

訳文を修正しながら辞書更新ができる翻訳支援インタフェース

(株) 富士通研究所 大倉清司

okura.seiji@jp.fujitsu.com

1. はじめに

自動翻訳エンジンを使った翻訳支援ソフトウェアが産業翻訳をターゲットに開発されている[1, 2, 3]。翻訳作業の効率化をうたっているものの、訳語の修正と辞書登録が別の作業になっているのが現状である。自動翻訳結果の訳語を修正すると、辞書を編集しないでも、それ以降の自動翻訳結果に修正した訳語が反映されるという流れが理想である。

本稿では、翻訳エンジン辞書とは別に辞書テーブルを導入することにより、翻訳作業を中断することなく、訳語修正の結果を自動翻訳結果にリアルタイムに反映する仕組みを提案する。自動翻訳結果の単語対応情報を利用して更新情報を生成し、更新のタイミングを制御できる。またネットワーク上での辞書の共有にも対応できる。この仕組みを使ってインタフェースを実装したので、そのデモンストレーションを行う。

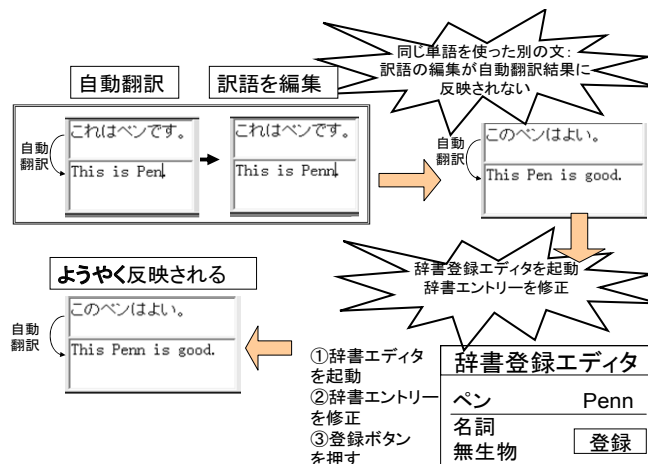


図1 翻訳作業と辞書更新作業の現状

2. 翻訳支援ソフトウェアにおける辞書更新の問題点

自動翻訳エンジンを使った翻訳支援ソフトウェアを使う場合、辞書更新や辞書登録の際に翻訳作業が中断されてしまう(図1)。訳文の編集作業の他に辞書エディタを呼び出す必要があるのに加え、辞書の登録自体、複雑な手順が必要な場合もある。産業翻訳においては、訳語の決定作業は全体の翻訳時間の中でも大きな割合を占めると考えられるため、翻訳者が時間をかけて調べた単語をすぐに辞書に登録する仕組みが必要である。辞書登録作業を意識することなく、翻訳をすすめるうちに自動翻訳が修正した訳語を自動的に反映した翻訳文を出力できることが望ましい。自動翻訳を使い訳語統一するには辞書登録が必須の作業となるが、訳文修正以外に辞書エントリの保守作業が発生してしまうのは問題である。

ることは仕組みとしては可能であるが、現実的でない。この単純な仕組みを実現するには、翻訳エディタがキータイプイベントを受け取るたびに(または単語区切りがわかっているため、カーソルが別の単語に移ったときに)翻訳エンジン辞書の更新(エントリを辞書から読み込み修正したエントリをその辞書へ書き込む処理)をすればよい。しかし、頻繁に発生する訳文の編集作業と同時に、辞書データベースを頻繁に更新しなければならないのは好ましくない。また、翻訳エンジン辞書そのものが更新されてしまうため、例えば誤って編集した訳語を後で元に戻したいときに戻せなくなる。

訳文を修正していくと自動的に辞書が更新される処理を現実的に実現するには、2つの問題を解決する必要がある。①翻訳エンジン辞書へのアクセスをなるべく少なくし、かつ②訳語の編集が次の翻訳結果にリアルタイムに反映されなければならない。

3. 辞書リアルタイム更新手法

この問題を解決するため、自動翻訳結果の訳文を修正すると辞書も更新されていくという手法(以下、辞書リアルタイム更新手法)を提案する(図2)。自動翻訳の過程で原文の単語と訳文の単語の対応が保存されるため、訳語の修正に応じて自動的に翻訳エンジン辞書を更新す

4. 提案する方法とその利点

1つ目の問題については、翻訳エンジン辞書へのアクセスを最小限にするため、辞書テーブルを翻訳エンジン辞書と別に用意し(図3)、訳語編集の際に辞書テーブルのみ更新されれば、翻訳エンジン辞書へのアクセスは不

