

翻訳支援のための自動翻訳技術の部品化およびサービス提供

大倉清司、長瀬友樹、潮田明

(株) 富士通研究所

{okura.seiji, nagase.tomoki, ushioda}@jp.fujitsu.com

1. はじめに

自動翻訳の研究は 30 年以上されており、PC 用製品も販売されている。近年はインターネット上で無料翻訳サービスも多数提供されている[1]。一般の消費者にとって、自動翻訳はかなり身近なものとなっている。また自動翻訳技術を使った様々な翻訳支援技術の研究がされている[2,3,4]。当社は従来より「翻訳支援ワークベンチ Cliché」をベースに様々な機能を提供してきた[5,6,7,8,9,10,11,12,13,14,15]。しかし、産業翻訳の現場では自動翻訳技術は依然としてあまり使われていないのが現状である。翻訳精度や初期コストの問題から導入を見送る翻訳会社も多い。

翻訳作業の効率化のためには辞書引き機能は必須と言える。翻訳支援ワークベンチ全体を大きな 1 システムとして提供する他、翻訳支援に必要な 1 つ 1 つの機能 (例えば辞書引き機能) を部品として提供できれば、産業翻訳界で自動翻訳技術の適用が進むと考える。Cliché はクライアント・サーバー型のシステムであるが、サーバーは SOAP 機能により構成されており、全機能は部品化されている。この機能をサービスとして使えば、初期コストを抑えて自動翻訳技術、翻訳支援技術を導入することが可能になる。本発表においては、「自動翻訳技術の部品化」という面から、辞

書引き機能、対訳文の単語対応づけ機能、単語対応つき翻訳機能などを例に、これらの部品を組み合わせることで開発したシステムのデモンストレーションを行う。

2. 翻訳支援ワークベンチ Cliché

富士通研究所では、産業翻訳における業務効率を向上させるための翻訳支援ワークベンチ (開発コード:Cliché [クリーシェ]) を開発している (図 1)。GUI 改良とユーザビリティテストとの繰り返しによる GUI の最適化を図った。

翻訳業界で見られる翻訳業務フローの環境を構築し、Cliché の評価を行った。評価にあたっては、品質を犠牲にすればいくらかでも早く翻訳できるため、翻訳時間と翻訳品質を合わせた「翻訳効率」を定義した。Web 上のオンラインヘルプマニュアルの英日翻訳で人手翻訳の約 3 倍、特許文書の日英翻訳で約 4 倍の効率化を実現した[8]。

Cliché の特徴は以下の通り：

1. 機械翻訳技術と訳例検索技術の統合

Cliché は、訳例検索と自動翻訳の統合を設計の段階から考えた、使えば使うほど賢くなるシステムである。自動翻訳技術により訳例内の単語対応を付与でき、自動語句対応付けシステムにより訳例中の対訳例文から自動翻訳用辞書を構築できる。翻訳していくにつれ、辞書や訳例が自動的に蓄積されていき、語句や訳例の再利用性が高まる仕組みになっている。つまり、辞書が使い込まれれば使い込まれるほど自動対応付けシステムの精度が向上し、訳例が蓄積されれば蓄積されるほど自動翻訳用辞書が高精度に構築される。

