

## みんなの翻訳

内山 将夫<sup>1</sup>

阿辺川 武<sup>2</sup>

隅田 英一郎<sup>1</sup>

影浦 峯<sup>2</sup>

1 情報通信研究機構 MASTAR プロジェクト

2 東京大学大学院教育学研究科

### 1 はじめに

ボランティアの翻訳者は、様々な文書を翻訳している。そのなかには、たとえば、オープンソースのマニュアルの翻訳や、Global Voices Online<sup>1</sup>のようなブログの翻訳などがある。

ボランティアの翻訳者は、翻訳により、世の中に貢献していると考えられる。たとえば、オープンソースのマニュアルの日本語訳は、日本人のユーザーにとっては、大変有難いものであるし、Global Voices Online の記事は、他のメディアが注目しない場所や人々について光を当てるものである。

したがって、ボランティアの翻訳者による翻訳を支援することは、世の中に貢献することであるといえる。そこで、それを実行することを考えた。

### 2 ボランティアの翻訳者を如何に支援するか

ボランティアの翻訳者を支援する既存のプロジェクトとしては、「椎茸プロジェクト」がある [4]。このプロジェクトでは、翻訳支援用のエディタである QRedit [1] など、様々な翻訳支援用のツールやアルゴリズムを提案・実装しており、それらのツールは、実際に、ボランティアの翻訳者に利用されている。しかし、これらのツールの支援対象は個々の翻訳者であるので、世の中の多数の翻訳者を一度に支援することはできない。

そこで、我々のプロジェクトにおいては、世の中の多数の翻訳者を一度に支援するためのシステムを構築することを考えた。具体的には、我々のプロジェクトにおいては、ボランティアの翻訳者をホスティングすることを目

標とし、そのための Web サイト<sup>2</sup>を構築した。図 1 は、我々のプロジェクトで構築した「みんなの翻訳」サイトのスクリーンショットである。

ホスティングにより、世の中の多数の翻訳者を支援できると考えた理由は以下のものである。

(1) まず、他分野における成功例として、オープンソースの世界においては、sourceforge.net のように、オープンソースプロジェクトをホスティングすることにより、オープンソースの開発や普及を促進している例がある。そのため、ボランティアの翻訳者をホスティングすることにより、同様な成功が望めるのではないかと考えた。

(2) 次に、文献 [1] からわかるように、オンラインのボランティアの翻訳者は、翻訳支援ツールをあまり使っていない。一方、みんなの翻訳サイトにおいては、椎茸プロジェクトのツールを組込むことにより、みんなの翻訳サイトの利用者が、自然に、翻訳支援ツールを使える環境を提供する。これにより、利用者が、特に意識せずとも、翻訳支援を受けることができる。

(3) また、みんなの翻訳サイトにおいては、ボランティアの翻訳者が翻訳したテキストは、原文と共に保存されているので、みんなの翻訳の利用者は、自分の翻訳結果だけでなく、他の翻訳者の翻訳結果を共有できる。そのため、他人の翻訳を自分の翻訳に利用できる。

(4) 最後に、みんなの翻訳サイトにおいては、ボランティアの翻訳者が翻訳したテキストは、サイト上に保存され、みんなに公開されるので、自分の翻訳を公開したい人にとっては、公開する場所を提供することになる。

