自然言語処理 87−11
（1992. 1. 17）

日本語校正支援システム「ＦｌｅＣＳ」

奥村薫，建石由佳，脇田早紀子，金子宏
日本アイ・ビー・エム㈱　東京基礎研究所

概要：本稿では，日本語校正支援システム「ＦｌｅＣＳ」について紹介する。ＦｌｅＣＳは，パーソナル・コンピュータ上で対話的校正環境を提供するシステムである。また対象分野への適応や，さまざまなユーザ・インターフェイスに対応できるような柔軟な構成を，その特徴とする。さらに計算機マニュアル向けのＦｌｅＣＳのインプリメンテーション，使用報告と評価についても述べる。

Japanese Critiquing System "FleCS"

Kaoru Okumura, Yuka Tateishi, Sakiko Wakita, Hiroshi Kaneko

Tokyo Research Laboratory
IBM Research

Abstract: A Japanese critiquing system, FleCS (Flexible Critiquing System), is described in this article. This system has been constructed on personal computer, and provides an interactive critiquing environment. The first target area of FleCS is computer manuals, and we here report on the actual use by manual writers.

1. まえがき

ワードプロセッサの普及・高性能化にしたがって，高品質な文書を作成するツールの一つとして，校正支援システムへの要求が高まってきている。しかし，英文についてはパソコン上でスペルやスタイルをチェックするプログラムが普及しているのに，日本語の校正支援の技術はまだそれほど普及していない。

日本語の自動校正や校正支援については，1980年代の半ばごろから研究が行われてきた[1,2,3]。単純な手法としては，文法解析に失敗したらその付近に間違いがある可能性が高いと推測する方法があげられる。既にワードプロセッサ上位機種の一部には，形態素の接続誤りに対して警告を行う機能が搭載されている[4]。一方，構文・意味解析の技法といった深い解析を大型機を用いて行い，特定目的(主にコンピュータ・マニュアル)用に様々な校正情報を付与するシステムも報告されている[5]。これらはいずれもバッチ方式または逐次方式のユーザ・インターフェイスを持っている。

我々はさらに高品質で実用に耐える校正支援を目指して，パーソナル・コンピュータ上の校正支援システムFleCS[1]を開発した[6,9]。FleCSは形態素解析と，エキスパート・システムの技法を用いて，入力文書から誤字・脱字，不適切な表現，表記の揺らぎ等の誤りを検出し，書き直し方を提案するものである。校正者はエディタやワードプロセッサの中から対話的にこの校正支援を利用できる。現在IBMパーソナル・システム/55上でOS/2のアプリケーションとして稼働している。[2]

2. 校正支援システムの要件

校正作業は大きく誤りの発見と修正とに分けられる。これら2つのステップはほぼ交互に行われ，発見→修正→発見→修正→発見……というサイクルをなす。校正支援システムには，誤りを的確に発見し，その情報を修正時に使いやすい形で示すことが要求される。また，どのように修正すべきかを提示し，ユーザが確認または選択するだけで，修正できることが望ましい。校正支援システムの要件をここに列挙する。

(a).校正機能
日本語では誤った文に対しても可能な文法的解釈が存在することが多いため，単に文法解析に失敗したところを挙げるのみでは誤りの発見率をある程度以上に上げることができない。また「解析失敗」というだけでは修正の役に立つ情報として不十分である。さらにどのようなヒューリスティクスを導入して発見率を上げ，かつ適切な誤り原因および書き直し方を推論できるかが，重要な問題となる。

誤り発見率を上げるのと同様に重要なことは，「過検出率」を下げることである。間違っていないところに警告を出すことを「過検出」という。既存の研究ではあまりそのデータが出ていないが，ユーザはこれに非常に敏感である。

(b).速度
速度において肝心なのは計算処理の実速度ではなく，ユーザを待たせずに校正支援を行うこと，すなわち見かけの速度である。このために，処理途中でも，終わったところまでの部分的な情報が取れるようにしておくことが実速度の向上とともに非常に重要である。

(c).インターフェイス

---

1 FleCS: Flexible Critiquing System の略称。

2 OS/2，パーソナル・システム/55 はIBMの登録商標です

校正情報をもっとも使いやすい形で提供するには，ユーザが通常文書を作成したりチェックしたりしている環境(＝エディタ/ワードプロセッサ)の中で校正支援を使えるようにするのが理想的である。そのためには，校正支援システム自体はエディタ等から独立して働き，様々な文書作成環境から呼び出せるインターフェイスを提供するのがよいであろうと考える。

