

日本語文章推敲支援ツール「推敲」の使用について

牛島, 和夫
九州大学工学部情報工学科

日並, 順二
九州大学工学部情報工学科

尹, 志照
九州大学工学部情報工学科

<https://doi.org/10.15017/1468111>

出版情報：九州大学大型計算機センター広報. 18 (1), pp.9-37, 1985-01-25. 九州大学大型計算機センター
バージョン：
権利関係：

日本語文章推敲支援ツール「推敲」の使用について

牛島 和夫* , 日並 順二*¹⁾ , 尹 志熙*

1. はじめに

日本語ワードプロセッサの普及は著しい。九大大型計算機センターでもJEF (Japanese processing Extended Feature) のもとでFDMS和文エディタと専用フォーマッタとを使ってこれを使用できる。ところで日本語ワードプロセッサの主な仕事は文章の入力、書式の設定、出力であって、文章の推敲を積極的に助けるわけではない。英語文章の推敲を助けるツールには、実用に供されているもの (UNIXのライターズワークベンチ^[1])、実験中のもの (IBMのEPISTLE^[2]) などが知られている。「推敲」とは、唐代の詩人が「推す」か「敲く」と語の選定に苦心した故事から発している。推敲支援ツールは文章中にそのような問題点があれば、それを指摘することができればよい。推敲するのは書き手であって計算機ではない。日本語文章にもその推敲を助けるソフトウェアを実現できないのか。

ワードプロセッサでつくられた文書は機械可読な形で蓄えられている。それにもかかわらず清書出力にしか利用されないというのではつまらない。計算機でその中に何とかふみ入って推敲に役立つ情報を取り出してみたい。分かち書きされていない日本語文章に対して英語文章と同じ方法をとったのでは、まず形態素解析のところで苦勞した挙句に、解析結果にあいまいさが生ずる可能性があること、解析時間がかかりすぎる事が予想される^[3]。そこで、辞書を使わず文法解析も行わずに入力文章を字面だけで解析する方法を探る。実用を考えれば、我々が普段書く程度の長さの文章 (例えば学会誌刷上がり8ページ程度) を待ち遠しくない時間内で処理してほしい。

推敲に役立つ情報とは何か、計算機によってどんな情報をとりだせるのか、計算機を用いることによって

* 九州大学工学部情報工学科

本稿は JEF による原稿を写真版にしたものである。

表2.1 入力文章の輪郭

文 章	A	B	C	D
文 字 数	5 0 0 7	1 0 3 6 7	2 3 5 8 3	3 6 6 0 5
刷り上りページ	4	8	18	28
レコード数	83	232	536	1072
文 の 数	130	253	542	1185
英 字 列	142	333	525	1524
漢 字 列	755	1369	3368	4187
ひらがな列	841	1576	3918	4970
カタカナ列	71	225	437	633

2. 1 日本語テキストファイルの解析

STATISTIC 入力文章の輪郭（文字数、文の数、字種毎の文字 [列] 数等の統計）を出力する。

SENTENCE 入力文章から文を切り出し、文頭、文末、文の長さ（ * 印1個20字および + 印1個10字で示す）を横に並べて文の出現順に列挙する。SENTENCEの出力結果の一部を図2.2に示す。ここで文とは句点（。またはピリオド）で終わるもののほか行の途中の改行によって終わるもの（例えば、見出し）も含む。日本語の文末は比重が重い。また、その文末の単調さ（「である」「である」の連続、

NO.	レコード	文頭	文末	文長
1	01003000	解析評価用ツール	-- 解析評価用ツール:	
2	01004000	牛島和夫	-- 牛島和夫:	
3	01005000	九州大学工学部情報工学科	--九州大学工学部情報工学科:+	
4	01006000	1. はじめに	-- 1. はじめに:	
5	01007000	製品としてのソフトウエ	--含む) などが挙げられる.	:*****
6	01008000	要求仕様や設計仕様の記述	--っているとはいいがたい.	:****
7	01009000	ソフトウェアの解析手段	--れた方法とに大別できる.	:*
8	01010000	前者、すなわちインスペク	--であるという見方もある.	:**+
9	01011000	これらはソフトウェア生産	--用することが可能である.	:**
10	01012000	形式言語で記述されたも	--析の対象とはなりえない.	:**+
11	01013000	従って本稿では、ソフトウ	--ールについてのみ述べる.	:**+
12	01014000	2章では事例研究を通じ	--場から主として接近する.	:**+
13	01014000	2章が本稿の中心を成す.	-- 2章が本稿の中心を成す.	:+
14	01015000	3章と4章では静的解析ツ	--がらその考え方を述べる.	:**
15	01016000	2. 事例研究	-- 2. 事例研究:	
16	01017000	あるFORTRANプロ	--とめたものを図1に示す.	:**
17	01018000	プログラムの解析は、入力	--る静的解析に大別される.	:***
18	01019000	図1を見ながらこれらの解	--解析の要点を概観しよう.	:*
19	01020000	NAME欄はサブルーチ	--ァベット順に並べたもの.	:**

