

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 8 回 翻訳ソフト活用のためのテキスト処理入門(2)

2009 年 2 月 小室誠一 (MT 研究会主宰 : <http://www.babel.co.jp/mtsg/>)

PC-Transer 翻訳スタジオを活用するには、翻訳メモリやユーザー辞書を充実させることが大切です。そのためにはインポートや一括登録機能を活用する必要がありますが、対訳ファイルインポートするには一定の形式に整形しなくてはなりません。そこで「検索・置換」の出番となります。テキストを柔軟に加工するためには正規表現を使った検索・置換機能を使いこなすスキルが必要です。翻訳スタジオにも「正規表現」に対応した検索・置換機能が搭載されているので、しっかり活用しましょう。

みなさんの中で、ワープロやテキストエディタの「検索・置換」を使ったことのない人はまずいないでしょう。でも、「正規表現」を利用したことのない人はいるかもしれません。

もしかしたら聞いたことがないという人がいるかもしれないので、念のために説明すると、正規表現 (regular expression) とは、メタキャラクタという特別な意味を持った記号文字列を使ってパターンを表記する方法です。これは、テキストエディタなどの検索・置換で利用できますが、ソフトウェアによって多少の違いがあります。と言っても、基本的な記述法はほとんど共通です。

例えば正規表現では、英単語を検索するとき、検索文字列を、

`receiv(e|es|ed|ing)`

のように入力すれば、`receive`、`receives`、`received`、`receiving` にマッチします。これは `receiv` に () 内の | で区切られた文字列のどれかを加えたものにマッチすることです。活用形を含めた検索をする場合には必須のテクニックと言えるでしょう。

また、`19[0-9][0-9]` と検索文字列を入力すれば、`1900`～`1999` までの数字にマッチします。これは `[]` で囲まれた文字のどれか 1 文字にマッチすることを表します。「`0-9`」は、`0` から `9` までの範囲を意味します。

以下のような、上が英語で下が日本語の対訳ファイルがあつて、これを上下入れ替えたい場合、どうしたら良いでしょうか。

「正規表現」を使えばあっという間にできてしまいます。

=====

--/H

We were flattered by the letter we receive
on the 15th of Jan. 2009.

2009 年 1 月 15 日付け貴書を拝受致しました。

=====

試しに「秀丸エディタ」を使って、以下のように入力して検索置換してみました。

検索文字列 : `--/H¥n(.*)¥n(.*)¥n`

置換文字列 : `--/H¥n¥2¥n¥1¥n`

「全置換」をクリックすると、一瞬で入れ替わりました。

=====

--/H

2009 年 1 月 15 日付け貴書を拝受致しました。

We were flattered by the letter we receive
on the 15th of Jan. 2009.

=====

ここでは、検索文字列の `.` (ピリオド) は、任意の一文字を表します。`*` は直前パターンの 0 回以上のくり返し、`¥n` は改行です。() は括られた範囲を記憶します。

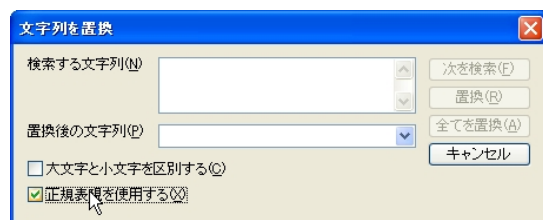
置換文字列の ¥2 は、検索文字列の 2 番目の () の中身を挿入することを表します。¥1 は、1 番目の () の中身を挿入します。結果として、日本語と英語が入れ替わったということです。これなどは、翻訳メモリにインポートするためのソースファイルを作成するときなどに利用できそうです。

このように「正規表現」を使うと、通常の検索・置換ではできないようなことが簡単にできてしまいます。

ちなみに、MS-Word のワイルドカードも正規表現の一種です。書き方がちょっと違いますが、同じようなことができます。Word では、改行は「¥n」ではなく、「^13」と書きます。また、任意指定の行区切りは「^11」です。

実は、このような「正規表現」を使った検索・置換は、PC-Transer 翻訳スタジオでも行うことができます。

翻訳エディタの「編集」メニューにある「検索」「置換」機能、「フレーズ/語」メニューにある「一括フレーズ」が正規表現に対応しています。さらに、「検索」では、改行で区切ることで複数の検索語を指定できます。



前処理フィルタ (ejpre.txt、jepre.txt) と後処理フィルタ (ejpost.txt、jepost.txt) でも正規表現が使えます。フィルタはテキストファイルなので自由に追加削除ができます。記述形式は以下のようになります。

s/検索文字列/置換文字列/g (改行)

最後の g はマッチする文字列がなくなるまで処理をする、つまり「すべてを置換する」という記号です。

例えば、マイクロソフトがカタカナの長音記号を付ける方式に変更しましたが、これに対応するには、以下のように記述します。

s/アクセサ/アクセサー/g
s/アクタ/アクター/g
s/アクティベータ/アクティベーター/g
s/アグリゲータ/アグリゲーター/g

このように改行を入れて検索文字列と置換文字列を複数記述しておくことができるので便利です。前処理フィルタは原文ファイルを開く時に、後処理フィルタは訳文を保存する時に実行されます。うまく工夫すれば、ある程度自動的な前編集や後編集ができます。活用してみましょう。

正規表現で利用できるメタキャラクタなどの詳細は、「スタート」ボタン→「すべてのプログラム」→「Cross Language」→「PC-Transer 翻訳 Studio2009」

→「ヘルプ」の『指定できる正規表現』をご覧ください。

翻訳エディタのメニューにある「ヘルプ」には書かれていません。

駆け足でしたが、テキスト処理の基本が正規表現を使った検索・置換だということがご理解いただけましたでしょうか。最初は難しそうですが、初歩的な使い方だけでも覚えれば、業務の効率化に大いに役立ちます。