

みんなの翻訳第 2 報

内山 将夫¹

阿辺川 武²

隅田 英一郎¹

影浦 峯³

1 NICT MASTAR プロジェクト

2 国立情報学研究所 3 東京大学

1 はじめに

「みんなの翻訳」(<http://trans-aid.jp/>) は、世界中の文書をみんなで協力して翻訳するサイトである [4, 7]。そのために、みんなの翻訳では、個人およびグループの翻訳者を支援する機能を実装することを目指している。みんなの翻訳は 2009 年 4 月にオープンし、それから 9ヶ月程度で、登録翻訳者が 1000 人、登録文書数が 3000 文書を越えた。さらに、2013 年までには、図 1 のような展開ができればと考えている。以下では、まず、みんなの翻訳の概要を物語的にまとめる。次に、みんなの翻訳の今後の展開について述べる。

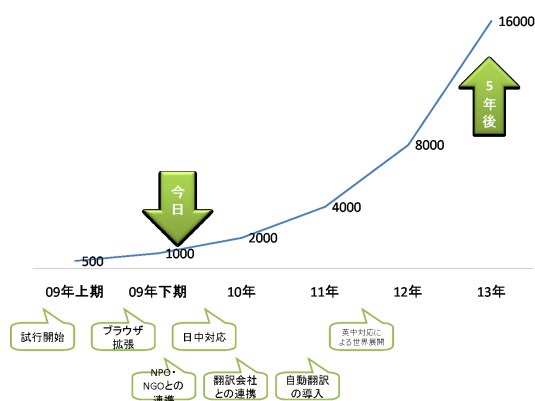


図 1: 利用拡大計画

そのためには、まず、個人の翻訳者の翻訳を支援するソフトウェアが必要である。そのようなソフトウェアは、既に、「椎茸プロジェクト」[6] で開発されていた。特に、翻訳支援エディタ QRedit [1] (図 2) は、高品質な三省堂「グランドコンサイス英和辞典 (36 万項目収録)」の辞書引きに加えて、Wikipedia や Web サーチを翻訳リズムを崩さずに利用できることから、翻訳時間を 20~30%削減することができる [7]。

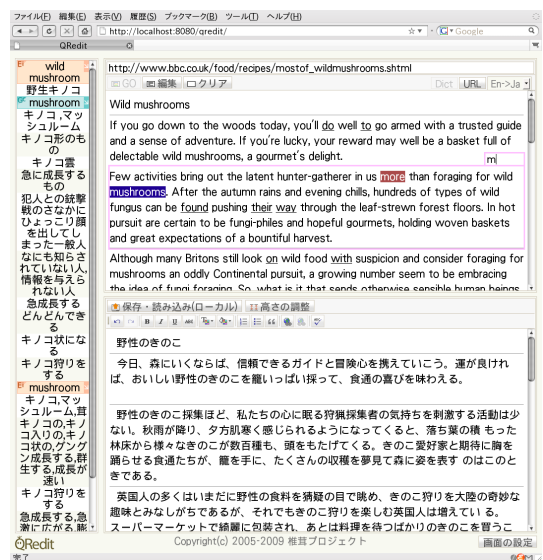


図 2: 翻訳支援エディタ: QRedit

2 みんなの翻訳の概要

2.1 翻訳支援

みんなの翻訳は、世の中の多数のボランティア翻訳者を支援することにより、翻訳者が楽しく翻訳ができ、その結果として、翻訳文書が世の中に流通することを目的としている。

その他にも、翻訳に便利な機能としては、過去の対訳を検索する機能があるが、それを実現するためには、原文と翻訳文とから文アラインメント [5] を適用して対訳例を作成し、図 3 のように検索すれば良い。なお、この対訳検索システムは、中央大学大羽良氏作成のシステムをベースとしている。

更に、専門用語自動抽出ソフトウェア TermExtract [2] を利用し、日英の用語候補を抽出し、抽出された日本語用語候補の英訳候補



図 3: 対訳検索画面

を要素合成法 [3] により抽出し、もし、その英訳候補が、TermExtract により抽出された用語候補にあったならば、英訳として採用するという手順により、原文と翻訳文から高精度で対訳用語を抽出できる (図 4)。

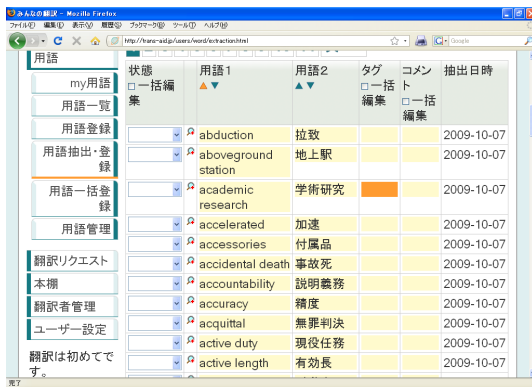


図 4: 用語抽出選択画面

このように、翻訳支援に必要な個々の機能については、既に、自然言語処理の分野で開発済の技術を利用すれば良い。

したがって、これらの技術を、誰もが利用できるようにすれば、世の中の多数の翻訳者を、翻訳という側面から支援することができる考えた。

2.2 翻訳共有

次に解決しないといけな問題は、翻訳文書を共有する仕組みである。そのためのツール

としては、Creative Commons¹によるライセンス CCL (およびそれと類似するライセンス) を利用した。CCL を利用することにより、著作権者は、自分の著作物に対して、その使用許諾条件をあらかじめ宣言しておくことができるので、他人は、その使用許諾条件に従うかぎりには、自由に、その著作物を利用することができる。