(d).文書の変化に対応すること
校正は誤りを見つけては書き直していく過程であるから，文章は校正途上で変化していく。校正過程で誤りを入れてしまうこともありうるし，直した時に本当にそれで問題が解消したかが自明ではないこともある。既存のシステムは，このような書き直しにうまく対応していない。すなわちもう一度新しい文書のように処理しなおしている。ユーザの修正を素早く結果に反映させるためには，校正支援システムは一回限りの解析を行うのみならず，文章の変化に適切に追随した解析を行う必要がある。

3. 校正対象の分類

校正の対象となる誤りの分類法には，原因による分類・現象による分類などいろいろあるが，ここでは，その誤りの発見に必要なコンテキストによる分類を述べよう。

カテゴリー1　単語内の誤り：単語にならないタイプミス，使ってはいけない語などがこれに属する。英語など単語間がスペースで区切られている言語では非常に簡単に発見できる種類の誤りであるが，日本語ではまず単語の同定から始めなければならない。ゆえにこの種の誤りを発見するためにも形態素解析は必須となる。

カテゴリー2　付近のコンテキストで誤りとなるもの：単語自体は正しいのだが，近くの単語との関係が正しくない場合である。例えば，助詞が抜けたり，漢字に化けることである。(『編集続ける』『修正野誤り』)。また，隣接単語ではないが1文内で完結している誤りとして，呼応(『決して正しい』)や冗長表現(『例えば～例である』)などの文体上の問題もこのカテゴリーに属する。

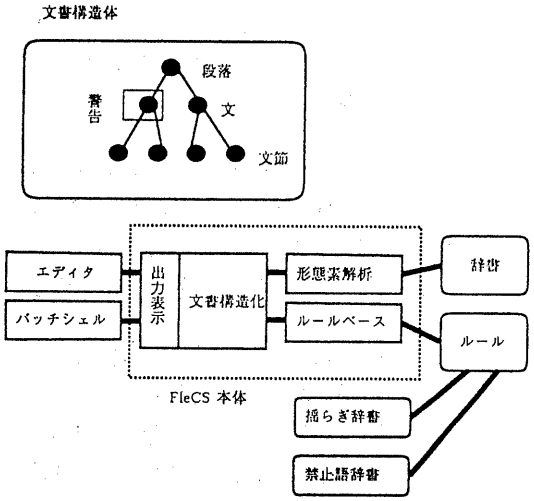
英語等の場合には，構文解析の失敗によって誤りを発見する手法が用いられる。しかし日本語の構文には自由度が高く，多くの場合何らかの解釈が成り立つので，この手法は手間の割に有効な手段とならない。

カテゴリー3　文章全体に係わる誤り：意味的な誤りには文全体の話題を理解しないと判定できないものがある。かな漢字変換の誤りにはこの種のものもある。また，意味処理は必要としないが1文内を見ただけでは発見できないものに，表記の「揺らぎ」がある。例えば，ある個所で『コンテクスト』と書き，別な個所では『コンテキスト』と書いていた場合である。いずれが誤りだと断定される訳でもないが，共存していると統一的でない印象を与える。これらの誤りを発見しようとすれば，文章全体にまたがる情報を保存する機構が必要となる。

日本語の入力は「かな漢字変換」によって行われることが多い。かな漢字変換は入力かな列に対して，もっとも確からしい文法的な漢字列を出力する。このためタイプミスの場合にも，誤った単語(カテゴリー1)として検出されるとは限らない。例えば，「次のとうりに行う」という文は「とうり」が「とおり」の間違いであると警告したい。しかし，通常の形態素解析プログラムでは，

「次-の-と-うり-に-行う」

と解釈されうる。「と」は接続助詞，「うり」は野菜のうりであり，形態素解析の誤りを引き起こさない。この場合構文解析も通ってしまう。したがって，入力が正しい日本語であると想定していた既存の文法解析を行うだけでは，誤り発見能力としては不十分である。的確なメッセージを出すためにヒューリスティクスによる誤りパターンの同定が必要である。



第1図．ＦｌｅＣＳの仕組み。

4. ＦｌｅＣＳの仕組み

ＦｌｅＣＳの構造を第１図に示す。本体は大きく分けて文書構造化，文法解析，ルールベースの３つの部分からなる。さらに，エディタやワードプロセッサ側に，本体との通信や制御を行う部分を組み込む。

