

翻訳教育向け「みんなの翻訳」

影浦峽[†], Martin Thomas[‡], 阿辺川武[‡], Bogdan Babych[‡],
内山将夫[§], 隅田英一郎[§], Anthony Hartley^{‡†}

[†] 東京大学大学院教育学研究科

[‡] Centre for Translation Studies, University of Leeds

[‡] 国立情報学研究所連想情報学研究開発センター

[§] 情報通信研究機構 MASTAR プロジェクト

1 はじめに

筆者らは、主としてオンラインで活動するボランティア翻訳者やNGOの支援を目的とする統合翻訳支援ホスティングサイト「みんなの翻訳」(<http://trans-aid.jp/>)を開発・運用している(Utiyama, et. al., 2009; 内山他 2011)。本発表では、これを翻訳訓練・翻訳学習のために拡張した「翻訳教育向け『みんなの翻訳』」の背景と基本的考え方を紹介する。

従来、翻訳学校の学生による翻訳と教員の修正やそれに伴う様々な議論はその場限りのものとして扱われ、学年が変われば同じプロセスをその都度繰り返すことが多かった。しかしながら、そのプロセスで作られたデータを体系的に蓄積し、新しい学生がこれまでの学生たちの経験を「再演」し、そこから学ぶようにできれば、翻訳教育の質的向上と能率化をはかることができる。そのようなデータはまた、翻訳論や機械翻訳の研究にとっても極めて有用である。「翻訳教育向け『みんなの翻訳』」は、翻訳修正プロセスに加えて翻訳チームのメンバー間のやりとりも含めて構造化したデータを蓄積し追跡可能にすることで、こうした要求に対応するものである。

2 前提

2.1 パイロットプロジェクト

「みんなの翻訳」に必要な要件を具体的に整理するために、神戸市外国語大学(KCUFS)とリーズ大学翻訳研究所(CTS)が、2009年末から2010年3月まで、「みんなの翻訳」を使った翻訳教育のパイロットプロジェクト(以下「神戸＝リーズ・プロジェクト」)を行なった。プロジェクトでは、4カ月間に2サイクルの翻訳タスクを定義し、参加学生は「みんなの翻訳」を使って翻訳を行ない、やりとりには「みんなの翻訳」が提供するコミュニケーション機能に加えてSNSと通常の電子メールも利用した。対象とした日本語文書は神戸の「阪神・淡路大震災記念人と防災未来センター」が

提供する、1996年の阪神淡路大震災を経験した個人の経験談書き起こし、英語文書はブロンテ博物館(リーズに近く日本人観光客も多い)のブロンテ家の歴史に関する常設展の説明である¹。「翻訳教育向け『みんなの翻訳』」の機能仕様は、主にこのパイロットプロジェクトを通して検討した。

2.2 「みんなの翻訳」のグループ機能

パイロットプロジェクト実行時に「みんなの翻訳」が提供していた基本的なグループ翻訳機能は、(1)共同編集・共訳のための文書の共有、(2)翻訳者間でのやりとりを行なうためのメッセージ機能と掲示板機能(特定文書や翻訳セグメントとは独立で紐付けはできない)、(3)最大10バージョンまでの修正バージョン保存とdiffによる任意のバージョン対の対比表示、である(その後、プロジェクト管理機能が追加された)。

3 関与する要素とプロセス

3.1 テキスト実体

基本テキスト実体は以下の3階層で定義される。

- (a) 文書集合 クライアントが提供するテキスト全てからなる集合あるいはその部分集合で、翻訳属性記述(想定対象読者、テキスト機能、スタイルガイドなど)は文書集合に対して規定され、用語の抽出と検証もこのレベルで行なわれる。プロジェクトは基本的にこのレベルで定義される。
- (b) 個別文書 翻訳作業と役割はこのレベルで割り当てられ、また、基本翻訳属性記述に対する例外既定も個別文書に割り付けられる。

¹ 翻訳教育の観点から言うと、神戸＝リーズ・プロジェクトはKirály (2000)の「社会構築論」的アプローチを採用した。これは、実際の翻訳作業の状況に身を置いて共同作業に協力し、専門知識とプロ意識を経験により育むことで、学生が自ら知識を構築することが可能になるという考えであり、文書の選択は、実際に翻訳を必要としているテキストのみが、このアプローチを具体化できるという考えに基づきなされた。談話の書き起こしやローカルな土地への言及といった難しい素材を使ったのも、学生間のやりとりを促すことを目的とした意識的な選択である。

