

QRedit: 英日ボランティア翻訳者向け統合エディタ

阿辺川 武 (東京大学大学院教育学研究科)

影浦 峯 (東京大学大学院教育学研究科)

1 はじめに

我々は、現在、オンライン文書を翻訳しオンラインで翻訳文書を公開するボランティア翻訳者を支援するシステムを開発している。

高精度な全自動の機械翻訳を実現することの困難さが認識されたこともあって [9]、これまでに、人間の翻訳者を支援する様々なシステムが開発され、提供されている [1, 6]。中でも、TRADOS[10]をはじめとする、翻訳者向けのデータ管理システムは、翻訳会社やその契約翻訳者により、現実によく使われている。しかしながら、我々が対象とするオンライン・ボランティア翻訳者は、こうしたシステムを使わない [7]。また、英国でも同様に、フリーランスの翻訳者はほとんどこうしたシステムを用いていない [4, 5]。ボランティア翻訳者がこれらのシステムを使わない理由として、そうしたシステムは、有料で金銭の取得を主目的としない翻訳者には高すぎる¹、 unnecessary機能まで付きすぎていて作業環境がかえって煩わしくなる、自分が必要な機能はついていない、データ・コントロールのために支援・管理システムを使えという翻訳会社やクライアントからの圧力がない、などがあげられる。しかしながらその一方で、これらボランティア翻訳者は、必ずしも自分たちの作業環境に満足しているわけではない。したがって、オンライン・ボランティア翻訳者の要望を反映した支援システムがあれば、それは活用される可能性が高い。

本稿では、こうした背景から我々が開発中のボランティア翻訳者支援システムのうち、翻訳エディタ・インタフェースについて紹介する。まず、我々が想定するボランティア翻訳者の位置づけを明確にしたのち、ボランティア翻訳者を支援するシステムの基本要件をまとめる。その後、それにしたがって実装した単純なエディタ・インタフェースについて紹介する。

2 翻訳者の二類型

翻訳者の類型には様々なものが考えられるが、ここでは議論を明確にするため、あえて二類型だけを考える。

第一の類型は、翻訳の専門的な訓練を受け、特化した

¹Omega-T[12]のようなフリーの翻訳メモリーもあるが、やはりそれほど浸透していない。

文書を対象に職業として翻訳を行うタイプであり、例えば2つの公用語を持つカナダの政府文書を扱う翻訳者などが典型的にこれにあたる。こうした翻訳者はしばしば、元文書の言語に対しても翻訳先の言語に対しても、ネイティブ・レベルの言語力を持ち、扱う文書の内容についても深い知識を有している。また、職業として翻訳を行っているため、「時は金なり」の原則で翻訳環境を評価する²。こうした翻訳者をターゲットとした場合、元文書の意味を理解するプロセスや関連する事実情報をチェックするプロセスの比重は極めて低く、翻訳文書入力の手間を省くことが翻訳支援システムに求められることになる。実際、こうした翻訳者を念頭に開発された翻訳支援システム TransType[3, 11]は、翻訳文書のコンプリーション機能を中心としたもので、どれだけ翻訳者の時間を削減できるかを評価の軸の1つとしている。

第二の類型は、オンライン・ボランティア翻訳者に代表される翻訳者で、関心のある文書を時間のあるときに翻訳するタイプである。こうした翻訳者は、プロの翻訳者である場合もあるが、一般に元文書の言語に対してネイティブ・レベルの言語力を持っているわけでもなく、元言語で表現された事実情報に対する造詣も特化した職業翻訳者ほど深くない場合が多い。拘束が少なく収入と直結しないが、面倒でない限り翻訳をしたいと考えており、「面倒を減らせれば」型とすることができる。これらの翻訳者をターゲットとした場合、元文書の意味や関連事実を理解するためのレファレンス作業の負担を軽減することの比重が極めて高く、翻訳文書入力の手間は相対的に問題にならない。また、レファレンス作業の負担は必ずしも時間の節約に還元できるものではなく、主観的な労力と煩わしさにも大きく関係する。我々が対象としている翻訳者は、この後者のタイプである。

