

マニュアル文章検査システムTECS/M — 検査機能と評価

山田洋志† 大山裕‡

日本電気(株)

† 情報メディア研究所 ‡ パーソナルC&C開発研究所

本稿では、マニュアル向けに開発した文章検査システムTECS/Mの検査機能を紹介する。また、送りかなや表記のゆれを検出する“送りかなゆれ検査”について、性能を評価した結果について報告する。

TECS/Mは、一般文書向け検査機能(20種)に加え、マニュアル作成部門のマニュアル作成ガイドラインをもとに追加したマニュアル向け検査機能(16種)を備えている。また、不要な検査結果を出力しないための検査除外箇所指定機能、 \LaTeX 文書を検査するためのフィルタを用意した。

“送りかなゆれ検査”機能の性能を評価した。評価は、人為的に表記ゆれを加えた文章で行い、再現率(もれなく検出できたか)と適合率(むだなく検出できたか)を測定した。結果、適合率は87-100%と高かったが、再現率が67-74.5%とやや低かった。検出もれの原因としては、形態素解析段階での品詞や読みの認定誤りの影響が大きい。

TECS/M : A Text Critiquing System for Manual

Hiroshi Yamada Yutaka Ohyama

NEC Corporation

This report describes system structure and functions of TECS/M (A TExt Critiquing System for Manual).

TECS/M has 36 proofreading functions for Japanese text strings. 20 functions are suitable for general text and 16 functions for manuals text. 16 functions for manual text are selected from guidelines to write manual. TECS/M also has a function which excludes unnecessary proofreading results in output files.

One of important function is 'YURE KENSA' (detection of Kanji usage inconsistency). This report also describes performance of 'YURE KENSA'. The function can detect 70% inconsistency of Kanji usage. And superfluous output is 0 - 15 % of all outputs. Morpheme analysis errors cause most of superfluous output.

1 はじめに

近年の日本語文書処理技術の向上につれ、計算機を用いた日本語文書作成が日常的に行われるようになって久しい。それに、誤字・脱字、用語の誤りの指摘など、校正・推敲作業の一部を自動的に行うという研究も数多く行われてきた[1, 2, 3, 4]。

筆者らは、日本語文章作成支援システムCOMET[5, 6]を試作し、文章校正支援の研究を進めていたが、今回、マニュアルを対象とした実用に耐える校正支援システムを開発することを目的としてマニュアル検査システムTECS/M(A TExt Critiquing System for Manual)を開発した[7, 8]。

本稿では、TECS/Mの概要と検査機能を紹介し、主要な検査機能のひとつである“送りがなゆれ検査”機能について性能を評価した結果を報告する。また、今後の課題についても述べる。

2 TECS/Mの開発

筆者らは、以前からワープロ組み込み型の日本語文章作成支援システムCOMET[5, 6]を試作し、文章校正支援に関する研究開発を行ってきた。また、1992年には、出版社の校閲部門を対象とする校正支援システムSt.WORDS[9, 10, 11]を開発した。

一方、近年マニュアルの品質に関する関心が高まっており、マニュアルを対象とした校正支援システムの要望が強くなっている。COMETおよびSt.WORDSは、多種多様な文書への対応を想定しており、特定の書式に適合しているかについては検査しない。そこで、COMETで試作した一般文書向け検査に加え、マニュアル向けに検査機能を強化してマニュアル文章検査システムTECS/Mを開発した。

TECS/Mは、以下の方針に基づいて開発した。

1. 社内のマニュアル作成部門での実用に耐えるシステムを作成する。
2. 一般的な文章検査項目の他に、マニュアル特有の書式に関する検査項目を備える。
3. 検査項目や結果の出力形式は、ユーザが必要に応じて選択できるようにする。
4. 利用者自身が検査項目を定義できるようにする。

マニュアル向けに追加する検査項目は、計算機用マニュアル作成部門で作成したマニュアル作成ガイドラインの中から、計算機処理の有効性と機能の実現難易度を考慮して決定した。

