

# ハイブリッド機械翻訳を柔軟に構成するための ソフトウェアアーキテクチャ設計

2014SE062 松岡秀樹

指導教員：沢田篤史

## 1 はじめに

コンピュータの処理性能と記憶容量の拡大により、機械翻訳が実用化に近づいている。記憶容量の拡大により扱うことができる言語データが増え、処理能力の向上により、翻訳できる言語の精度が上がった。今後も翻訳速度の向上や、機械翻訳ツールの使用性の向上などの技術が次々に追加されることが予想される。ハイブリッド機械翻訳 [1] はそのような技術の一つで、既存の翻訳手法を適材適所に組み合わせ、より良い翻訳結果を得ようとするものである。このような状況では、新たに追加される技術も含めて、翻訳ツールを柔軟に構成する技術が必要である。我々はそれを実現することのできる共通のアーキテクチャを構築することで、この要求に応えられると考えている。

本研究の目的はハイブリッド機械翻訳を統一的に説明でき、かつ実装の基盤となるアーキテクチャを作成することである。

本研究では、パイプとフィルタのアーキテクチャスタイル [4] に基づいて翻訳ツールの組み合わせを耐故障アーキテクチャの考え方を取り入れることで構成し、本研究室で提案されている自己適用のためのソフトウェアパターンである PBR パターン [3] を用いて設計する。

このアーキテクチャにより、既存のハイブリッド機械翻訳で用いられるツールの組み合わせを網羅し、またツールの入れ替えや組み合わせを柔軟に変更することができる。

## 2 ハイブリッド機械翻訳アーキテクチャ構築 対しての課題

### 2.1 機械翻訳の手法

機械翻訳の主な手法には統計的機械翻訳 (Statistical Machine Translation, SMT) とルールベース機械翻訳 (Rule-Based Machine Translation, RBMT) と用例ベース機械翻訳 (Example-Based Machine Translation, EBMT) の3つがある。SMT は多くの文法を翻訳することができ、RBMT はより精度の高い翻訳が、EBMT は用例さえそろえば、完璧な翻訳ができる。

ハイブリッド機械翻訳は SMT, RBMT, EBMT の3つの翻訳手法のそれぞれの長所を生かすために、用途にあわせて組み合わせで作られる。

### 2.2 ハイブリッド機械翻訳の翻訳手法の組み合わせ

ハイブリッド機械翻訳内で、翻訳手法は主に原言語の簡略化、原言語から目的言語への翻訳、目的言語の精度向上の場合に用いられる。

翻訳手法の組み合わせの例を図 1 に示す。

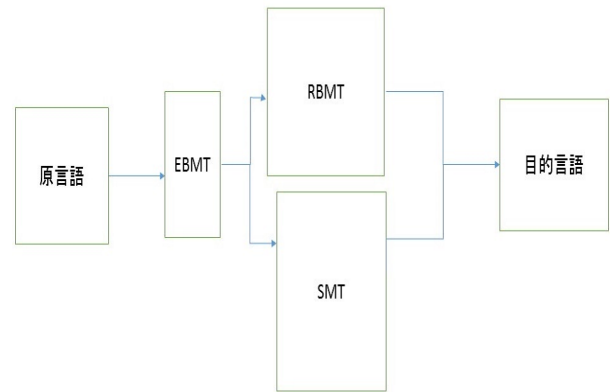


図 1 ハイブリッド機械翻訳の結合の組み合わせの例

図の例では、原言語を EBMT を用いて翻訳し、出力された目的言語を SMT と RBMT の両方で精度向上のための翻訳を行い、何らかの基準で評価を行った結果、より適切な方の出力を翻訳結果として採用する。

ハイブリッド機械翻訳の翻訳手法の組み合わせは簡略化、翻訳、精度向上の3つの場面で直列または並列つなぎで構成される。

### 2.3 ハイブリッド機械翻訳の課題

現在ハイブリッド機械翻訳には様々な翻訳技術の結合方法があるが、今後はより新しい形のハイブリッド機械翻訳が作られていくと予想される。それらの新たな技術に柔軟に対応できる環境を整えることが必要である。

## 3 ハイブリッド機械翻訳を柔軟に構成するための アーキテクチャ設計

### 3.1 ハイブリッド機械翻訳のためのアーキテクチャ設計 技術

本研究で提案するアーキテクチャでは、相互作用のパターンが連続するデータを変換するという特徴があるパイプとフィルタアーキテクチャスタイルを用いる。

翻訳手法の組み合わせと耐故障性の類似点から耐故障アーキテクチャの N バージョンプログラミングとリカバリブロックの技術を用いる。

パイプとフィルタに耐故障アーキテクチャを組み合わせるにあたり、翻訳結果の評価に応じてツールを組み合わせることが必要である。それを可能するソフトウェアパターンとして江坂らが提案している PBR パターン [3] を用いて設計した。

