

ネットワーク接続される翻訳システムの辞書環境

6E-2

長谷部浩一 伊藤悦雄 武田公人 天野真家
(株)東芝 総合研究所

1 はじめに

我々は、ワークステーション(ASシリーズ)上で動作する機械翻訳システム AS-TRANSAC (TRANslation ACcelerator)を開発した。また、現在、ホスト計算機と、パソコンをネットワークで接続する機能分散型の機械翻訳システムの開発を進めている。

これらの機械翻訳システムは、複数の対訳辞書(標準辞書・専門用語辞書・ユーザ登録辞書)と学習辞書を用いて翻訳を行う。これらの辞書環境の整備/利用方法が、翻訳効率に大きく影響する。

本論文では、ネットワークを介して複数ユーザ間で共有される翻訳辞書の利用形態について述べる。

2 翻訳辞書環境

2.1 AS-TRANSACの辞書環境

AS-TRANSACの知識は、文法と辞書に大別され、さらに、辞書は、標準辞書・専門用語辞書・ユーザ登録辞書・学習辞書に分けられる(図1)。ユーザは、翻訳する文書に合わせて適切な辞書を選択すること

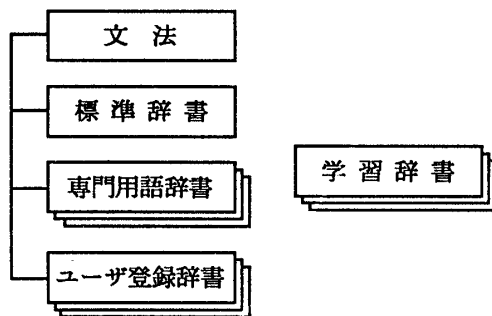


図1 AS-TRANSACの辞書環境

翻訳環境を設定してください。

専門用語辞書 : 【使用する】

information 他分野

ユーザ辞書 : computer_manual 他分野

図2 AS-TRANSACの辞書環境設定

とができる。これらの辞書の組合せを辞書環境と呼ぶ。

図2にAS-TRANSACの辞書環境の選択画面を示す。これは、情報処理の専門用語辞書と、計算機のマニュアル用のユーザ登録辞書を選択した場合の表示例である。[他分野]を選択すると、利用可能な辞書の一覧表が表示され、簡単な操作で切替えることができる。

ここで選択された辞書環境は、文書毎に記憶され、原文編集で辞書にない単語をユーザ登録辞書に登録すると、その文書を翻訳するときに、原文編集で選択した辞書環境が継承されて翻訳される。

学習辞書は、翻訳する文書の分野にあった言回しや訳語等の知識を記憶し、機械翻訳システムの中で、最も重要な役割を果たしている。訳語の学習にはいくつかの方法が考えられるが、きめこまかな学習効果を得られることに重点を置き、選択された専門用語辞書とユーザ登録辞書の組合せ毎に学習辞書を持つ方法を採用した。

2.2 ネットワークで共有される辞書の利用形態

複数のユーザが、ネットワークを介して機械翻訳システムを利用する場合の辞書の利用形態には、次の3つが考えられる。

- (1) ユーザ毎に別々の辞書をすべて持つ。
 - ◆ 大きなディスク容量が必要である。
 - ◆ 共通の知識(単語、訳語の学習等)が蓄積できない。
 - ◆ 辞書の一元管理が難しい。
- (2) ユーザが使用する辞書環境をすべて共有する。
 - ◆ 辞書アクセスの相互排他制御が必要であり、アクセス効率が低下する。
- (3) 翻訳環境の学習辞書のみユーザ毎に持つ。
 - ◆ 翻訳を開始するときに学習辞書の情報を対訳辞書に移す必要があり、辞書の管理が複雑になる。

一般に、ユーザ毎にすべての辞書環境を持つという(1)の方法は、ディスク容量から考えると実際には実現困難である。また、(3)の方法も、複数のユーザ

が同時に翻訳しているときに、あるユーザが単語の登録 / 削除を行った等という情報を、同一辞書環境で翻訳実行中の全てのユーザに伝えることが必要になり、実現が難しい。

