

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4回 フレーズ辞書の効率的の良い作成法

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

前回、「文型翻訳メモリ」と「フレーズ辞書」を十分に蓄積すれば、驚くほど生産性が向上すると書きましたが、実はフレーズ辞書はそれだけではなく、機械翻訳の逐語訳を自然な直訳に改善する効果もあるのです。今回は、PC-Transer 翻訳スタジオの付帯機能を活用して素早くフレーズ辞書を作成する方法を検討してみましょう。

He emphasized that the U.N. would play a positive role in supporting the reconstruction of Iraq.

いきなりですが、この英文を PC-Transer 翻訳スタジオ 2008 で訳してみると以下のようになります。

「彼は、国連がイラクの再建を支持することで積極的な役割を演ずると強調した。」

「positive=ポジティブな」を「積極的な」に訳語変更しましたが、それ以外は手を加えていません。意味がよくわかる正しい訳文になっているものの、逐語訳のためにどこか不自然です。

そこで「play a positive role (動詞)」

「supporting the reconstruction of Iraq (名詞)」をユーザー辞書に登録して再翻訳してみました。

「彼は、国連がイラク復興支援に積極的な役割を果たすことを強調した。」

最終的に「would」のニュアンスを付加するために、「す」を「していく」と後編集して「～役割を果たしていくこと～」とすれば「翻訳」の完成です。

このようにフレーズ辞書には、逐語訳を後編集しやすい自然な直訳にする効果もあります。

●対訳ファイルからフレーズ辞書を作成する

すでに対訳ファイルが手元にある場合は、手作業でフレーズ辞書を作成しましょう。

事前に書き込み用のユーザー辞書を作成しておいてください。
対訳ファイルを翻訳エディタに読み込んでから以下の作業を行います。

- (1) 訳語を範囲選択して、「Ctrl+C」キーを押してクリップボードにコピー
- (2) 原文の見出しにする部分を範囲選択して「辞書登録」ボタンをクリック
- (3) 「辞書登録」ダイアログが開く（すでに見出し語が入力された状態になっている）
- (4) 「訳語」のテキストボックスにカーソルを置いて、「Ctrl+V」で訳語を貼り付ける
- (5) 「品詞」を選択し、必要であれば活用形を修正して「登録」をクリック

「辞書登録」にショートカットキーを割り当てておけば一々「辞書登録」ボタンをクリックしなくて済みます。

ショートカットの割り当ては、「ツール」→「カスタマイズ」→「アクセラレーターキー」をクリックします。デフォルトで「辞書参照」が「Ctrl+D」になっているので、筆者は「Alt+D」を「辞書登録」に割り当てています。

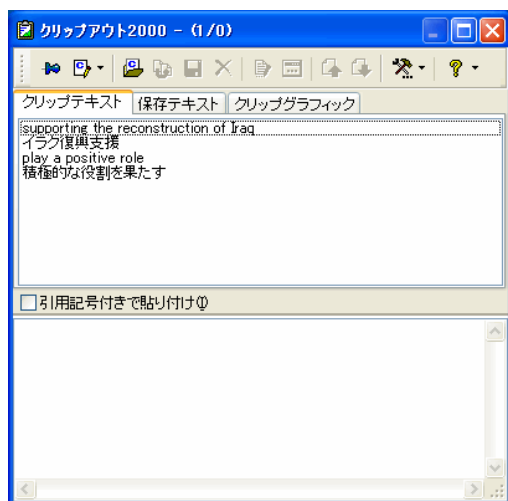
原文と訳文が別々のファイルになっていて、対訳ファイルを作成する時間がない場合は、訳語と見出し語をコピーして貼り付けて行く

ことになります。

しかし、テキストエディタなどで原文と訳文のファイルを開いたうえに、さらに貼り付け用のファイルを開いて作業するのは面倒です。これなどは「クリップボード監視ソフト」を使うと楽に作業できます。

いろいろなソフトがありますが、筆者は「クリップアウト 2000」（フリーソフト）を使っています¹。

範囲選択して「Ctrl+C」を押せば一行ごとに「クリップテキスト」ウインドウに貼り付けられます。あとはテキストエディタや Word の検索置換で辞書ソースの形式に整形するだけです。



●出現頻度リストから辞書を作成する

それでは原文だけしかない場合にはどうしたらいいでしょうか。

PC-Transer には「頻度リスト出力」という便利な機能があります。この機能を活用してみましょう。

まず、原文（英文）を翻訳エディタに読み込んでおきます。翻訳はしなくて結構です。

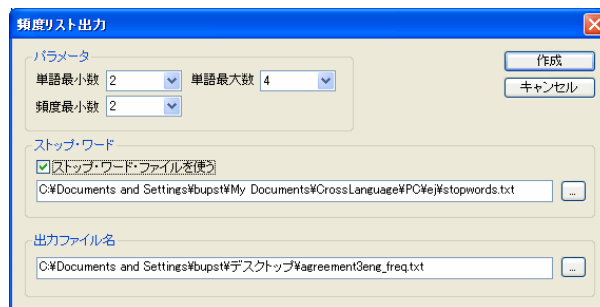
- (1) 「ツール」→「頻度リスト出力」をクリック
- (2) 「頻度リスト出力」ダイアログが開いたらパラメータを設定

