

Cliché: さらなる翻訳効率化のための翻訳支援インターフェース

大倉清司、富士秀、徐国偉、長瀬友樹、潮田明
(株)富士通研究所

okura.seiji@jp.fujitsu.com

1. はじめに

機械翻訳は近年広く知れわたり、いろいろなところで使われている。Web 上では無料の翻訳サービスを提供しているところも多い[1]。しかし、質の高い翻訳が要求される産業翻訳の現場においては、翻訳対象文の複雑性ゆえになかなか自動化が進んでいない。富士通研究所では、産業翻訳すなわち「書くための翻訳」に対して翻訳作業効率を上げることを目的に、翻訳支援システム Cliché[2,3,4,5,6]を開発した。今回、特に長く複雑な文の作業ステップ数を従来に比べ大幅に削減する、すなわちより効率よく翻訳するためのインターフェースを実装した。

2. 翻訳支援システム Cliché

Cliché は機械翻訳技術と訳例検索技術を統合したシステムである(図 1)。その基本的な思想は"divide-and-conquer"である[2](図 2)。
"divide-and-conquer"とは、文の大まかな構造をつかみ訳文の骨格を決め、そして骨格の主要要素をそれぞれ翻訳し、最後に各要素の翻訳を骨格にはめ込んで訳文を「組み立て」て完成させるという考えである。要素が長く複雑な場合はその要素を更に骨格と要素に分割して翻訳を進める。この分割の過程で、同一の表現が訳例中に見つかるか、あるいは機械翻訳で要素が正しく訳されるようになる。ユーザは最後に作成

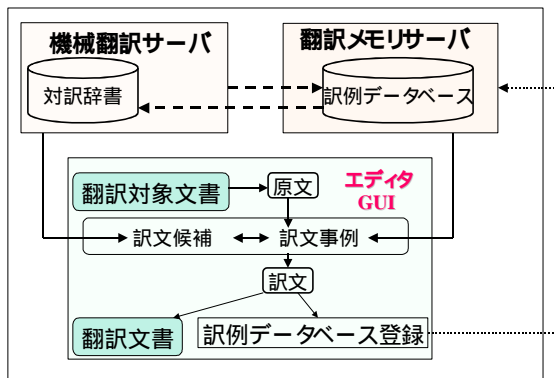


図 1 Cliche 構成図

された訳文をととのえて完成する。

3. 改善すべき問題点

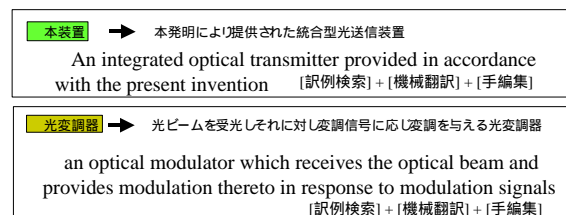
この考えは、分割した単位でうまく翻訳作業がすすむときに効果的である。つまり、分割した通りに要素が翻訳でき、組み立てて訳文が完成する場合である。しかし、特に長く複雑な文を翻訳する場合、分割要素の訳をつくったところで分割を修正したり、組み立てた訳文を修正したりすることがあり、その作業に時間がかかってしまう。

原文：

〔本発明により提供された統合型光送信装置〕は、〔光ビームを受光しそれに対し変調信号に応じ変調を与える光変調器〕によって特徴付けられる

骨格文：〔本装置〕は、〔光変調器〕によって特徴付けられる。

→ This device is characterized by an optical modulator. [機械翻訳]



最終訳文

[An integrated optical transmitter provided in accordance with the present invention is characterized by an optical modulator which receives the optical beam and provides modulation thereto in response to modulation signals.]

(a) 骨格文-主要素の分割と翻訳結果の結合

光ビームを受光しそれに対し変調信号に応じ変調を与える光変調器

- 1 光ビームを受光し
receives the optical beam [訳例検索]
- 2 変調信号
modulation signals [機械翻訳]
- 3 変調を与える
provides modulation [訳例検索]
- 4 光変調器
an optical modulator [機械翻訳]

an optical modulator which receives the optical beam and provides modulation thereto in response to modulation signals [手編集]

(b) [訳例検索]+[機械翻訳]+[手編集]の例

図 2 Cliche における翻訳作業の流れ([2]から引用)

このような作業での問題には、以下の4点がある。

- (a) 分割の表示に関する問題
- (b) 訳文の組み立てに関する問題
- (c) 分割の修正および削除に関する問題
- (d) 分割要素の訳の修正に関する問題

(a)に関しては、分割して翻訳をすすめていくと、どこを翻訳しているのか、どこまで翻訳したのかが判断できなくなるという問題が起こる。そもそも、各分割部分は独立して翻訳可能という前提をしていたが、実際の翻訳においては他の分割部分との関係や、全文の中の位置などを参照したい場合も出てくることわかった。

