

次世代翻訳支援エディタ「Teddy」

松井 亜希子 安達 久博

株式会社サン・フレア 研究開発センター 自然言語処理研究所

{matsui_a, adachi_h}@sunflare.co.jp

1 はじめに

企業のグローバル化が加速化している現在、多様な技術文書を相手先言語へ翻訳する需要も年々増加傾向にある。このため産業翻訳を担う翻訳業界では、大量の文書群を短納期で仕上げる必要性があると同時に、訳質の向上も求められている。機械翻訳 (MT) という概念は、ある意味においてこの大量・短納期という制約条件を満たしている。しかしながら、翻訳品質という制約条件を現状の市販システムの多くが満たしていないという指摘がある。そのため、計算機が翻訳する部分を捨て、翻訳対象原文に類似する翻訳用例を翻訳メモリ (TM) から最適検索し、それらを翻訳者に提示し、翻訳処理に再利用する翻訳支援システムが利用されている。さらに、機械翻訳技術と用例検索技術を融合する形の翻訳支援システムなども提案されている [2]。本稿では、現在、我々が開発中の次世代翻訳支援システム「Teddy」について、その基本構想と一実現例を基に次世代翻訳支援システムの方向性について、デモを通して広く参加者と議論したい。

2 背景と問題点

最近の市販機械翻訳システムの多くは、翻訳支援システムの要素技術である対訳用例 (翻訳メモリ) と、その最適照合検索、提示機能を搭載している。すなわち、機械翻訳技術と用例検索技術の融合化と捉えることができる。この枠組みは大きく2つに分類できる。(1) 機械翻訳が主であり、用例検索が従という構成の機械翻訳システム、(2) 用例検索が主であり、機械翻訳が従という構成の翻訳支援システム

しかしながら、この両者の枠組みに共通する機械翻訳技術の利用という点に、利用者側の視点から以下の問題点があると我々は考える。

利用者にその仕組みが見えないブラックボックス化している部分に解析誤り、単語の対応関係や原文と用例との照合誤り等があった場合、そのリカバリーに複雑な操作を必要とする。翻訳現場での翻訳原稿の多くは文法的な誤りを含むものが少なくなく、またメタ言語的な引用や長い並列句の多用、新語 (造語) などもあり、非文と捉えられる文といえる。そのため未定義語の対処や隣接条件など非 (文法的な) 文に対する柔軟性が形態素解析、構文解析、用例検索技術に求められるが、現状の解析技術等はこれらを受け入れる構成法とはいえない。

3 Teddy の特徴と構成

図.1 に Teddy¹ の概観を示す。我々は前述の問題点を回避する意味で、形態素解析による単語分割、品詞同定に基づくアプローチではなく文字列の最適照合をベースとする方式を採用することにした [1]。その基本的な考え方は翻訳メモリ内に蓄積された例文と句単位で完全一致したもののみを置換候補とする。なお、正規表現の形式で蓄積されたものは、その限りではない。必要に応じて使用者は句単位の区切りを変更できる等、基本的な操作は仮名漢字変換における操作性と同じインタフェースと捉えることができる。

¹ Teddy「テディ」は「Translation EEditor, but Do it Yourself」に由来し「翻訳は自分でしてね!」の意味を込めて命名。

また、本システムは原文を上書きする方式の翻訳支援エディタである。この原文を上書きする方式は訳文の訳抜け、訳漏れを防止する上で重要であり、原文と訳文との対応関係を保持した状態で置換処理が適用される。また、置換された結果は自動的に翻訳メモリに登録される。すなわち、原文と訳語が混在する形で翻訳処理が進む。さらに使用者である翻訳者の操作履歴が管理される構成のため、訳語を含む形で句単位のメモリ登録がなされる場合に、その訳語の置換履歴を基に対応する言語に還元された形式で登録される利点がある。

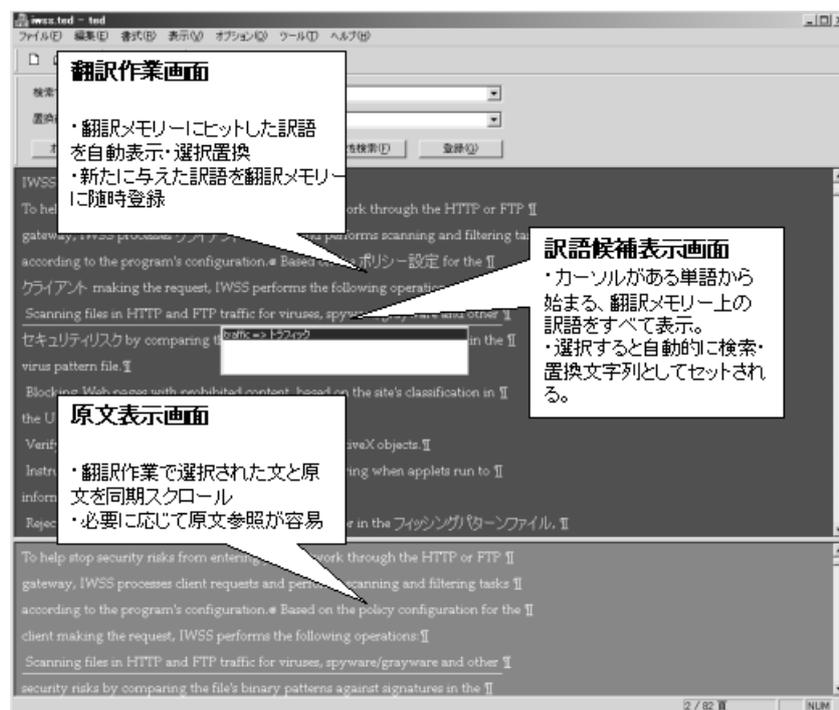


図 1: Teddy の概観

4 おわりに

本稿では、次世代翻訳支援システム「Teddy」の基本構想について述べ、その一実現例を示し翻訳支援システムの方向性について述べた。今後、本エディタの定量的、定性的な評価実験を行い、翻訳支援効率の向上をユーザインタフェースなどの使い勝手も考慮し、検証する必要があると考える。

参考文献

- [1] Satoshi SATO. CTM:An Example-Based Translation Aid System. Proc.of COLONG-92, pp.1259-1263. 1992.
- [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. pp.413-416 2003.