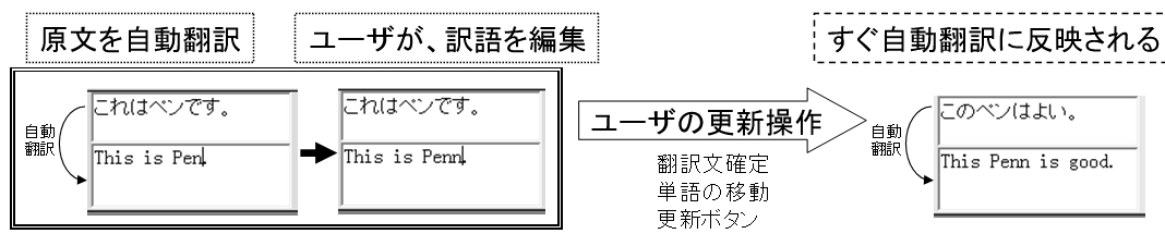


図2 本稿で提案する仕組みのイメージ図

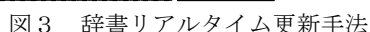
提案する辞書リアルタイム更新手法は、更新のための翻訳エンジン辞書へのアクセス数を減らせるだけでなく、以下2つの利点もある：

辞書テーブルをネットワーク上におき、複数の翻訳者で共有する仕組みをつくることも可能である。辞書テーブルには登録する単語の他、更新対象の辞書、更新日付、更新者を記録しておく（図 4）。例えば優秀な翻訳者 A さんが作成したエントリについて自動翻訳結果の訳語を編集不可能にすることができる（A さんだけは編集可能）。これにより複数の翻訳者間で訳語統一作業が効率化できる。もちろん、A さんが訳語を更新したときは、その訳語を含む訳文全てを（半）自動修正することが可能である。

今回は更新対象の品詞を名詞に限ったが、他の品詞にも適用していく。この手法を使うことによりどれだけ翻訳の効率化がはかれるかの評価も行っていく。

なお、辞書テーブルは、UTX-Simple[6]に準拠することもできる。

1. 翻訳エンジン辞書を直接更新すると、エントリーは上書きされ、更新の履歴が残らないが、辞書テーブルでは更新したエントリーを更新日時とともに追加することにより更新履歴を保存することも可能である(図4)。今回開発したシステムは、辞書テーブルが参照されて翻訳されたという情報も保存される。辞書テーブルに更新履歴を保存することにより、開発したシステムの仕組みを使えば、例えば2008年3月1日時点の辞書に準じた訳語に戻したい、という従来不可能だったことも可能となる。
2. 現状のサーバー・クライアント型翻訳システムにおいては、翻訳中に辞書更新ができないため、クライアントで辞書を作成し、サーバーの翻訳プロセスを停止した後に辞書をサーバーに登録する必要がある。



登録単語		辞書名	更新日付	更新者
ペン	Pen	ユーザ辞書2	20071010	User2
ペン	Penn	ユーザ辞書2	20071201	User1
自動翻訳ソフトATLAS バージョン14	machine translation software ATLAS version 14	ユーザ辞書2	20071201	User1
.	.			

図4 更新履歴の保存

- [1] 富士通. 英日・日英翻訳ソフト ATLAS.
<http://software.fujitsu.com/jp/atlas/>
- [2] Seiji Okura, Tatsuo Yamashita, Masaru Fuji, Guowei Xu, Akira Ushioda. Integrating MT and TM: Cliché. 20th International Conference on Computational Linguistics (COLING), Exhibition.
- [3] 大倉清司、富士秀、長瀬友樹. オートコンプリートによる翻訳支援. 言語処理学会第 13 回年次大会予稿集, 2007.
- [4] 大倉清司、徐国偉、山下達雄、富士秀、潮田明. 多言語翻訳プラットフォーム Cliché. 言語処理学会第 10 回年次大会予稿集, 2004.
- [5] 長瀬友樹、大倉清司、富士秀、徐国偉、潮田明. 多言語翻訳プラットフォームによる翻訳サービスの実装. 言語処理学会第 12 回年次大会予稿集, 2006.
- [6] 大倉清司、山本ゆうじ、村田稔樹、内元清貴、加藤マikel、孝仁、島津美和子、鈴木次良. 共有ユーザー辞書仕様 UTX の現状と今後の展開. 言語処理学会第 14 回年次大会予稿集, 2008.