

多言語翻訳統合プラットフォーム Cliché

大倉清司、徐国偉、山下達雄、富士秀、潮田明

(株)富士通研究所

okura.seiji@jp.fujitsu.com

1. はじめに

産業翻訳における業務効率を向上させるために開発した統合型翻訳支援システム Cliché[1,2]をベースに、多言語翻訳ができるプラットフォームを構築した。Clichéは機械翻訳システムと翻訳メモリシステムを内包する翻訳支援システムだが、従来の英日・日英翻訳だけでなく、他の言語の翻訳エンジンを組み込むことで従来と同じインターフェースからその翻訳エンジンおよび、その翻訳エンジンで解析された翻訳メモリを検索しながら翻訳することを今回実現した。機械翻訳、翻訳メモリ共に英日・日英翻訳と同等の機能を持つ多言語版 Cliché をデモンストレーションにより報告する。

次節では Cliché の概要を説明する。3 節ではプラットフォームとしての Cliché の構成を説明する。4 節では具体的なシステム開発について説明する。5 節では拡張性について述べる。

2. Cliché の概要

Cliché は機械翻訳技術と訳例検索技術を統合した翻訳支援システムであり、これまでの英日・

日英方向の定量的翻訳評価で、人手翻訳の 3 ~ 4 倍の効率化を達成している[3]。従来の訳例検索を超えた翻訳支援用インターフェース[1,4]および類似度計算方法[4]を実現し、ネットワークによる翻訳メモリ、機械翻訳辞書の共有が可能である[1]。

3. プラットフォーム Cliché の構成

多言語翻訳統合プラットフォームの構築において、既存のシステムをモジュール化・再構成した(図1)。Cliché のコア機能ともいえる「Cliché システム」は、機械翻訳エンジンおよび翻訳メモリエンジンを内包する。このうち翻訳メモリエンジンについて言語に依存しないよう多言語対応を行った[5]。インターフェースとしては従来から機械翻訳結果表示 GUI、翻訳メモリ結果表示 GUI、および Cliché 画面の GUI があるがこれに関してもモジュール化した。機械翻訳部分に関しては、翻訳エンジンの違いを意識することなく任意の翻訳エンジンをプラグインして Cliché で使用できるように、翻訳エンジンの出力を設計した。

4. Cliché の翻訳エンジン組み込み対応

任意の翻訳エンジンを Cliché で使えるようにするため、図2に示す開発を行った。従来、英日・日英翻訳エンジン ATLAS[6]を改造し翻訳メモリとの統合のための内部情報を出力するようにしていた。この情報をベースに、多言語対応できるように再設計し、内部文字コードを Unicode ベースにすることにより、任意の翻訳エンジンの出力を扱えるようにした。今回、中日翻訳エンジンを改造し、Cliché に組み込めるようインターフェースを統一した。これは翻訳メモリエンジンからも直接参照し処理できる

