

翻訳者のための Linux

エラリー ジャンクリストフ
OmegaT プロジェクト現地化担当
株式会社 DOUBLET 代表取締役

翻訳業界では他の業界と同じように「ベンダーロックイン」を防ぐ様々な動きがここ数年見えてきた。しかし、90年代の半ば頃から Linux というデスクトッププラットフォームの可能性が十分にあると思った翻訳者たちによって典型的なプロプライエタリーソフトウェアである Windows から離れて、Linux のみで翻訳環境を整える活動も始まっていた。

ここでいう「翻訳」というのは、自由なソフトウェアの現地化やそのドキュメンテーションの翻訳などの活動ではなく、いわゆる「産業翻訳」を指している。

「産業翻訳」がカバーしている分野は非常に広くて、電化製品などの取扱説明書からマーケティングの文書や PR 活動で使われている多言語サイト、契約などの翻訳を意味している。

21世紀になって XML や JAVA、python などのテクノロジーの基づいて自由なソフトウェアを使用する翻訳者がやっとプロプライエタリー環境と比較しても恥ずかしくない「工具箱」を作ることが可能となり、翻訳業界もその周辺のツールやプロセスを認めつつあるといえよう。

しかし、その産業翻訳において使われている主なプロセスや自由なツール、標準規格などは自由なソフトウェアの世界では未だに広く知られていないため、テキストエディタなどを使って「ベタ打ち」で自由なソフトウェアの現地化やドキュメンテーションの翻訳を行っている方が大勢いる。

そういった曖昧なプロセスに基づいて作業環境の中では、既存翻訳データの再利用や参照は不便で、用語などの統一を図りきれない。そのため、成果物の品質や専門家（翻訳者）のボランティアの取得にも限界が見えているし、日本における自由なソフトウェアの普及に直接影響する問題だといえよう。

産業翻訳者が利用する自由ツールを使うからといって必ずしも品質が上がると言い切れないが、翻訳業界で検証された概念に基づいたプロセスを対応するツールだからこそ、自由なソフトウェアの翻訳初心者でも自分の能力を全て発揮できるし、複数人による作業も簡単となることによって翻訳の「バグ」も見つけやすくなるだろう。

ここでは、説明したいのは、まず、「翻訳メモリ」（翻訳既存データ）という概念とその関連の概念や規格だ。代表としては下記のようなものが挙げられる。

- ・ TMX（翻訳メモリ形式） LISA
<http://www.lisa.org/Translation-Memory-e.34.0.html>
- ・ XLIFF（翻訳対象データ媒介形式） OASIS
http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xliff
- ・ SRX（翻訳原稿分節化規則定義形式） LISA
<http://www.lisa.org/Segmentation-Rules-eXchange-SRX.40.0.html>
- ・ ITS（翻訳対処内容指定形式） W3C
<http://www.w3.org/International/its/ig/>
- ・ TBX（用語集形式） LISA
<http://www.lisa.org/Term-Base-eXchange-TBX.32.0.html>

そして、それぞれの概念や規格を適用する主な道具も紹介したい：

- ・ OmegaT（翻訳メモリベースのエディター）
<http://www.omegat.org/>
- ・ Okapi Framework（様々な形式の変換フィルターと検索・置換などの便利道具箱）
<http://okapi.opentag.com/>
- ・ Anaphraseus（OpenOffice.org のマクロで、翻訳メモリベースのエディター）
<http://anaphraseus.sourceforge.net/>
- ・ OpenLanguageTools（XLIFF エディター）
<https://open-language-tools.dev.java.net/>
- ・ OpenOffice.org（変換フィルターとして使われる Office スイート、スペル・

文法チェック機能)

<http://www.openoffice.org/>

参考：

翻訳者を支えるツール（翻訳支援ツール）の短い歴史をみると、自由化への動きも見えてくる：

- IBM の OpenTM（翻訳メモリベースのエディター）
<http://www.opentm2.org/>
- Sun の OpenLanguageTools（XLIFF エディターとフィルター）
<https://open-language-tools.dev.java.net/>

それと同時に、自由なソフトウェアの世界も少しずつ翻訳メモリという概念を取り入れて開発が進んでいるし、その関連の規格に対応しようとする動きも強い。

自由なソフトウェアの世界でよく使われている道具は下記のようなのはある：

- gettext（po 形式の変換フィルター、po 構造チェック機能など）
<http://www.gnu.org/software/gettext/>
- po4a（po 形式の変換フィルター）
<http://po4a.alioth.debian.org/>
- Poedit（po 形式の専用エディター）
<http://www.poedit.net/>
- TranslateToolkit（po 形式の変換フィルター、po 構造チェック機能など）
<http://translate.sourceforge.net/wiki/toolkit/index>
- Pootle（サーバー側の po データ管理ツール、オンライン翻訳エディター）
<http://translate.sourceforge.net/wiki/pootle/index>
- Lokalize（翻訳メモリベースの XLIFF/po エディター、KDE 専用）
<http://userbase.kde.org/Lokalize>
- virtaal（翻訳メモリベースの XLIFF/po エディター）
<http://translate.sourceforge.net/wiki/virtaal/index>