

機械翻訳システムにおける原文入力機能

2P-3 佐藤 牧子*、白井 成樹*、鎌田 敏弘*、伊藤 悦雄**、武田 公人**、天野 真家**

*(株)東芝 府中工場 **(株)東芝 総合研究所

1 はじめに

近年とみに機械翻訳システムの需要が高まってきている。そして更には翻訳精度向上に加えて、多量の文書の翻訳効率向上、誰でも使えるユーザインタフェースの操作性向上のニーズも非常に高まってきている。当社でも従来よりエンジニアリングワークステーションASシリーズを用いた英日・日英の2方向の機械翻訳システムASTRANSAC⁽¹⁾を販売している。今回手軽に、誰でも使え、かつ翻訳効率を大幅に向上させる英文ソフトOCR AS-Readerの開発を行い、機械翻訳システムASTRANSACと組み合わせて使うことにより、効率の良い、ユーザインタフェースに優れた機械翻訳環境の提供を可能にした。ここではこのAS-Readerの操作性、ユーザインタフェースをどの様に工夫し開発したのかを紹介する

2 システムの概要

図1で示すとおり、AS-Readerは、イメージ入力機能、認識機能の2つのプロセスから構成される。

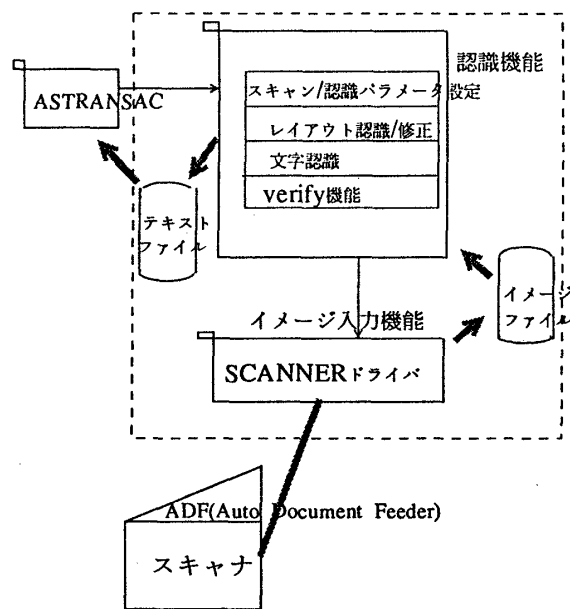


図1 AS-Readerのソフトウェア構成図

3 開発のポイント

AS-Readerは、当社で既開発のExpress Reader⁽²⁾の文字認識機能を利用し、ワークステーション環境下で、汎用的にテキスト文書の認識を行うソフトとすることにした。Express Readerでは、文字認識率は、99.7%、認識速度は、ハードを使用し、120字/秒である。AS-Readerでは、文字認識率は、Express Readerと同じ99.7%、認識速度は、ソフトでの処理であるが、60字/秒が設定できた。本報告では、ユーザインタフェースを中心に述べる。

AS-Readerを開発する際に留意した点。

- 使いやすいユーザインタフェース
- ASTRANSACの操作性との統一
- DTPシステムとの接続性

3.1 使いやすいユーザインタフェース

(1)verify機能の提供

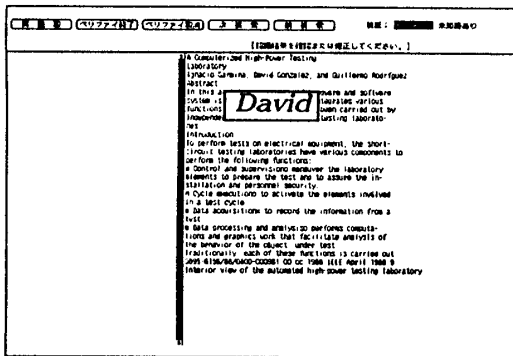
文字認識後、認識誤りの修正を容易に行うために、suspicious候補(各文字の認識における誤りの可能性のあるもの)の検索機能を設けた。これがverify画面である。修正時には、認識前のイメージやその文字の原稿中の位置、および、修正候補を提示する。suspicious文字の検索の他に、未知語の検索(認識した後の語で、辞書に無い語を検索する)も行えるようにした。これは、翻訳の高速辞書を組み込むことによって実現させた。verify機能の提供によって、翻訳に渡す文書の正確性を高める事ができる。

