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 2 回　多段階翻訳処理って何？

小室誠一（MT 研究会主宰：http://www.babel.co.jp/mtsg/）

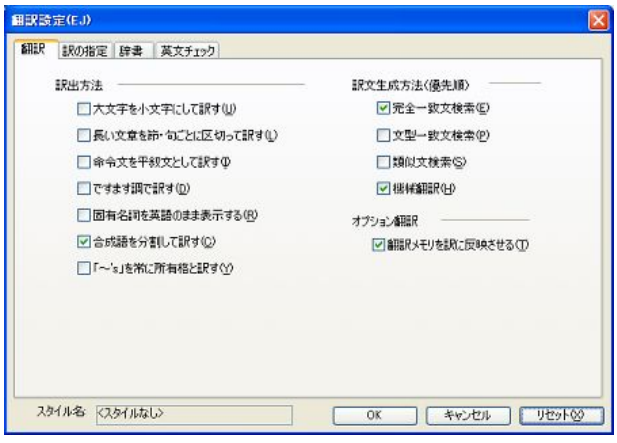
前回、翻訳生産性を向上させるために「多段階翻訳処理」が有効であると書きましたが、実際にどのようなものかピンとこないという方も多いかと思います。
そこで、今回は PC-Transer を使って具体的に基本原理を説明します。翻訳メモリ機能が付いているものであれば、PC-Transer 翻訳スタジオに限らず、PAT-Transer、MED-Transer、Legal Transer でもできますので、お持ちの方は実際に操作してみましょう。

TM（翻訳メモリ）と MT（機械翻訳）をシームレスに連動して訳文を生成する方式を、筆者は「多段階翻訳処理」と呼んでいますが、現在、翻訳ソフトでこの機能を搭載しているのは PC-Transer だけといってもいいでしょう。他にも、翻訳メモリと連携する機能を持った翻訳ソフトもありますが、使いやすさは今一つです。

それでは早速 PC-Transer を起動してみましょう。ここでは PC-Transer 翻訳スタジオ 2008 を使用します。

●翻訳設定（訳文生成方法）

翻訳エディタは「英日」モードにします。
メニューバーの「翻訳」→「翻訳設定」をクリックして「翻訳設定(EJ)」を開きます。
「翻訳」の設定画面の右側に「訳文生成方法」があるのがおわかりでしょうか。



（「翻訳設定」画面）

ここで、この設定に関するヘルプを見てみましょう。

ヘルプから引用----------

■ 訳文生成方法（優先順）

ここでは、翻訳メモリを使用した翻訳の場合の検索方法を選択します。
チェックされた検索方法による翻訳メモリ検索が上から順番に実行され、最初に見つかった検索結果が訳文として採用されます。
選択する検索方法は、次のとおりです。

完全一致文検索　－ 翻訳メモリで[完全一致文検索]を実行する
文型一致文検索　－ 翻訳メモリで[文型一致文検索]を実行する
類似文検索　　－ 翻訳メモリで[類似文検索]を実行する
機械翻訳　　－ Transer EJ が翻訳する

設定画面ではこれらの検索方法のどれを使うか選択できるようになっています。
初期設定では「完全一致文検索」と「機械翻訳」にチェックが入っています。

ここでは「文型一致文検索」にもチェックを入れます。つまり、「類似文検索」以外はすべてチェックを入れるということです。

「多段階翻訳処理」の訳文生成ステップではあくまでも自動生成を行うのが目的ですので、最初から原文の意味と違った訳文になることがわかっている「類似文」に置き換えることはしません。(ただし、類似文検索機能は、訳文修正ステップで活用します)。

●翻訳メモリ機能

例文を W3C の WEB ページから拝借することにします。
（http://www.w3.org/TR/html4/types.html）
「6 Basic HTML data types」というページの後半「6.12 Link types」のセクションに、以下のような同じパターンのセンテンスが 14 個ほど出てくるところがあります。

Start
　Refers to the first document in a collection of documents.　～

箇条書き項目が文の主語になっています。これを翻訳エディタに読み込むと、主語と述語が分割されてしまいますので、機械翻訳しても正しい訳文は出力されません。この場合、幸いにも動詞 Refers に三人称単数現在の「s」が付いているおかげで、訳文はなんとか命令形にならずに済みますが...。

さて、このような場合は翻訳メモリを活用したいところです。
メニューバーの「翻訳メモリ」→「設定」をクリックし、「翻訳メモリ設定」画面の「翻訳メモリ」タブをクリックして使用する翻訳メモリを設定します。「追加」ボタンをクリックして適当な名前を入れて「開く」をクリックすると新規作成できます。さらに「登録用」ボタンをクリックするとこの翻訳メモリに登録できるようになります。これ以外の翻訳メモリは解除してください。設定方法がわかりやすくて簡単ですね。

設定が済んだら、対訳エディタに以下の原文と訳文をコピー・アンド・ペーストします。

Refers to the first document in a collection of documents.
文書群の中の最初の文書を指す。

メニューバーの「翻訳メモリ」→「登録」→「カレント文を登録」でこの対訳を翻訳メモリに登録します。

（「翻訳メモリ登録」画面）

さて、それではもう一度上記の文を翻訳してください。「翻訳」ボタンをクリックすると訳文の色がグリーンに変わったのがおわかりでしょうか。完全一致文検索にマッチしたことを表しています。

それでは、以下の文はどうでしょうか。

Refers to the previous document in an ordered series of documents.

