

「翻訳パレット」を用いた翻訳支援の提案

熊野 正 西脇 正通 田中 英輝

E-mail: {kumano.t-eq, nishiwaki.m-hk, tanaka.h-ja}@nhk.or.jp

NHK 放送技術研究所

〒157-8510 東京都世田谷区砧1-10-11

概要

我々は従来より、ユーザの要求に応じて過去のニュース原稿の翻訳結果を検索・提示することで翻訳作業に必要な調査を支援する「翻訳用例提示システム」をNHKのニュース翻訳現場に提供してきた。本稿では、より高度な支援を目指した「自発的情報提示」による翻訳支援を提案し、そのプラットフォームとして「翻訳パレット」を考案した。翻訳パレットは、ユーザが翻訳元文書を与えられた際に、翻訳文書の執筆に先立って翻訳元文書中の各種表現の対訳（「翻訳部品」）を準備するツールであり、ユーザはその準備後に翻訳部品を用いて翻訳先文書を執筆する。コンピュータは、自発的な翻訳部品の提案やユーザとの協調によって、ユーザの翻訳部品準備作業を支援する。本稿では、翻訳パレットを説明するために実装したコンセプト・プレビューを紹介する。

1. はじめに

NHKでは日々、国内向け二カ国語放送や国際放送のために、英語など21外国語のニュース原稿を翻訳作業によって作成している。

放送現場におけるニュース翻訳作業は、一般に時間的制約が厳しい。ところが、ニュース原稿には人名や組織名といった翻訳の正確さに注意を払わなければならないような表現が含まれていることが多く、十分に調査を行う必要がある。また、ニュース原稿の翻訳は背景知識やこれまでの経緯などを理解しないと適切に行えないことが多いが、ニュース翻訳現場の翻訳者は日々さまざまな

ジャンルの原稿を翻訳するため、その背景知識を常に十分に持っているとは限らず、調査を行わなければならないことが多い。このような各種の調査に時間と手間がかかることが、従来より問題であった。

我々は、翻訳者がこのような調査を行うことを支援するために、「翻訳用例提示システム」を開発してNHKのニュース英訳現場に提供してきた。またその好評を受けて、英語以外の外国語への翻訳現場に対しても導入を進めているところである[1,3]。翻訳用例提示システムは、過去のニュース原稿の翻訳結果を対訳原稿対として蓄積し、ユーザが単語や表現・文などの検索を要求すると、それを含む原稿とその対訳原稿の全体を提示する。例えば人名や組織名を入力すると、過去にその固有表現が含まれる原稿がどのように翻訳されたかが提示されるので、その表現の承認された翻訳を知ることができる。また本システムは対訳原稿対全体を提示するので、ニュース特有の言い回しの訳され方を調査したり、ある話題に関する一連の原稿とその翻訳を眺めることで過去の事実関係を調査したり新しい翻訳の参考にする、といった用途にも用いることができる。

翻訳用例提示システムや、その他電子辞書やWebなど、各種調査がコンピュータ検索によって手軽に行えるようになった結果、翻訳者の多くは、翻訳作業に必要な調査を行う時間や手間が軽減し、結果として翻訳品質の向上にもつながりうる、と報告している[3]。しかし、単に各種検索システムが導入されている現状では、翻訳元原稿中のどの部分が調査を要し、どのような情報源を調査し、調査結果のどの部分を翻訳先原稿に取り入れるかについては、未だユーザにゆだねられており、コンピュータによる更なる支援を行う余地は大きいと考える。そこで本稿では、より積極的に

Collaborative translation through ‘Translation Palette’.
Tadashi Kumano, Masamichi Nishiwaki and Hideki Tanaka.
NHK Science and Technical Research Laboratories.
1-10-11 Kinuta, Setagaya-ku, Tokyo, 157-8510 JAPAN.

