

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7回 翻訳ソフト活用のためのテキスト処理入門(1)

2009年1月 小室誠一 (MT 研究会主宰 : <http://www.babel.co.jp/mtsg/>)

新年を迎えたところで、初心に帰ってテキスト処理の基本を復習することにしましょう。高機能な PC-Transer 翻訳スタジオを使いこなすには、初歩的な PC 操作スキルが不可欠だからです。今年こそは、PC-Transer を十分に使いこなして、この厳しい時代を乗り切りましょう。

翻訳生産性を劇的に向上させるためには、翻訳支援ツールの活用が有効です。しかし、翻訳支援ツールを使うには基本的な PC 操作ができることが前提であることは言うまでもありません。無駄のない操作がしっかりできるようになって、さらに効率化を求めたときに翻訳支援ツールを活用するのが物事の順序というものです。ところが、意外に基本ができていない人が多いのが実情です。

例えば、ファイルの拡張子を表示する方法、Word の編集記号を表示する方法、フォルダの作成方法、ファイルの関連付けの方法などを知らない人や、ショートカットキーを使わない人などがいます。このような最低限の操作をマスターしておかなくては生産性の向上どころではありません。

もっとも、このメルマガの読者のみなさんは、このあたりの基本スキルは十分にクリアしているものと思います。それを前提に話を進めて行きます。

●翻訳支援ツールを活用するために必要なスキル

翻訳支援ツールを十分に活用するために必要なスキルを挙げてみます。

- (1)迅速に対訳ファイルを作成するスキル
- (2)迅速に用語集を作成するスキル
- (3)翻訳メモリ、ユーザー辞書を活用するスキル
- (4)自動生成された訳文を迅速に修正するスキル

翻訳支援ツールを使う場合、(3)と(4)が重視される傾向がありますが、実は、大幅な効率化を達成するには、(1)と(2)が極めて重要です。回り道のようにですが、対訳ファイルと用語集を十分に準備することが結果的に効率化につながります。これらのスキルは突き詰めればテキスト処理のテクニックと言うことになり、さらに言えば、検索・置換の機能を自由に使いこなして思い通りのテキスト整形ができるということです。

●素材の収集

対訳ファイルや用語集を作成するには、電子化されたテキストが不可欠です。今では、インターネットを使って大量のテキストデータを簡単に収集することができます。Google や Yahoo!などの WEB 検索エンジンを利用して、自分の専門ジャンルの文書をどんどん集めておきましょう。ファイル形式を指定して Excel や pdf ファイルを検索すれば、役に立つ用語集そのものが見つかることもあります。

たまに、印刷物は入力面倒だと敬遠する人がいますが、イメージスキャナと OCR を使えば楽に電子化できます。今ではイメージスキャナもずいぶん安価になっています。複合プリンターならスキャナ機能が付いています。また、最近の OCR は認識率が非常に高くなっているので、適切に設定すればほとんど修正する必要はありません。

PC-Transer 翻訳スタジオには「CROSS OCR」が付属しています。この OCR は単なるオマケではなく、十分に実用的です。みな

さんは活用していますか？

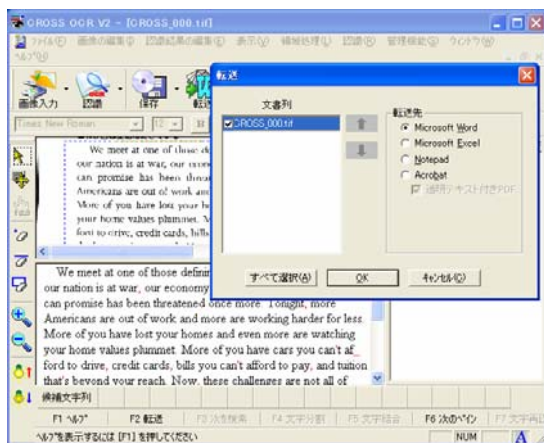
試しに手元にある雑誌を「CROSS OCR」で文字認識してみました。

We meet at one of those defining moments, a moment when our nation is at war, our economy is in turmoil, and the American promise has been threatened once more. Tonight, more Americans are out of work and more are working harder for less. More of you have lost your homes and even more are watching your home values plummet. More of you have cars you can't afford to drive, credit cards, bills you can't afford to pay, and tuition that's beyond your reach. Now, these challenges are not all of government's making. But the failure to respond is a direct result of a broken politics in Washington and the failed policies of George W. Bush.

まったく手を加えていません。きれいな印刷物ならこのようにほとんどそのまま使用できるレベルです。

「Ameri_can」「af_ford」「gov_ernment's」は原本ではハイフネーションの部分です。誤認識ではありません。このようにハイフンがアンダーバーに変換されているので、検索・置換を使えば簡単に修正できます。

原文のレイアウトをおおまかに再現したい場合は、「転送」機能を使って Microsoft Word に転送すると良いでしょう。また画像や PDF ファイルから文字認識することもできますので、インターネットで収集した文字画像も手入力する必要はありません。



これまで翻訳した訳文はたくさんあるものの、原文が印刷物なので、翻訳メモリとして再利用するのをあきらめていた人は、是非 OCR を使って原文を電子化しておきましょう。

「CROSS OCR」は英語と日本語に対応しており、原稿の種類によってどちらかに切り替えて認識しますが、一つの文書に日本語と英語が混ざり合っているような場合は、日本語モードで認識してみてください。ほとんど問題なく読み取れます。これまで無理だと思っていた、愛用の用語集なども簡単に電子化できます。

さて、このようにして集めたテキストは、そのままでコーパスや簡易用語集として十分に利用できます。テキストエディタの GREP 検索やコンコーダンサーなどを使って効率良く検索したり文章を分析したりしてみましょう。

これだけでもかなり役に立ちますが、さらに活用するために PC-Transer にインポートするには一定の形式に整える必要があります。

ここで登場するのが検索・置換機能です。Word のオプション検索やワイルドカード、テキストエディタの正規表現による検索・置換を使うことで、どのような形式にも自在に整形することができます。このテクニックをマスターすることが翻訳支援ツール活用の第 1 歩となります。

サバイバル・ツールとしての PC-Transer を使いこなすには、思い切って既成概念を捨て去る必要があります。これからの時代は人より少し「翻訳力」があるだけでは不十分です。新しいアイディアをしぼり出して PC-Transer を活用することで、他の翻訳者には追従できない高度な翻訳生産力を実現しましょう。

次回から、具体的にテキスト処理のテクニックを説明して参ります。ご期待ください。