図2.2 文章BをSENTENCEで処理した結果の一部

「・・た」「・・た」の連続) はよく指摘される。図2.2はこれを際立てる。出力には日本語ディスプレイ画面1行に1文の情報がおさまるよう工夫している。文の数=出力行数だからA、B程度の文章なら比較的短時間で出力することができる。ディスプレイを見ながら同じ文末が連続するのを発見してもよい。文の長さは長すぎる文への注意をうながす。

R-SENTENCE SENTENCEと同様に文を切り出し、文末から整列させる(逆順ソート)。これによって同じ文末語を用いる文が近くに寄り集まるので、それぞれの文章における文末語の使い方の特徴を知ることができる。このコマンドの実行結果から表2.2に四つの文章の特徴的な文末を挙げる。「である」と「だ」の比率など、それぞれの文章の文体について何らかの指標になっているかもしれぬ。木下^[4]は、同じ文末が続くことを心して避け「である」に代って「だ」の登場する機会をふやすと述べている。表2.2から、四つの文章では「だ」をほとんど使っていないことがわかる。このほか「思われる」「考えられる」といった、使うには要心のいる言葉が使用されているのを容易に指摘することもできる。

KOSO A 11個の指示詞「これ」、「それ」、「あれ」、「この」、「その」、「あの」、「ここ」、「そこ」、「あそこ」、「前者」、「後者」を入力文章中から探し出し、それにアンダラインを引いたものを改めてフォーマットで出力させる。これらの指示詞は文章中で比較的近い範囲の指示関係を示すために使われるものであるが、それらが何を指しているか不明な文章によく行き当たる。アンダラインによりその存在を目立たせて、書き手に指示関係の確認を一つ一つ求める。また、日本語では複数の概念を明確に表わすことが必ずしもできないが、明らかに「これ」よりは「これら」と受けた方が文意が明快になるといった場合を発見することもできる。

STRUCT 入力文章をパラグラフ毎に分解し、その第一文章と最終文章はそのまま、中間の文章は、

表2.2 文末の種類と頻度

種 類	A	B	C	D
である	7	27	46	65
だ	0	1	8	0
た	3	12	23	50
ている	8	7	34	61
ない	7	16	46	48
ならぬ	0	7	7	0
する	18	16	12	64
受身	3	9	14	31
全 体	130	253	542	1185

句読点と括弧を除く全ての文字を「-」文字で置き換えたものをフォーマットで出力させる。見出しや小見出しはそのまま出力する。出力文章の大きさや形はそのままであるが、パラグラフ単位で文章を把握する助けにすることをねらっている。

PASSIVE 入力文章から受動態の候補を探し出し前後の文脈とともに列挙する。文章Bの場合の一部を図2.3に示す。文章Bでは候補が71個抽出された。視察によりその中に60個受動態があった。抽出精度の向上を計画しているが、現状のままでも、読んで受動態か否かを判断し、そのままにしておくか能動態に変えるべきか、との推敲を助けることができる。

RENUMBER 入力ファイルを1文/1レコードに変換する。これはFDMS和文エディタによる後の編集作業を容易にする。JEFの和文エディタで作ったファイルは必ずしも1文/1レコードではなく、また1文が二つ以上のレコードにまたがっていることもあるためである。表2.1で文の数とレコード数が異なっていることに注意。

HEADING, PAREN HEADING は文章中から見出し、小見出しを抜き出して列挙する。PARENは文章中の括弧の対応を調べる。

2. 2 字種に基づく文字列単位の処理

英語文章中に現われる単語をABC順に整列させるだけで、辞書と照合しなくても、綴りの誤りを容易に発見できる。日本語文章では単語の切り出しには辞書を何回も引きにゆかねばならぬので、英語と同じ方法