TECS/Mは当初パソコン用のシステムとして開発していたが、開発・評価に協力してもらったマニュアル作成部門でUNIXワークステーション上で \LaTeX を使用してマニュアルを作成していたことと、検査する文

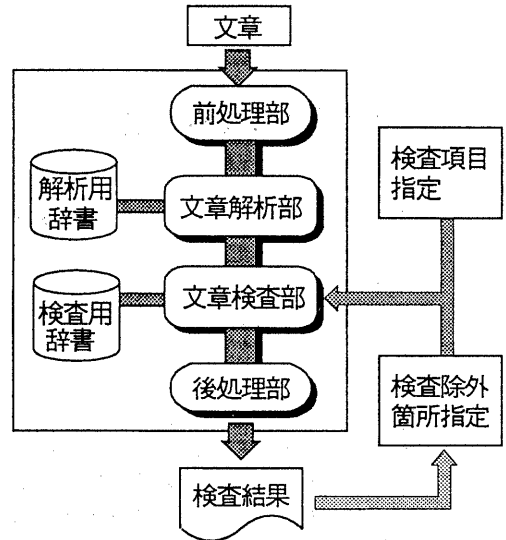


図1: TECS/M システム構成

章の量が多いためパソコンでは速度や容量が不足気味だったことから、ワークステーション上で動作するようにした。全体の構成は、エディタなどの種類を問わないバッチ型システムとした。

現在は、TECS/M Ver.2の協力部門での試用段階を終え、複数部門で実際に業務に使いながら改善を進めていく段階である。

3 システム概要

システム構成について説明する(図1)。

検査対象文章 UNIXのプレーンテキスト。 \LaTeX で書かれたテキストを検査する場合は、 \LaTeX コマンドの削除・置き換えを行うフィルタを用いてプレーンテキストに変換する(後述)。

前処理 文字コード変換、罫線削除、文の途中の改行を削除する。

文章解析 入力テキストを形態素解析する。

解析用辞書 50万語の単語辞書。

文章検査 形態素解析結果に対して各検査を行う。どの検査を行うかの指定は、検査項目指定ファイルで行う。

検査用辞書 誤用語検査、書換え語検査など特定の単語を対象とした検査項目のための辞書。現在約3800語登録しており、ユーザが編集して追加削除できる。

検査項目指定 検査項目およびオプションを指定したテキストファイル(プロパティシート。図2)。

誤用表現検査	辞書="../dic/goyou"
長段落検査	文数=10 行数=20
送りがなゆれ検査	マーク対象=少数
受動態抽出	

図 2: プロパティシート

011 字種連続検査	シミュレーションシステム
014 漢字書換検査	推敲
044 類語検査	作成

図 3: 検査除外箇所指定ファイル

ユーザが目的に応じて編集できる。

後処理 文章検査結果を、所定の書式に編集して出力する(例を付録Bに示す)。出力形式は起動時のオプションで指定する。

検査除外箇所指定 検査結果のうち修正の必要のない部分については、検査除外箇所指定ファイルに登録することで次回から出力されなくなる。検査コード、検査種別、対象語を指定する形式(図3)で、ユーザが検査結果を編集して作成する。除外指定は文章全体にわたって有効である。

LaTeXテキストは、文や単語の途中に各種のコマンドを含むため、そのまま検査にかけると形態素解析に失敗して不必要な検出が多くなる。そこで、LaTeXコマンドを削除・置換し、プレーンテキストに変換するフィルタを用意した。TeXに付属するdetexに似た機能を持ち、不要なコマンドの除去、章番号の数値への置き換え、検査対象外の環境(tabular、figure、verbatimなど)の削除を行う。また、特定部門のローカルマクロの処理も行っている。