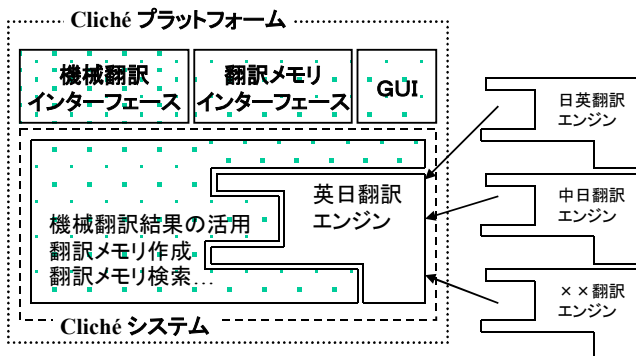


図1: 多言語統合プラットフォームCliché

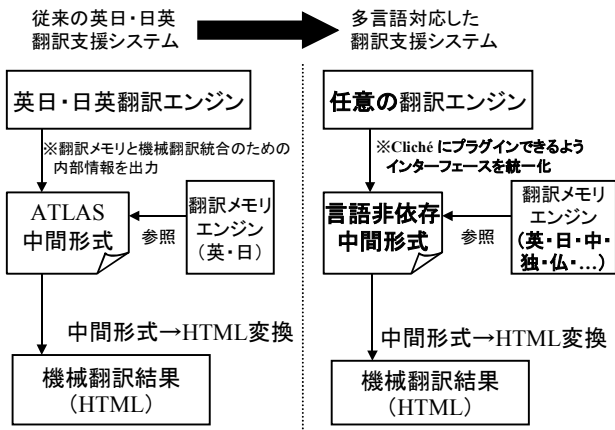


図2: 機械翻訳エンジンのプラグイン対応

ようになっている。Cliché は、翻訳エンジンから出力された内部情報をもとに機械翻訳結果を出力したり、翻訳メモリを検索・作成などする。つまり、Cliché 内部で扱うデータ形式を言語に依存しない形で設計・開発することにより多言語翻訳支援システムのプラットフォームを構築した。Cliché の翻訳要求はhttpベースでネットワークを通じて行われるため、翻訳エンジンからの出力フォーマットさえ同じにすれば任意の翻訳エンジンを組み込める。実際のネットワーク構成は図3のようになっており、Cliché 側の切り替えによって再起動することなく、その場で翻訳エンジンを選択可能となっている。

このようにしてできた Cliché プラットフォームからは、任意の翻訳エンジンの結果から機械翻訳結果を表示したり、翻訳メモリを作成・検索したりすることが可能である。翻訳メモリは機械翻訳により出力された対応情報も同時に出力されるため、再現率が低い翻訳メモリでも有効活用できる[3]。

5. 拡張性

Cliché プラットフォームは拡張性に優れている。まず、翻訳エンジンと Cliché 本体はhttpベースで通信するので翻訳エンジンの変更は Cliché に影響しない。機械翻訳結果、翻訳メモリ検索結果、翻訳作業領域など GUI を変更するのも、Cliché システムとは別に修正できる。現在、図4のインターフェースで英日、日英、中日翻訳支援が可能となっている。

参考文献

[1]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.
 [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]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.
 [4]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.
 [5]山下達雄,大倉清司,徐国偉,富士秀,Stefan Schmidlin,潮田明.単語対応情報付き訳例検索システムの多言語対応.言語処理学会第10回年次大会予稿集,2004.
 [6]富士通.英日・日英翻訳ソフト ATLAS.
<http://software.fujitsu.com/jp/atlas/>.

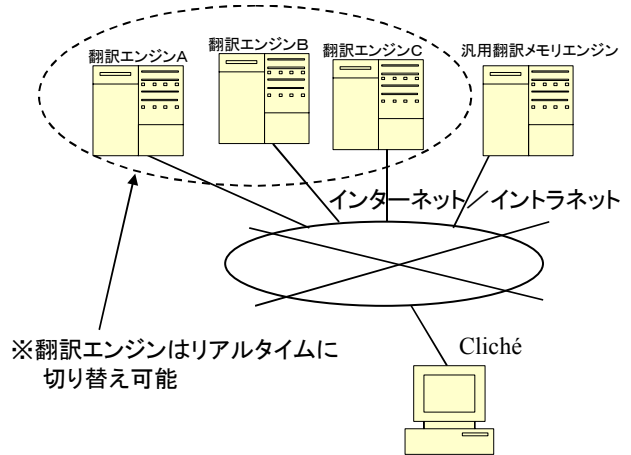


図3: Cliché と翻訳エンジンのインターフェース

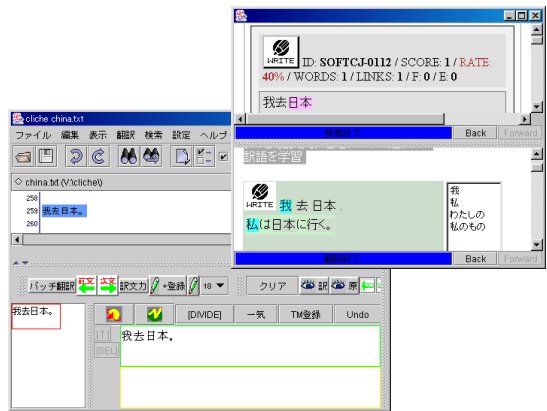


図4: 多言語翻訳インターフェース