(c) テキストスパン 個別の修正はこのレベルでなされる。一文内の場合（複合表現や単純語など）も複数文にわたる場合（照応的指示など）もある。

3.2 参加者の役割

現実の翻訳ワークフローはタスクの見積りから請求まで様々なステップからなるが（Van der Meer 2006）、神戸＝リーズ・プロジェクトでは翻訳の専門技術が必要なフェーズに集中し、ターミノロジスト、翻訳者、修正者、アドバイザを定義した（「みんなの翻訳」に追加されたプロジェクト管理機能では、「翻訳」「修正」「レビュー」「完成」に対応する役割をデフォルトの設定としている）。

3.3 やりとり

テキスト実体のレベルと参加者の役割から、翻訳教育プロセスで異なるタイプのやりとりが行なわれる。例えば、以下のようなものがある。

- ターミノロジスト／翻訳者：用語の揺れなどを同定し解決する。
- 翻訳者／アドバイザ：実在物や場所など、文化的含意の説明を求めたり提供する。
- 修正者／翻訳者：下訳に対する変更の勧告やそれをめぐる質問を行なう。

オンライン協調学習の研究から、やりとりのタイプを事前に定義することの有用性が示されており（Singley et al. 2000; Soller 2000）。特に異文化コミュニケーションを伴う場合、これは、発言の意図を明らかにする点でも有用である。我々は、パイロットプロジェクト参加者による（構造化されない）コミュニケーション行為を複数階層対話行為マークアップ（DAMSL）分類（Allen and Core 1997）に基づいて分類し、そこから、「要求」、「通知」、「説明」、「解決」、「激励」、「感謝」という6つの基本対話行為タイプを同定した。

3.4 修正タイプと動機

翻訳学習プロジェクトでは、翻訳に加えられた修正を類型化して参照することが重要になる。これまで、誤りと修正の分類がいくつか提案されている（Secară 2005; Castagnoli, et. al. 2006; Mossop 2001; Abekawa & Kageura 2008a; Abekawa & Kageura 2008b; Shih 2006; Robert 2008）。神戸＝リーズ・プロジェクトで実際に加えられた修正の分析から、Castagnoli et. al. (2006) の枠組みに日本語に対応した変更を施すことで、修正タイプと変更の動機を示すほぼ十分な分類メニューを提供することができるとわかった。この情報を体系的に構築・提供することで、学習者が修正を検討するよう仕向けるだけでなく、タグ付きの翻訳

修正コーパスデータベースを構築することができ、これによって例えば日英間で翻訳の難しい部分をさらに明確にすることができる。

4 システムの仕様と開発

翻訳教育向け「みんなの翻訳」では、前節で説明した要素からなる翻訳学習プロセスに応じて発生する修正、やりとり等の情報を、再利用・追跡可能なかたちで体系的に管理することが課題となる。このために、(1) 文書と情報を管理するためのメカニズムと、(2) オンラインでの修正とやりとりを支えるメカニズムを設計した。

4.1 文書と情報の管理

文書と情報を紐付けて管理するために、Translation Memory eXchange (TMX) (LISA 2011) を基本とし、それと整合性を保ったかたちで必要な拡張を定義する。文書と修正・やりとり情報の管理は、以下の2つの要素によって行なう。

- (1) 標準的な TMX 仕様に従い、翻訳単位要素（<tu>）で原言語テキストのテキストスパンを指定し、最初に作られた目標言語の訳をそれに対応付ける。基本的に、テキストの翻訳単位として、段落単位を設定されるが、修正のために翻訳単位の任意の下位スパンを修正対象として選択することができる。
- (2) TMX の枠組みに、修正に関わる情報（基本翻訳単位に対して加えられた修正およびそれに関連するメタデータ）の履歴を追加する。そのために、別枠のタグを導入した。例えば、目標テキストにおいて、一貫性を維持しクライアントの要求を遵守するためにある用語を別の用語で置き換えたとき、置き換えが行なわれた翻訳単位の正確な位置とともに、置き換え情報が記録される。さらに、修正提案を行なったユーザの ID、役割²、修正を正当化するために利用者がメニューから選んだ修正の動機、必要に応じて利用者が書き込む自由記述が記録される。また、チームメンバー間のやりとりがある場合には、タイムスタンプも記録される。

情報の管理を2つのフェーズに区別して行なうことには、いくつかのメリットがある。メタデータの共通要素を、タイムスタンプを中心に、機能要素も取り入れて定義することで、文書とやりとり・修正関連情報という二つの相補的なデータセットを橋渡しする操作要求に対応することができる。また、最新バージョンからそこに至るまでのやりとりを復元し、各ステージでの翻訳を再現できる。また、ある翻訳単位に対する