NO. 1	用データを含む) などが挙げられる	要求仕様や設計仕様の記述	01007001
NO. 2	用いて行う提案がいくつかな	されているもの、現状ではそれら	01008000
NO. 3	フトウェア生産の場で広く使	われているようになっているとはい	01008000
NO. 4	い手動による方法と、自動化	された方法とに大別できる。前者、	01009000
NO. 5	能である。形式言語で記述	されたもの以外は、自動解析の対象	01012000
NO. 66	ある。拡張された論理式に現	われる変数の値が変化する度に動的	01167000
NO. 67	題点をまとめる。(1) 挿入	されたコードによるソースコードの	01169000
NO. 68	て、どんな解析ツールが求め	られるのか、それは言語がその機能	01181000
NO. 69	、実際にどんなツールが用意	されているのか、そのツールをどう	01181000
NO. 70	用した。このプログラムに現	われな	い事項についてはほとんど触
NO. 71	ない事項についてはほとんど	触れな	かった。本特集の他の解説に

受け身の候補 = 71

図2.3 文章BをPASSIVEで処理した結果の一部

研究開発

をここではとれない。

COUNT, LENGTH 同じ字種の文字列の長さがあまりに長いと、読み易さをそこなうおそれがある。COUNTは字種を指定して文字列を切り出し、文字列長に関する頻度表を作る。文章B中のひらがな列に対して実行した結果を図2.4に示す。これの結果を見ながら、指定した長さの文字列を、前後の文脈とともに出力させるのがLENGTHである。文章B中のひらがな列に対して長さ13以上を指定して得られた出力を図2.5に示す。

種類：ひらがな列		
長さ	キーワード数	割合 (%)
1	806	51.14
2	248	15.73
3	195	12.37
4	138	8.75
5	63	3.99
6	44	2.79
7	20	1.26
8	30	1.90
9	7	0.44
10	10	0.63
11	3	0.19
12	3	0.19
13	3	0.19
14	3	0.19
16	1	0.06
18	2	0.12
合計	1576	100.00

図2.4 文章Bのひらがな列の長さ (COUNT)

種類：ひらがな列

キーワード長：13文字以上

NO.	1	を用いて行う提案：がいくつかなされてい-るものの、現状で	01008000
NO.	2	生産の場で広く使：われるようになってい-るとはいいがたい	01008000
NO.	3	またはN>102：となるようなことがな-かったためであり	01035000
NO.	4	る仮引数が入力用：であることがわかって-いるので、M、N	01057000
NO.	5	に登録された手続：きなどはみなこのよう-になっている。使	01133000
NO.	6	的解析ツールに教：えてやることができれ-ばよい。 解析	01134000
NO.	7	大な警告を読み落：としてしまうこともあ-りうる。そこで筆	01135000
NO.	8	外から会話的に与：えることができるよう-にした [9]。教	01139000
NO.	9	挿入されるコード：はやはりかさばったも-のになる。さらに	01166000

図2.5 文章B中の長さ13以上のひらがな列 (LENGTH)

これから、長さ18のひらがな列は

- ・・・われるようになっているとはいいいがたい
- ・・・となるようなことがなかったためであり

とわかる。このままでよいかどうか推敲の手掛りとする。

XREF 字種別に作成する。文字列を見出しとし、それが現われたレコード番号を列挙する。

KWIC 字種を指定して文字列を切り出し、KWICリストを作成する。事後の処理のため、このKWICリストをファイルに保存することもできる。英字列、カタカナ列は漢字列、ひらがな列に比べて出現数が少ない(表2.1参照)ので、そのまま出力して全数検査を行い、綴りの誤りを正すのに用いてもよい。それぞれABC順、五十音順に並んでいる。カタカナ列は表記のゆれのため微妙に違う文字列を書いてしまうことがよくある(例エディタ、エディターなど)^[5]。これらが容易に発見できる。日本文の中に現われる英字列の綴りも、もう一度誤りがないか確認する。ABC順に並んでいることがこのことを容易にする。

R-KWIC 切り出した文字列を末尾文字から整列させ同様のKWICリストを作成する。科学技術文書ではひらがなから非ひらがなに移るところに文節の切れ目があることが多いという調査がある^[6]。したがってR-KWICはひらがな列に対して特に有効である。

