

3. ユーザ辞書を素早く作成する方法

Excelなどで作成された既存の対訳用語集を、ユーザ辞書に一括登録して使用することは、だれでも行っていると思いますが、それではそのような用語集がない場合はどのようにしているのでしょうか。

翻訳ソフトを使用する際にはできるだけ翻訳対象文に特化した過不足のないユーザ辞書が不可欠です。たとえ既存の用語集があったとしても、あまりマッチしなければ気休めにしかありません。適切な辞書を作るためには、翻訳対象の原文を解析して見出しを抽出し、適切な訳語をつける作業を行う必要があります。それも迅速にです。

翻訳ソフトを使用する際にはできるだけ翻訳対象文に特化したユーザ辞書を作成する必要があります。PC-Transer 翻訳スタジオ 2007 には効果的なユーザ辞書を素早く作成するための機能が搭載されています。

「未知語リスト出力」機能を使えば製品名などの固有名詞を抽出することができます。また、「頻度リスト出力」機能でキーワードを抽出し、翻訳エディタで訳語を取得し、ユーザ辞書の一括登録を行えば、過不足のないユーザ辞書を効率よく作成できます。今回も英日翻訳について説明します。

未知語リスト出力

翻訳スタジオ 2007 には基本語辞書だけでも英日 105 万語(日英 124 万語)が登録されているので、ほとんどの英単語や熟語は未知語としてリストアップされることはないでしょう。ただ、会社名や製品名などは辞書に登録されていないことが多いので、この機能でリストアップできます。

ユーザ辞書を作成する際には真っ先に未知語リスト出力を行いましょう。翻訳時点で訳語が決定していない場合でも、訳語を原語のまま登録しておくで管理しやすくなります。

リストの形式は、「原語：」です。拡張子は「.opt」となります。コロンの右側に訳語を入れて上書き保存するだけで辞書登録のソースファイルになります。固有名詞は通常、名詞で複数形にならないので、この形式で十分です。

未知語リストを出力するには、ツールバーの「ツール」 「未知語リスト出力」をクリックします。

頻度リスト出力

翻訳対象の原文からキーワードを抽出して辞書を作れば、理論的には過不足のない辞書ができます。全文を単語に分解し、重複を除いて辞書の見出しにするのは簡単ですが、あまり有効な辞書とは言えません。というのも、英語の場合は語と語の相性が強い言語なので、できるだけ連語で登録するのが望ましいからです。特に「形容詞 + 名詞」や「名詞 + 前置詞 + 名詞」といった名詞句をできるだけたくさん登録すると訳文の品質が格段に向上します。

連語の抽出はなかなか難しいのですが、「頻度リスト出力」機能を使えば、単語の頻度だけではなく、指定した語数の頻度もカウントしてくれるので、簡単に連語を抽出することができます。

辞書の検索は最長一致なので、より語数の多い見出し語にマッチすればその訳語が採用されます。そこで、先ず単語レベルの頻度リストを出力し、さらに3語から5語程度の連語レベルの頻度リストをして、それぞれの辞書を作成してみると良いでしょう。

頻度リストを出力するには、「ツール」「頻度リスト出力」をクリックします。

頻度リストの出力には、パラメータを指定する必要があります。「単語最小数」と「単語最大数」を指定することで、何語から成る連語をリストアップするか決定できます。また「頻度最小数」によって低頻度の語句を除くことができます。「ストップ・ワード・ファイル」を使用して、前置詞や代名詞などの基本語句をカウントの対象からはずすことで、より重要なキーワードをリストアップすることもできます。

一括訳語出力

頻度リスト出力した結果は、英語の単語（連語）がリストアップされているだけなので、辞書にするには訳語をつける必要があります。一つ一つ手入力しても結構ですが、数が多くなるとそうもいきません。そこで、見出し語を翻訳エディタで開いて機械翻訳し、一度に訳語を付けてしましましょう。

翻訳の際には、基本語辞書だけでなく専門語辞書も使用すると良いでしょう。どの辞書を選択すれば良いか分からないときは、「専門語辞書自動選択」機能を使います。専門語辞書には必ずそれぞれフォントの色を指定しておき、どの辞書の訳語が採用されたか分かるようにしておきます。

「専門語辞書自動選択」を行うには、「翻訳辞書」「専門語辞書自動選択」をクリックします。

「解析」すると、専門語辞書一覧の中の適切であると推定された辞書にチェックが入ります。ここで「設定」ボタンをクリックすると翻訳設定の辞書画面が開き簡単に設定できます。

また、翻訳設定の翻訳画面で訳出方法の「命令文を平叙文として訳す」にチェックを入れておくと、動詞の訳語が「～する」という形になり、辞書の訳語にふさわしくなります。

翻訳エディタで訳語を出力したら、適切な訳語になっているかざっと確認します。専門語辞書ごとに色分けしてあるので参考になります。適切でない場合は、訳語リストを開いて変更しても良いでしょう。クリックだけで済みます。確認が完了したら「対訳ファイル(カンマ区切り)」で保存して辞書ソースとします。

ユーザ辞書の一括登録

辞書ソースを一定の形式で作成すると、一括でユーザ辞書に登録できます。ここからは、みなさんが普段行っている通りです。

原語と訳語がカンマで区切られた「CSV フォーマット」、タブで区切られた「TSV フォーマット」は

一番簡単な辞書ソースで形式ですが、すべて「名詞」として登録されます。この形式は、名詞以外の品詞には不適切です。

翻訳スタジオ 2007 の辞書機能の素晴らしいところは、詳細情報を辞書ソースに記述できることです。たとえば、見出し語と訳語の間に「(n,*)」を挿入すれば、名詞で規則活用するということになり、「(v,*)」を挿入すれば、動詞で時制が規則活用することになります。

名詞以外の品詞の語句が含まれる場合、テキストエディタや Excel を使って品詞情報を挿入するか、そのまま登録しておいて、「ユーザ辞書ブラウザ」を使って修正するか、いくつか対処法があります。とりあえず荒っぽく作成したユーザ辞書をどうやって修正していくかは次回取り上げる予定です。

記事の内容は筆者自身のノウハウに基づいております。記事の内容によって万一損害を被ることがあっても一切責任を負いません。また、この記事の内容に関して発売元の株式会社クロランページへの問い合わせはご遠慮ください。(小室誠一)