(a). 文書構造化部
ＦｌｅＣＳは，エディタのバックグラウンドで働く非同期の別プロセスとして起動され，編集中の文書の解析を進める。ゆえに，ユーザはＦｌｅＣＳの解析処理中も編集・校正を続けられる。文書構造化部の役割は次のようなものである。

1. 対象の文書に対して形態素解析を行ったのち，その結果を段落・文・文節をノードとする木構造として文書構造体を保存する。
2. 文書構造体のノードを順次ルールベースに投入し，警告メッセージなどを文書構造体に付加する。
3. エディタからのコマンドに応じて文書構造体を検索し警告情報等をエディタに知らせる。エディタとのコマンドのやり取りはパイプを通じて行われる。

(b). 文法解析部
どれだけ深い文法解析をするかは，処理量とのトレードオフでもある。ＦｌｅＣＳでは，形態素解析のみを文法解析部で行っている。入力誤り，不適当な文体などは形態素のならびのパターンとして検出できる。前節にも記したように，日本語の場合，誤った文であっても構文解析に失敗することはあまりない。また，ＦｌｅＣＳはルールベース・システムを持っているので，係り受けや意味的な情報が誤り発見に必要な場合にのみ，局所的により深い解析を行うようルールで指示できる。

(c).ルールベース部
誤りを発見し，的確にその理由を推定するには，文法解析だけでなくヒューリスティクスによる誤りパターンの同定が必要である。ＦｌｅＣＳでは誤り発見パターンはルールベース化することによって追加・変更を容易にしている。また，校正の段階やユーザの好みによって使いたいルールが違うことが多い。ここでは，各ルールを使うかどうかを実行時に設定できる仕組みを持つ。

一般に，前向き推論を行うエキスパート・システムは「あるパターンにマッチするものが見つかったらある動作をせよ」という形のルールの集合からなる。ＦｌｅＣＳのルールでは，その条件節は見つけるべきパターン，即ち誤りと推測されるものであり，実行節には書き直し方を推定して，警告を付ける動作が記述されている。ＦｌｅＣＳのルールは，一般の推論システムと異なり，実行節を２つ持つ。一つは条件成立時の動作を記述するものである。もう一つは条件消滅時の動作を記述する。ＦｌｅＣＳではRETEアルゴリズム[８]に消滅時の動作を加えて拡張したものを用いている。

第２図に「揺らぎ」を発見するルールを示す。このルールでは，同一語の別表記があった時に，それぞれに対応する文書構造体のノードに，相手の表記を置き

換え候補として「揺らぎ」の警告を付加することが書かれている。また片方が消滅した(条件が成り立たなくなった)時に，両方のノードから警告を除去することを示している。

5. 再解析メカニズム

前述のようにＦｌｅＣＳはエディタとは非同期に動くプロセスで，解析は起動時の文書に対して行われるのでテキストの修正時にはユーザは陽に再解析を指示する必要がある。（この指示は最初の解析の終了を待たずに行える。）再解析時の文書構造化部は次の作業を行う。

1. 現在の処理を中断する。
2. 修正前の文書の，文書構造体に対応する部分テキストと，新しいテキストとの差分をとり，削除された部分と追加された部分にわける。
3. 削除部分に対して，文書構造体から削除するとともに，削除したノードをルールベースに投入する。ルールベースでは投入されたノードをルールベース内のメモリから削除する。該当ノードがパターンにマッチしていれば，条件消滅時の実行部が起動される。
4. 挿入部分の処理を形態素解析し，文書構造体に付加するとともにルールベースに投入する。該当ノードがパターンにマッチしていれば，条件生成時の実行部が起動される。

ある種のエラーを修正すると修正された部分以外にも影響を及ぼすことがある。「揺らぎ」の場合，片方の表記をすべて修正してしまえば，残りはもはや揺らぎではなくなる。このようなとき単純に考えれば文章全体を解析しなおさないと正しく修正されたかどうか判定できない。ＦｌｅＣＳではルールベース・システムに，修正時の差分に基づくデータだけで全体を解析しなおしたのと同じ結果が得られるような機構を組み込み，修正時の再解析の時間を短縮している。「揺らぎ」の場合では「消滅時には相手のノードにある警告メッセージも除去する」という動作を指定しておくことによって，局所的な解析で全体の整合性を保つことができる。

6. インプリメンテーション

ＦｌｅＣＳはさまざまな分野に適用できるように設計されているが，はじめにマニュアル用校正支援システムとしてのルールやインターフェイスを実装した。現在，ＩＢＭパーソナル・システム／５５上にＯＳ／２のアプリケーションとして実現されている。処理速度は毎秒43文字（ＰＳ／５５７０Ｖモデル）であった。この速度は人が文章を読むよりは早く，また始めの方からＦｌｅＣＳの警告を見ては訂正している時には全く問題は無い。置き換えについては候補を示唆することにとどめている。最終判断はユーザが下すべきだと考えるからである。ＦｌｅＣＳを用いて校正を行っている画面例を第３図に示す。