²2011 年 1 月現在の仕様では役割の扱いは確定していない。

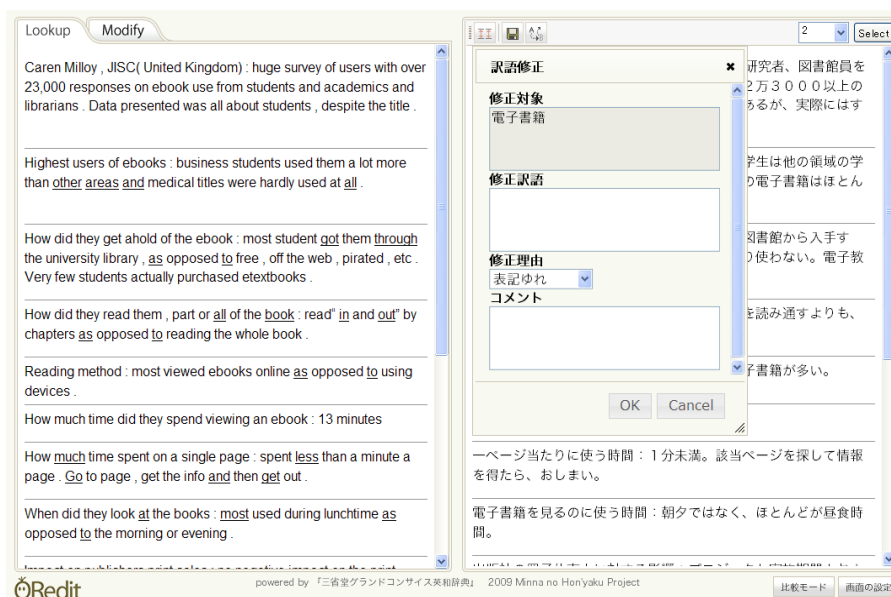


図 1. QRedit の修正用ポップアップ

修正のステップを、それぞれのステップにおけるユーザのやりとりを参照しながら、再現することもできる。最後に、TMX 標準に対して別枠のタグ情報を付けることで、複数バージョンのファイルを、修正プロセスの各ステージでの原テキストと目標テキストの対になる一つの正当な TMX インスタンスに変換することができる。それによって、再利用可能な翻訳データのグローバルなコレクションに貢献することができる。

4.2 QRedit インタフェース

オンラインでのやりとりの中核を構成するのは、文書に対する修正とその動機付け、その参照と確認などである（やりとりの多くは、修正をめぐって、あるいは修正を参照して行なわれる）。「みんなの翻訳」では、翻訳及び修正作業はすべて翻訳支援エディタ QRedit 上で行なわれる。QRedit は、原文書領域と翻訳文書領域からなる 2 ペインの翻訳支援エディタで、原文領域から高品質辞書の他、ウィキペディア検索、Google 検索、対訳検索などをシームレスに起動できる総合的翻訳支援環境である（Abekawa & Kageura 2007）。

「翻訳教育向け『みんなの翻訳』」の仕様に合わせて、QRedit を以下のように拡張した。

- (1) 4.1 で述べた拡張 TMX に QRedit を対応。
- (2) 基本翻訳単位（段落）内の任意のテキストスパンを指定し、それに対して修正、修正理由、自由記述を加えるポップアップメニューの追加（図 1）。
- (3) 任意の修正バージョンに対する、3 ペインでの原文書、目標言語文書 1、目標言語文書 2、の対照表示機能・履歴参照機能の追加（図 2）。

これによって、修正と修正の動機に関する利用者のアクションは、QRedit を通してすべて一貫して文書と情報管理につなげることができる。

5 課題

現在残っている大きな設計上の課題は、「やりとり」をどう位置づけるかである。以下のような点を、現在、検討中である。

- やりとりをどこに、どのように紐付けるか、翻訳学習（翻訳、修正等々）のプロセスで、どこから「やりとり」のトリガをかけるか。
- どんなプラットフォームでやりとりを行なうか。
- ある一つのやりとりの途中で、対話行為タイプが変わるようなシナリオをどこまで吸収するか。やりとりの中で対話行為タイプはどのように動くのか。
- 修正の対象となっていないテキストスパンに対してどのようにやりとりのトリガをかけ紐付けるか。

また、修正をめぐって、基本翻訳単位（段落）を横断する修正をどう扱うかも検討課題として残っている。

6 おわりに

本論文では、「翻訳教育向け『みんなの翻訳』」の基本的な設計方針と現状を紹介した。2011 年 7 月までに、文書と情報の管理プラットフォームを含めたプロトタイプの実装を終え、翻訳教育の現場で試験的に導入し検証を進めていく予定である。現在のところ、検証はリーズ大学翻訳研究所を中心に行い、その他に、中国政法大学政法翻訳研究所でも検証を行なうべく、調整

