流暢性に欠けるという問題点がありました。

従来技術 (PBSMT) のベースであるSMTによる翻訳

注意：説明のために大幅に簡略化しています



コーパス：データベース化された大規模な言語資料のことを指します
機械翻訳に用いる対訳コーパスとは様々な言語のコーパスを対応付けしたものです

en	ja
I'll stay at a hotel.	ホテルに泊まります。
A little water, please.	水を少し下さい。
:	:

対訳コーパスを学習し、最もそのように訳される確率が高いものを統計的に計算して語順を組み替えます

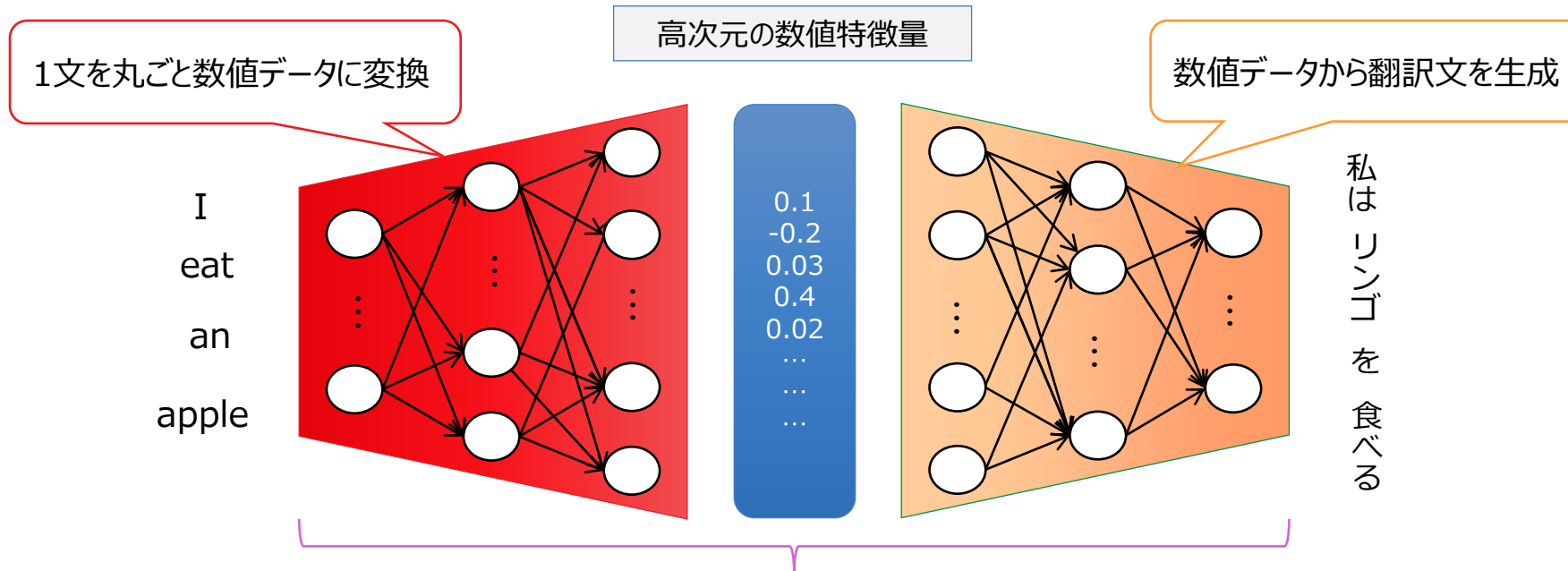
統計計算のイメージ

	studyの訳	Iと併用された時に訳される確率
1	勉強する	80%
2	研究する	●%
3	考慮	▲%

→ I study は“私は・・・勉強します”という訳になると統計的に判断

1. ニューラル機械翻訳の性質

みらい翻訳Plusが採用しているのはニューラル機械翻訳（Neural Machine Translation, NMT）です。NMTは単語列単位でなく、1文丸ごと数値データに変換した後、この数値データから新しい翻訳文を生成します。これによりPBSMTと比べて、より流暢な翻訳結果を生成できることが特長です。