3 翻訳エディタの基本要件

オンライン・ボランティア翻訳者の以上のような性質を考慮し、また翻訳者の翻訳行動を考慮して³、翻訳エ

²カナダで職業翻訳者を経て翻訳支援システム TransType の開発に関わった Elliott Macklovitch 氏との個人的議論による。

³15名の翻訳者にアンケートを、5名の翻訳者に簡単な聞き取りを行った。

ディタの基本的な設計理念は、最も抽象的なレベルでは、(1) 新たな情報・機能を提供するのではなく、翻訳者が現に行っている作業の手間を省く、(2) システムが決めるのではなく翻訳者が決めるのに必要な情報を提供する、(3) 翻訳者の発想を助ける情報を表示する、(4) できるだけシンプルにする、という四点に集約される。

より具体的には、次の諸点が重要な要件としてあげられる。このうち1と2はインタフェースに、3と4はインタフェースとレファレンス情報の内容の双方に関係する⁴。

1. 翻訳者は翻訳文書作成のリズムを重視する。通常、あまり長くない文書を扱っているため、文書単位がリズムの基本単位であり、文単位でセルに区切られた翻訳インタフェース等は不適切である⁵。
2. 翻訳文書作成のリズムを壊すという理由でオンライン辞書を用いない翻訳者もいる。したがって、キーボードは基本的に翻訳文書作成エリアに割付られている必要がある。
3. 翻訳プロセスで翻訳者はさまざまな翻訳表現候補を検討し、必要ならば複数のレファレンス・ツールをチェックする。翻訳支援システムは、複数の情報源へのアクセスに際して負担を減らすことで翻訳者の作業ステップを減らすことが求められる。
4. この際、システムが提示する「良い情報」とは、翻訳者がそれに基づいて自らの翻訳表現を確定することができるような情報であり、最終的に翻訳者が選ぶ情報をシステムが選ばばよいわけではない。
5. 「洗練された」手法を用いて翻訳者に代わって情報をしぼることはできるだけしないため、情報を参照するインタフェースが重要になる。参照は、大体わかっていて簡単に確認するためだけの参照と、フルに情報をチェックして考え決断する参照とに大きくわかれるため、このインタフェースを考慮する必要がある。

4 QRedit: 翻訳者向けエディタ環境プロトタイプ

以上の概念と要件を考慮し、我々は、オンライン・ボランティア翻訳者向け翻訳支援エディタ環境プロトタイプ QRedit を実装している。システムは現在、主に Java

で実装されて TOMCAT 上で稼働し、ユーザは Web ブラウザを通して利用する。

図 1 に、QRedit の画面イメージを示す。エディタは元文書領域と翻訳文書領域とに分割される。利用者は環境により縦の分割と横の分割から作業しやすい分割を選ぶことができる。元文書領域と翻訳文書領域のスクロールは同期しており、元文書のパラグラフごとに同期ポイントが設定されている。

4.1 元文書領域の機能

翻訳者は元文書の URL を指定する（その場合システムはタグ解析を行いテキスト領域を抽出する）かテキストを貼り込み、辞書引きをオンにする。システムは単語列に対し辞書エントリとの照合を行なう。

辞書引き機能

翻訳者が元文書の単語にマウス・カーソルをあわせると、システムは翻訳候補をポップアップで表示する。辞書は現在三省堂『グランドコンサイス英和辞典』[13]と『英辞郎』[2]とが組み込まれている。辞書引きの際には、単純語の辞書引きだけでなく、挿入や置換を考慮したイディオムのさまざまな異形まで引けるようになっている [8]⁶。

翻訳表現候補の提示

システムは、二種類の候補表示機能を備えている。一つは、原文書領域の中でのポップアップによる簡略化された表示であり、ここでは訳語候補だけが表示される。もう一つは、ウィンドウ内の大半を占有するポップアップ領域を生成し、辞書に含まれる見出し語の情報をすべて表示するものである。前者はちょっとした確認に、後者は訳語や意味を深く検討する際に適している。

