

Translators' Workbench:

対訳エディタを中心とした統合翻訳支援環境

熊野 正[†]

後藤 功雄

江原 暉将

NHK 放送技術研究所

Abstract

翻訳者が効率的に翻訳作業を行うための、統合翻訳支援環境 — *Translators' Workbench* — を提案する。これは、ユーザが作業を行う「対訳エディタ」が各種情報資源との仲介能力を持つことで、多様な情報資源、例えば

- 過去の翻訳例
- グループで共有される翻訳メモ
- WWW
- 機械翻訳

などの横断的な利用や、ユーザと環境との協調を実現するものである。

本論文では、対訳エディタのプロトタイプと、我々が従来より研究・開発してきた過去の翻訳例の検索機能などへのアクセスを実現する対訳エディタのプラグインについて概説し、将来の可能性について論じる。

1 はじめに — 統合翻訳支援環境の必要性 —

我々は、NHKの国際放送および2ヶ国語放送のために日本語のニュース原稿を英語に翻訳する業務をより効率的に行えるための、翻訳支援技術の研究・実用化を進めている。これは、単に翻訳作業にかかる時間や労力の削減を目指すだけでなく、その結果として生ずる時間的なゆとりを各翻訳作業者が翻訳品質の向上やより創造的な仕事に振り向けられるよう、発想支援や翻訳者間のコミュニケーション支援をも目指すものである。

“Translators' Workbench: an integrated translation-aiding environment using Bi-text Editor as a mediator between users and computer resources”

KUMANO Tadashi[†],

GOTO Isao, EHARA Terumasa

NHK Science and Technical Research Laboratories
1-10-11, Kinuta, Setagaya-ku, Tokyo, JAPAN
157-8510

[†] 現在は (株) ATR 音声言語通信研究所に所属
currently belong to ATR Spoken Language Translation
Research Laboratories, Kyoto, JAPAN
(E-mail: tkumano@slt.atr.co.jp)

これまでに、以下の機能を Web ブラウザ経由で利用できる「用例提示型翻訳支援システム」を開発した。

日英記事検索機能:

過去に実際に翻訳され放送に使用された英語のニュース原稿を、元となった日本語の原稿と対にして検索・閲覧することができる [2, 3]。

用語集検索機能:

翻訳者間で共有管理される備忘録的なメモ書きを、検索・閲覧や追加・編集できる。

これは NHK 国際放送局英語センターにて実用システムとして運用を開始しており、ユーザからは作業時間の短縮など業務効率の改善につながっていると声を得ている。

しかし、現状のシステムには以下のような限界がある。

限界 1:

これらの検索機能は、単純に与えられた表現に類似した表現を含む記事や用語集を検索するだけであり、その表現を与えたユーザの「暗黙の意図」を考慮することができない。

例えばユーザが珍しい固有名詞を入力したとき、ユーザはまさにその固有名詞の訳を知りたいと思っている可能性が高いが、非常に一般的な単語や表現を入力したときには、その表現が含まれている文脈と似通った部分を持つ記事を参照したい可能性が高い。例えば、「(カンボジアの) ラナリット下院議長」という入力を行ったユーザは、おそらく「ラナリット」の綴りやカンボジアの下院議長の公式な英訳を知りたいと思っているであろうが、「小淵首相」という入力にそれと同じ意図があるとはまず考えられない。一般にこのような検索を適切に指示するようなクエリをユーザが作成するには、検索エンジンの仕組みについて深い知識を必要とするため、これを支援する仕組みが必要である。