INDEX 漢字列やひらがな列は非常に数が多くなるので(表2.1参照)全数出力は必ずしも実際的ではない。必要な情報を取り出すためにいくつかの手掛かりが必要である。INDEXは項目(字種を指定して切り出した文字列)とその頻度を並べて表示する。図2.6にINDEX(漢字列)の一部を示す。例えば、漢字列「行」は、「行<ぎょう>」「行く」「行(な)う」に亘る。INDEXの結果から「行」の文脈だけを全て出力してこれらを確認する(図2.7)。さらに送り仮名が不統一となっていないか調べることができる。人間の不注意もあるが同一文章を何度も中断しながらワードプロセッサで入力するとその時の状態で送り仮

NO. 296	呼	(9)	NO. 330	構成物	(1)
NO. 305	呼出	(10)	NO. 331	構成要素	(2)
NO. 315	後処理	(1)	NO. 333	構造	(5)
NO. 316	誤	(8)	NO. 338	構築	(1)
NO. 324	効果的	(1)	NO. 339	考	(2)
NO. 325	向	(1)	NO. 341	考慮	(1)
NO. 326	広	(1)	NO. 342	行	(16)
NO. 327	拘	(1)	NO. 358	行以下	(1)
NO. 328	更	(1)	NO. 359	行数	(2)
NO. 329	構成	(1)			

図2.6 文章B中のINDEX(漢字列)の結果の一部

研究開発

名がゆれても気付かないことが多い。図2.8から「呼び出す」「呼出す」と二様に使われていることを発見できる。

このほかにも、複数人で分担した文章について、それらを併わせてKWICリストを作ると、同じ概念に別の漢字列を当てているのを発見したり、カタカナ表記のゆれを発見したりといった効用を経験している。

NO. 342	形式言語を用いて:行	:う提案がいくつか	01008000
NO. 343	を実際に動かして:行	:う動的解析と, ソ	01018000
NO. 344	N手続きが260:行	:, TOKU手続き	01026000
NO. 345	下である. 260:行	:というようなかさ	01026000
NO. 346	U手続きが127:行	:. 他の手続きは6	01026000
NO. 347	ムは全体で661:行	:. うちHOKAN	01026000
NO. 348	トケースを選んで:行	:うテストを構造テ	01046000
NO. 349	-1)とテストを:行	:った. これにより	01061000
NO. 350	ればデバッグを:行	:わねばならぬ. 我	01073000
NO. 351	解析し意味付けを:行	:って目的プログラ	01105000
NO. 352	このような検査を:行	:えないわけではな	01113000
NO. 353	える作業は人間が:行	:うのだからその際	01140000
NO. 354	に基づいて解析を:行	:ったのではかえっ	01141000
NO. 355	スコードに修正が:行	:われてしかもイン	01144000
NO. 356	管理はどのように:行	:えばよいのかとい	01144000
NO. 357	めの細かい計測を:行	:うことが可能にな	01152000
NO. 358	他の手続きは60:行以下	:である. 260行	01026000
NO. 359	S欄は各手続きの:行数	:. EXECUTA	01021000
NO. 360	. ソースコードの:行数	: (表1 LINES	01073002

図2.7 文章BのKWIC (漢字列) の結果の一部 (その1)

NO. 296	数 (以下手続きと:呼	:ぶ) の名前をアル	01020000
NO. 297	とめて分岐述語と:呼	:ぶ. 表1参照) に	01039000
NO. 298	トを構造テストと:呼	:ぶ. 構造テストの	01046000
NO. 299	. MとPは手続き:呼	:び出しCALL	01057000
NO. 300	きはまた3万回も:呼	:び出されている.	01086000
NO. 301	れている. 1回の:呼	:び出しに費やされ	01087000
NO. 302	からこの手続きを:呼	:び出しているかも	01089000
NO. 303	位 (以下手続きと:呼	:ぶ) がコンパイラ	01107000
NO. 304	実行時ルーチンの:呼	:び出しでもよい.	01151000
NO. 305	ED欄は手続きが:呼出	:された回数. EX	01023000
NO. 306	される時間をc. :呼出	:し回数をnとすれ	01087000
NO. 307	. 特定の手続きを:呼出	:している手続きと	01090000
NO. 308	い. ある手続きが:呼出	:している手続きは	01090000
NO. 309	いる手続きとその:呼出	:し場所を知るには	01090000
NO. 310	出し関係を調べて:呼出	:しグラフを作成す	01119000
NO. 311	は. まず手続きの:呼出	:し関係を調べて呼	01119000
NO. 312	続き)からそれを:呼出	:す手続きへとデー	01120000
NO. 313	き (他の手続きを:呼出	:していない手続き	01120000
NO. 314	(A, B, C)と:呼出	:せば, 手続きPの	01130000