単語が 3 つ異なっているので完全一致文検索にはマッチせず、自動的に機械翻訳に切り替えられて、以下のように出力されました。

「きちんとした一連の文書で、前の文書を参照する。」

訳文は出力されたものの、Refers to の訳にばらつきが出ました。

試しにこの文と翻訳メモリとのマッチ率を調べると「45%」でした。デフォルトでの一致率は 80%になっていますので、たとえ類似文検索をしたとしてもマッチしません。

このように、翻訳メモリと機械翻訳を連動すればとりあえず訳文は出力されるものの、これでは翻訳メモリの有効性が極めて低いとい

うことになってしまいます。

●文型一致文検索の活用

翻訳メモリの有効性を高めてくれるのが「文型一致文検索」です。
この機能こそ、TM と MT の統合技術で、翻訳メモリソフトにはまねのできない訳文生成方法です。PC-Transer ユーザーでこの機能を活用したことがない人はずいぶん損をしていることになります。

文型一致文検索を使うには、翻訳メモリに文型登録する必要があります。といっても難しくはありません。文の共通要素はそのままにして、変動部分にタグを付けるだけです。

ここでは、以下のようにタグを付けます。

Refers to <$1>.
<$1>を指す。

このパターンに一致した場合、<$1>の部分の英文を機械翻訳エンジンで日本語に訳し、訳文の<$1>に挿入して出力してくれます。

さて、これで先ほどの文を訳してみましょう。
Refers to the previous document in an ordered series of documents.

「きちんとした一連の文書の前の文書を指す。」と出力されました。
しかも「を指す。」の文字が青色になっています。文型一致がマッチしたことを表しています。少なくとも「Refers to」の訳は統一されるようになりました。

あとは機械翻訳の際に参照されるユーザー辞書を整備すれば訳文の完成度がアップします。この場合、ユーザー辞書に「previous document in an ordered series of documents（続き物の、前の文書)」が名詞として登録されていれば理想的です。

このようにユーザー辞書に登録し、文字色を赤色に設定して訳せば、

「続き物の、前の文書を指す。」

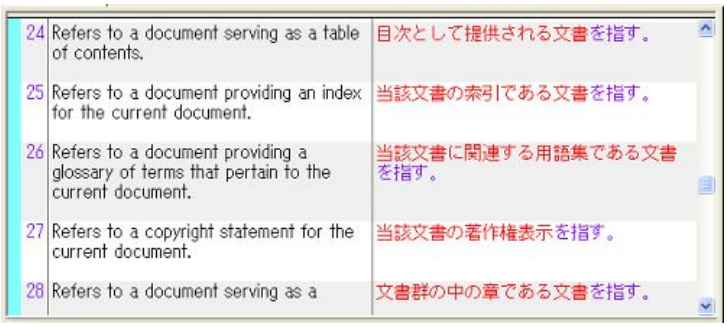
と出力され、「続き物の、前の文書」が赤色、「を指す。」が青色で表示され、これが翻訳メモリとユーザー辞書が反映された訳文であり、基本的に修正不要であることが一目でわかります。

以降の同じパターンの文も、ユーザー辞書を整備してから自動処理してみましょう。

Refers to the next document in a linear sequence of documents.
続き物とされている文書群の、次の文書を指す。

Refers to a document serving as a table of contents.
目次として提供される文書を指す。

Refers to a document providing an index for the current document.
当該文書の索引である文書を指す。



24	Refers to a document serving as a table of contents.	目次として提供される文書を指す。
25	Refers to a document providing an index for the current document.	当該文書の索引である文書を指す。
26	Refers to a document providing a glossary of terms that pertain to the current document.	当該文書に関連する用語集である文書を指す。
27	Refers to a copyright statement for the current document.	当該文書の著作権表示を指す。
28	Refers to a document serving as a	文書群の中の章である文書を指す。

みなさんは、翻訳メモリはセンテンス単位で対訳を登録するもの（もちろん見出しなどのフレーズの場合もありますが)、ユーザー辞書は単語あるいは数語からなる連語を登録するものと思いこんでいませんか？

実は、翻訳メモリはフレーズを格納するフレームであり、ユーザー辞書はフレームにはめ込むフレーズ（チャンク）を蓄積するものであると考えることが活用への第一歩なのです。

「多段階翻訳処理」を行うには、適切な翻訳メモリと十分なユーザー辞書の蓄積が必要です。そう考えると、事前準備を効率よく行うテクニックが不可欠であることがわかるでしょう。