TECS/Mの検査実行速度は、検査の項目数に依存するが、200-250文字/秒程度(形態素解析の時間—350文字/秒—を含む)。一度に処理できる文章の量は、100ページ程度(検出箇所数に依存する。メモリに余裕があればさらに増やせる)。

4 検査機能

TECS/Mでは、文章検査機能を以下の3種類に分類している。

- 一般文書向け検査機能
マニュアル文章に限らず、通常使用される日本語文章を検査する際に有効な項目。日本語文章作成支援システムCOMET[5, 6]の検査機能をベースに、20検査項目を実現している。
- マニュアル向け検査機能
TECS/Mの開発にあたって追加した機能。16検

査項目を実現している。

- ユーザ定義機能
利用者自身が新規作成するための検査機能。ユーザが作成した検査用辞書を参照してユーザが定義したエラーメッセージを表示する(最大9種)。

付録Aに検査機能の一覧を挙げた。

マニュアル向け検査のいくつかを簡単に説明する。

細別記号参照1

“1章”、“図2.1”のように、テキストの一部として出現する細別符号(参照符号)を抽出する。ただし、章題自体は抽出対象とならない。

何を抽出対象とするかは、プロパティシート内で指定する。

細別記号参照2

参照用細別符号のうちで、参照対象となる細別符号自体が存在しないものを抽出する。たとえば、“3.4節では……”という文があるのに3.4節がない。この機能は、LaTeXで作成した文章には必要ない。

のみ検出・おいて検出・など検出・受動態検出

意味が曖昧になりやすいといった理由により、マニュアル作成ガイドブックでなるべく使用しないように指示している表現を検出する。

受動態検出では、敬語表現・可能表現との区別をしない(できない)。

並列連続表現抽出・動詞化検査

“並列連続表現抽出”は、“[や]”、“[と]”、“[・]”で指定回数以上に列挙している箇所を抽出する。“動詞化検査”は、“○○を行う”、“○○を実行する”、“○○ができる”といった表現を抽出し、“○○する”、“○○できる”のような表現を勧める。

いずれもわかりやすく簡潔な文章を書くための指針である。

年号抽出・時制表現抽出

“年号抽出”は、年号(昭和・平成など)を抽出する。

“時制表現抽出”は、過去・推量の助動詞を抽出する。

いずれも、マニュアル一般というよりも特定部門対応の検査機能である。

類語検査

マニュアル向けではないが、TECS/Mの開発にあたって追加した。

同音類語を有する単語(「愛惜/哀惜」、「合う/会う/逢う/遇う/遭う」など)で、用法に注意が必要なものを検出する。

5 ゆれ検出機能の評価

TECS/Mには、送りがなや表記のゆれを検出する、「送りがなゆれ検査」機能がある。人間がゆれを検出するには多大な労力を要するため、自動化の期待が大きい。そこで、現状でのゆれ検出性能を調査した。

指標として、以下で定義される再現率(もれのなさ)と適合率(むだのなさ)を用いた。

$$\text{再現率} = B / (A + B), \quad \text{適合率} = B / (B + C)$$

A: 検出したかったが、検出できなかった箇所

B: 検出したいくて、実際に検出できた箇所

C: 検出したいくなかったが、検出した箇所

5.1 表記ゆれの検出方法

送りがなゆれ検査では、形態素解析結果から自立語をすべて取り出し、ゆれの関係になっているかどうかを検査する。今回の実験では、ゆれかどうかの判断に以下の2通りを用いた。

方式 a: 読みと品詞による判定

単語の読みと品詞が同じで表記が異なればゆれと見なす。この方式では、同音異義語も検出する。

方式 b: 読みと品詞と漢字による判定

単語の読みと品詞が同じという条件に加えて、以下のいずれかが成り立つときにゆれと見なす。

1. 一方の語の表記がすべてひらがなである。
2. 表記に含まれている漢字が同一、同順序。
例) 「取り上げ」「取上げ」「取上」

5.2 ゆれ検出実験結果

ゆれ検査機能は一般文章向け検査なので、硬い文章と通常の文章の2種類を用いた。いずれも同じ内容を2回繰り返してあり、一方に表記のゆれを加えた。

- 日本国憲法の一部 (4,918字、検出対象箇所 204)
- 小説の一部 (9,092字、検出対象箇所 615)