図2.8 文章BのKWIC (漢字列) の結果の一部 (その2)

図2.9に文章Bのカタカナ列に対するINDEXの結果を示す。表記の揺れ「インターフェース」「インタフェース」、綴りの誤り「プログラミング」が見出される。

なお、漢字列KWICリストの順序はJISの漢字コードの順であって読みの順ではない。読みの順にするにはまず辞書を引きにゆく必要がある。しかし漢字コードの順でもそれなりに役に立つ。

「情報システム」、「計算機センター」などのように漢字列とカタカナ列が結びついて一つの複合語を形成することが多い。したがって、漢字とカタカナの混在した文字列を漢字列の代わりに解析の対象とする方法もありえよう。なお、本文中で使われている「漢字列」とは、JISの第一水準と第二水準に含まれるものだけを指し、拡張漢字や特殊文字（ユーザ定義文字を含む）は含まない。

FREQUENCY 文字種を指定して切り出した文字列を出現頻度の順に出力する。

2. 3 その他のコマンド

「推敲」には、これらの他にFACOM OS IV/F4のTSS環境を利用するためのコマンドをいくつか用意している（図2.1その他の欄を参照）。

NO. 1	アルファベット	(1)	NO. 96	ツール	(30)
NO. 2	インスペクション	(1)	NO. 126	テキストエディタ	(1)
NO. 3	インターフェース	(1)	NO. 127	テスト	(22)
NO. 4	インタフェース	(2)	NO. 149	テストケース	(1)
NO. 6	ウォークスルー	(1)	NO. 150	テストデータ	(1)
NO. 7	エラー	(3)	NO. 151	データ	(16)
NO. 10	オプション	(1)	NO. 167	データフロー	(2)
NO. 11	カウンタ	(7)	NO. 169	デバッグ	(1)
NO. 18	ギャップ	(1)	NO. 170	デモンストレーション	(1)
NO. 19	グラフ	(7)	NO. 171	ハードウェア	(1)
NO. 26	ケース	(1)	NO. 172	バイト	(2)
NO. 27	コード	(9)	NO. 174	バウンド	(1)
NO. 36	コンパイラ	(8)	NO. 175	プリプロセッサ	(3)
NO. 44	コンパイル	(4)	NO. 178	プログラム	(30)
NO. 48	サブルーチン	(1)	NO. 208	プログラムグラフ	(2)
NO. 49	サポート	(1)	NO. 210	プログラムデータペー	(1)
NO. 50	ステップ	(1)	NO. 211	プログラムミング	(1)
NO. 51	センタ	(1)	NO. 212	メカニズム	(1)
NO. 52	センター	(1)	NO. 213	ユーザ	(1)
NO. 53	ソースコード	(29)	NO. 214	ライブラリ	(3)
NO. 82	ソースプログラム	(1)	NO. 217	ラベル	(1)
NO. 83	ソフトウェア	(8)	NO. 218	ルーチン	(1)
NO. 91	ソフトウェア・サイエ	(1)	NO. 219	ループ	(7)
NO. 92	チェック	(4)			

図2.9 文章BのINDEX (カタカナ列) の結果

3. 使用法

3. 1 起動と終了

FACOM 6650ディスプレイ端末（通称日本語端末）でセッションを開き、コマンドSUIKOUを入力する。

READY

SUIKOU

*** 「推敲」 開始 MM/DD/YY HH:MM:SS ***

処理対象日本語テキストファイルのデータセット名: XXXX.XXXX

ここで処理の対象となる日本語テキストの入ったデータセット名を入力すると、次にプロンプト「推敲:」が出力されるので、以後、各コマンドを入力していく。終了するにはコマンドEXITを入力する。