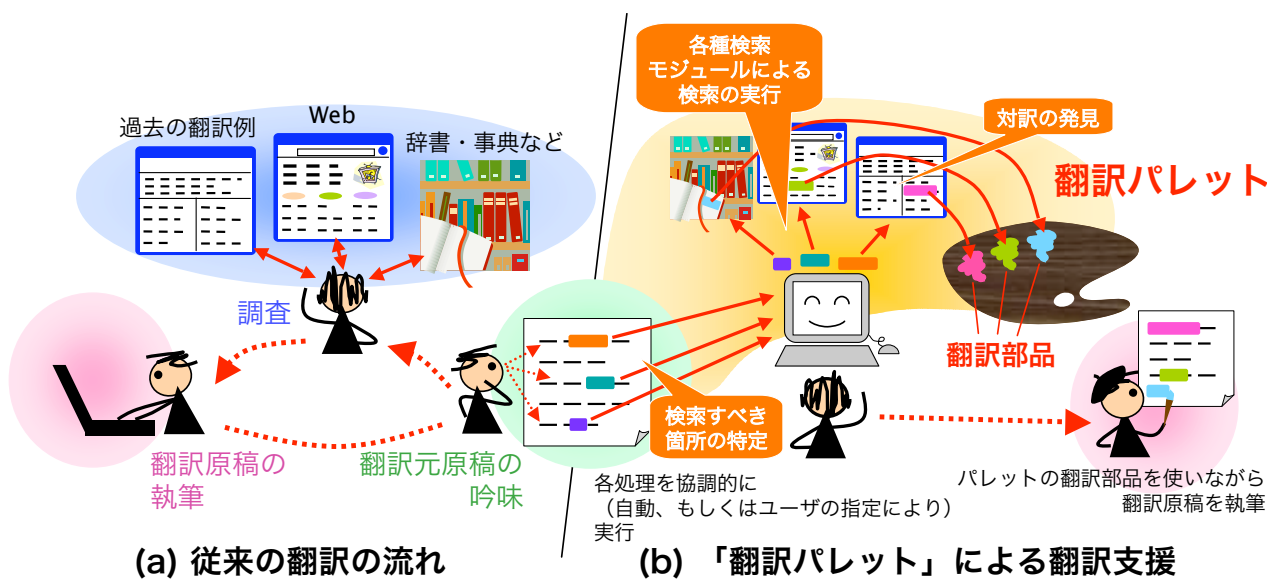


図1 「翻訳パレット」による翻訳支援のイメージ

翻訳に必要な調査を支援することを目指し、「自発的情報提示」による翻訳支援を提案する。これは、コンピュータがユーザに先立って翻訳元原稿を自動処理して調査すべき表現を推定し、それらの調査結果から翻訳先原稿に利用可能な表現の候補である「翻訳部品」を生成してユーザに提案するものである。我々は、自発的情報提示による翻訳支援を行うプラットフォームとして、コンピュータとユーザが協調しながら翻訳部品を準備する「翻訳パレット」と呼ばれる支援ツールを考案した。本稿では、翻訳パレットを用いた翻訳作業の流れについて述べ、そのコンセプト・プレビューの実装を紹介して将来の展望を述べる。

2. 「翻訳パレット」による翻訳支援

以後の節では、日英翻訳の支援を一例に、「翻訳パレット」による支援を説明する¹。

前節で述べたように、翻訳用例提示システムを用いた従来の翻訳支援においては、ユーザは翻訳元の日本語原稿が受け取ると、その中から不明な固有表現や難しい言い回し、その他調査したい箇所を見つけ出し、翻訳用例提示システムや電子辞書、Web 検索ページなどの情報検索ツールに入力して検索結果を得て、その検索結果を参考にして

翻訳を進める。あるいはより直接的に、検索結果中の英語表現をコピー&ペーストで翻訳先の英語原稿に導入する (図 1(a))。

これに対して、本稿で提案する「翻訳パレット」システムにおいては、従来完全にユーザの分担であった、

1. 日本語原稿中から調査したい表現を発見
2. 1. をさまざまな情報資源 (電子辞書、過去の翻訳用例、Web、etc.) から検索
3. 2. の結果中から有用な英語表現を抽出

という処理の各々を可能な限りシステム側で先立って実行する。システムは、最終的に抽出された英語表現を「翻訳パレット」と呼ばれる領域に配置し、翻訳先の英語原稿の一部として利用可能な「翻訳部品」としてユーザに提案する。また、ユーザは、自ら日本語原稿中から調査したい表現を指定したり、検索結果中から翻訳部品として抽出したい表現を指定する、といった、システムの動作を補完する操作を行うこともできる。このように、システムとユーザの協調による準備作業によって必要な翻訳部品を準備した後に、ユーザはあたかもパレットに並べられた絵の具を使って絵を描くように、翻訳パレットに並べられた翻訳部品を、各自のエディタにて編集集中の英語原稿中に適宜導入していくことで翻訳作業を進める (図 1(b))。