最新の人工知能タイプの翻訳エンジンであり、原文を数値データに置き換えてから翻訳結果を生成する

1. ニューラル機械翻訳の性質 翻訳精度向上実例



従来技術であるPBSMTと比較して、NMTでは飛躍的に翻訳精度が向上しています。

原文	夕食は部屋食、朝食は食堂でという旅館もよくあります。
----	----------------------------



PBSMT 翻訳結果	Dinner is a Breakfast it is common in japanese inn that dinner at the dining room.
---------------	--

意味が分からない

初期：みらいNMTエンジン (2017) 翻訳結果	There are many Japanese-style inns in the dining room for dinner.
---------------------------------	---

原文と意味が異なっている

みらいNMTエンジン (2018) 翻訳結果	There are many ryokans that have dinner in the room and breakfast in the dining room.
------------------------------	---

Good

現在：みらいNMTエンジン (2022) 翻訳結果	There are many ryokans where dinner is served in the room and breakfast is served in the dining room.
---------------------------------	---

Very good!

1. ニューラル機械翻訳の性質 留意すべき特徴



高い翻訳精度が大きな特徴で、従来技術に比べ圧倒的に自然な文章を生成します。
一方で以下のような特徴もあるため、その性質を理解した上でご利用することをおすすめします。

留意すべき特徴

- ✓ うまくいく理由が十分に分かっていません。
→ 翻訳エラーがあった場合に修正する方法が明確ではありません。
- ✓ 原文を過不足なくカバーして翻訳しないことがあります。
→ **訳抜け**、**湧き出し**、**過剰生成**などが発生します。
- ✓ **入力の違いに敏感**で、表記揺れに弱く、句読点の有無によっても翻訳結果が大きく変わることがあります（不安定性）。
- ✓ 辞書を十分に制御して当てるのが難しい。
→ 用語集を使用する際は **4. 用語集** の項目をよく読んだ上でお使いください。

1. ニューラル機械翻訳の性質 不安定性



NMTは入力の違いに敏感で、表記揺れに弱く、句読点の有無によっても翻訳結果が大きく変わることがあります。

• 句点の有無

例) 原文	翻訳結果
8月20日には一時140円26銭と2016年3月以来、3年5ヶ月ぶりの高値圏まで上昇した。	On August 20, the price rose to 140.26 yen, the highest level in 3 years and 5 months since March 2016.
8月20日には一時140円26銭と2016年3月以来、3年5ヶ月ぶりの高値圏まで上昇した	On August 20, it momentarily rose to 140.26 yen, reaching the highest level in 3 years and 5 months since March 2016

• 表記揺れ

例) 原文	翻訳結果
振替輸送 を行います。	There will be a transfer of transportation.
振り替え輸送 を行います。	We will provide alternative transportation.

2. 翻訳しやすい原文をつくるために



機械翻訳は、私たち人間のように文脈を理解することができないため、正しく翻訳するためには、シンプルで分かりやすい原文であることが必要です。
日本語を英語に翻訳する場合は、日本語にありがちな曖昧な表現や内容の省略、間接的な言い回し等を避けることにより、機械翻訳の精度を高めることができます。

➤ 具体的には、以下のポイントがあります。（日→英の場合）

1. 理解しやすい語順にする

- 修飾するもの・修飾されるものは近くに置く

2. 「省略」を避ける

- 砕けた言い回しは避け、書き言葉として適切なものを選ぶ
- 数値の省略表現に注意する
- 主語や目的語を省略しない

3. 曖昧さを避ける

- 修飾先を明確にする
- 複数の意味をもつひらがなは漢字にする

4. 難解な表現を避ける

- 慣用句やことわざは避ける
- あまり使わない表現は避ける

2. 具体例



1. 理解しやすい語順にする

- 修飾するもの・修飾されるものは近くに置く

	例) 原文	翻訳結果
修正前	弊社では、本年度から安定した高品質の商品の提供に力を入れております。	Starting this fiscal year, our company will focus on providing stable, high-quality products.
修正後	弊社では、本年度から高品質の商品の、安定した提供に力を入れております。	From this fiscal year, our company has been focusing on the stable supply of high-quality products.