推敲: EXIT

*** 「推敲」 終了 MM/DD/YY HH:MM:SS ***

READY

3. 2 システムのモード

「推敲」には、3つのモードがある。

(i) 推敲モード（プロンプト「推敲:」を出力）

→ 「推敲」の各コマンドを実行する。

(ii) TSSモード（プロンプト「TSS:」を出力）

→ 任意のTSSコマンドを実行する。

(iii) HELPモード（プロンプト「HELP:」を出力、フル・スクリーン形式）

→ 「推敲」の各コマンドの説明を行う。

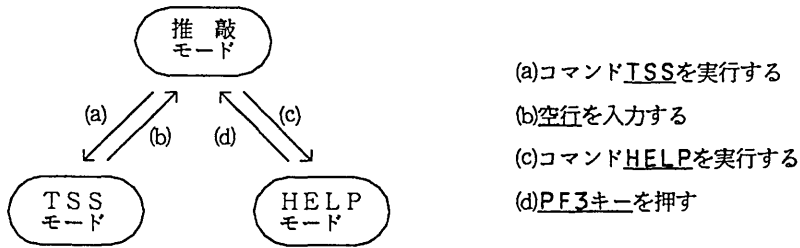


図3.1 システムの状態遷移図

- (a)コマンドTSSを実行する
- (b)空行を入力する
- (c)コマンドHELPを実行する
- (d)PF3キーを押す

《注意》推敲モードで実行するコマンドの多くは、出力結果が多量となるため、一画面に収まらない。従って結果をディスプレイに出力する際には、サブコマンドを用いる。

ディスプレイの最下行にプロンプト「==>」が出力されるので、そこに各サブコマンドを入力する。サブコマンドは実行中のコマンドによって数が異なる。そのコマンド中でどのようなサブコマンドがあるかを調べるにはH(ELP)を入力する。

表3.1 サブコマンド一覧

T	結果の先頭から表示する。
B	結果の末尾を表示する。
P	一画面分前の部分を表示する。
空行	次の画面を表示する。
+ (-) 数字	数字で指定した行だけ後 (前) の行から表示する。
I *	索引の先頭から表示する。
F 数字*	数字で指定したNO. の行から表示する。
S	現在表示されている部分をファイル**に保存する。
E	コマンドを終了する。

* コマンドKWIC、R-KWIC、INDEXでのみ有効

** データセット名は「SUIKOU.SAVE.RES」。尚、以後「ファイル」と「データセット」は区別しない。

3. 3 コマンドの具体的説明

図2.1のHELPコマンドのメニューを見てもわかるように、「推敲」のコマンドは(1)日本語テキストファイルを処理するコマンドと(2)その他のコマンドの2つに分かれている。

※ 以下の説明は、HELPコマンドを実行することによってディスプレイに表示できる。尚、付録Aにコマンド一覧(以下の説明順に並べたもの)を、付録Bにコマンド一覧(アルファベット順に並べたもの)を載せている。

3. 3. 1 日本語テキストファイルを処理するコマンド

◆ STATISTIC コマンド ◆

(1) 入力形式

<u>S</u> TATISTIC	なし
-------------------	----

(2) 機能

・入力文章の輪郭(文字数、レコード数、文の数、文の平均長、各字種毎の文字[列]数等の統計情報)を出力する。

(3) オペランドの説明

(4) 注意

・このコマンドで表示する文の数の中には、表の部分(@TB=...@TE@)は含まれない。

◆ SENTENCE コマンド ◆

(1) 入力形式

<u>S</u> ENTENCE	<u>N</u> ODISP	<u>L</u> IST
------------------	----------------	--------------

(2) 機能

・日本語テキストファイルから文を抽出し文頭、文末(各12文字)と文長(*:20字,+:10字)を横に並べて文の出現順に表示する。但し、文章中の表の部分(@TB=...@TE@)は解析の対象としない。

(3) オペランドの説明

a. NODISP

このオペランドを指定すると結果をディスプレイに表示しない。

b. LIST

このオペランドを指定すると結果をプリンタ（あらかじめNPRコマンドで指定しておく）に出力する。

◆ R - SENTENCE コマンド ◆

(1) 入力形式

<u>R</u> - <u>S</u> ENTENCE	<u>N</u> ODISP	<u>L</u> IST
-----------------------------	----------------	--------------

(2) 機能

- 日本語テキストファイルから文を切り出し文末からソートする。但し、文章中の表の部分 (@TB=...@TE@) は解析の対象としない。表示形式はコマンドSENTENCEと同様である。

(3) オペランドの説明

a. NODISP

このオペランドを指定すると結果をディスプレイに表示しない。

b. LIST

このオペランドを指定すると結果をプリンタ（あらかじめNPRコマンドで指定しておく）に出力する。

◆ P A S S I V E コマンド ◆

(1) 入力形式

<u>P</u> ASSIVE	<u>N</u> ODISP	<u>L</u> IST
-----------------	----------------	--------------

