

DTPシステム上の文章作成支援システム(1)

3J-1

-全体構成-

上田隆也 八木沢津義 柵木孝一 川端洋一
 キヤノン(株) 情報システム研究所

1 はじめに

最近デスクトップ・パブリッシング(DTP)が日本でも普及しはじめています。これによって一般ユーザーでも高度な編集機能や高品質の印刷機能を利用することができるようになった。しかし、文書を作成する過程を考えると、「文章を作る」ことに関する機能も充実させる必要がある。いくら仕上がりがきれいでも中身が伴わなければ、人に見せる文書として意味がないからである。我々は、このような観点からDTPシステム上に文章作成支援システムを構築している。

DTPシステム上で最初から文章を作ることを考えると、さまざまな支援機能が必要になってくる。我々は文章作成支援システムの機能について検討し、

- (a) 入力ミスを指摘する。
- (b) 表現に関して、情報を提供したり注意を促したりする。
- (c) 読みやすく、かつ意図が正しく伝わるような文章を書く支援をする。
- (d) アイデアから文章をまとめる手助けをする。

という四種類の機能が必要になると考えた。ここでは(a)(b)(c)を推敲支援機能、(d)を作成支援機能と呼ぶ。現在は推敲支援機能を中心に開発している。

これらを実現するにあたっては、

- ワークステーション環境を利用してインタフェースを工夫することにより操作性の向上を図る。

という点に留意した。

本稿では、まずシステムの全体構成を示し、次に推敲支援機能について、とくに入力ミス検出機能を中心に述べる。

2 システム構成

本システムの構成を図1に示す。この図は現在サポートしている機能を表している。これらの機能の一部は日本語解析の結果を利用している。

推敲支援機能はエディタの中から使うこともできるし、バッチ的に使うこともできる。また、出力されるメッセージはディスプレイ上に表示されるだけでなく、プリントアウトして確認することもできるようにしてある。

なお、全体のシステムはUNIXベースのDTPシステム上に構築されている。

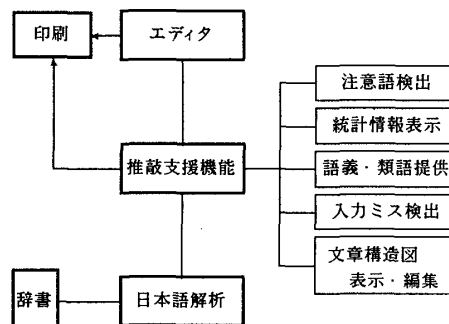


図1: システム構成

3 推敲支援機能

本節では本システムの推敲支援機能について述べる。

1. 注意語検出機能

注意語の辞書を用いて文章中の注意語を指摘する。ここで注意語というのは、

- (a) 用法を誤りやすい同音異義語 (ex. 追究・追及・追求)
- (b) よくある用字の誤り (ex. 先入感→先入観)
- (c) 使用しないほうがよい単語・表記 (ex. 意欲→意欲)

の三種である。これらは、誤りだと気づかずに使ってしまうがちなものである。それぞれの注意語に対して、置き換え候補の語を提示してそれを選択できるようにしてある。

2. 統計情報表示機能

文の長さや各字種の割合などの統計情報をグラフで表示する。これは文章の読みやすさに関する一つの指標である。

3. 語義・類語提供機能

ユーザーが指定した単語について、その語義や類語を表示する。これは文章を作る際、表現を考える上で欠かせない機能である。また、単語の使い分けを調べるのにも役立つ。

4. 入力ミス検出機能

文解析を行って、文章中で入力ミスの可能性がある箇所を指摘する。これについては4節で述べる。

5. 文章構造図表示・編集機能

文の構造を図で表示して係り受けの関係を示す。これを見れば、誤った解釈をされそうな箇所を知ることができる。また構造図の上で簡単に係り受けの変更を行えるようにしてあるので、エディタで文を修正しなくても文の書き換えが容易にできるようになっている[1]。

4 入力ミスの検出

本節では、文章中の入力ミスを検出する機能について述べる。ここで入力ミスとは、ユーザーの意図していた文字列と入力・表示された文字列が異なるものをいい、次の二種類がある。