4.2 原文書領域と翻訳文書領域の関係

翻訳者は、マウスクリックによって、原文書領域中で表示させた訳語候補を翻訳文書領域に自動的にコピーできる。システムはまた、数値表現を自動的に日本語の表現に変換する機能（“Seventeen thousand three hundred fifty”や“1.3 billion”を「1万7350」や「十三億」等に）、原文書の HTML タグをコピーする機能を備えている。一般に Web ブラウザ内では、マウスクリックを行なうとキーボードのカーソルもそのオブジェクトへと移動するが、本システムでは原文書領域内でどのようなマウス操作を行っても、キーボードカーソルは常に翻訳文書領域に留まるような工夫をしている。そのため、マウスとキーボードの独立性に慣れれば、翻訳文書作成作業のリズムを崩さずに辞書引きやコピーを行うことができる。

⁴レファレンス・ツールと情報の要件については、[7]を参照。

⁵Macklovitchによると、元文書の一文が対象文書の一文中に訳されるという対応は、英仏のように近い言語で公式の硬い文書を扱っている場合でさえ破られる。文単位でセルに区切るインタフェースが有効なのは、ゲームのセリフやソフトウェアのボタンの翻訳のように、翻訳ユニットの対応が最終的な産物の形式に厳密に縛られる場合である。

⁶この機能を十分なかたちで組み込んだ機械翻訳システム・翻訳支援システムは我々が2006年夏までに調べた限りでは存在しなかった。



図 1. QRedit の画面イメージ

4.3 翻訳文書領域

翻訳文書領域では、オープンソースの WISWIG 型 Web エディタ FCKeditor⁷ を用いている。文書は簡単な文字装飾も含めた HTML 形式あるいはテキスト形式で保存することができる。あらかじめ元文書と同じパラグラフ構造に分割されているため、翻訳者は好きな箇所から翻訳できる。

5 統合レファレンス機能

現在のところ、QRedit プロトタイプに組み込まれているレファレンス機能は、高品質のまとまった辞書への参照に限られる。しかしながら、翻訳者は一般に、翻訳プロセスにおける情報のチェックのために、辞書だけでなく「図書館」も活用している（オンラインで作業する場合には、図書館をウェブ上の情報で代替することが多い）。

参照する言語単位の側から参照情報源を類型化すると、一般語・熟語慣用句には既存の高品質辞書が、固有名・専門語・引用句等にはそれに加えて「開かれた」オンライン情報源を含む「図書館」が用いられる。現在、固有名や専門語の対訳を開かれた Web 情報から取得するシステムの開発が進められており [14, 15]、そこから提供される情報を自動的に参照できる機能の組み込みを検討中である。

人間が編集した高品質レファレンス・ツールと比べて

の本質的な違いは、開かれた情報源からシステムが提示する情報には、それ自体としては候補範囲の信頼性があまりないことである。これをカバーするためには、システムがある結果を提示したとき、翻訳者がシステムが用いた元の情報に遡ってその善し悪しを検証する機能を組み込む必要がある。したがって、QRedit 環境の観点からは、開かれた情報源から固有名や専門語の対訳を提示するシステムは、システムが提供する結果の精度や再現率からではなく、開かれた情報源から関連情報を入手するという、翻訳者が現に行っている行動を基準にして、そこで得られる結果を損なわずに、けれどもそれに伴う負担をどこまで軽減することができるかに依存している。

したがって、これら開かれた Web 資源からの情報を参照する機能は、翻訳者が、インタラクティブにシステムの結果とそれが元とした情報源、そして翻訳文書作成環境とを行き来できるようなインターフェースとして組み込まれる必要がある。これらについては、今後の実装課題である。

6 おわりに

英日ボランティア翻訳者向け統合エディタ環境 QRedit のプロトタイプについて紹介してきた。既に述べたが、本プロトタイプの基本設計理念は、以下の四点に集約される。

- (1) 新たな情報・機能を提供するのではなく、翻訳者

⁷<http://www.fckeditor.net/>

が現に行っている作業の手間を省く、

- (2) システムが決めるのではなく翻訳者が決めるのに必要な情報を提供する、
- (3) 翻訳者の発想を豊かにする情報を表示する、
- (4) できるだけシンプルにする。