2. 効果的な GUI および類似度計算方法による訳例検索機能

従来の訳例検索において類似度が低い訳例が使われることはまれだったが、個々の

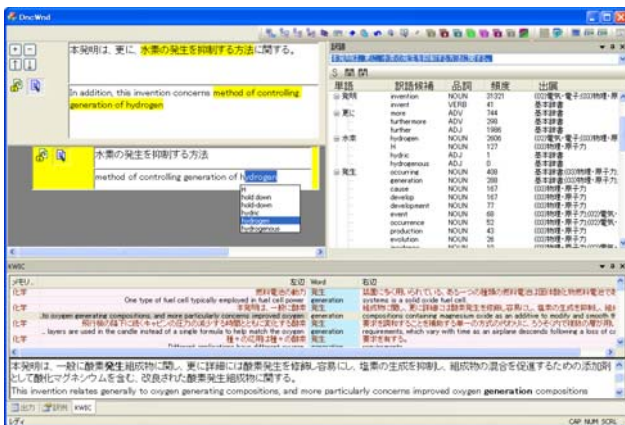


図 1 翻訳者ワークベンチ Cliché

文同士が文全体としては類似度が低くても、部分的に良く使われる表現を多く含む場合は組み合わせることで効率的な再利用が可能ではないかという発想で Cliché の訳例検索機能を設計した[9]。その結果、従来の訳例検索では効率化がはかれなかった、特許文書や契約書などの文書にも Cliché が有用であることが実証された。

3. ネットワークによる訳例データベースおよび機械翻訳辞書の共有

ネットワークを通じて機械翻訳、訳例検索をすることができ、リアルタイムに共有メモリを更新できる[9]だけでなく、自動翻訳辞書もリアルタイムに更新可能である[14]。これにより、Web サービスで提供される自動翻訳結果を、用語集によりチューニングすることが可能となった。

4. オートコンプリートによる入力支援

訳文入力時に、入力予測により自動的にオートコンプリート機能が働き、正確な綴りを素早く入力できる(図 2)[12]。

5. 訳語信頼度の表示による訳語選択支援

辞書引き画面において、訳例中の、原文の単語と訳文の単語が同時に出現する頻度が表示されるため、訳語の確認時間が大幅に削減される。辞書の出展も表示される[13]。

3. 翻訳支援技術の部品化

Cliché は大規模な統合的翻訳支援ワークベンチであり、翻訳規模が大きければ大きいほど効果を発揮する。一方で、100 文ほどの英語を日本語に翻訳するのに Cliché 一式をインストールし、辞書引き機能だけを使って翻訳するのは現実的ではない。システムの利便性という観点から考えると、Cliché が持つ翻訳支援機能を部品化し、インターネットを通じて提供できれば、より多くの翻訳者に利益をもたらせる。また、利用者が多ければ多いほど、利用者にかかるコストが少なくなる。

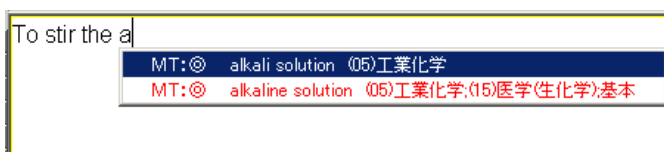


図 2 オートコンプリート機能

ここで、部品化とは単なるライブラリの提供ではなく、「Web サービス」としての提供を意味する。サーバーと通信するライブラリを開発者に提供するだけでは、ライブラリを更新するたびにダウンロードが必要になる他、開発マシン・クライアント実行マシン上の他のライブラリとの整合性を考慮しなければならない。Web サービスとして提供できれば、クライアント開発者は WSDL を理解して SOAP インタフェースを構築するだけでよく、開発コストも大幅に低減される。また、利用したい機能だけに絞って開発できるのも特徴である。

4. サービスとしての翻訳支援技術の提供

以下は、Cliché の翻訳サーバー、訳例検索サーバーにより提供される機能であり、サービスとして提供される。これらのインタフェースは全て SOAP として提供されているため、Java、C++、C# や JavaScript など、SOAP に対応する全ての言語でクライアントを開発することができる。

1. 翻訳機能

- (ア) **テキスト翻訳**: テキスト文を入力とし、自動翻訳して、訳文テキストを出力する機能。
- (イ) **単語対応つき翻訳**: テキスト翻訳に加え、原文と訳文の単語対応情報も出力する機能。
- (ウ) **HTML 翻訳**: HTML を入力とし、自動翻訳して、訳文 HTML を出力する機能。
- (エ) **辞書引き**: 単語を入力とし、その訳語候補を出力する機能。
- (オ) **辞書名取得**: 自動翻訳で使用可能な辞書を取得する機能。取得された辞書名は翻訳時に指定できる。
- (カ) **対訳文の単語対応づけ機能**: 2 言語 (例えば英語と日本語) の対訳文を入力とし、対訳文中の対応する単語を自動的に対応づけて出力する機能。