以上の理由により、もし、多数のボランティアの翻訳者がみんなの翻訳を利用すれば、翻

<sup>1</sup><http://jp.globalvoicesonline.org/>

<sup>2</sup><http://trans-aid.jp/minna/ja/> で公開予定



図 1: 「みんなの翻訳」サイト

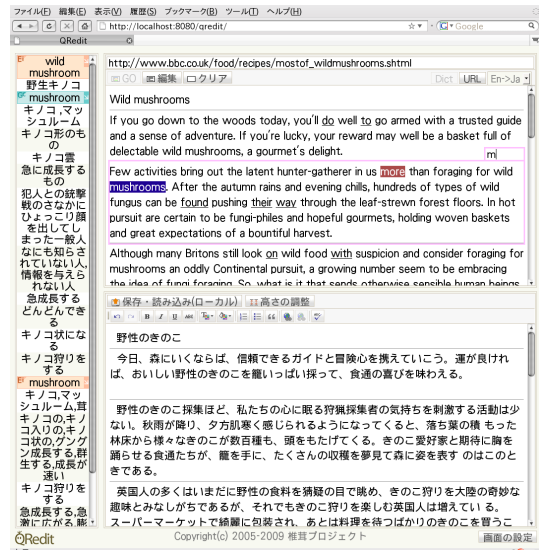


図 2: 翻訳支援エディタ: QReddit

訳支援ツールの提供や、翻訳の共有と再利用という利点により、多数のボランティアの翻訳者をサポートできると考えた。

### 3 みんなの翻訳のコア

みんなの翻訳において、上記の理由を実現するために、以下のことを考慮した。

- 翻訳結果を共有する仕組みを準備すること
- 良い翻訳支援エディタを提供すること

なお、見栄えや使い勝手も重要であるので、それらは継続的に改善する予定である。

### 4 翻訳結果を共有する仕組み

翻訳結果を共有するためには、原文と翻訳文の使用許諾について考慮する必要がある。たとえば、当然であるが、原文の著者が翻訳文の公開を許可していない場合には、翻訳文は公開できないので、翻訳結果を共有することはできない。

そのため、みんなの翻訳の利用者には、原文と翻訳文の使用許諾について確認を求めた。とくに、みんなの翻訳の利用者には、各自が翻訳した文は、2 次的利用ができるように許可

することを求めた。そのために、システムは、みんなの翻訳の利用者が翻訳文を保存するときに、以下のようにして、使用許諾などを確認することにした。

(1) まず、システムは原文の使用許諾条件を確認する。

あなたが翻訳の対象とした文書 (原文) は、原文の著者が明示的に許可した場合を除いて、私的な利用等しかできません。原文の著者は、その翻訳を公開しても良いと (あなたや他の人に) 許可をしていますか? (はい/いいえ)

もし、「いいえ」が選択された場合には、翻訳文は非公開となる。

(2) 次に、システムは翻訳文の使用許諾条件を確認する。

「みんなの翻訳」では、原著者や訳者の著作権を保護したうえで、その作品を、みんなが、翻訳などに利用できるようにしたいと思っています。そのため、以下の項目より最も適した使用許諾条件を選択してください。

ここで、システムは、いくつかの使用許諾条件を提案するが、その代表的なものは、Creative Commons<sup>3</sup>によるライセンスである。また、いずれの使用許諾条件を選択した場合でも、「あなたの文書を翻訳し、その文書を公開してもよい」という条件に矛盾しないようになっている。

このようにして、みんなの翻訳では、原著者や翻訳者の著作権を尊重しつつ、翻訳を共有できる仕組みを準備している。

以上により、著作権上の問題はクリアできるので、次に、共有された翻訳をどのように利用するかを述べる。

利用のためのツールとしては、まず、2言語検索がある。みんなの翻訳サイトには、原文と翻訳文が保存されているので、それらに対して文アラインメント[3]を適用することにより、対訳文対を作成する。この対訳文を検索することにより、どのような文字列がどのように訳されたかを検索できる。用語や翻訳文なども検索できる。また、みんなの翻訳の利用者が互いにコミュニケーションする手段として、質問および回答の機能も設けている。さらに、まだ実現してはいないが、用語の自動抽出ということも考えている。翻訳において、用語は重要であるので、それを自動抽出できれば、翻訳支援としての価値が高いと考えている。

## 5 翻訳支援エディタ: QRedit

実際の翻訳にあたっては、翻訳支援ツールQReditを用いる。QReditの概要については、文献[1, 2]に詳しい。本節では最初にQReditの基本設計理念を述べ、次に本エディタの機能の1つである「注目度別強調表示機能」について簡単に説明し、最後に今年度新たに実装された機能を紹介する。

