

翻訳生産性向上のテクニック

－ PC-Transer 翻訳スタジオ活用法

第9回 翻訳メモリを日英翻訳で活用する方法

2009年3月 小室誠一 (MT研究会主宰 : <http://www.babel.co.jp/mtsg/>)

さて、今回は日英翻訳での利用法を考えてみましょう。

日英翻訳で最も難しいのが最終的な英文チェックです。ネイティブに頼らずにチェックできるようにするために「翻訳メモリ」を活用してみましょう。

日本人が英文を書く際に、英語らしい構文や語句を選択することはなかなか難しいものです。どうしても、日本語に引きずられた英文になってしまいます。

そのために、日英翻訳作業では、日本人が下訳した英文をネイティブがチェックして完成させるという方法が一般的になっています。

最近では、Web 検索エンジンを使って、Web 上のテキストをコーパスとして利用し、表現のチェックをすることが流行しています。皆さんも、出現回数を単純に比較して、回数の多い表現を選択する、というような使い方をしたことがあるのではないかと思います。

しかし、このような雑多な WEB 上のテキストよりも、自分が得意とする専門ジャンルの、正しく自然な英文を翻訳メモリに大量に蓄積しておいて検索したほうが、表現辞書としての有効性が高まります。

PC-Transer の翻訳メモリは、英日と日英のどちらで登録しても、そのまま双方向で利用できます。この特長を活用しましょう。

日英で利用する翻訳メモリを作るには、Web や印刷物 (OCR で電子化) などから英文を収集して、機械翻訳し、英日方向のまま翻訳メモリに一括登録します。

それをそのまま日本語から検索します。

日本語の訳文は単に検索のためのキーワードとして利用できさえすれば良いので、訳文全体の完成度は問いません。直訳で十分です。訳文作成に時間をかけているのは、大量に処理できません。大量になれば「参照用」とし

て役に立ちません。

機械翻訳する際に専門語辞書を適切に設定しておけば、訳文全体として見れば不適切でも訳語そのものは十分に使えます。PC-Transer には、専門語辞書を自動的に選択してくれる機能がありますので設定も簡単にできます。

要するに、「英文を検索するために機械翻訳を使って日本語のインデックスを付ける」と考えてください。

実際に日英翻訳で使用する場合は、翻訳メモリペインで「キーワード検索」を行います。これにより、日本語キーワードの英訳とその使い方が確認できます。

これを英文チェックに生かします。「Web 検索」機能を併せて使えば、さらに柔軟なチェックができるでしょう。

PC-Transer を日英翻訳に活用したい人は、とにかく英日方向で翻訳資源の蓄積を行っておくことが大切です。これは、英語学習と通じるところがあります。

たくさん英文を読んだり聞いたりしてインプットしておかなければ、英文を書いたり話したりすることができないのと同じです。

PC-Transer なら文句も言わずにどんどん学習してくれます。

<翻訳設定のヒント>

インデックスとして利用するために訳文出力する場合は、専門語辞書やユーザー辞書を適用し、単語やフレーズがそれなりに正しく訳

出されるように設定することが重要です。

翻訳設定は以下のように指定すると良いでしょう。

訳出方法

- 大文字を小文字にして訳す(L)
- 長い文章を節・句ごとに区切って訳す(L)
- 命令文を平叙文として訳す(L)
- ですます調で訳す(D)
- 固有名詞を英語のまま表示する(R)
- 合成語を分割して訳す(C)
- 「～'s」を常に所有格と訳す(V)

- (1) できるだけ日本語に訳出されるように、「大文字を小文字にして訳す」にチェックを入れる。
- (2) 長い文章がある場合は、訳出に時間がかかる上に正しく訳せないことが多いので、「長い文章を節・句ごとに区切って訳す」にチェックを入れる。
- (3) 辞書としての利用にふさわしい訳語にするために、「命令文を平叙文として訳す」にチェックを入れる。
- (4) 「固有名詞を英語のまま表示する」ほうが不適切に訳されるよりも無難なのでチェックを入れる。
- (5) 「合成語を分割して訳す」かどうかは、出力結果を見て判断する。

このような参照用としての翻訳メモリは、文書の種類やジャンルをできるだけ絞り込んで作成することをお勧めします。

翻訳メモリは、同時に 10 個まで使用できます。絞り込んで作成した翻訳メモリを、翻訳する文書に合わせて適切に選択すれば、効率の良い検索ができます。

<チェック作業の例>

抽象的な話だけでは退屈なので例を挙げましょう。(あくまでも操作手順の参考例です)

以下の日本語を英文に翻訳することにします。

【原文】しかし、これらの資源を手作業で構築することは高価である。

スタイルは「標準」、専門語辞書自動選択により「電気電子」「コンピュータ」専門語辞書を設定して、訳文を出力します。

【訳文】 However, it is expensive to build these resources by manual operation.