2. 訳例検索機能

- (ア) **訳例検索**: 入力キーに基づき、指定された訳例データベース中の訳例を

串刺し検索し、結果をソートして出力する機能。

- (イ) **訳例登録**: 訳例を登録する機能。
- (ウ) **訳例削除**: 訳例を削除する機能。
- (エ) **訳例データベース名取得**: 使用可能な訳例データベース名を取得する機能。
- (オ) **訳例コメント付与**: 訳例データベースに格納されている各訳例ごとに、コメントを付与する機能。
- (カ) **訳語信頼度のための頻度計算**: 指定された訳例データベース中で、指定された原文単語と訳文単語が同時に出現する訳例数を計算する機能。
- (キ) **KWIC 検索**: 指定された訳例データベース中で、指定された原文単語と訳文単語が同時に出現する訳例を出力する機能。

Cliché の翻訳ワークベンチは、これらのサービスを組み合わせることで全ての機能を実装している。これらのサービスを使い、必要な機能に絞って翻訳支援クライアントを開発することが可能である。

例えばオートコンプリート機能 (図 2) を実装したいとしよう。上記機能の中で最低限必要なのは、1-(エ)のみである。例えば「アルカリ溶液を攪拌する」という日本語を英語に翻訳する場面を考える。翻訳者が入力するであろう訳語は、「アルカリ溶液」「攪拌」「アルカリ」などの訳語 (英語) である。よって、これらの単語に対して、辞書引き機能を使って訳語候補一覧を作ればよい。このとき、自前の用語集などがあればそれを検索するシステムを別個につくれば、ユーザ定義の辞書の訳語をもオートコンプリート候補に入れることができる。また 2-(カ) と組み合わせると、原文語句と訳文語句のそれぞれの組み合わせごと

に訳例データベース中の同時出現頻度を取得し、訳語信頼度を付与することも可能である。表 1 は、このようにして取得されたオートコンプリート機能で使う情報の例である。この表があれば、例えば翻訳者が「a」と入力したとき、「alkali solution」「alkaline solution」というオートコンプリート候補を表示できる。オートコンプリート機能に関する詳細は[12]を参照されたい。

また、訳例検索において単語対応づけをしたい場合 (図 3) には、1-(カ) を呼び、その結果を表示するところだけを作りこめばよい。GUI は自由にカスタマイズできるため、簡単に翻訳支援アプリケーションを開発できる。

現在 Web 上で公開されている翻訳サービスは、ただ翻訳結果を返すだけで、指定の辞書を使った翻訳はできない。このため、用途は一般向けに限られている。上記 1-(イ) の機能を使い、指定した辞書で翻訳する枠組みを構築することが可能である[14]。ローカルまたは、ネットワーク上に辞書 (辞書テーブルと呼ぶ) を格納しておくことが前提である。辞書テーブルには、原文単語と訳文単語を 1 レコードとして格納しておく。原文を、単語対応つき翻訳サービスを使って翻訳すると、出力として単語対応がついた形で翻訳結果が返ってくる。この結果をもとに、原文の単語で辞書テーブルを検索し、マッチした場合、対応づけられた訳文の単語を得られた訳語に置換する。このとき、英語の名詞や動詞、日本語の動詞などは活用処理を行う必要があるが、日本語名詞は活用しないため、活用処理は不要である。辞書エントリを名詞に限れば、Cliché の翻訳サービスを使って、辞書によるカスタマイズは簡単にできる。辞書テーブルは UTX-Simple 形式[16]で作成すれば使い回しがきくだろう。