マニュアル向けＦｌｅＣＳのルールベースは現在約１００個のルールを用意している。それらが検出する誤りの代表的なものを以下に列挙する。

1．入力の誤り
　誤：どんあ時　　正：どんな時
2．かな漢字変換の誤り
　典型的な誤変換と
　幾つかのヒューリスティック・ルール
　誤：人口知能　　正：人工知能
　　２０万を越える　　２０万を超える
　　始めて会った人　　初めて会った人
　　学校二行く　　学校に行く
3．修正の誤り
　誤：　学校行く．　　正：学校に行く
4．禁止語，誤用語．
　盲　　→　　目の不自由な人
　ビルマ　　→　　ミャンマー
5．表記法の揺らぎ
　送り仮名：行う／行なう
　片仮名：チェーン／チェイン

```
rule ″ 揺らぎ ″
/*   条件判断   */
   @p1: PHR [ ]
   @p2: PHR [ @p1.spelling != @p2.spelling, areSameWord(@p1,@p2)]
{ /*   上の条件が成り立つようになった時の動作   */
   makereplace(@p1,@p2.spelling);
   warnON(@p1,″ 表記の揺らぎ ″);
   makereplace(@p2,@p1.spelling);
   warnON(@p2,″ 表記の揺らぎ ″);
}
{ /*   上の条件が成り立たなくなった時の動作   */
   warnOFF(@p1);
   warnOFF(@p2);
};
```

第2図．揺らぎ発見のルール。

中黒：　テキストファイル／テキスト・ファイル

6．典型的な重言

再検討し直す　→　再検討する

7．文体に関するもの

二重否定

「です・ます」体と「だ・である」体

長い文

受動態の多用

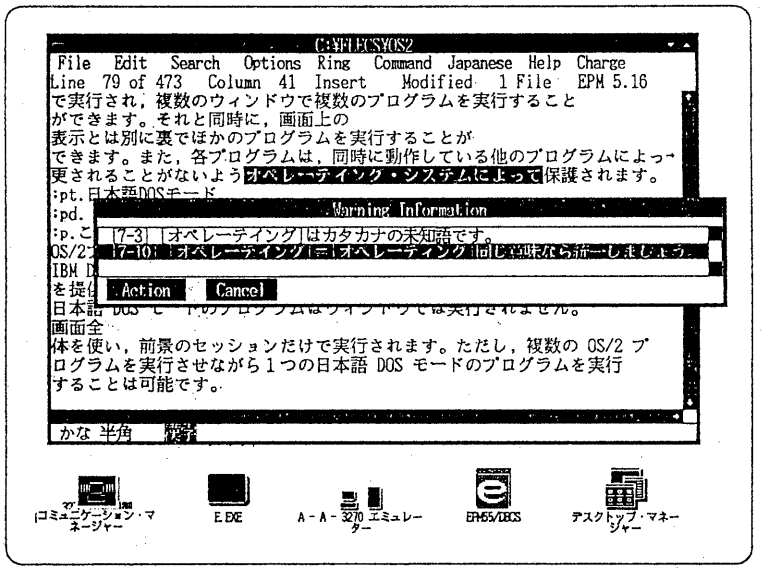
繰り返し：「～の～の～の」，

「～が，～が，」など

これらの中には，一般的なものと，マニュアル向けに作られたものとがある。1～3はどのような文書にも共通して誤りといえる性質のものだが，4～7は筆者の方針や文書の性格によって必ずしも誤りとはならない。このような多様性に対処するためＦｌｅＣＳでは，実行時にルールベース中のルールの中から使いたいもののみに絞ったり，ルールの選択をプロファイルから読み込んだりすることができるようにしている。これによって，過剰なメッセージをおさえることができるとともに，メモリの節約や，高速化にもつながる。

ＦｌｅＣＳは，形態素解析とルールのほかに，誤り発見のために，「禁止語辞書」「揺らぎ辞書」の2つの辞書を持つ。禁止語辞書は，禁止語，誤用語を登録しておき，そこに登録された語をすべて警告の対象とするルールを持っておく。「揺らぎ辞書」は「揺らぎ」の検出のために同一語の異なる表記のペアを登録したもので，辞書中のペアが同一文書に現れた場合に警告するルールを持っておく。ただし，新しい外来語などの表記の揺れをチェックするには辞書登録が間に合わないこともありうるため，カタカナ語については表記の類似度を計算し，類似度の高い語の組が同一文書中に現れた場合に警告するルールを設けている。禁止語辞書，揺らぎ辞書ともユーザがカスタマイズできる。