結果を表1に示す。A,B,Cは前述の説明に対応する。

5.3 結果の分析

5.3.1 検出過剰

適合率が87-100%と、ゆれ以外のものを検出することは少ない。主な原因を挙げる。

同音異義語の検出 方式aでは、読みと品詞だけで判別するため、同音異義語を検出した。方式bでも、かな書き語と漢字語の組み合わせを検出した。

文章	方式	もれ	成功	過剰	再現率	適合率
		A	B	C	B/(A+B)	B/(B+C)
憲法	a	52	152	16	74.5%	90.5%
	b	58	146	0	71.6%	100.0%
小説	a	185	430	62	69.9%	87.4%
	b	204	411	22	66.8%	94.9%

表 1: ゆれ検出実験結果

原因	方式a	方式b
品詞の不統一	32	32
読み付け誤り	9	9
漢字の異なるゆれ	—	6
その他	11	11
計	52	58

表 2: 検出もれの原因別内訳(検査対象:「憲法」)

読み付けの誤り 形態素解析段階で単語の読み付けを誤り、別の単語をゆれとして検出した。

例) 「あいだ」 = 「間」(正しい読みは“ま”)

品詞認定の誤り 形態素解析段階で品詞の認定を誤り、本来別の単語がゆれとして検出された。

例) 「言った」 = 「いった」(“行った”と対応)

5.3.2 検出もれ

再現率は70%前後とやや低く、表記ゆれのうち約3割は検出できなかった。主な原因を挙げる。表2には「憲法」の検出もれの件数を原因別に示す。

別の品詞として認定 表記ゆれでありながら、別の品詞の語として解析された。

例) 「ひとしく」(副詞) ≠ 「等しく」(形容詞)

読み付けの誤り 形態素解析段階で、単語の読み付けを誤り、別の単語と判断した。

例) 「いっさい」 ≠ 「一切」(ひとときれ)

漢字の異なるゆれ(方式b) 一部だけがかな書きされたり、漢字が異なると検出できない。

例) 「引っ張る」 ≠ 「引っぱる」

その他 形態素解析での単語区切り誤り、付属語のゆれ(自立語だけを検出対象とした)、未登録語など。

5.3.3 対応の誤り

「小説」での検出成功例には、検出はできたがゆれの対象を間違えたものが9例あった。

例) 「あい(会い)に行く」 = 「藍」: 「あい」を名詞と解析

5.4 ゆれ検出に関する考察

5.4.1 検出性能改善の方法

どちらの方式も、適合率は87-100%と高いが、再現率は67-74.5%と十分とはいえない。現在の再現率では、いずれの方式でも、検出もれの見直しが必要になるので、適合率が非常に高いという点から方式bを使用する方が有効である。

形態素解析の誤りと単語登録の不備が、検出誤りの大きな原因となっており、改善する必要がある。辞書の不備については、登録もれや品詞の不統一、読み順位を見直すことで現状のシステムで性能を上げることができる。形態素解析誤りについては、付属語の見直しや用例の利用を検討する。

5.4.2 その他の検出もれについて

現在の検出アルゴリズムでは、「踊・る」と「おどり(名詞)」や、「ある・く」と「歩き回・る」のように別の単語・品詞間での表記ゆれは検出できない。

これらを機械的に検出すると、むだな検出が増えて、人間がゆれと見なすべきかを判断しなければならなくなる。精度の維持と実現性を考えると、出現頻度の多い特定の単語の組み合わせについてあらかじめ登録しておくのが現実的である。