- タイプミス (誤字・脱字・^{えんじ} 衍字)
- かな漢字変換における誤変換

これらは、修正を要するにもかかわらず文章の作成者がとくに見落としやすい誤りである。

4.1 形態素解析による検出

タイプミスがあると文が文法的におかしくなることが多い。したがって、これは形態素解析によって検出できる。

ex. かすみは本当 の 妹ではない。

誤変換の場合は、その結果できあがった文字列が単語として辞書に存在しないときや、辞書にあっても品詞が異なるときに限って形態素解析で検出できる。

ex. 加奈子には 道場 (→同情) できない。

ここで用いている形態素解析アルゴリズムでは、文中に解析不能箇所があると名詞の未知語として処理しようとする。これができないときは、そこに何らかの誤りがあるとしてスキップする。そこで、形態素解析で未知語とされたところやスキップされたところを、入力ミスの可能性がある箇所として提示する。実際には、誤っている箇所を形態素解析で正確に抽出できるとは限らない。しかし、その場合でも指摘された場所の近くに誤りが存在するので、ユーザーに注意を促すことはできる。

この方法で問題となるのは、正しい単語でも辞書に入っていないと入力ミスとしてしまうということである。これに対処するために、入力ミスだと指摘されたときにユーザーがその単語を辞書登録できる機能を設けている。

4.2 結合価パターンと意味素性による検出

誤変換があってもその結果が文法的に正しければ、意味を考慮しないと誤りがあることがわからない。しかし意味解析はまだ確立した技術ではないので、ここでは結合価文法を利用して、比較的簡単な処理で誤変換の一部を検出している。

まず、各形態素を文節としてまとめて、それらの間の係り受け関係を解析する。次に、用言に係る各名詞について格助詞の置き換え[2]を元に戻し、さらに省略されている格助詞を推定した後に結合価パターンと比較する。このとき意味素性が合わない名詞があったら、名詞か用言かが誤っている可能性があるとして指摘する。

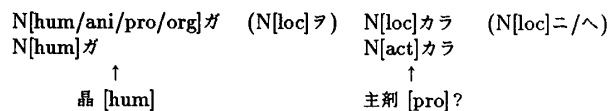


図 2: 「帰る」の結合価パターン

例えば、

ex. 晶が 主剤 (→取材) から帰る。

という文では、名詞「主剤」の意味素性proが「帰る」の結合価パターン(図2)における「カラ格」の意味素性loc/actと一致していないので、誤変換の疑いがあることがわかる。

結合価パターンは、IPAL[3]を参考にして動詞・サ変名詞・形容詞・形容動詞について作成している。また、名詞の意味素性はIPALの20分類に従って割り当てている。

図3に入力ミス検出の実行例を示す。

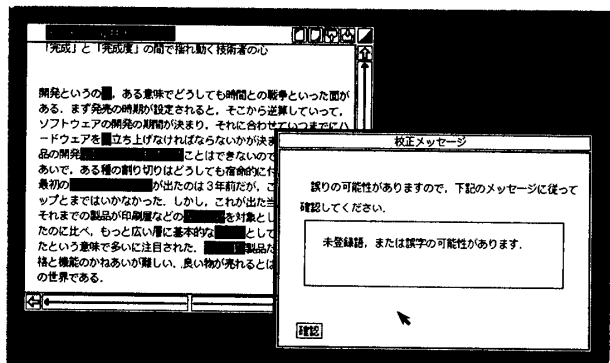


図 3: 入力ミス検出の実行例

5 おわりに

DTPシステム上の文章作成支援システムの全体構成を示し、推敲支援機能、とくに入力ミス検出機能について述べた。現在はこれらの機能の有効性を評価しているところである。今後の検討項目は、

- 入力ミス検出の精度を向上させる。
- DTPシステムとどのように統合するのが望ましいか。
- どのような作成支援機能が役に立つか。

の三点である。

参考文献

- [1] 八木沢他：DTPシステム上の文章作成支援システム(2)－文章構造図に基づいた支援機能－，本大会予稿。
- [2] 木村・空閑：態による格助詞変換，計量国語学，Vol.15，No.2，1985。
- [3] 情報処理振興事業協会：計算機用日本語基本動詞辞書 IPAL (Basic Verbs)－辞書編－，1987。