(2) 機能

- 日本語テキストファイルから受動態を捜す。但し、文章中の表の部分 (@TB=...@TE@) は解析の対象としない。

(3) オペランドの説明

a. NODISP

このオペランドを指定すると結果をディスプレイに表示しない。

b. LIST

このオペランドを指定すると結果をプリンタ（あらかじめNPRコマンドで指定しておく）に出力する。

◆ K O S O A コマンド ◆

(1) 入力形式

<u>K</u> O <u>S</u> O <u>A</u>	SY (O K S U)
--------------------------------	--------------------

(2) 機能

- ・日本語テキストファイル中の指示詞（これ、それ、あれ、この、その、あの、ここ、そこ、あそこ、前者、後者）に下線を引いて日本語ラインプリンタにフォーマット出力する。

(3) オペランドの説明

a. SY (O | K | S | U)

出力クラスを指定する。省略するとSY (O) を指定したとみなす。

◆ S T R U C T コマンド ◆

(1) 入力形式

<u>S</u> T <u>R</u> U <u>C</u> T	SY (O K S U)
----------------------------------	--------------------

(2) 機能

- ・日本語テキストファイルを段落の最初の文と最後の文はそのまま、その他の文は各文字を「-」で置き換えて（特殊文字はそのまま）日本語ラインプリンタにフォーマット出力する。

(3) オペランドの説明

a. SY (O | K | S | U)

出力クラスを指定する。省略するとSY (O) を指定したとみなす。

◆ H E A D I N G コマンド ◆

(1) 入力形式

<u>H</u> E <u>A</u> D <u>I</u> N <u>G</u>	<u>N</u> O <u>D</u> I <u>S</u> P <u>L</u> I <u>S</u> T
---	--

(2) 機能

- ・日本語テキストファイル中から見出しとなる文を抽出する。

(3) オペランドの説明

a. NODISP

このオペランドを指定すると結果をディスプレイに表示しない。

b. LIST

このオペランドを指定すると結果をプリンタ（あらかじめNPRコマンドで指定しておく）に出力する。

◆ PAREN コマンド ◆

(1) 入力形式

<u>P</u> AREN	<u>N</u> ODISP <u>L</u> IST
---------------	-----------------------------

(2) 機能

- ・日本語テキストファイル中の括弧の対応を調べる。対象となるのは () [] { } < > ◊ ◊ 「 」 『 』 “ ” ‘ ’ の 26 個である。

(3) オペランドの説明

a. NODISP

このオペランドを指定すると結果をディスプレイに表示しない。

b. LIST

このオペランドを指定すると結果をプリンタ（あらかじめNPRコマンドで指定しておく）に出力する。

◆ RENUMBER コマンド ◆

(1) 入力形式

<u>R</u> ENUMBER	OUT (日本語テキストファイル名)
------------------	--------------------

(2) 機能

- ・FDMSで作成した日本語テキストファイルを1文/1レコードに変換して出力する。

(3) オペランドの説明

a. OUT (日本語テキストファイル名)

変換後の日本語テキストファイルのデータセット名を指定する。

このオペランドを省略すると「RENUMBER.JEF」という名前のデータセットに出力する。

◆ K W I C コマンド ◆

(1) 入力形式

<u>KWIC</u>	字種	OUT (データセット名)	<u>NODISP</u>	<u>LIST</u>
-------------	----	---------------	---------------	-------------

(2) 機能

- ・日本語テキストファイルから、指定した字種のKWICリストを作成する。

(3) オペランドの説明

a. 字種

KWICを行う際の字種 (A : 英字, C : 漢字, H : ひらがな, K : カタカナ, N : 数字のいずれか) を指定する。

b. OUT (データセット名)

KWICリストを保存したい時に、このオペランドにデータセット名を指定する。

c. NODISP

このオペランドを指定すると結果をディスプレイに表示しない。

d. LIST

このオペランドを指定すると結果をプリンタ (あらかじめNPRコマンドで指定しておく) に出力する。

◆ R - K W I C コマンド ◆

(1) 入力形式

<u>R-KWIC</u>	字種	<u>NODISP</u>	<u>LIST</u>
---------------	----	---------------	-------------