Good

2. 具体例



2. 「省略」を避ける

- 砕けた言い回しは避け、書き言葉として適切なものを選ぶ

	例) 原文	翻訳結果
修正前	そのお客様には午後一で電話をかけてください。	Please call the customer at one p.m.
Good 修正後	そのお客様には午後一番で電話をかけてください。	Please call the customer first thing in the afternoon.

- 数値の省略表現に注意する

	例) 原文	翻訳結果
修正前	2-30万	2 -300000
Good 修正後	20万-30万	200000 -300000

- 主語や目的語を省略しない

	例) 原文	翻訳結果
修正前	明後日の14時半にお願いします。	14:30 the day after tomorrow, please.
Good 修正後	明後日の14時半に打ち合わせをお願いします。	Please have a meeting at 14:30 the day after tomorrow.

3. 曖昧さを避ける

- 修飾先を明確にする

例文：私は通勤中に彼女がメールしたのを知った

これは“通勤中”がどこを修飾するのかによって解釈が二通りあります。 1)私が通勤中に知った 2)彼女が通勤中にメールした 2)の意味を意図している場合、修飾先が明確な文章になるよう原文を調整して翻訳する必要があります

	例) 原文	翻訳結果
修正前	私は通勤中に彼女がメールしたのを知った。	I learned that she emailed me while I was commuting.
Good 修正後	私は彼女が通勤中にメールしたのを知った。	I learned that she emailed me on her way to work.

- 複数の意味をもつひらがなは漢字にする

	例) 原文	翻訳結果
修正前	あめがふるといけないので傘をもってってください。	Please take an umbrella in case the candy drips.
Good 修正後	雨が降るといけないので傘をもってってください。	Take an umbrella with you in case it rains.

4. 難解な表現を避ける

- 慣用句やことわざは避ける

	例) 原文	翻訳結果
修正前	彼女に何度も連絡したが梨の礫だ。	I have contacted her many times, but it is pear pebbles.
Good 修正後	彼女に何度も連絡したが返事がない。	I have contacted her many times, but there is no answer.

- あまり使わない表現は避ける

	例) 原文	翻訳結果
修正前	小生までご連絡ください。	Please contact my elementary school.
Good 修正後	私までご連絡ください。	Please contact me.

3. テキスト翻訳とファイル翻訳



翻訳を行う方法として「テキスト翻訳」と「ファイル翻訳」の2種類の方法があります。用途に応じて使い分けると便利ですが、それぞれ使用する際には注意点があります。

- 翻訳を実行する前に
 - **言語方向**が合っているか確認をしましょう。
 - **翻訳設定**を使って語調を統一すると読みやすいです（日本語への翻訳の場合）。※デフォルトは「です・ます」調に統一されます。
- テキスト翻訳の注意点
 - 改行は文区切りとみなすため、**文の体裁を整える（レイアウト調整の）**ために文の途中で改行を入れていると翻訳精度が劣化します。（特にPDFからテキストをコピーして翻訳する場合や、メール文で所々に改行を入れていたりすると注意が必要です。）
- ファイル翻訳の注意点
 - **装飾（太字や色、ハイパーリンクなど）**があると、翻訳結果がテキスト翻訳と異なることがあります。
 - PDFファイルの翻訳では、ファイルによっては**OCRによって文字認識**をする必要があり、うまく認識できない場合翻訳精度も劣化する可能性が高くなります。PDFにする前のファイルをお持ちであれば、そちらをファイル翻訳にかける方が精度は高くなります。
 - テキスト翻訳と同様に、改行は文区切りとみなします。特にPDFファイルの翻訳では、解析時に意図していないところに改行が入る可能性があります。

3. テキスト翻訳とファイル翻訳



テキスト翻訳で改行を入れている例は以下のようなものがあります。改行は文区切りとみなして訳すため、文途中の不要な改行は削除した上で翻訳されることをおすすめします。

例：メール文の体裁を整えるために改行している

お世話になっております。〇〇商事のXXです。

先日ご案内いたしました新製品について、大変貴重なご意見をいただき、誠にありがとうございました。

このような文であると、改行されているため1行ずつ翻訳してしまいます

Good

先日ご案内いたしました新製品について、大変貴重なご意見をいただき、誠にありがとうございました。

改行を削除し、1文1行で翻訳すると適切に訳すことができます

例：PDFからコピー&ペーストした時、原文にレイアウト調整のための改行が含まれている

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

PDFからコピー & ペースト

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

各行の終わり等、文の途中に改行がないかを確認してください。改行があればそれを削除することで適切に訳すことができます

3. テキスト翻訳とファイル翻訳



ファイル翻訳では装飾の有無によって翻訳文が変わることがあります。もし装飾を含むファイル翻訳の翻訳精度が低いと感じた場合、一度装飾を削除して翻訳してみることをおすすめします。