そこで、我々は (2) の方法を取ることにした。

3 辞書の相互排他制御によるアクセス効率の変化

3.1 辞書アクセス

翻訳システムにおいて、辞書は、次の処理のときにアクセスされる。

◇ 翻訳	辞書引き	読込み
◇ 原文編集	辞書登録	書込み
	未知語検索	読込み
◇ 2か国語編集	辞書登録	書込み
	優先度変更	書込み
	辞書参照	読込み
◇ 辞書構築ツール	一括辞書登録	書込み

このとき、翻訳・辞書参照・未知語検索等、辞書の読込みは、複数のユーザが同時に要求しても問題がないが、辞書登録等、辞書の書込みは禁止する必要がある。一方、あるユーザが辞書の書込みを行っている間は、他のユーザの辞書の書込み・読込みともに禁止する必要がある。

そこで、辞書をアクセスするときの相互排他制御を行うことにした。このとき、翻訳プロセスは図3に示すようなステップに分かれているので、実際に辞書にアクセスする期間のみ、他のユーザのアクセスを禁止することとし、できるだけ翻訳速度に影響がでないようにした。この条件の下で、相互排他制御により生ずるアクセス効率の低下が、どの程度翻訳速度に影響を与えるか実験を行った。

3.2 実験結果

前節で述べた方法を検証するために、ネットワークに接続された複数のワークステーション (ディスクサーバ 1 台 + クライアント 数台) で、同一の辞書環境を同時に使用する実験を行った。

図4に、翻訳速度とワークステーションの台数の関係を示す。

ワークステーション 4 台以下で同時に翻訳を行った場合は、翻訳速度にほとんど影響が見られなかった。5 台以上になると、サーバの速度が数 % 低下した。これは、ディスクの I/O が多くなり、サーバの CPU の実行効率が下がったことが原因と考えられる。

一方、辞書の書込みを頻繁に要求する一括辞書登録を翻訳と同時に実行した場合は、翻訳速度の低下は見られず、一括辞書登録の速度が 10% 程度低下した。

しかし、一括辞書登録は頻繁に利用する機能ではないことから、実用上大きな問題はないと考えられる。

4 終りに

機械翻訳システム AS-TRANSAC を、ネットワークに接続されたワークステーション (ASシリーズ) 上で、複数のユーザが同時に利用する際の、翻訳辞書環境について述べた。今後、ワークステーション間だけでなく、パソコンやホスト計算機で構築された知識を共同で利用するシステムの開発を進めていきたい。

参考文献

1. 天野 他 『ワークステーションで実用速度を実現した翻訳システム』, archive No.8, 1988
2. 武田 他 『ホスト/パソコン機能分散型機械翻訳システム (1), (2)』, 情報処理学会 第37回全国大会, 1988

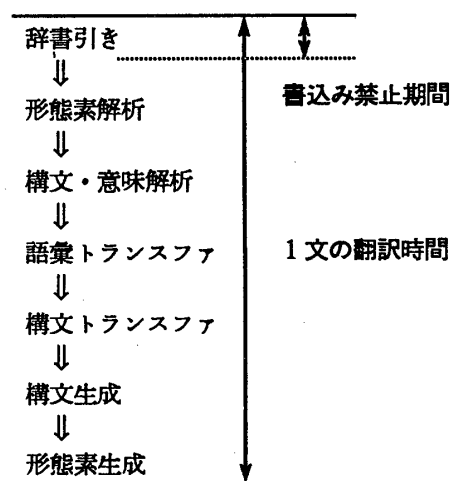


図3 翻訳プロセス

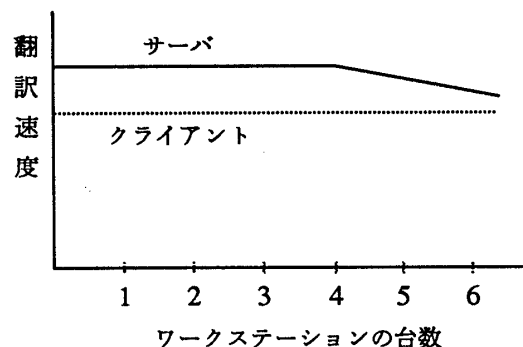


図4 実験結果