(2)わかり易い画面設計

AS-Readerが使いやすい様に、以下に気を配った画面設計とした。

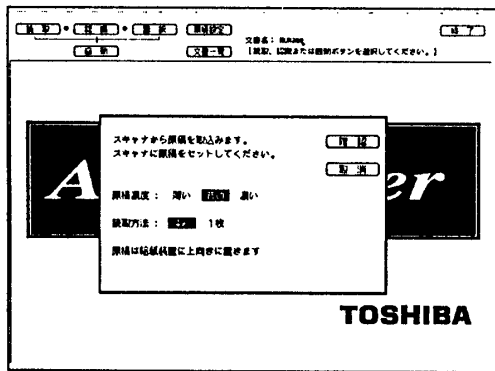
- 編集する際のマウスの移動距離を最小にする
- 最小限のボタンだけを画面に表示する
- ボタンを見れば、操作がわかる

例えばマウスの移動距離を最小にするために、verify画面では、右側にある編集画面と頻繁に行き来することを考えて、suspicious characterを取り出す次検索と前検索のボタンを右側に持ってくるなどの工夫をした。(図2のボタン周りを参照)



認識誤りの訂正を行うverify編集
図2 verify

又、例えば、図3のボタンを見れば、読取り・認識・翻訳という処理が、どういう流れで運ばれるかが分かるようになっている。



読取り認識翻訳のボタン
図3.読取りの図

3.2 ASTRANSACとの操作性の統一

以下の点でASTRANSACとの操作性の統一を図った。

- ボタンの選択は、翻訳システムもAS-Readerもマウスの左ボタンを押す。
- ボタンとチョイスで、操作する。
- ASTRANSACの翻訳環境をAS-Reader側でも設定出来る。

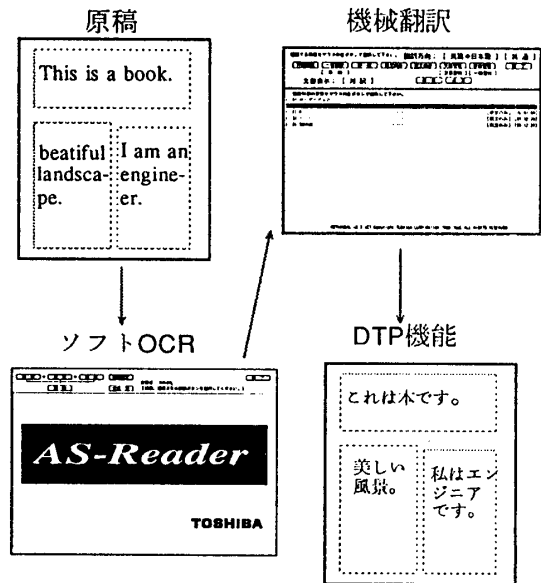
翻訳環境とは、ユーザーの目的とするユーザー辞書や、専門辞書を選んだり、英訳の際は主語のない文や冠詞の付け方、箇条書きの表現などを設定する機能である。これにより、AS-Readerから容易にASTRANSACによる翻訳処理が可能になった。

3.3 DTPシステムとの接続性

近年、コンピュータを用いたドキュメントの電子化は、急速な勢いで広がっている。ドキュメンテーションシステムの一貫として翻訳システムを位置付け、その全体として効率化を図ってゆくという考え方は、今後ますます重要になっていくと

思われる。そこで、AS-Readerを機械翻訳を通してASシリーズ上で稼働するDTPシステムAS-Documentsと接続出来るようにした。

ASTRANSACは、テキストを用いて翻訳しているので、レイアウト記述言語を持つ必要性はなかった。しかし、AS-DocumentsにはMIF(Maker Interchange Format)というドキュメント記述言語がある。AS-Reader,ASTRANSACで、MIF形式を採用することによって、お互いのやり取りを可能にした。その結果、多段組みの英文レターを読みとって、同じ段組みの日本語レターにすることが可能となった。(図4参照)



4 おわりに

以上、述べてきたようにAS-Readerを開発したが、スキャナによる文書入力から翻訳、印刷、出力までの一連の処理を容易に行う環境が実現できた。今後の課題としては、読み誤りの学習などのサポート機能の充実、日本語入力機能の開発等を行い、システムをより使いやすいものとしていきたい。

[参考文献]

(1)Amano,S. et al "English / Japanese Toshiba Machine Translation Systems",Informant, Vol.4 No.2 pp86-108, 1987
(2)Tsujiimoto, S., and Asada, H.: "Understanding Multi-Articled Documents", in Proc. 10th Int. Conf. Pattern Recognition, Atrantic City, New Jersey,pp.701-709, 1991.