• ファイル翻訳の注意点

- ファイル翻訳では、文中の装飾（太字や色、ハイパーリンク等）を保持することが可能^{*1}ですが、一方で、装飾の有無によって訳文が変わることがあります。

- 例：赤字にしている単語があると、対応する訳文の単語も赤字になります。
赤字にしている単語があると、対応する訳文の単語も赤字になります。



If there are words in red, the corresponding words in the translation will also be in red.

If there is a word in red, the corresponding word in the translation will also be in red. ← 翻訳文が変わっている

- 装飾が1文中に大量にあったり、意味の区切りに沿わない装飾箇所があったりすると、翻訳精度が劣化することがあります。

- 例：赤字にしている単語があると、対応する訳文の単語も赤字になります。
赤字にしている単語があると、対応する訳文の単語も赤字になります。



Word of doing in red becomes yes, and word of translation becomes red as well .

If there is a word iru in red, the corresponding translation word will also be in red.

← 精度が劣化している

ファイル翻訳の結果、精度が低いと感じたら、一度装飾を削除して翻訳してみてください。

*1 2021年9月現在で、文中の装飾保持に対応しているのは PDF, Word, PowerPoint で、Excel には未対応です。

4-1. 用語集とは



用語集とは、固有名詞や専門用語をユーザが指定した通りに翻訳させることができる機能です。社内で共有されている用語集や、**みらい翻訳Plus**をご使用いただく中で正しく翻訳出来なかった単語を登録していくことで、翻訳の精度を向上させていくことができます。

例

サービス開発部に所属している東海林と申します。

辞書なしでは・・・

This is Tokairin from the service development department.



人名「東海林」を誤訳している・・・
部署名「サービス開発部」はうちでは“SDD”と訳したい・・・

辞書を使うと

This is Shoji from SDD.



「東海林」が正しく訳せるようになった
うちの用語が反映された

※ 用語集を使うと、登録した単語以外の訳も変化する可能性があります。

4-1. 用語集とは



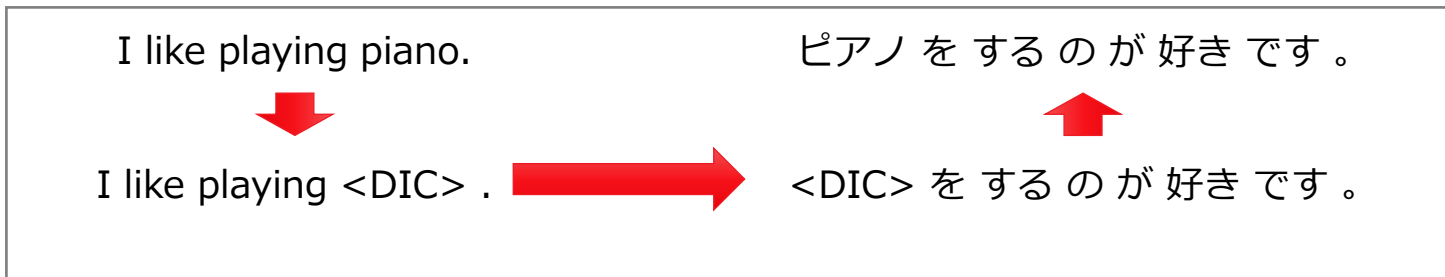
用語集機能を適用する場合、原文の一部をマスキングした状態で翻訳することになるため、訳質の低下や誤訳につながる可能性があります。これを回避するため、1) 一般的な名詞の登録はなるべく避けること、2) 用語集なしのデフォルト設定で翻訳し、上手く訳出できない単語を順次登録することをおすすめします。