みんなの翻訳では、翻訳者に対して、自分の翻訳文書を公開するときには「その文書を翻訳し、その文書を公開してもよい」という条件に矛盾しないライセンスを付けてもらうようにしているため、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。

このライセンスがあることにより、文書の公開をしたり、対訳検索をしたり、用語抽出をしたりすることが自由にできる。

2.3 翻訳コミュニティの支援

みんなの翻訳では、グループによる翻訳を支援するための機能として、同一の文書をグループで編集する機能、用語をグループで共有する機能、編集結果の履歴を比較する機能がある。

用語をグループで共有する機能は、NPO/NGO などで、複数人が、異なる文書を翻訳しているときに、文書間で同一の用語を使用するとき必要である。

次に、NPO/NGO などでは、1次翻訳者が翻訳した文書を、2次翻訳者が修正するというをしているが、そのときに、みんなの翻訳の同一文書共同編集機能を利用し、さらに、編集結果の履歴を比較する機能を利用すると、1次翻訳者の翻訳が、2次翻訳者によりどのように修正されたかを確認できる。また、翻訳や英文和訳の授業などでは、生徒が訳した文を先生が修正することがあるが、そのときに、生徒は、先生がどこを修正したかを確認できる。したがって、1次翻訳者や生徒は、自分の翻訳結果と2次翻訳者や先生の修正結果とを

¹<http://creativecommons.org/>

比べることにより、自分の翻訳技能を向上する機会がある。

また、みんなの翻訳では、翻訳者間でメッセージを伝達したり、文書にコメントをつけたり、質問応答をしたり、掲示板を作成するなど、翻訳者間の交流を促進する機能がある。これらは、グループで翻訳をするときに、必須な機能である。

以上、みんなの翻訳には

- 自然言語処理技術を利用した翻訳支援
- CCL による翻訳共有
- グループ翻訳支援

の機能があることを述べた。

3 みんなの翻訳の今後の展開

まず、自然言語処理の観点からみんなの翻訳の展開について述べる。

約 10ヶ月の運用と改良により、みんなの翻訳は、個人の翻訳者および小規模なグループによる翻訳については、必要な機能は、一通り網羅できていると考えている。今後は、各機能を向上させるとともに、その評価をする予定である。つまり、翻訳支援機能の実装のフェーズは終わったので、次は、みんなの翻訳をフィールドとして、それら进行评估する予定である。

次に、翻訳者のための翻訳支援機能は実装したが、読者のために読みたいものが読める機能の実装は不完全である。みんなの翻訳で公開されている文書は既に 1300 以上なので、これらの文書に簡単にアクセスできるような機能が必要である。

また、現状、みんなの翻訳は、日英と英日にしか対応していないので、それを多言語に拡張する予定である。

次に、みんなの翻訳の利用者の拡大という観点について述べる。もともとみんなの翻訳はオンラインのボランティア翻訳者に向けたものであり、その面では、以下のように拡大をはかる。

- NPO や NGO . 現在もいくつかのグループが利用しているが一層の拡大をはかる。
- ブログなどで関心のある文章を部分的に翻訳している個人。こうした人々に対しては、翻訳したい文章を指定してからみんなの翻訳を呼び出せるブックマークレットを作成することで展開をはかりたい。

それ以外に、みんなの翻訳が提供する枠組みと親和性が高いユーザとして、以下が考えられる。

- 共訳による出版翻訳、及び一定の主題を継続的に翻訳出版している出版社。有用性はすでに実証されている。
- 大学のゼミにおける文献購読と翻訳。現在、複数の大学・大学院のゼミで活用されている。
- 高校レベルでの英文和訳学習。翻訳と英文和訳は異なるし、高校英語には到達目標があるので、みんなの翻訳を活用できるかどうかはさらなる検討が必要である。

最後に、みんなの翻訳の運営という観点からは、永続的な運営を可能とする枠組を確立したいと考えている。

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] Hiroshi Nakagawa and Tatsunori Mori. Automatic term recognition based on statistics of compound nouns and their components. *Terminology*, Vol. 9, No. 2, pp. 201–209, 2003.
- [3] Masatsugu Tonoike, Mitsuhiro Kida, Toshihiro Takagi, Yasuhiro Sasaki, Takehito Utsuro, and Satoshi Sato. A comparative study on compositional translation

estimation using a domain/topic-specific corpus collected from the web. In *Proc. of the 2nd International Workshop on Web as Corpus*, pp. 11–18, 2006.

- [4] Masao Utiyama, Takeshi Abekawa, Eiichiro Sumita, and Kyo Kageura. Hosting volunteer translators. In *MT Summit*, 2009.
- [5] Masao Utiyama and Hitoshi Isahara. Reliable measures for aligning Japanese-English news articles and sentences. In *ACL*, pp. 72–79, 2003.
- [6] 影浦峽, 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.
- [7] 内山将夫, 阿辺川武, 隅田英一郎, 影浦峽. みんなの翻訳. 言語処理学会第15回年次大会, 2009.