表 1 オートコンプリート機能で使う情報例

原文語句	訳文語句	種別	辞書名	訳語信頼度
アルカリ溶液	alkali solution	自動翻訳	辞書 1	高
アルカリ溶液	alkaline solution	自動翻訳	辞書 2	高
攪拌	stir	自動翻訳	辞書 1	高
攪拌	mix	自動翻訳	辞書 1	低
アルカリ	alkali	ユーザ定義	---	高
アルカリ溶液	alkali solution	ユーザ定義	---	高

the prior art systems without very good reasons mentioned
In any case the video camera system 60 of this invention is a lower cost alternative to the prior art systems for the reasons mentioned and even with the losses, if a Gen III tube used in the system, operation even on dark nights is more than fully satisfactory.
損失があったとしても、本発明のビデオカメラシステム60は、上記の理由から、従来技術のシステムとは異なり低コストであり、夜間であっても、GenIII型管が使用されていれば、その動作は、より十分にされる。

図 3 単語対応つき訳例検索の例

5. 今後の課題

Cliché サーバーは数十人～数百人のユーザ数を想定して開発されているが、Web サービスとしてインターネット上で公開すれば、使用するであろうユーザ数は計り知れない。Cliché と同等の機能を使うには、ユーザごとに辞書や訳例を管理する必要があり、数千～数十万規模のユーザ管理に対応する必要がある。

辞書・訳例は翻訳会社や翻訳者にとっての資産であり、セキュリティを確保しなければならない。

まずは、著作権に支障がない範囲で翻訳支援サービスを提供することが考えられる。4節で述べた方法では、ローカルに辞書を持って翻訳サービスの結果をカスタマイズできるため、辞書をサーバーにアップロードする必要がない。しかし、翻訳方向によっては活用処理が必要となるため、この方法も万能ではない。

今後は、セキュリティ面での対応を進めなければならない。

6. まとめ

Cliché の翻訳サーバー、訳例検索サーバーが提供する SOAP サービスにより、様々な形の翻訳支援ツールが開発できることを実証した。また、SOAP サービスをクラウドとして提供できれば、サーバー資源を多くのクライアントで共有できるため、初期コストを抑えて翻訳支援技術を提供できる。今後、さらなる翻訳の効率化のために、様々なサービスを研究開発していく他、大規模なユーザ管理にも対応し、翻訳支援サービスの提供を検討していきたい。

参考文献

- [1]excite 翻訳.
<http://www.excite.co.jp/world/>
- [2]石田亨. 言語グリッドと異文化コラボレーション, 電子情報通信学会誌, Vol. 91, No. 6, pp. 515-517, 2008.
- [3]影浦峽・佐藤理史・竹内孔一・宇津呂武仁・辻慶太・小山照夫. 翻訳者支援のための言語レファレンス・ツール高度化方針. 言語処理学会第 12 回年次大会論文集, 2006.
- [4]田中英輝, 木下明德, 小早川健, 熊野正, 加藤直人. 構文情報を使ったリード文の具体化による要約支援. 第 15 回言語処理学会年次大会発表論文集, 2009.
- [5]Seiji Okura, Tatsuo Yamashita, Masaru Fuji, Guowei Xu, Akira Ushioda. Integrating MT and TM: Cliché. 20th International Conference on Computational Linguistics (COLING), Exhibition.
- [6]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.
- [7]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.
- [8]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.
- [9]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.
- [10]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.
- [11]大倉清司, 富士秀, 徐国偉, 長瀬友樹, 潮田明. 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. 言語処理学会第 13 回年次大会予稿集, 2006.
- [12]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.
- [13]富士秀, 大倉清司, 長瀬友樹. 翻訳支援における訳語信頼度. 言語処理学会第 13 回年次大会予稿集, 2007.
- [14]大倉清司. 訳文を修正しながら辞書更新ができる翻訳支援インターフェース. 言語処理学会第 14 回年次大会予稿集, 2008.
- [15]大倉清司, 長瀬友樹, 潮田明. フィードバック対応型翻訳メモリシステムの試作. 言語処理学会第 15 回年次大会予稿集, 2009.
- [16]大倉清司, 山本ゆうじ, 村田稔樹, 内元清貴, 加藤マイケル孝仁, 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.