6 今後の課題

LaTeX文書の直接検査 現在、LaTeX文書にフィルタをかけ、プレーンテキストに変換してから検査しているため、検査結果の位置情報がもとの文書と一致しない。そこで、LaTeX形式文書での位置を出力したい。また、roffやSGMLなど他の形式のコマンドやタグの入った文書にも対応したい。

インタフェースの改良 現在、利用形態拡大のひとつとして、電子メールを用いた文章検査サービスを部門内で試行している。その他、Emacsからの起動や、検査機能をライブラリとして他のアプリケーションからの利用できるようにしたい。ゆれ検査の出力形式などの細かい点に関しても改善の余地がある。

エキスパートシステム化 現システムでは、検査用の辞書はユーザが編集できる。しかし、大部分の検査機能はシステムに固定で、実行するかどうかを選択できるだけである。検査機能についても、ユーザがある程度構築できるようにしたい。

動作機種が多様化 TECS/MはNEC製のEWS4800シリーズで動作しているが、他のUNIXマシンでも動作させたい。また、最近の機種なら十分な速度が期待できるのでパソコンでも動作させたい。

7 おわりに

マニュアル文章検査システムTECS/Mについて、開発方針とシステム構成、検査機能について紹介した。TECS/Mは、ワークステーション上で稼動するバッチ形式の文章検査システムで、46種類の検査機能、7種類の出力形式をもつ。ユーザ支援機能として、プロバティシートによる検査項目の選択、ユーザ定義可能な検査用辞書、LaTeX文書検査用のフィルタを備える。

また、2種類の文章を対象に、2通りの表記ゆれ検出方法による性能評価結果を報告した。両方式とも適合率は87-100%と高い(余計なものは検出しない)が、再現率は67-74.5%と不十分(検出もれが生じる)という結果になった。

検査結果を調べた結果、検出性能を向上させるためには、検査の基礎となる形態素解析の性能向上と、単語辞書の品質向上が必要であることが分かった。

今後は業務での利用を通して、改良を進めていく予定である。

謝辞 TECS/Mの開発・評価に協力していただいた、NEC情報メディア研究所ならびにNEC基本ソフトウェア開発本部の方々に深謝致します。

参考文献

- [1] 牛島他，“日本語文章推敲支援ツールの試作とその作成環境”，情処SW研,35-2,1984
- [2] 鈴木他，“日本語文書校正支援システムCRITAC”，情処JDP研,8-5,1986
- [3] 高橋他，“計算機マニュアル推敲・査読システムMAPLEの開発と運用”，情処論文誌,Vol.31,No.7,1990
- [4] 奥村他，“日本語校正支援システムFleCS”，情処NL研,87-11,1992
- [5] 福島他，“日本語文章作成支援システムCOMET”，信学技報,OS86-21,1986
- [6] 福島他，“日本語文章作成支援システムCOMET—文章解析応用の統合化方式を中心に—”，情処DPHI研,20-2,1988
- [7] 大山，“マニュアル検査システムTECS/M—基本構成—”，情処43全大,6H-3,1991
- [8] 大山，“マーク付けテキストを対象とした文章検査”，情処44全大,5C-2,1992
- [9] 福島他，“日本語文書校正支援システムSt.WORDS”，情処45全大,6C-1,1993
- [10] 福島他，“日本語文書校正支援システムSt.WORDSの文書検査機能”，情処46全大,3L-3,1993
- [11] 竹元他，“口語的表現を含む日本語文の形態素解析”，情処46全大,1B-2,1993