翻訳パレットに準備する翻訳部品には、人名・

¹ 当然ながら、「翻訳パレット」は任意の言語対の翻訳支援に対しても適用可能である。

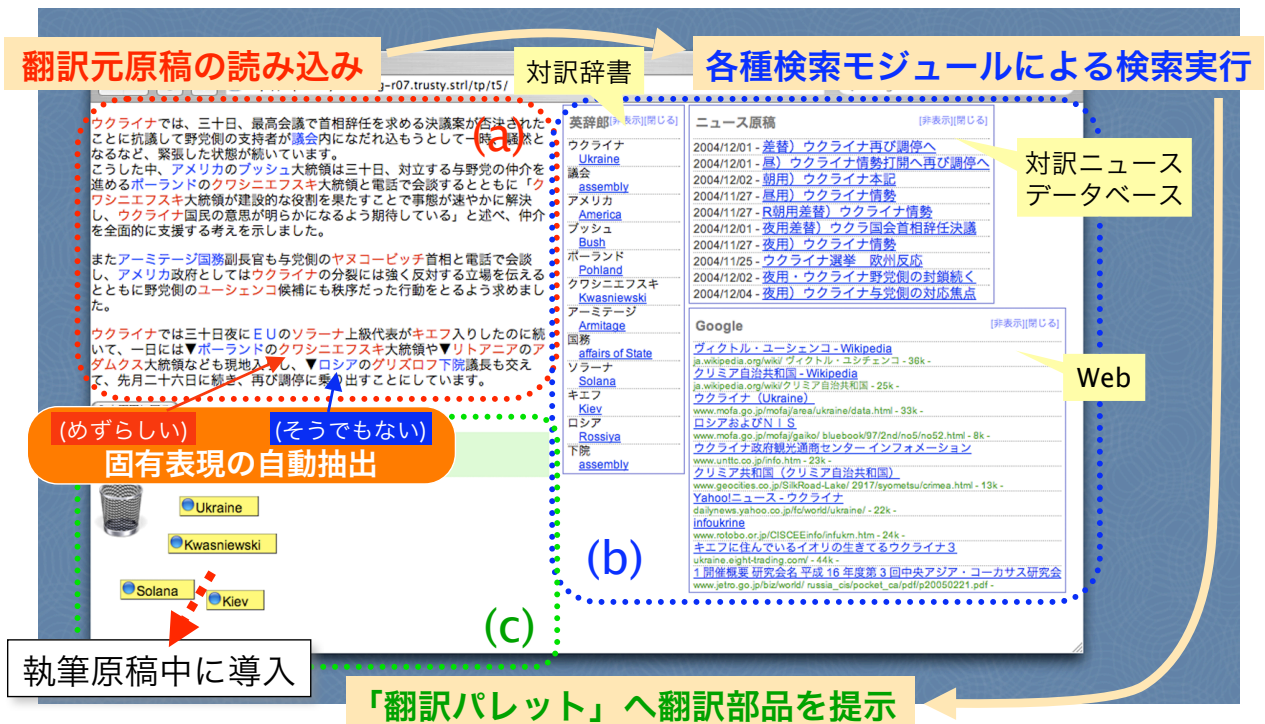


図2 コンセプト・プレビュー

地名・組織名などの固有表現のように非常に短いものから、ニュース特有の言い回しのような少し長い表現、さらには、1文のようなより長い表現など、任意の表現を考えることができる。システムが固有表現などの翻訳を自動的に提案できれば、ユーザのニュース翻訳への熟練度を問わず、翻訳作業の速度向上に寄与することが期待できる。一方、より長い表現の翻訳を適切に提案するのは容易ではないが、実現すれば、特にあまりニュース翻訳への熟練度の高くないユーザに対して、よりよい翻訳を作成する支援になる可能性がある。

3. コンセプト・プレビュー

翻訳パレットの使い方を説明する目的で、最低限の機能を持つコンセプト・プレビューを実装した。図2にその外観を示す。以下ではプレビューの各部の機能と実装の現状について説明する。

(a) 日本語(翻訳元)原稿表示部

読み込んだ翻訳元の日本語原稿を表示する。また、日本語原稿中の調査対象表現を自動抽出する。現状では、CRF++²を用いて実装した日本語固有

表現抽出器を用いて抽出した固有表現(人名/組織名/地名/固有物名)を、調査すべき表現としている。この時、後段の処理のために、これらの固有表現を、めずらしいものとそうでないものとに分類しておく。分類の基準には、日本語ニュース原稿データベースを各固有表現で検索して求めた逆文書頻度(idf)を用いる。すなわち、日本語ニュース原稿 N 記事中、 df 記事にその固有表現が含まれる時、 $idf = \log(N/df)$ となる。

さらに、表示されている日本語原稿の任意の表現をマウスで選択することで、ユーザが調査対象表現を追加指定することができる。

(b) 各種検索モジュール部

(a)で用意した調査対象表現の各々について、各種情報資源に対して横断的に検索を実行し[2]、検索結果を表示する。現在利用可能な検索モジュールは、以下の情報資源にアクセスするものである。