専門語辞書は次の 2 件がマッチしました。

資源 = resources

手作業 = manual operation

これらの訳語はとりあえず信頼できると考えて、その他の語句を翻訳メモリで調べてみます。

「構築する」をキーワード検索します。(図 1)

筆者が作成した「参照用ユーザー翻訳メモリ」(マッチ率 8%) で以下の検索結果が表示されました。([]は筆者が付加)

英語の大きな注釈がついているコーパスを [構築する] こと : [Building] a Large Annotated Corpus of English:

日本語の文章は不自然ですが、「構築する = build」が参考になります。

「高価」も調べてみましょう。(図 2)

「参照用ユーザー翻訳メモリ」(マッチ率 2%) で以下の検索結果が表示されました。

しかしながら、異なる領域にカスタマイズするのが困難であることに加えて、個々のスコアリングは、極めて [高価] およびエラーに従属する全機械翻訳システムを通る遅い、手コード化流路をすることによって実行された。
However, the individual scoring was

performed by doing a slow, hand-coded pass through an entire machine translation system which is extremely [expensive] and subject to errors, in addition to being difficult to customize to different domains.

日本語文はすさまじいものがありますが、「高価=expensive」は確認できました。

さて、「手作業」は専門語辞書で「manual operation」となっていますが、前置詞を確認したいので翻訳メモリを検索してみます。

(図 3)

残念ながら「参照用ユーザー翻訳メモリ」にはヒットしませんでした。翻訳スタジオに付属している「科学技術翻訳メモリ」を検索したら 9 件ほど見つかりました。

このように、初期画像位置合わせのための各パラメータは、[手作業で]決定され、次に、「理想的」減算画像が得られる。

Thus, each parameter for the initial image matching is determined [manually] and "ideal" subtraction images can then be

obtained.

これを見ると、「手作業で=manually」となっています。これなら前置詞が不要ですっきり表現できそうです。さらに「Web 検索」機能を使って調べても良いでしょう。

こんな調子でチェック作業を進めることができます。これを参考にして訳文を完成させてみましょう。

今回説明したことを実行するには、Web 等から必要な原文を収集したり OCR で印刷物を電子化したりするスキル、その原文ファイルを適切な形式に整形するスキル、PC-Transerでの「翻訳設定」、「専門語辞書自動選択」機能の実行、翻訳メモリの一括登録、翻訳メモリの設定、翻訳メモリペインでの検索、「Web 検索」機能の実行などのスキルが必要です。

これらの操作をスムーズにできるようにするには適切なトレーニングが必要ですが、少しずつ練習すれば誰でも役に立つテクニックを身に付けることができます。役立ちます。

しかし、これらの資源を手作業で構築することは高価である。					
However, it is expensive to build these resources by manual operation.					
件数	原文	訳数	訳文	メモリファイル	類似度
1	英語の大きな注釈がついているコーパスを構築すること:	1	Building a Large Annotated Corpus of English:	MT参照用(英日)	16% 2009

(図 1)

原文	しかし、これらの資源を手作業で構築することは高価である。					
	However, it is expensive to build these resources by manual operation.					
	件数	原文	訳数	訳文	メモリファイル	類似度
訳文	79	しかしながら、異なる領域にカスタマイズするのが困難であることに加えて、個々のスコリングは、極めて高価およびエラーに從属する全機械翻訳システムを通る遅い、手コード化流路をすることによって実行された。	79	However, the individual scoring was performed by doing a slow, hand-coded pass through an entire machine translation system which is extremely expensive and subject to errors, in addition to being difficult to customize to different domains.	MT参照用(英日)	2% 2009
	検索結果					

(図 2)

原文	しかし、これらの資源を手作業で構築することは高価である。					
	However, it is expensive to build these resources by manual operation.					
	件数	原文	訳数	訳文	メモリファイル	類似度
訳文	6	このように、初期画像位置合わせのための各パラメータは、手作業で決定され、次に、「理想的」減算画像が得られる。	6	Thus, each parameter for the initial image matching is determined manually and "ideal" subtraction images can then be obtained.	科学技術翻訳メモリ	6% 2008
	検索結果					

(図 3)