

文構造を有する日本語テキストエディタ

4J-2

J E S S

福島 俊一¹, 首藤 友喜¹, 柳生 弘之²

(日本電気株式会社 C&Cシステム研究所¹, 日本電気技術情報システム開発株式会社²)

1. はじめに

日本語ワードプロセッサが広く普及し、文章が電子化されて扱われるようになってきた。電子化された文章には、修正・保存・流用などが容易に行えるというメリットがあるが、それだけにとどまらず、近年さらに、電子化された日本語文章について、機械的校正を行うことが試みられてきている[1-4]。

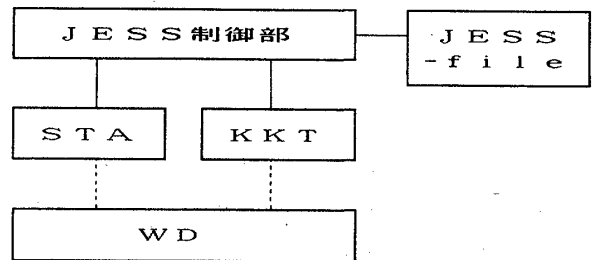
従来、文章は単なる文字コード列の形態で記憶・処理されているが、機械的校正では、文章の構造を扱う必要がある。そこで、文章の構造を備えた、機械的校正に適した電子化文章の形態を検討する必要がある[3,5]。

筆者らは、文構造を備えた文章の管理方式を検討し、文構造に基づいた文章の編集を可能とする日本語テキストエディタ J E S S (Japanese Text Editor with Sentence Structure) を試作したので、報告する。

2. J E S S - f i l e

J E S S は、文章について、表記文字列だけでなく、読み仮名や文構造情報もファイルとして持つ(これを J E S S - f i l e と呼ぶ)。J E S S - f i l e は次のような4ファイルから構成される。

- ① 表記文字列ファイル
- ② 読み仮名ファイル
- ③ 文構造ファイル
 - 文・文節・単語の先頭位置
 - 単語の品詞・アクセント
 - 読み仮名との対応
- ④ 書式情報ファイル



STA: Syntactic Text Analyzer
 KKT: Kana-Kanji Translator
 WD: Word Dictionary

図1 J E S S の構成

3. J E S S の構成

J E S S の構成を図1に示す。J E S S はパソコンPC98XA上に試作されている。

J E S S は、文章解析部STAと仮名漢字変換部KKTとをサブユニットとして備えている。このSTAとKKTは、単語辞書WD(単語の表記・読み仮名・品詞・アクセントなどの情報を格納している)を参照した次のような処理を行い、J E S S - f i l e の内容を生成する。

文章解析部 STA

表記文字列を解析し、文・文節・単語の単位や単語の品詞の認定、単語の読み・アクセント付けを行う[6]。

仮名漢字変換部 K K T

読み仮名列を解析し、文・文節・単語の単位や単語の品詞の認定、単語の表記の決定、アクセント付けを行う。

Japanese Text Editor with Sentence Structure: JESS
 Toshikazu FUKUSHIMA¹, Tomoki SHUTOH¹ and Hiroyuki YAGYU²

¹ C&C Systems Research Laboratories, NEC Corporation

² NEC Scientific Information System Development, Ltd.

従って、JESSは、表記文字列（従来の文章ファイル）の読み込みと、仮名漢字変換の2通りの入力が可能である。

4. JESSの機能

従来の表記文字列に加えて読み仮名や文構造情報を持つJESS-fileをもとに、JESSでは次のような編集機能を実現した。

文構造単位カーソル移動機能

文単位、文節単位、単語単位にカーソル移動を行うことが可能である。

文構造単位編集機能

削除・移動・複写などの編集の対象として、文字列の範囲（範囲の先頭位置、末尾位置）を指定するのではなく、文・文節・単語などの文構造単位の指定が可能である。カーソル位置を含む文構造単位を自動的に抽出する。

再変換機能

読み仮名ファイルを持っているので、仮名漢字変換結果が確定した部分についても、随時、再変換を行うことが可能である。

単語検索機能

単なる文字列照合による検索でなく、品詞を考慮した単語の検索を行うことが可能である。ユーザによって入力された被検索文字列について、文章解析部STAによる品詞認定を行い、表記（用言については語幹の表記）と品詞の一致する単語を検索する。従って、活用した用言の検索も可能である。

また、現在まだインプリメントしていないが、置換についても、同様の機能が実現できる。

JESS-fileをベースにすることによって、上記の編集機能以外にも、校正機能をはじめとする様々な日本語文章作成支援機能の実現が容易になる[4]。

5. 検討項目

(1) 文章解析部STAや仮名漢字変換部KKTによる解析誤りにより、JESSの処理する文構造とユーザの意図する文構造とに食い違いが生ずることがある。

(2) 部分的な挿入・削除を行った場合、それまでの文構造情報が壊れるので、文章解析部STAを再起動し、文構造情報を生成し直す必要が生ずる。

6. おわりに

文構造を有する日本語テキストエディタJESSの試作を行った。今後、文構造に基づく編集機能を強化してゆくとともに、JESSを核とした日本語文章作成支援システムCOMETの構築を行ってゆく予定である[4]。

また、文・文節・単語などの文構造単位だけでなく、章・節・段落などの論理構造単位の編集も検討してゆきたいと考えている。

最後に、本研究の機会を与えて下さった首藤正道ターミナルシステム研究部長、並びに有益な助言を与えて下さった関係者の皆様方に深く感謝いたします。

[参考文献]

- [1] 牛島・他：“日本語文章推敲支援ツールのプロトタイピング”，情処・ソフトウェア工学研究会 40-8, (1985)
- [2] 空閑・他：“日本語校正支援用のOANERS”，情処全大 60年前期
- [3] 武田・他：“日本語文書校正支援システムCRITAC”，情処全大 61年前期
- [4] 福島・他：“日本語文章作成支援システムCOMET”，情処全大 61年後期
- [5] 角田：“多層テキスト構造を持つ日本語エディタ”，第27回プログラミングシンポジウム, (1986)
- [6] 福島・他：“盲人用読書器における文音声変換のための文章解析”，情処・日本語文書処理研究会 2-4, (1985)