¹ 「クリップアウト 2000」

<http://www.vector.co.jp/soft/dl/win95/util/se125672.html>

例えば、

- ・単語最小：2、単語最大：4、頻度最小：2（2語～4語で2回以上出現するものをリストアップする）
- ・「ストップワード・ファイルを使う」にチェックを入れる
- ・出力ファイル名を指定して「作成」をクリック

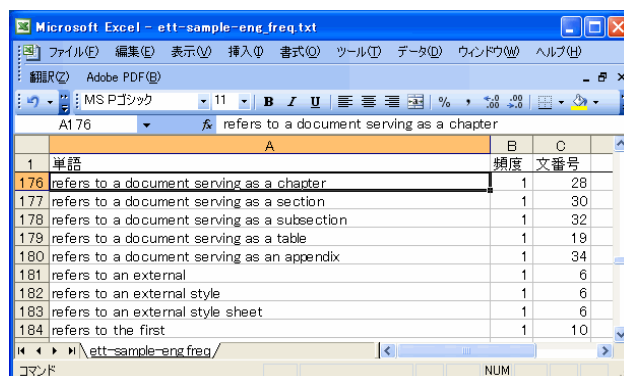


- (3) 「頻度リストが作成されました」というメッセージが表示されたら「OK」をクリックしてダイアログを閉じる

- (4) 出力された頻度リストファイル（～freq.txt）を Excel で開く

- ・「ファイルを開く」ダイアログでファイル形式を「テキスト ファイル」にする
- ・ウィザードが開いたら元データの形式を「カンマやタブなどの区切り文字によってフィールドごとに区切られたデータ」にチェックを入れる
- ・区切り文字の指定画面が表示されたら「タブ」と「カンマ」にチェックを入れる

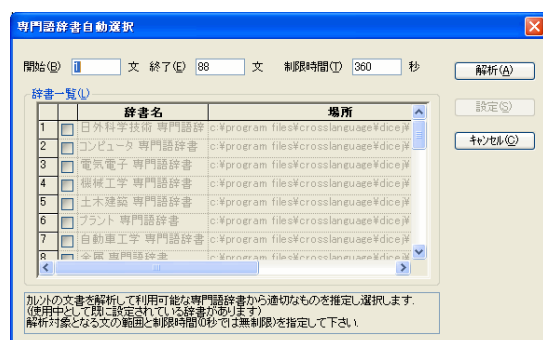
- (5) Excel で開いたら、「単語」の列で並び替えて不要な見出しを削除する



- (6) 見出しにする単語（フレーズ）をコピーする

- (7) 翻訳スタジオで「ファイル」→「新規作成」をクリック

- (8)原文側のウインドウにカーソルを置いて、「編集」→「1行1文貼り付け」をクリック
 (9)見出しとなる語句が原文側のウインドウに読み込まれたら「翻訳辞書」→「専門語辞書自動選択」をクリック
 (10)「専門語辞書自動選択」ダイアログが開いたらそのまま「解析」をクリック



- (11)辞書名にチェックが入ったら「設定」をクリック
 (12)「翻訳設定」ダイアログが開く（すでに選択された専門語辞書が「使用中の辞書」に追加されているのを確認する）
 (13)さらに、「翻訳」タブをクリックして「訳出方法」の設定で「命令文を平叙文として訳す」にチェックを入れて「OK」をクリック
 (14)翻訳エディタの翻訳ボタンをクリックして全文を翻訳する
 (15)出力結果をチェックし、不適切な訳語を修正する（「訳語変更」を利用）
 (16)「名前を付けて保存」で、ファイルの種類を「対訳ファイル（カンマ区切り）(*.csv *.txt)」にして保存する

頻度リストは英文だけでなく和文でも出力できます。日本語の場合は「単語」ではなく文字の出現頻度から語句を抽出します。

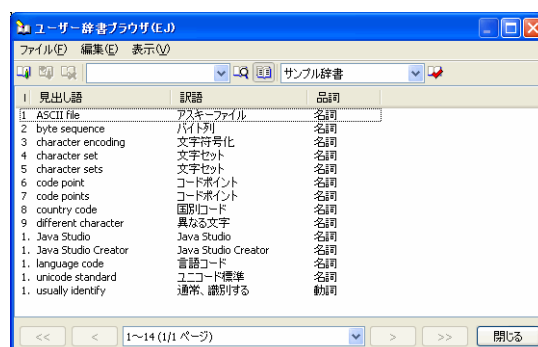
●ユーザー辞書を修正する

カンマ区切りやタブ区切りのファイルからユーザー辞書に一括登録すると「名詞」扱いになります。辞書ソースの中に名詞以外の語句が含まれる場合は、ヒットしなかったり副作用があったりしますので修正する必要があります。

そのような時に役立つのが「ユーザー辞書ブラウザ」です。

「翻訳辞書」→「ユーザー辞書ブラウザ」を

クリックするとブラウザが表示されます。ブラウザには見出し語、訳語、品詞が一覧表示されます。修正したい語句をダブルクリックすると、「辞書登録」ダイアログが開いてすぐに修正できます。



今回は、PC-Transer の「頻度リスト出力」「1行1文貼り付け」「専門語辞書自動選択」「ユーザー辞書ブラウザ」機能や、Excel、クリップアウト 2000などを組み合わせて、フレーズ辞書を迅速に作成する方法を検討しました。詳しい使い方はそれぞれのマニュアルやヘルプをご覧ください。