

改版文書の翻訳を支援する差分翻訳システム

5H-6

介弘達哉 下畑さより 松下久明

沖電気工業株式会社 研究開発本部 関西総合研究所

1. はじめに

現在、機械翻訳の実用化に関する様々な研究が行われているが、我々はその一つとして差分翻訳システムの開発に着手している。過去に翻訳した文書の一部を変更した文書を翻訳しようとする場合新しい文書の翻訳と同様に、初めから機械翻訳を実施し、その後翻訳チェックを行う方法は多くの部分で無駄な作業時間を費やすことになる。そこで、人手により新しい原文と古い元の原文との相違点を見つけ出し、その部分を翻訳して元の訳文を修正するといった方法をとるが、文書量が膨大になると相違点を見つける作業は容易ではない。

差分翻訳システムは、文書の旧版の翻訳結果(対訳)を利用して、文書の新版の翻訳を効率良く行うことのできるシステムである。旧版と新版の原文を比較し、同じところは旧版の訳文を利用し、違いがある文だけ翻訳を行うことによって、翻訳作業の軽減が可能となることから、改版が頻繁に行われるマニュアルなどの文書の翻訳業務などに適用される。

本稿では差分翻訳システムの構成を述べる。また、実用化のポイントとなる原文と訳文の対応付けの部分に関する速度評価を行う。

2. システム概要

差分翻訳システムの構成を図1に示す。本システムは大きく分けて以下に述べる3つのステップにより実現される。

2.1. 新旧原文対応

文書比較部は旧原文ファイルと新原文ファイルとの相違する部分を抽出して、差分情報を生成する。具体的には、それぞれの文書の対応する部分を取り出して、単位量のデータ(単語、文字など)ごとにその有無を比較し、その文を構成するすべてのデータ数に対する他方の文書にないデータ数の割合(以下、「差分率」とする)を計算する。

2.2. 原文訳文対応

対訳文書の文の対応付けとは、ある言語で記述された文書(原文)とそれを他の言語に翻訳した

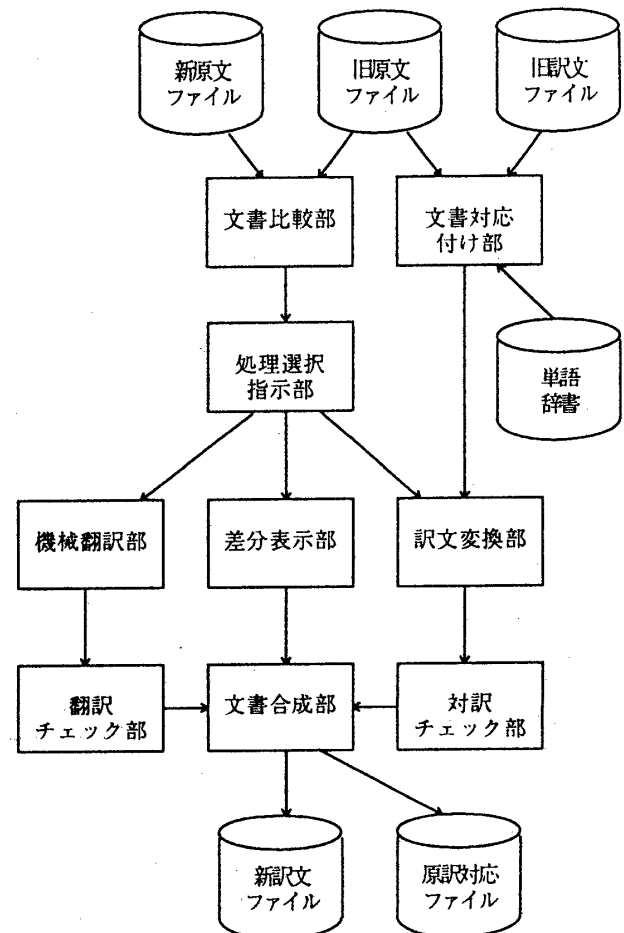


図1 差分翻訳システム

The difference translation system for revised documents
 Tatsuya Sukehiro, Sayori Shimohata and Hisaaki Matsushita
 Kansai Laboratory, Oki Electric Industry Co., Ltd.
 E-mail: sukehiro@kansai.oki.co.jp