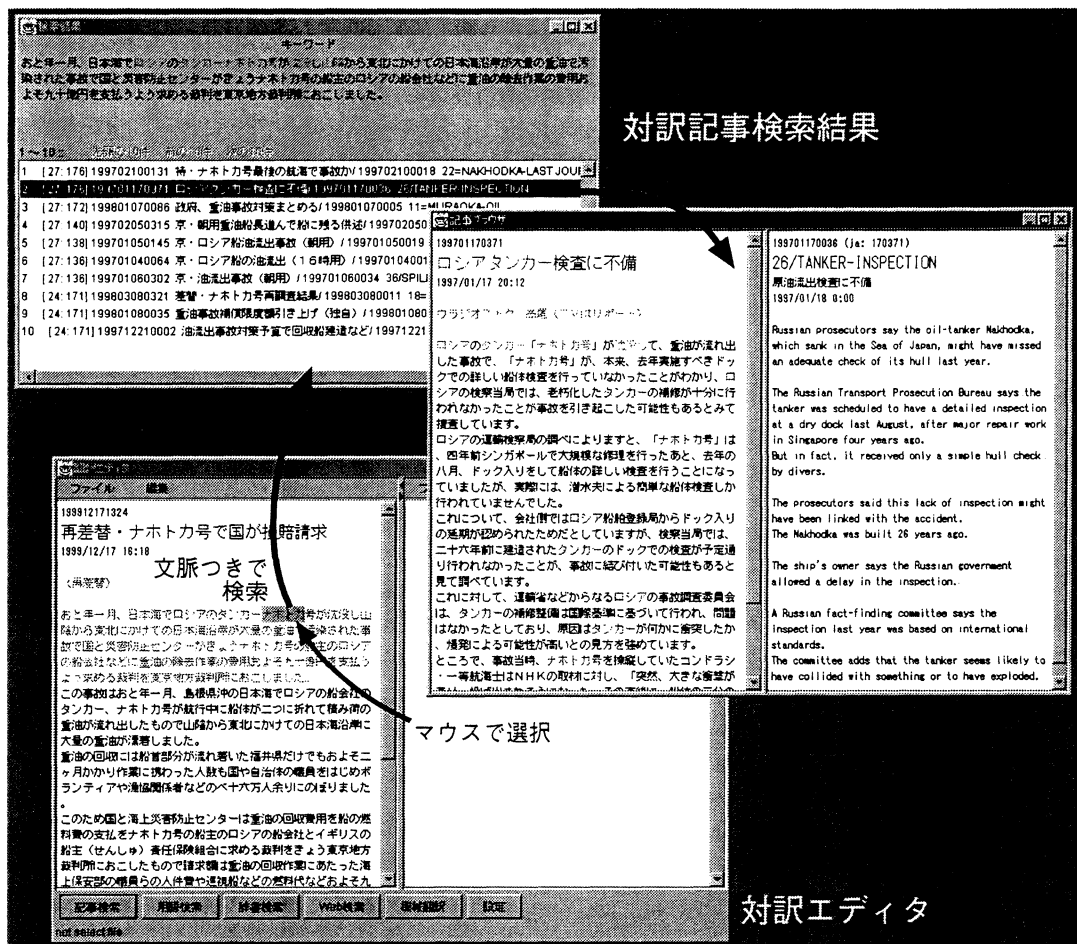


図 1: 対訳エディタと、日英記事検索へのアクセスウィンドウ

限界 2:

他の情報資源にアクセスするには別の手段が必要である。また現在提供されている機能についても、ユーザ作業との協調は十分でなく、より高度な支援を行うように拡張していくことが困難である。2つの資源を Web ブラウザから検索できるようにしたことで、他のインターネット上の資源と同じアプリケーションでアクセスできるようにはなったが、他の資源を検索するためには、あいかわらず他の Web ページにアクセスし、違った tips を駆使して検索を行わなければならない。結果として、ある内容についていろいろな資源に「横断的に」アクセスし内容を比較することは、非常に労力を要する。

これらの問題を解決していくためには、現状のようにつぎはぎ状に新しい情報資源を提供していくの

ではなく、統一的な「統合翻訳支援環境」の構築を目指していく必要がある。

2 Translators' Workbench

我々は、前節で述べたような統合的な翻訳支援環境として、“Translators' Workbench”の開発を進めている。これは、「対訳エディタ」を翻訳者が翻訳に必要な作業を行うための「作業台 (= Workbench)」として中心に据え、ユーザと各種資源との仲介を行わせることによって、統合的な支援を行うものである。基本的なアイディアは古くからあるものである [1] が、近年のネットワーク環境の充実とコンピュータの処理能力向上に伴い、ユーザが必要とする各種の情報資源に対して電子的にアクセスできる環境が整ってきたため、本当の意味での実用化が可能になってきた。

