

3R-7 機械翻訳系と放送系とのインターフェイス

住吉 英樹¹ 斎藤 聡² 西城 秀雄² 相沢 輝昭¹

¹NHK放送技術研究所

²NHK技術局

1 はじめに

NHKでは、1989年8月から、衛星放送「ワールドニュース」の一部で、英語ニュースに対する日本語テロップ(字幕)の作成を、機械翻訳システムにより行っている[1]。

現在は、従来の放送系に機械翻訳系が単純に組み込まただけなので、例えば、機械翻訳結果の日本語訳文をテロップ送出装置に再入力しなければならないなど、作業手順に重複や無駄が多く、効率の悪いものになっている。

そこで、機械翻訳系と放送系とを統合し、必要な作業が単一の端末で柔軟に実行できるシステムを開発した。このシステムは、第1バージョンの開発・テストを終え、現在試用の段階に入っている。

2 機械翻訳によるテロップ作成の現在の流れ

衛星放送における、ニュースの聴取からテロップ送出までの一連の作業は、現在、3人の担当者によって次のように行われている。

トランスレーター

- ・英語ニュースの聴取
- ・テロップを想定した内容の要約
- ・英語原文の入力と前編集
- ・機械翻訳の実行
- ・日本語訳文の後編集

デスク(放送責任者)

- ・日本語訳文の校閲

テロッパー

- ・テロップ作成装置への日本語訳文の入力
- ・テロップの作成

作業環境としては、トランスレーターは機械翻訳ワークステーションを、デスクは文字通り単なる机を、テロッパーはテロップ送出装置の作業端末を使ってお

り、その間にコンピュータ・ネットワークなど存在しないので、3人の間の情報の授受は専らハードコピーによっている。これも問題の1つである。

3 統合システムの考え方

そこで、機械翻訳系とテロップ送出系を Ethernet TCP/IPで結合した統合システムを開発した。ネットワークには当面10台までの作業端末が接続でき、そのいずれからでも、次の全ての作業が並行して実行できるよう設計した。

- ・英文および日本語の入力と編集
- ・機械翻訳の実行
- ・放送用日本語の校閲
- ・テロップの作成

その他、次の点を考慮した。

- 1) テロップの作成・送出システムとしても汎用性を持たせ、オンライン化を図る。
- 2) テロップごとに1つのファイルを設け、作業端末から各種情報が順次添加されていく方式とする。
- 3) 機械翻訳系は、英語原文を入力されたファイルに日本語訳文を添加するフィルターとして働くよう組み込む。

4 システム構成

4.1 作業単位とファイル構成

衛星放送ワールドニュースの送出に当たっては、番組全体がコーナーと呼ばれる10分程度の作業単位に分けられ、それぞれのコーナーに翻訳者が付く。現在、英日機械翻訳の対象になっているのは、その内の1コーナーである。以下、このコーナーを項目と呼ぶ。

1項目には、通常、数10枚のテロップが作られる。テロップ1枚をイベントと呼び、イベントごとに1つのファイルを用意する。イベント・ファイルの内容は次のものから成る。

- ・イベント番号
- ・テロップの表示位置
- ・コメント(最大22バイト)

- ・日本語文データ（不定長）
- ・英語文データ（不定長）

イベント番号によって、1項目を構成する全イベント、およびその順序が把握できる。

4.2 ファイルの排他制御

イベント・ファイルは複数の作業端末からアクセスされ、作業内容も入力、校閲、テロップ作成等と多岐にわたる。これに伴って、各イベント・ファイルは、入力済み、校閲済み等、様々な中間状態を持つ。これらを適切に処理するため、以下のような排他制御プログラムを開発した。

イベント・ファイルはイベント番号をファイル名として持つが、さらに、中間状態に対応する、次の6つの拡張子を持たせる。

- ・ NH …… 入力中（翻訳対象ファイル）
- ・ N …… 入力済
- ・ KH …… 校閲中
- ・ K …… 校閲済
- ・ TH …… テロップ作成中
- ・ T …… テロップ作成済

作業端末からの作業要求は、基本的に、イベント・ファイルの拡張子の書き換え要求であると考える。

排他制御全体は、1つの制御プロセスEPにより管理され、その手順の概略は以下の通りである。

1) 作業端末Aからの作業要求は、要求ファイルRを作成することで行われる。要求ファイルの内容は、

- R1: 作業対象イベント・ファイル名、
- R2: 作業種別、
- R3: 作業端末名、など。

2) 制御プロセスEPは、常に要求ファイルの作成を監視しており、作成されると同時に、ファイルR1が使用中かどうかを調べ、使用中でなければ、R2の内容に応じてR1の拡張子を書き換える。それ以後、R1は使用中とする。さらに、要求ファイルRを削除すると共に、結果ファイルSを作成する。Sには、作業の実行が可能か否かが書き込まれる。

3) 作業端末Aは、自分の作成した要求ファイルRがある限り待ち状態となるが、Rが無くなると、結果ファイルSを読みに行く。作業可能ならば実行し、作業が終了すると、そのことをやはり要求ファイルの形で制御プロセスEPに知らせる。それを受けてEPはイベント・ファイルR1の使用を解く。

複数の作業端末から、同じイベント・ファイルに作業要求が出された時には、制御プロセスが一定の順序（端末番号順）で要求ファイルを読み出すので、排他制御は確実に行われる。

4.3 機械翻訳システムとのインターフェイス

作業端末からの機械翻訳の実行は、拡張子「.NH」を持つイベント・ファイル中の英文データに対してのみ可能である。4.2で述べた制御プロセスが、イベント・ファイルからの英文データの切り出し、翻訳結果のイベント・ファイルへの書き込み、必要な漢字コード変換等を含めて、翻訳システムを起動する。

ただし、作業端末から実行できるのは標準的な機械翻訳機能だけで、それ以外の機能、例えば、多数の翻訳候補の表示や、辞書への新語の登録等は、機械翻訳ワークステーションから行う。それらの実行結果の中には、作業端末からの操作により、イベント・ファイルとして取り込むことのできるものもある。

4.4 テロップ作成装置とのインターフェイス

テロップ作成装置は、もともとパソコンを用いたオフラインの装置である。この装置にEthernet対応ボードを付加し、本システムと結合した。拡張子「.K」を持つイベント・ファイルがテロップ作成装置で処理可能な文字列ファイルである。テロップ作成装置は、これに排他をかけ、「.TH」にした後、文字列を読み込み、従来のテロップ作成ソフトで処理して、テロップの画像ファイルを生成している。

この画像ファイルとテロップ送出順を管理するファイルもEthernet上で共有化されており、複数のテロップ作成が可能になっている。

5 試用状況と残された問題

現在は、ユーザーに慣れてもらうための試用の段階である。人間系の対応も含めて、問題点の洗い出しを行っている。

本システムは、機械翻訳からテロップ作成までの文字列の流れをオンライン化したのが、テロップ送出のタイミングや、VTRとの同期制御情報の取り扱いが今後の課題である。

参考文献

- [1] 相沢ほか「衛星放送ワールドニュースの英日機械翻訳」情報処理学会第40回全国大会2F-1(1989).