### 5.1 基本設計理念

ボランティア翻訳者向け統合エディタ環境QRedit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。意思決定を翻訳者に委ねるならば、システムの意義は、正解を決めるのではなく、翻訳者の意思決定を促し、翻訳者の発想を豊かにする情報を表示することになる。四点めの「できるだけシンプルに」は、翻訳者が翻訳作業に集中するための基本要件であり、同時に、情報環境に対しては保守的であるという多くの翻訳者の性格を考慮したものである。

### 5.2 注目度別強調表示機能

本エディタでは入力された原文に対し、複数の辞書や翻訳者が登録した用語を対象に辞書引きを行ない、翻訳者は単語をクリックすることで簡単にその訳語を把握することができる。また、高度なイディオム検出機能[5]を備えており、語間に挿入を許した異形イディオムを検出できる。辞書引きの際に翻訳者が見落としがちな用語に対して、エディタ側から

<sup>3</sup><http://creativecommons.org/>

何らかの警告を与えることは、誤訳を防ぐためにも有用である。ただ、すべての用語に同一の警告を与えたのでは意味がなく、何らかの基準を設けることが必要である。

そこで我々は、翻訳者がその用語にどのくらい気づいているかという度合を *awareness level* という尺度として定義した。具体的には用語について「構成要素」「難易度」「専門性」「出典」の値を定義し、それらの値の和を基に *awareness level* を算出する。原文領域では、この尺度を基に「常に強調表示」「マウスオーバー時のみ強調」「強調表示なし」と3段階で単語の表示法を分けている。詳しくは [2] を参照されたい。

### 5.3 新しい機能

今年度 QRedit に新たに実装された機能のうち、ユーザーインタフェースの改善について2つ挙げる。

原文表示位置と辞書引き表示位置の選択機能

翻訳文の入力環境の良し悪しは作業効率に大きく影響する。また、どのような環境が最適であるかは翻訳者より様々であり、なるべく多くの翻訳者に対応できるよう柔軟にカスタマイズできる仕組みが望ましい。QRedit では、原文と翻訳文の位置関係を縦分割・横分割にでき、原文は上下左右どこでも配置できる。また、辞書引き表示も、狭い画面では原文領域内にポップアップ表示を、広い画面の場合は原文の左右に指定できる。図2は、原文表示「上」、辞書引き表示「原文の左」に設定した例である。

キーボードによる訳語表示機能

従来の QRedit では、訳語を表示するためにその単語をマウスでクリックする必要があった。これは翻訳文を入力している際には、手をキーボードからマウスへ移す必要があり、効率的な翻訳作業を妨げる要因となっていた。そこで、訳語表示をキーボードのみで実行できる機能を開発した。翻訳文入力中にショートカットキーの入力により単語選択モードに遷移し、そこではカーソルキーによる単語選択

のほか、インクリメント検索により素早く目的の単語を選択することが可能である。

## 6 おわりに

ボランティアの翻訳者による翻訳を支援することにより、世の中に貢献することを考え、そのために、ボランティアの翻訳をホスティングする Web サイト「みんなの翻訳」を構築した。みんなの翻訳は、翻訳を支援するツールに加えて、翻訳した文書をみんなで共有できる枠組を備えている。我々は、みんなの翻訳により、ボランティアによる翻訳の発展を助けることができればと思っている。

## 参考文献

- [1] Takeshi Abekawa and Kyo Kageura. QRedit: An integrated editor system to support online volunteer translators. In *Digital humanities*, pp. 3–5, 2007.
- [2] Takeshi Abekawa and Kyo Kageura. A translation aid system with a stratified lookup interface. In *ACL Demos and Posters*, pp. 5–8, 2007.
- [3] Masao Utiyama and Hitoshi Isahara. Reliable measures for aligning Japanese-English news articles and sentences. In *ACL*, pp. 72–79, 2003.
- [4] 影浦峯, 阿辺川武. 『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』(椎茸プロジェクト)について. *AAMT Journal*, Vol. 40, pp. 23–27, 2007.
- [5] 蛭浜康雄, 金平昂, 平尾一樹, 竹内孔一, 阿辺川武, 影浦峯. Wordnet と同音異義語を利用した異形イディオム検索. 言語処理学会第14回年次大会発表論文集, pp. 1045–1048, 2007.