現在、対訳エディタには「過去の翻訳例」と「共有メモ」へのアクセス機能を実装している。これらの機能は対訳エディタに対するプラグインモジュールとして実現されており、他の資源へのアクセスを実現するモジュールを開発し組み込んでいくことにより機能を追加していくことができる。また、モジュールがユーザのアクションによらずに自ら翻訳作業の内容を取得し外部とやりとりをすることができる枠組を備えており、これを利用して、ユーザからの要求によって駆動される受動的な機能にとどまらない、能動的な機能の実現を可能にしている。

対訳エディタをはじめとするユーザ側のクライアント群はすべて Java を用いて実装されており、ユーザの使用するコンピュータの種類を問わず利用できる¹。

以下では、クライアントの各コンポーネントの実現している機能について概説する。各々のコンポーネントがアクセスする各種情報資源の詳細については、文献 [3] など、および今後の別稿を参照されたい。

2.1 対訳エディタ

対訳エディタは、翻訳元の文書を読み込んで表示する部分と、ユーザが実際に文書を記述していく部分、それから各種情報へのアクセスを起動するボタン群からなる (図 1)。

ユーザは翻訳元文書や作業中の翻訳先文書の任意の部分をマウスで選択し、各種資源への問い合わせのキーとする。このとき、対訳エディタに組み込まれた各種資源へのアクセスモジュールは、ユーザが選択した領域とともにその前後の文字列を取得することができる。各モジュールはこれを用いることによって、ユーザに選ばれた文字列がどのような文脈の中に存在しているのかを考慮して、ユーザの「暗黙の意図」にできるだけ合致した検索結果がより高い優先順位で得られるような検索クエリを、各資源の検索の能力や特性にあわせて作成し、検索を要求することができる。これは前節の 限界 1 を克服しようとするものである。

翻訳元文書として、ブレインテキストの他、我々が開発してきた日英記事検索システム [3] のデータベースの文書構造記述に採用している XML タグによって文書構造 (段落や文など) がタグづけされた文書を読み込み、表示することができる。

¹ Java 2 を用いて実装しているため、現時点では Java 2 VM が提供されていない Mac OS 上では動作しない。現時点で動作を確認している環境は、Solaris 7 と Windows 98、Windows NT 4.0 である。

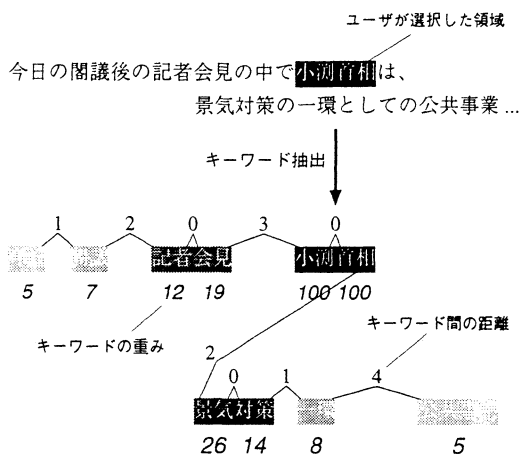


図 2: ユーザに選択された領域と、その周囲の「文脈」から取り出されたキーワード列

2.2 対訳エディタから各種資源へのアクセス

対訳エディタから各モジュールが取得した前後の文脈付きの選択文字列は、キーワード抽出サーバに送られ、各語間の距離付きの内容語キーワードの AND 列に分解される (図 2)。この距離情報は、対訳記事や用語集の検索においては、類似検索に直接的に利用される。また、その他のモジュールにおいても、文脈情報の取捨選択に任意に利用できる。

2.2.1 対訳記事検索モジュール

前述の「用例提示型翻訳支援システム」の日英記事検索機能に対する接続を実現する。これは、入力された任意の表現を形態素解析して抽出した、各語間の距離付き内容語キーワードの AND 列をクエリとし、キーワードの一致個数とキーワード間の距離の変位を用いて類似表現検索を行うものである [4]。検索結果は「用例提示型翻訳支援システム」と同様に、対訳記事対全体を左右に並べて提示し、その表現がどのような文脈に出現して、結果としてどのような文脈中に訳出されているかを眺めることを可能にしている (図 1)。