(b)に関しては、訳文を組み立てたときに、どの分割要素の原文が組み立てた訳文のどこに対応するかということに関する明示的なインターフェースがなかった。また分割要素も文の骨格も同じ色・フォントで表示されていたため、その区別もできなかった。

(c)に関しては、要素の分割を多数回、あるいは再帰的にすすめていった結果、ある時点で分割が違っていった、という場合に、分割の修正や削除が容易にできる必要があった。

(d)に関しては、分割要素に相当する部分の訳を、訳文を組み立ててから修正したい場合、どこを対象に修正すればよいか(分割要素の部分か、組み立てた文か)わからないという問題があった。組み立てた訳文を最終形とするなら、その訳文を修正すればよいはずだが、例えば他の分割要素の修正もしたい場合や、分割を修正してから再度、文を組み立てたい場合などは全体の整合性が保たれていなくてはならない。また、分割要素の中の骨格部分を修正して訳文を再度組み立てる場合も同様である。

4. ユーザインターフェースの改良

これらの問題に対処するために、Cliché 内部のデータ構造も見直し、インターフェースの拡充を行った(図3)。

(a)に関しては、分割構造を木構造により表示する他、常に翻訳対象文の全体を画面下に表示するようにした。木構造においては、翻訳者が今どここの要素の翻訳をしているかを示す他、その要素の訳文が完成したか、していないかがわかるように表示するようにした。

(b)に関しては、対応する分割要素の原文・訳文を同じ背景色で強調表示し、多数回分割をしても、どの原文箇所がどの訳文箇所に相当するか即座に判断できるようにした。編集によって、色表示がずれないようなデータ構造で実装した。

(c)に関しては、分割の修正、削除が容易に行えるインターフェースを実装した。訳文を組み立ててからも分割の修正ができるが、その場合は修正後の分割と、訳文の対応する部分との整合性がなくなるため、

修正後の分割部分の訳文をつくりあげてから再度訳文を組み立てる必要がある。

(d)の部分に関しては、訳文の修正は、該当する分割要素を修正する、というインターフェースにした。修正が終わったところで、訳文を組み立てて、全体の整合性をとることができる。

5. これからの展望

翻訳支援システムにおいては、人間が主体となって翻訳作業をすすめていくが、自動化できるところはかなりある。例えば、文を要素に分割するのは訳例や学習によったりして自動化できる可能性がある。訳文の組み立てに関しても、同様に自動化できる可能性がある。翻訳支援の1つ1つの翻訳フェーズに対して正確に自動処理できるようになれば、まさに究極の機械翻訳が完成できると考えている。そのためには、まずは翻訳支援のインターフェースをつきつめ、自動化できるところは自動化して、産業翻訳の効率化をはかっていきたい。

参考文献

- [1]excite 翻訳 <http://excite.co.jp/>
- [2]潮田明,富士秀,大倉清司,山下達雄.機械翻訳と訳例検索を統合した翻訳支援システム.言語処理学会第9回年次大会予稿集,2003.
- [3]山下達雄,富士秀,大倉清司,潮田明.翻訳支援に有効な訳例検索の類似度計算方式と検索結果提示方式.言語処理学会第9回年次大会予稿集,2003.
- [4]富士秀,潮田明,大倉清司,山下達雄.翻訳支援システム導入による効率化の評価.言語処理学会第9回年次大会予稿集,2003.
- [5]大倉清司,山下達雄,富士秀,潮田明.機械翻訳と訳例検索を統合した翻訳支援システムのインターフェース.言語処理学会第9回年次大会予稿集,2003.
- [6]Seiji Okura, Tatsuo Yamashita, Masaru Fuji, Guowei Xu, Akira Ushioda. Integrating MT and TM: Cliché. 20th International Conference on Computational Linguistics(COLING), Exhibition.

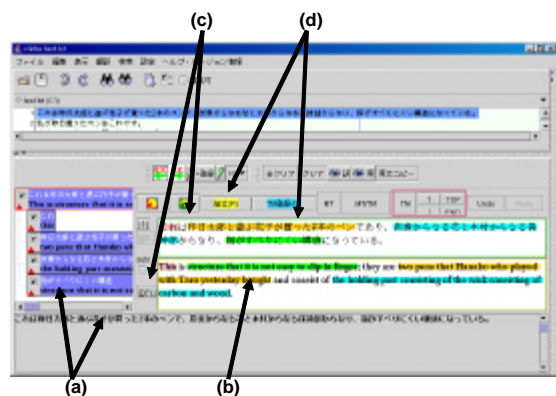


図3 改良したユーザインターフェース