この方針は、翻訳者の要望だけでなく、翻訳支援システム及びコンピュータの現状における位置づけを理論的に考慮した結果の選択肢でもある。すなわち、(a) 単純に技術的な観点から翻訳プロセスにおける諸作業の中間結果および最終的な結果である翻訳文書を見たとき、コンピュータはまったく人間に及ばない、(b) 仮にコンピュータによる出力結果が改善されて人間の翻訳結果に近くなったとしても、独立した翻訳文書として受け入れられるために必要な人間の側のコンピュータに対する認識（「私たちはコンピュータの出力結果を一応信頼できるものとしてみよう」）が決定的に欠けている、という二点である。

実際のところ、翻訳支援における翻訳者と支援システムの作業配分において、我々は今回実装したよりはもう少しシステム側の処理機能を信頼した設計にしてもよいと考えている。たとえば、新たな情報・機能を提供することで、翻訳者がこれまで気付かなかった改善を実現する可能性は少なからずあるだろう。しかしながら、外部的な強制がない限り、ボランティア翻訳者に限らず多くの翻訳者が翻訳支援システムを使わない現状を鑑みるに、上記の基本設計理念は出発点として必須の条件であった。

ここで紹介した QRedit プロトタイプは、実際にオンラインで翻訳作業を定期的に行っているボランティア翻訳者 2 名に実験的に使ってもらい、極めて主観的で制約のない使用報告を提出してもらった。報告内容は、(a) 辞書引き機能について、(b) 辞書情報の内容について、(c) インタフェースについて、(d) ソフトウェア的な問題について、に分けることができるが、インタフェースについては、元文書の HTML の論理構造の維持などいくつかの要求があったものの、大枠としては好意的に受け取られている。

付記

本研究の一部は、日本学術振興会科学研究費補助金基盤 (A) 「翻訳者を支援するオンライン多言語レファレンス・ツールの構築」(課題番号 17200018) の支援を得て行われた。

参考文献

- [1] Bowker, L. 2002. *Computer-aided Translation Technology: A Practical Introduction*. Ottawa: University

of Ottawa Press.

- [2] 英辞郎. <http://www.eijiro.jp/>, checked October 31, 2006.
- [3] Foster, G., Langlais, P. and Lapalme, G. 2002. "User-friendly text prediction for translators," *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*. p. 148–155.
- [4] Fulford, H. and Granell Zafra, J. 2004. "The uptake of online tools and web-based language resources by freelance translators: implications for translator training, professional development, and research," *Proceedings of the Second International Workshop on Language Resources for Translation Work, Research and Training*. p. 37–44.
- [5] Fulford, H. 2001. "Translation tools: An exploratory study for their adoption by UK freelance translators," *Machine Translation*, 16(3), p. 219–232.
- [6] Gow, F. 2003. *Metrics for Evaluating Translation Memory Software*. Ottawa: University of Ottawa. PhD Thesis.
- [7] 影浦峽, 佐藤理史, 竹内孔一, 宇津呂武仁, 辻慶太, 小山照夫. 2006. 「翻訳者支援のための言語レファレンス・ツール高度化方針」言語処理学会第 12 回年次大会. p. 707–710.
- [8] 竹内孔一, 金平昂, 平尾一樹, 阿辺川武, 影浦峽. 2007. 「置換・挿入を考慮した異形イディオム検索システムの構築」言語処理学会第 13 回年次大会. (発表予定)
- [9] Kay, M. 1997. "The proper place of men and machines in language translation," *Machine Translation*, 12(1), p. 3–23.
- [10] TRADOS. <http://www.trados.com/>
- [11] Macklovitch, E. 2006. "TransType2: the last word," *LREC'2006*. p. 167–172.
- [12] Omega-T. <http://www.omegat.org/>
- [13] 三省堂, 2006. 『グランドコンサイス英和辞典』. 三省堂.
- [14] 辻慶太, 佐藤理史, 影浦峽. 2005. 「対訳人名検索における翻字・サーチエンジンの有効性評価」言語処理学会第 11 回年次大会発表論文集. p. 352–355.
- [15] 外池昌嗣, 宇津呂武仁, 影浦峽, 佐藤理史, 阿辺川武. 2007. 「ウェブを用いた専門用語翻訳支援における多様な情報源からの信頼度情報の提示」言語処理学会第 13 回年次大会. (発表予定)