用語集機能の仕組みと注意点

- 用語集機能は、辞書に登録されている単語を含む原文に対し、**該当部分をいったん特殊な単語（クラストークン）に置き換えてから翻訳**を行い、翻訳後に**クラストークンを辞書に登録されている訳語に書き戻す**ことで実現しています。（以下の図は「東海林 → Shoji」を辞書登録した場合）



- このため、辞書が適用されると、**原文の意味が一部欠落した状態で翻訳**をすることになり、登録する単語によっては以下の例のような誤訳が発生します。（「piano → ピアノ」を登録した場合）



* <DIC>が楽器であることが認識できていないため、play を“弾く”と訳すことが出来なくなっています（現在の機械翻訳エンジンでは、「<DIC>」は固有名詞のようなものとして認識しています）

→ 一般的な名詞はなるべく登録しないようにすることが重要です

用語集を効果的に使うためには、以下の4つのポイントがあります。

1. マッチングの優先順位を理解する

- 文中で先に登場する単語が優先される
- 文字数の多い単語が優先される
- 指定順番が先の辞書の単語が優先される

2. 固有名詞や専門用語以外の単語（普通名詞や抽象名詞, 名詞以外の品詞）の登録はなるべく避ける

- 単数形と複数形の訳し分けは出来ない
- 大文字と小文字の訳し分けは出来ない
- 活用形の区別は出来ない
- 複合語である（他の単語に含まれる）ことは考慮出来ない

3. 反映されないケースを理解する

- 単語の一部には反映されない場合がある
- 前方から優先的に適用されるため、後方にあると反映されない
- 登録する原語と完全一致しないと反映されない（特に英→日）

4. その他

- the がつく固有名詞を登録する際は、the も含めて辞書登録する（日→英）

1. マッチングの優先順位を理解する

➤ 辞書は基本的には指定した順番で優先されますが、場合によっては後に指定している辞書の単語が優先してマッチすることがあります。辞書のマッチングの優先順位は以下の順で決定されます。(一意に決まるまで、i. → ii. → iii. の順に条件をあてていきます)

- i. 文中で先に出現する
- ii. 文字数が多い
- iii. 指定順番が先

辞書1

No.	原語	訳語
1	東京	Tokyo

東京タワー大展望台は東京都港区にあります



辞書2

No.	原語	訳語
1	東京	TOKYO
2	東京タワー	TOKYO TOWER
3	タワー大展望台	TOWER LARGE OBSERVATORY

辞書2_2 大展望台は 辞書1_1 都港区にあります



TOKYO TOWER main observatory is located in Minato Ward, Tokyo.

※辞書1, 辞書2 の順で指定しているとします。

【“東京タワー大展望台”の辞書マッチングの考え方】

①まずは条件i. で判断
“東京”と“東京タワー”が同時に出現している (“タワー大展望台”は出現順では後)

- ・東京 (辞書1-1)
 - ・東京 (辞書2-1)
 - ・東京タワー (辞書2-2)
- がマッチング

②次に条件ii. により
文字数は東京タワー > 東京
→東京タワー (辞書2-2)
と一意に決まる

- ① 辞書1に「東京」が登録されているため、文中の「東京タワー大展望台」の「東京」に、辞書2の「東京」よりも優先してマッチングされますが、ii. のルールにより、辞書2にある「東京タワー」の方が文字数が多いため、さらに優先してマッチングされます。
- ② このとき、ii. のルールで「東京タワー」よりもさらに文字数の多い「タワー大展望台」の方が優先されるわけではなく、i. のルールにより、文中に先に出現している「東京タワー」が優先されています。
- ③ 文中の「東京都港区」の「東京」に対しては、iii. のルールに従い、優先して指定している辞書1の「東京」がマッチングされます。

2. 固有名詞や専門用語以外の単語 (普通名詞や抽象名詞, 名詞以外の品詞) の登録はなるべく避ける

登録してしまうと以下のような制限事項により、翻訳精度が悪くなることがあります。

- 単数形と複数形の訳し分けは出来ない

例) 普通名詞 猫 → cat を登録した場合

	例) 原文	翻訳結果
○	猫が1匹います	There is one cat .
×	猫がたくさんいます。	There are many cat .