A 検査機能一覧

一般文書向け検査(20種)	
誤用表現検査	誤用表現辞書を参照して、誤った表現を検出し、正しい表現を対応づけて表示する。
括弧対応検査	文中の括弧対応が不正な部分を検出する。全角、半角の括弧は別のものとして検査する。
句読点重複検査	文中で、句読点が2つ以上連続している部分を検出する。
形態素解析検査	形態素解析に失敗した部分を検出する。
長段落検査	長すぎる段落を検出する。
長文検査	長すぎる文を検出する。
句読点間隔検査	句読点間隔が長すぎる部分を検出する。
字種連続検査	同一の字種(ひらがな、カタカナ、漢字)が一定数以上連続している部分を検出する。
かな書換検査	かな書きが望ましい語を検出し、推奨する表現を対応づけて表示する。
漢字書換検査	別の漢字への書換が望ましい語を検出し、推奨する表現を対応づけて表示する。
文体検査	文体が統一されているかを検査する。文体として常体と敬体の2種類を識別する。
指示語検査	指示語辞書に登録されている指示語を検出する。
句読点ゆれ検査	句読点が混在して使用されているときにエラーと見なす。
送りがなゆれ検査	送りがなやかなづかいなど、表記のゆれを検出する。
カナゆれ検査	カタカナ表記のゆれを検出する。
英字ゆれ検査	英字表記のゆれを検出する。
章節番号検査	章題(節題)に対して、番号が連続しているかを検査する。
平均文長検査	平均文長を計測し評価する。
漢字含有率検査	テキスト中の漢字含有率を評価する。
章題抽出	テキスト中の章題(節題)部分を抽出する。
マニュアル向け検査(16種)	
年号抽出	西暦でない年号文字列(明治・大正・昭和・平成)を抽出する。
文字列抽出	カタカナ、英字、記号で、単語辞書に登録されていない文字列を抽出する。
受動態抽出	助動詞「れる」「られる」を検出する。
時制表現抽出	過去の助動詞(だ、た)および未来(推量)の助動詞(う、だろう、でしょう)を抽出する。
句読点かぎ検査	かぎ括弧と、句読点の位置関係を検査し、“[”、“、”、“、”などの組み合わせを検出する。
句読点括弧検査	括弧と、句読点の位置関係を検査し、“(”、“、”、“、”などの組み合わせを検出する。
並列連続表現抽出	表現を曖昧にする可能性のある並列連続表現を抽出する。
細別記号参照1	テキストの一部として出現する細別符号を抽出する。
細別記号参照2	参照用細別符号のうちで、参照対象となる細別符号自体が存在しないものを抽出する。
動詞化検査	「○○を行う」という表現を抽出し、「○○する」という表現を推奨する。
のみ検出	助動詞「のみ」を検出し、「だけ」を推奨する。
おいて検出	連語「おいて」を検出し、「で」を推奨する。
など検出	副助詞「など、なぜ」を検出し、曖昧にする表現なので用法の注意を喚起する。
接続詞検査	接続詞で直後に読点がないものを検出する。
類語検査	同音類語を有する単語で、用法に注意が必要なものを検出する。
細別符号抽出	段落の先頭で章節の番号を示す細別符号部分を抽出する。
ユーザ定義検査	
ユーザ辞書参照	利用者が作成した検査用辞書(最大9個)を使用する。

B 出力形式一覧

B.1 ローカルエラー出現箇所情報

入力テキストの全文および、エラーコードをエラー検出箇所の下部に付与したものの。

[検査結果の例]

最近、ペンを用いて文字の入力やコンピュータの操作を行うペンコンピューティ
#007 #008 #038
ングへの期待が急速に高まってきており、また製品も次々に発表されている。ペ
#030 #007
ンを用いて文章を入力する場合、漢字文字認識によって、かな漢字変換を使用せ
#008
ずに漢字かな混じりの文章を入力することができる。しかし、漢字の認識率や複
#044 #007

B.2 ローカルエラー出現箇所情報

エラーを検出した行だけを、ページ番号・行番号およびエラーコードとともに出力したものの。

[検査結果の例]

P.001 L.04 || 最近、ペンを用いて文字の入力やコンピュータの操作を行うペンコンピューティ
|| #007 #008 #038
P.001 L.05 || ングへの期待が急速に高まってきており、また製品も次々に発表されている。ペ
|| #030 #007
P.001 L.06 || ンを用いて文章を入力する場合、漢字文字認識によって、かな漢字変換を使用せ
|| #008
P.001 L.07 || ずに漢字かな混じりの文章を入力することができる。しかし、漢字の認識率や複
|| #044 #007