今回対訳エディタからの検索に使用するにあたっては、

1. ユーザが選択した領域中のキーワードをできるだけ多く含み、
2. かつその周辺文脈中のキーワードの出現が似ている

表現を含む記事ほど高い提示順位になるような検索が行える必要がある。このため、これまでの実装では各キーワードの存在の重みは等しいものとしてき

たが、ユーザが実際に選んだ領域中のキーワードと周辺文脈中のキーワードとで重みに差をつけることができるよう、検索エンジンの拡張を行った。

実際の重みづけは、以下の戦略で行う（図 2）。

1. ユーザ選択領域中の全てのキーワードに等しい重みを付与
2. 周辺文脈中のキーワードについては
 - (a) ユーザ選択領域の両端からの距離に反比例
 - (b) 周辺文脈中のキーワードの重みの総和は、ユーザ選択領域中のキーワード 1 つの重みを越えない範囲での最大値

定量的な比較は行っていないが、拡張前の検索エンジンに対してユーザが選択した領域に含まれるキーワードのみを与えて検索を行った場合に比べて、拡張後の検索結果は、より有用な記事が上位に提示される傾向があった。

2.2.2 用語集検索モジュール

前述の「用例提示型翻訳支援システム」の用語集検索機能に対する接続を実現する。検索エンジンは前項の対訳記事検索モジュールと共用しており、同等の検索能力を実現している。

3 今後の展開 — より横断的に、能動的に —

現在、対訳エディタから利用できる情報資源は限られており、支援環境としては不十分なものであるが、今後、ユーザのリクエストに応じて、より多種の資源に対してアクセス可能なモジュールを作成して拡張していく予定である。例えば以下のようなモジュールの開発を計画している。

- WWW 対訳ページの検索
- 電子辞書・辞典

また、単なる静的な情報ではないようなものに対しても、まったく同様の操作手順でアクセスが可能であれば、静的な情報と同列に比較検討が可能になる。例えば以下のような機能が考えられる。

- 機械翻訳
ユーザが選択した領域を含む文を機械翻訳で翻訳し、ユーザの選択した領域に対応する訳出部分を示す。
- ユーザ間のコミュニケーション機能
ユーザが選択した領域を含む文を掲示板に載せ、広く意見を募る。

アクセス可能な資源が増え、ユーザが必要とする情報を広くカバーできるようになると、横断的に資源

にアクセスできることのメリットが生きるようになると期待できる。

また、自発的に動作するモジュールの実現により、支援環境が能動的にユーザを支援することが可能になる。例えば、

- 新しい翻訳元原稿が読み込まれたときに、直ちに各文を記事検索モジュールに渡して検索を行い、似た記事があれば提示する。
- 翻訳作業が終わったときに、翻訳元の記事中に用語集に対訳として載っている表現があるのに翻訳先にその訳がないときには警告する。
- 新しい翻訳元原稿が読み込まれたときに、直ちに各文を機械翻訳モジュールに渡して翻訳を行い、翻訳機が翻訳のクオリティが高いと判断したした翻訳結果のみを提案する。

4 まとめ

本論文では、統合翻訳支援環境 “Translators’ Workbench” の機能について概説し、将来の可能性について論じた。

今後は、各種モジュールの開発を進め、より統合された環境を実現するとともに、実際に翻訳現場のユーザによるテストを行うことで、その有用性を評価していく予定である。

参考文献

- [1] Martin Kay. The Proper Place of Men and Machine in Language Translation. Working Paper CSL-80-11, Xerox Palo Alto Research Center, 1980. (reprint in *Machine Translation*, pp. 12:3–23, 1997).
- [2] 熊野正, 田中英輝, 松田伸洋, 浦谷則好, 江原暉将. 日英放送原稿翻訳者のための類似用例提示型翻訳支援システム. 第 55 回情報処理学会回全国大会, pp. 2:582–583, Sep. 1997.
- [3] Tadashi Kumano, Hideki Tanaka, Noriyoshi Uratani, and Terumasa Ehara. Translation examples browser: Japanese to English translation aid for news articles. In *Natural Language Processing and Industrial Applications*, pp. I:96–102, Moncton, New-Brunswick, Canada, Aug. 1998.
- [4] 田中英輝. 長い日本語表現の高速類似検索手法. 自然言語処理研究会, 121–10, pp. 69–74, Sep. 1997.