“cats”と複数形にすべきものも、辞書登録した単数形で訳されてしまいます

- 大文字と小文字の訳し分けは出来ない

例) 化学 → chemistry を登録した場合

	例) 原文	翻訳結果
○	彼は 化学 が好きだ。	He likes chemistry .
×	化学 は面白い学問である。	chemistry is an interesting study.

文頭に来ても大文字にはなりません

4-2. 用語集 登録のヒント（基礎編）



2. 固有名詞や専門用語以外の単語（普通名詞や抽象名詞, 名詞以外の品詞）の登録はなるべく避ける

- 活用形の区別は出来ない

例) 抽象名詞 解決 → solution を登録した場合

	例) 原文	翻訳結果
○	私はそれが最善の 解決 につながると思います。	I think that will lead to the best solution .
×	私たちには 解決 すべき問題がたくさんあります。	We have a lot of problems to do in solution .

活用形を区別できていない文章になってしまいます

- 複合語である（他の単語に含まれる）ことは考慮出来ない

例) 理学部 → Faculty of Science を登録した場合

	例) 原文	翻訳結果
○	理学部 に所属しています。	I belong to Faculty of Science .
×	理学部長	Faculty of Science Chō

複合語であるかを判断できないため、用語集登録しているものをそのまま適用してしまい、適切でない訳になってしまいます

※この場合、「理学部長」も登録すれば誤訳は防げます。（4-3. 用語集 登録のヒント（実践編））

3. 反映されないケースを理解する

- 単語の一部には反映されない場合がある

例) 理学科 → the department of science を登録した場合

	例) 原文	翻訳結果
○	ここは理学科です。	This is the department of science.
×	ここは数理学科です。	This is the department of mathematics.

用語集は誤った訳になるのを防ぐため、単語の一部分にはマッチさせないようにしています。上記の例では、理学科は数理学科という単語の一部であるため、辞書は反映されません。

* どこからどこまでを単語と見なすかは内部のシステムが統計的に判断します。この処理はサービス上で確認することが出来ないため、もし辞書に登録したのに反映されなかった場合は、その単語を含むもっと長い単語を登録するとそちらが反映される可能性があります。

3. 反映されないケースを理解する

- 原文の中で先に出てくる単語を優先するため、後に出てくると反映されない

例) 「都庁舎 → Metropolitan Government Office building」を辞書登録したい場合

例) 原文	翻訳結果
東京都庁舎	Tokyo Government Office

このとき、仮に「東京都 → Tokyo」も辞書に登録されていると、用語集は先に出てくる単語を優先するため、「東京都庁舎」に対しては「東京都」が反映されてしまい、「都庁舎」は反映されません。

例) 原文	翻訳結果
東京都庁舎	Tokyo Metropolitan Government Office building

この場合は、「東京都庁舎 → Tokyo Metropolitan Government Office building」を登録すると、正しく翻訳することができます。

3. 反映されないケースを理解する

- 登録する原語と完全一致しないと反映されない (特に英→日)

例) 「neural machine translation → NMT」を辞書登録した場合

	例) 原文	翻訳結果
○	Our neural machine translation engine is highly accurate.	当社の NMT のエンジンは非常に正確です。
×	Neural machine translation is good at fluent translation.	神経機械翻訳 は流暢な翻訳が得意です。
×	You can try various neural machine translations for free.	さまざまな ニューラル機械翻訳 を無料で試すことができる。

4. その他

- the がつく固有名詞を登録する際は、the も含めて辞書登録する (日→英)

	辞書	例) 原文	翻訳結果
×	エッフェル塔 → Eiffel Tower を登録	エッフェル塔に登りました。	I climbed Eiffel Tower.
○	エッフェル塔 → the Eiffel Tower を登録	エッフェル塔に登りました。	I climbed the Eiffel Tower.