B.3 エラー内容の詳細情報

ローカルエラーを、発生位置順に出力したものの。代替候補を同時に出力する。

[検査結果の例]

P.001 L.04 #007 文中の文字数が40を超えています。(長文検査)
P.001 L.04 #008 句読点間隔が30文字を超えています。(句読点間隔検査)
P.001 L.04 #038 [サ変名詞]+[する/できる]の形式に直せませんか。【操作を行う】(動詞化検査)
P.001 L.05 #030 受身とみられる表現を検出しました。(受動態抽出)
P.001 L.05 #007 文中の文字数が40を超えています。(長文検査)
P.001 L.06 #008 句読点間隔が30文字を超えています。(句読点間隔検査)
P.001 L.07 #044 同音類語を検出しました。【混じり：くまじ・る 交じる・混じる・雑じる】(類語検査)

B.4 エラー内容の詳細情報

ローカルエラーを、エラーコード順に出力したものの。代替候補を同時に出力する。

[検査結果の例]

#008 句読点間隔が30文字を超えています。(句読点間隔検査)
P.001 L.04
P.001 L.06
#013 かな書きすべき語です。(かな書換検査)
P.001 L.34 【例えば→たとえば】

- #030 受身とみられる表現を検出しました。(受動態抽出)
P.001 L.05
- #038 [サ変名詞]+[する/できる]の形式に直せませんか。(動詞化検査)
P.001 L.04 【操作を行う】
- #044 同音類語を検出しました。(類語検査)
P.001 L.07 【混じり：くまじ・るゝ交じる・混じる・雑じる】

B.5 ローカルエラー出現箇所・詳細情報

入力テキストの全文および、エラーコードをエラー検出箇所の下部に付与し、エラー内容の詳細情報を付記したものの。

[検査結果の例]

最近、ペンを用いて文字の入力やコンピュータの操作を行うペンコンピューティ

#007 #008

#038

#007 文中の文字数が40を超えています。(長文検査)

#008 句読点間隔が30文字を超えています。(句読点間隔検査)

#038 [サ変名詞]+[する/できる]の形式に直せませんか。(動詞化検査)

ングへの期待が急速に高まってきており、また製品も次々に発表されている。ペ

#030 #007

#030 受身とみられる表現を検出しました。(受動態抽出)

#007 文中の文字数が40を超えています。(長文検査)

ンを用いて文章を入力する場合、漢字文字認識によって、かな漢字変換を使用せ

#008

#008 句読点間隔が30文字を超えています。(句読点間隔検査)

ずに漢字かな混じりの文章を入力することができる。しかし、漢字の認識率や複

#044

#007

#044 同音類語を検出しました。【混じり：くまじ・るゝ交じる・混じる・雑じる】(類語検査)

B.6 グローバルエラー情報

グローバルエラーを出力したもの。

[検査結果の例]

#045 平均文長(句読点对応) 48.60文字 評価：長い

#046 漢字含有率(句読点对応) 47.86% 評価：非常に多い

#047 章題抽出

1. はじめに
2. 混ぜ書きかな漢字変換

B.7 エラー埋め込み型ファイル

もとのテキストの中に、検出したエラーコードを直接挿入したもの。

[検査結果の例]

#007最近、#008ペンを用いて文字の入力やコンピュータの#038操作を行うペンコンピューティ
ングへの期待が急速に高まってきており、また製品も次々に発表さ#030れている。#007ペン
を用いて文章を入力する場合、漢字文字認識によって、#008かな漢字変換を使用せずに漢字かな
#044混じりの文章を入力することができる。#007しかし、漢字の認識率や複雑な漢字を書く手
間を考えると、かな漢字変換と漢字の直接入力とを#044使い分けるのが実用的である。