文書（訳文）がある場合に、原文中の文が訳文中のどの文に対応しているかを、自動的に判別することである。対訳文の対応付けに関する研究は各所で行われており^{[1][2]}、95%程度の正解率が報告されている。本システムでは、文書対応付け部において、旧原文ファイルと旧訳文ファイルの対応付けを行う。まず、各文書を1文ごとに分割し形態素解析を行い、次に、単語辞書を用いて原文と訳文の類似度を計算する。文の分割、形態素解析は商用の機械翻訳システムの機能を利用し、単語辞書は自立語とその訳語情報のみを抽出し再構築を行ったものを使用する。

2.3. 新訳文生成

2.1 のステップで得られた差分情報を調べ、あらかじめユーザによって設定された差分率の閾値（変更の割合がどの程度大きければ機械翻訳を行うかの指標となる）と差分情報中の差分率の値により以下のいずれかの処理を行う。

(1) 差分率 = 0% の場合

原文と訳文の対応付けが正しく行われているかどうかをチェックするために旧原文と旧訳文の対応する部分を画面上に表示する。ユーザは必要があれば修正を行う。旧訳文ファイルから、原文に対応するデータを新訳文ファイルに追加する。

(2) $0% < \text{差分率} < \text{差分閾値}$ の場合

旧原文、新原文、旧訳文の対応する部分を画面上に表示し、ユーザに新訳文の入力を促す。入力した訳文を新訳文ファイルに追加する。

(3) $\text{差分閾値} \leq \text{差分率}$ の場合

新原文を機械翻訳部に送り機械翻訳を行う。翻訳結果等を表示し、ユーザがチェックを行い、必要であれば修正し新訳文ファイルに追加する。

これらは対話的に行ってもよいし、バッチ処理的に行って、後で一括修正をかける方法を探ってもよい。以上の処理により、改版文書の翻訳を効率よく行うことができる。また、同時に原文と訳

文の対応をとったファイルを生成しておけば、さらに改版が行われた場合に 2.2 のステップは省略可能となる。

3. 評価

差分翻訳システムは実際の翻訳業務に使用され、大量の文書を扱うことから、高速に動作する必要がある。我々は、実際の文書の特徴であるフォーマット情報を利用して対訳文の対応付けを行うことによって処理速度を落とさずに正解率を上げることに成功した^[3]。また、単語辞書を本システム専用に作成したことにより、原文と訳文の単語のマッチングを高速に行えるようになった。

対訳文の対応付けの速度を SPARCserver1000上 で測定したところ、1時間でA4の文書600ページ程度の対応付けを行うことができるという結果が得られた。

4. まとめ

本稿では、差分翻訳システムの概要を述べ、本システムが改版文書の翻訳作業の負荷の軽減に役立つことを示唆した。また、対訳文書の対応付けに関する速度評価を行い、実用に耐える速度で動作することを示した。

今後の課題として、新旧原文の対応付けを正確に行う手法の研究があげられる。今後は、システムを完成させ、実際の翻訳業務にどの程度有効であるかを評価していく予定である。

参考文献

- [1] 宇津呂他：対訳辞書を用いた対訳文対応および未知訳語の推定、「自然言語処理における実動」シンポジウム論文集, 信学会, 1993.
- [2] 井ノ上他：対訳テキストを用いた日英対訳辞書の自動生成, 信学技法, NLC93-39, 1993.
- [3] 介弘他：差分翻訳システムにおける対訳文書の文の対応付け, 信学会総合大会, 1995.