用語集登録時には以下のような観点でチェックすると効果的です。

1. 用語集は、登録する前に以下のような観点をチェックしてください。
 - **品詞**をチェック
 - 名詞以外は取り除きます。文脈によって様々な訳になりうる一般名詞も極力取り除きます。
 - **文字数**をチェック
 - 文字数が少ない単語は他の単語と部分一致する可能性が高いため、取り除きます。
 - おおよその目安として、漢字だと1~2文字、ひらがなカタカナだと1~4文字程度の単語は取り除いておくことをおすすめします。
 - **重複**をチェック
 - 同じ用語に対して複数の訳語は登録出来ないため、取り除きます。
2. 一旦**登録してみて、翻訳結果がどう変わるかをチェック**するとさらに悪影響が少なくなります。
 - ある程度の量の文章を、用語集適用前と適用後それぞれ翻訳し、結果が変わった文をチェックします。
 - 頻出する一般名詞を登録していると、多くの場合訳質が低下しています。その際は、一般名詞を辞書登録から削除してください。
3. 専門用語を登録する場合、**文脈により別の訳語に訳し分ける必要があるかをチェック**。
 - 用語集は文脈に応じて訳語を選択することが出来ないため、訳し分けが必要な単語は登録を避けたほうが無難です。
 - ただし、辞書を複数用意し、別々にプロファイルに登録して使い分けることは可能です。
4. 登録しなくても**ある程度正しく訳せる単語は登録しない**。
 - 用語集は、登録した単語以外の周辺の訳も変化します。全体的な訳質は辞書を登録しないで訳す方が良くなることが多いので、ある程度翻訳出来ている単語は登録を避けてください。（もちろん文によって正しく訳せたり間違ったりするのを防ぐ意味でも辞書登録したいケースはあるかと思いますが、正しく訳せるケースの方が多く場合は、登録を避けたほうが無難です。）

• 効果的な辞書の登録例

例1	原文	翻訳結果
	The simulation allowed me the opportunity to prioritize nursing assessments and care.	シミュレーションの結果, 看護評価 とケアを優先することができた。

- ✓ nursing assessment は **看護評価**ではなく**看護アセスメント**と訳すべき。辞書登録で対応したい。

好ましくない登録のしかた

- 「assessment → アセスメント」を登録
 - ✓ 登録する原語と完全一致しないと反映されない (登録のヒント (基礎編) 3. 反映されないケースを理解する)
→ assessments と assessment は別の単語なので反映されない
- 「assessments → アセスメント」も登録
 - ✓ 文脈により別の訳語に訳し分ける必要があるかをチェック (登録のヒント (実践編) 3. より)
→ assessments は「評価」と訳すケースも多いため、登録すると想定しない影響が出る可能性があります

好ましい登録のしかた

- 「nursing assessments → 看護アセスメント」を登録
 - nursing assessments を 看護アセスメント 以外の訳にする必要性は低い
(ただし、文全体としての翻訳精度が悪くなる可能性はあるので注意が必要です (用語集機能の仕組みと注意点より))
- さらに、以下のパターンも登録
 - 「nursing assessment → 看護アセスメント」 (単数形)
 - 「Nursing assessments → 看護アセスメント」 (文頭大文字)
 - 「Nursing assessment → 看護アセスメント」 (単数形+文頭大文字)

• 効果的な辞書の登録例

例2	原文	翻訳結果
	マイナカード マイナポイント	Mina Card MyNaPoint

- ✓ 「マイナ」は「マイナンバー」の意味だが、正しく訳せていないので辞書登録で対応したい。

好ましくない登録のしかた

- 「マイナ → My Number」 を登録

- ✓ 登録のヒント(実践編) 1. 文字数をチェック

→ カタカナ 3文字は避けた方がよい

- ✓ 単語の一部には反映されない場合がある (登録のヒント (基礎編) 3. 反映されないケースを理解する)

→ マイナカードやマイナポイントはシステム内ではそれだけで1単語 (マイナとカードに分かれない) として処理されるため、このままでは「マイナ」は適用されません。

* 1単語なのか2単語に分かれるのかを事前に確認することは出来ません。

一旦登録してみて、翻訳結果が変わったかどうかで判断する必要があります。

好ましい登録のしかた

- 「マイナカード → My Number Card」、 「マイナポイント → My Number Point」 の2つを登録



- *1 文字正規化では、NFKC正規化を採用しており、全角の数字やアルファベットを半角に変換したりします。
- *2 翻訳前処理では、文分割処理等を行います。
- *3 トークナイズとは、文章を単語に分割することを言います。辞書のマッチングはこのトークナイズの後で行われます。
- *4 辞書に「東京タワー → Tokyo Tower」を登録している場合