### 3.2 ハイブリッド機械翻訳のためのアーキテクチャ設計

図2にハイブリッド機械翻訳の言語から言語への翻訳の流れの基本構造を示す。

External：機械翻訳の利用者（システム外部の実体）

Filter：機械翻訳ツール（SMT, RBMT, EBMT のいずれか）

Source：Filter へ入力データを渡す利用者や機械翻訳ツール

Destination:Filter の出力するデータを受け取る機械翻訳ツールまたは利用者

MT Factory:翻訳結果の受け入れ評価に基づいて、Filter を生成するファクトリ

Hybrid MT Policy：データの受け取り

Evaluator：翻訳の評価

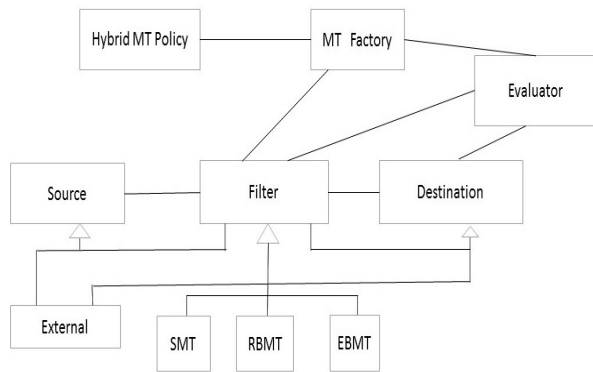


図2 静的構造

図2, 図3では日英翻訳の日本語から日本語への原言語の簡略化を動的振る舞いで記述した。図3では、リカバリブロックの直列な動きを J to J MT Factory で create と send をフィルタごとに行うことで表した。図4の Nバージョンプログラミングの並列な動きを J to J MT Factory で同時に send, 評価することで表した。

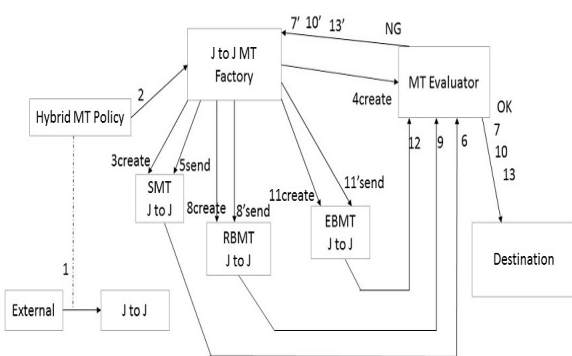


図3 リカバリブロック

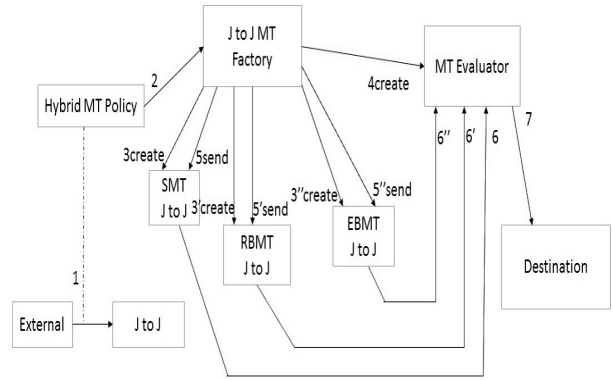


図4 Nバージョンプログラミング

## 4 考察

前章で設計したアーキテクチャにより、ハイブリッド機械翻訳の新しい組み合わせも統一的に説明できる。

例えば、[1]で提案されているハイブリッド機械翻訳(図1)については、SMTとRBMTが並列に、EBMTが直列に組み合わせられるがこれをNバージョンプログラミングとリカバリブロックを用いて構成できることが分かる。

このことから本研究で提案したアーキテクチャは妥当であるといえる。

## 5 おわりに

コンピュータの処理能力が向上に伴い機械翻訳ツールは進化を続けている。これからの機械翻訳は新技術の導入もふくめた翻訳ツールを柔軟に開発する環境が必要になる。

本研究ではこの問題を解決するためにPBRパターンを用いてハイブリッド機械翻訳を柔軟に構成するためのアーキテクチャ設計した。

今後の課題は、Evaluatorの検証を行い、本研究で提案したアーキテクチャに基づいて翻訳ツールを構成し、提案の実用性を示すことである。

## 参考文献

- [1] 張孝飛:ハイブリッド戦略ベースの特許機械翻訳に関する研究, Japio YEARBOOK 2015,pp334-341, 2015.
- [2] 知野哲朗, 釜谷聡史:ハイブリッド機械翻訳技術による日中英音声翻訳システム, 東芝レビュー 64巻2月号, 2009.
- [3] 野呂昌満, 沢田篤史, 江坂篤侍:インタラクティブシステムのための共通アーキテクチャの設計, ソフトウェア工学の基礎 24日本ソフトウェア科学会 FOSE2017, pp129-134, 2017.
- [4] Paul Clements, Felix Bachmann, Len Bass, David Garlan, James Ivers, Reed Little, Paulo Merson, Robert Nord, Judith Stafford: Documenting Software Architectures, Addison Wesley, 2004