日英対訳辞書(英辞郎):

日英対訳電子辞書(英辞郎)に対して各調査対象表現の検索を行い、一致するエントリがあれば英訳を表示する。

² 工藤拓氏が公開している、条件付き確率場(Conditional Random Field)に基づく汎用 tagger。

<http://www.chasen.org/~taku/software/CRF++/>

日英対訳ニュース原稿データベース:

翻訳用例提示システムで用いている日英対訳ニュース原稿データベースに対して、(a)で自動抽出した検索対象固有表現のうちめずらしいもののみをキーワードとして or 検索を行い、検索結果の原稿の上位一覧を表示する。検索結果の順位づけは、キーワードが多く、かつ作成日時の新しい原稿が上位となる。一覧の各項目は日本語原稿の題名であり、クリックすることで日本語原稿とその翻訳である英語原稿の対全体を見ることができる。ユーザが調査対象表現を追加指定した時には、自動抽出した表現とは別個にその表現単独で検索を行い、検索結果を表示する。

Web (Google) :

Web 検索エンジン (Google) に対して、(a)で自動抽出した検索対象固有表現をキーワードとして and 検索を行い、検索結果の上位一覧を表示する。一覧の各項目をクリックすることで、その Web ページを開くことができる。

(c) 翻訳パレット部

(b)の各検索モジュールには、自動的、もしくはユーザの指定に応じて、検索結果中の英語表現の一部を翻訳部品として抽出する機能を持たせることができる。翻訳パレット部は、それら抽出された翻訳部品をユーザに提示する部分である。

現時点では、(b)の検索モジュールのうち、自動的に翻訳部品の抽出を行うのは日英対訳辞書検索モジュールのみであり、めずらしい固有表現の英訳を翻訳部品として出力する。また、ニュース原稿データベース検索モジュールはユーザ指定による翻訳部品抽出機能を持ち、検索結果の日英対訳原稿対の表示中にユーザが英語原稿中の任意の表現を選択すると、それを翻訳部品として出力する。

ユーザは、パレットに配置された翻訳部品を適宜並べ替えたり、不要なものを廃棄したり、あるいは (b)とは関わりなく新規に任意の表現を翻訳部品として配置することができる。また、(b)の検索モジュールから出力された翻訳部品をクリックすることで、その表現がもともと含まれていた検索結果を参照することができる。

4. まとめ・今後の展開

本稿では、コンピュータとユーザが協調して翻訳部品を整備する「翻訳パレット」と呼ぶ考え方を提案し、翻訳パレットの説明として実装したコンセプト・プレビューを紹介した。

本プレビューは、翻訳パレットの挙動の説明に必要な最低限の機能しか持っていない。実際にこのシステムを有用なものにするためには、前節の各部分がより実用的な機能を持つ必要がある。一例として、以下のような機能が必要であると考えている。

- 翻訳元の日本語原稿から、例えばニュース特有の言い回しのような、固有表現以外の有用な表現を発見する機能
- 対訳ニュース原稿データベースの検索結果から翻訳部品を自動的に抽出する機能
- 日本語キーワードを用いて関連する英語 Web ページを検索する機能

対訳文書間の頑強な表現対応推定や言語横断 Web 検索など、各種自然言語処理技術の研究を進めることで、実用的なシステムの構築を目指したい。

また、今回実装したプレビューを翻訳現場の翻訳者に試用してもらい、翻訳パレットが放送ニュース翻訳の支援に有効であることを検証したい。

謝辞 コンセプト・プレビューの開発にあたっては、(株)漢字情報サービスの小鶴友樹氏より示唆をいただき、また実装にも協力いただきました。感謝いたします。

参考文献

- [1] Isao Goto, Naoto Kato, Noriyoshi Uratani, Terumasa Ehara, Tadashi Kumano, and Hideki Tanaka. A multi-language translation example browser. In *Proceedings of the MT Summit IX*, pp. 463–466, 2003.
- [2] 熊野正, 後藤功雄, 江原暉将. Translators' Workbench: 対訳エディタを中心とした統合翻訳支援環境. 言語処理学会第 6 回年次大会発表論文集, pp. 143–146, 2000.
- [3] 熊野正, 後藤功雄, 田中英輝, 浦谷則好, 江原暉将. 翻訳用例提示システムの設計・開発・運用. 電子情報通信学会論文誌, Vol. J84-D-II, No. 6, pp. 1175–1184, 2001.