誤りを発見するルールのいくつかは経験的知識によっている。「はじめて～する」のように「はじめて」の後に動詞が続く場合，副詞の「初めて」である確率が高いので「始めて～する」という並びは誤りの候補とする，かな漢字変換で誤ったかな列を入力して変換した場合無理に漢字に直そうとして一文字の漢字語の列となる場合が多いので一文字語の連鎖は誤変換の候補とする，などである。ほかのルールは，マニュ



第3図 . ＦｌｅＣＳの画面例

アル向けのガイドなど，文章の書き方のガイドによった。

6. 使用報告と評価

ＦｌｅＣＳは１９９１年2月から，ＩＢＭ社内でマニュアルの校正用に使用されている。集中的に使用された一ケ月余りの期間の統計によれば，ＦｌｅＣＳは約10,000ページ分のマニュアルの原稿をチェックしてその警告をもとにして約5,000か所以上が実際に修正された。また，一部を人手による厳密な校正と比較した結果，表現の改善も含めた人手校正個所のうち約4/5についてはＦｌｅＣＳの警告が出されていることがわかった。

使用者のコメントでは，特に役に立った点として未知語，禁止語，送り仮名及び片仮名語の揺らぎ等の比較的単純な規則が挙げられた。反面，文体に関するチェックは，自分の書きかたに自信を持っている場合には煩わしいという意見もあった。

将来的には校正支援システムの使い方は，熟練した書き手には未知語・揺らぎ等機能を絞って，逆に不慣れな書き手には，その分野における文体関連の規則などを備えて十分に支援を行うといった分極化が起こるであろうと推定される。ＦｌｅＣＳではその双方を十分にサポートできると考える。

また，校正の機能のみならず周辺機能が充実していることが，実際の使用者の所感を大きく左右する。今回は以下のようなものが挙げられた。

- 辞書のメンテナンス・ツール。
- 校正規則の出典，詳しい解説，対処法等のマニュアル。
- 大量の文書をバッチで予め解析した後での，効率的な修正法。

7. おわりに

本稿では，日本語校正支援システムとしてほぼ実用段階に入ったＦｌｅＣＳの概要と，マニュアル向けＦｌｅＣＳインプリメンテーションの使用報告を行った。

また，マルチ分野対応のために次は新聞・出版向け校正支援を目的とし，すでに新聞社用縦書きエディター上でＦｌｅＣＳが使えるようになっている。

今後さらにいろいろなニーズが出てくると考えられるが，新しい分野やユーザ・インターフェイスについてもこの柔軟な枠組みで対処できると考える。

謝辞

本研究を遂行するにあたり，ＦｌｅＣＳを使用して多くの有用なコメントを下さった日本アイ・ビー・エム製品情報設計・開発の方々に謝意を表します。また，本研究に常に理解を示し援助下さった大河内正明氏に感謝致します。

参考文献

1. 牛島ほか： 日本語文章推敲支援ツールのプロトタイピング，コンピュータソフトウェア Vol.3-1，pp.35-46，1986.
2. 大山ほか：日本語文書作成支援システムＣＯＭＥＴ，情報処理学会第36回全国大会5U-7，1987.
3. 武田ほか：日本語文書校正支援システムＣＲＩＴＡＣ，情報処理学会第32回全国大会，1986.
4. 日経バイト：次世代ワープロの決め手となるか校正支援/可読性評価ツール，日経バイト 1988年 3月号，pp.96-104.
5. 高橋ほか：計算機マニュアル推敲・査読システム MAPLEの開発と運用，情報処理学会論文誌Vol.31-7，pp.1051-1062，1990.
6. 清水：日本語校正支援システムＦｌｅＣＳの誤り検出処理，情報処理学会第39回全国大会 3J-4，1989.
7. K.Okumura, et al., "Dual-action rule base system for keeping the consistency", in preparation.
8. C.L.Forgy, "Rete: a fast algorithm for the many pattern/many object pattern match problem", *Artificial Intelligence*, Vol.19, pp.17-37, 1982.
9. Y.Tateishi, K.Okumura, S.Shimizu, "FleCS: a flexible critiquing system for Japanese", to appear.