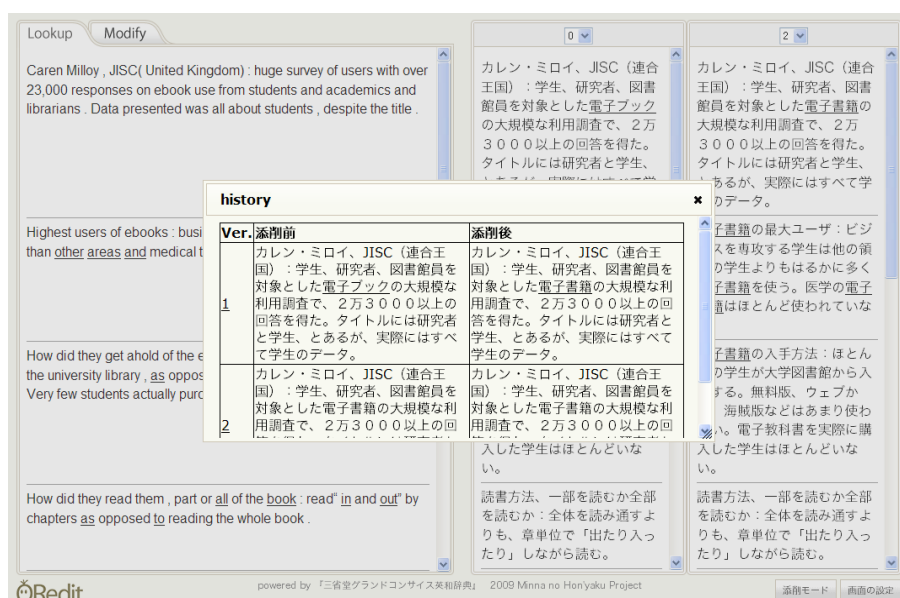


図 2. QRedit の 3 ペイン対照表示と履歴参照

を図っている。なお、開発したシステムは世界中の翻訳学校に導入できるよう、一般公開する予定である。

謝辞

本研究は、日本学術振興会科学研究費補助金挑戦的萌芽研究「翻訳における下訳・修正訳と機械翻訳出力の分析」(課題番号 20650020) および The Great Britain Sasakawa Foundation の補助を得て行なわれた。

参考文献

- Abekawa, T. & Kageura, K. 2007. "A translation aid system with a stratified lookup interface," *45th ACL Poster & Demo Session*, pp. 5-8.
- Abekawa, T. & Kageura, K. 2008a. "Constructing a corpus that indicates patterns of modification between draft and final translations by human translators," *LREC 2008*.
- Abekawa, T. & Kageura, K. 2008b "What prompts translators to modify draft translations? An analysis of basic modification patterns for use in the automatic notification of awkwardly translated text. *IJCNLP*, pp.241-248.
- Allen, J. & Core, M. 1997. *Draft of DAMSL: Dialog Act Markup in Several Layers. The Multi-party Discourse Group*. Rochester: University of Rochester.
- Castagnoli S., Ciobanu D., Kübler N., Kunz K. & Volanschi, A. 2006. "Designing a Learner Translator Corpus for Training Purposes," *TALC2006*.
- Kiraly, D. 2000. *A Social Constructivist Approach to Translator Education. Empowerment from Theory*

to Practice. Manchester: St. Jerome.

- LISA. 2011. <http://www.lisa.org/standards/tmx/>
- Mossop, B. 2001. *Revising and Editing for Translators*. Manchester: St Jerome.
- Secară, A. 2005. "Translation evaluation – a state of the art survey," *Proc. of the eCoLoRe/MeLLANGE Workshop*, pp. 39-44.
- Shih, Y. C. 2006. "Revision from translators' point of view. An interview study," *Target*, 18(2), pp. 295-312.
- Singley, M., Singh, M., Fairweather, P., Farrell, R., Swerling, S., 2000. "Algebra jam: supporting teamwork and managing roles in a collaborative learning environment," *CSCW 2000*, pp. 145-154.
- Soller, A. L. 2001. "Supporting social interaction in an intelligent collaborative learning system," *International Journal of Artificial Intelligence in Education*, 12(1), pp. 40-62.
- Robert, I. 2008. "Translation revision procedures: An explorative study," In Boulogne, P. ed. *Translation and Its Others. Selected Papers of the CE-TRA Research Seminar on Translation Studies*.
- Utiyama, M., Abekawa, T., Sumita, E. & Kageura, K. 2009. "Hosting volunteer translators," *MT Summit XII*.
- 内山将夫・阿辺川武・隅田英一郎・影浦峽. 2011. 「みんなの翻訳第 3 報」言語処理学会第 17 回年次大会.
- Van der Meer, J. 2006. Different approaches to translation workflow. Report on the TAUS Round Table, May 30 2006, Barcelona.