(2) 機能

- ・日本語テキストファイルから、指定した字種の逆KWICリスト (キーワードを末尾順からソートしたKWICリスト) を作成する。

(3) オペランドの説明

a. 字種

逆KWICを行う際の字種 (A : 英字, C : 漢字, H : ひらがな, K : カタカナ, N : 数字のいずれか) を指定する。

b. NODISP

このオペランドを指定すると結果をディスプレイに表示しない。

c. LIST

このオペランドを指定すると結果をプリンタ（あらかじめNPRコマンドで指定しておく）に出力する。

◆ COUNT コマンド ◆

(1) 入力形式

<u>COUNT</u>	字種
--------------	----

(2) 機能

- 日本語テキストファイルから、指定した字種のキーワードの長さ、数、割合の一覧表を作成する。

(3) オペランドの説明

a. 字種

一覧表を作る際の字種（A：英字，C：漢字，H：ひらがな，K：カタカナ，N：数字のいずれか）を指定する。

◆ LENGTH コマンド ◆

(1) 入力形式

<u>LENGTH</u>	字種	SIZE (i1-i2)	<u>LIST</u>
---------------	----	--------------	-------------

(2) 機能

- 日本語テキストファイルから、指定した字種における、指定した範囲の長さの文字列を表示する。

(3) オペランドの説明

a. 字種

字種（A：英字，C：漢字，H：ひらがな，K：カタカナ，N：数字のいずれか）を指定する。

b. SIZE (i1-i2)

i1-i2には表示したい文字列の長さの範囲を整数で指定する。省略すると長さ10文字以上を出力する。

c. LIST

このオペランドを指定すると結果をプリンタ（あらかじめNPRコマンドで指定しておく）に出力する。

◆ X R E F コマンド ◆

(1) 入力形式

<u>XREF</u>	字種	<u>NODISP</u>	<u>LIST</u>
-------------	----	---------------	-------------

(2) 機能

・日本語テキストファイル中の指定した字種の文字列とレコード番号とのクロスリファレンスを作成する。

(3) オペランドの説明

a. 字種

字種 (A : 英字, C : 漢字, H : ひらがな, K : カタカナ, N : 数字のいずれか) を指定する。

b. NODISP

このオペランドを指定すると結果をディスプレイに表示しない。

c. LIST

このオペランドを指定すると結果をプリンタ (あらかじめNPRコマンドで指定しておく) に出力する。

◆ I N D E X コマンド ◆

(1) 入力形式

<u>INDEX</u>	字種	<u>NODISP</u>	<u>LIST</u>
--------------	----	---------------	-------------

(2) 機能

・日本語テキストファイルから、指定した字種の「文字列-頻度一覧 (KWICファイルの索引部分に相当)」を作成する。

(3) オペランドの説明

a. 字種

「文字列-頻度一覧」を作る際の字種 (A : 英字, C : 漢字, H : ひらがな, K : カタカナ, N : 数字のいずれか) を指定する。

b. NODISP

このオペランドを指定すると結果をディスプレイに表示しない。

c. LIST

このオペランドを指定すると結果をプリンタ (あらかじめNPRコマンドで指定しておく) に出力する。

◆ F R E Q U E N C Y コマンド ◆

(1) 入力形式

<u>F</u> R <u>E</u> Q <u>U</u> E <u>N</u> C <u>Y</u>	字種	<u>N</u> O <u>D</u> I <u>S</u> P	<u>L</u> I <u>S</u> T
--	----	----------------------------------	-----------------------

(2) 機能

- ・日本語テキストファイルから切り出した、指定した字種の文字列を、頻度順に出力する。

(3) オペランドの説明

a. 字種

頻度表を作る際の字種 (A : 英字, C : 漢字, H : ひらがな, K : カタカナ, N : 数字のいずれか) を指定する。

b. N O D I S P

このオペランドを指定すると結果をディスプレイに表示しない。

c. L I S T

このオペランドを指定すると結果をプリンタ (あらかじめNPRコマンドで指定しておく) に出力する。

3. 3. 2 その他のコマンド

◆ H E L P コマンド ◆

(1) 入力形式

<u>H</u> E <u>L</u> P	[コマンド名]
-----------------------	---------

(2) 機能

- ・「推敲」の各コマンドの説明を行う。

(3) オペランドの説明

a. コマンド名

内容を知りたいコマンド名を指定する。省略するとコマンドのメニューを表示する。

(4) 注意

- ・このコマンドを実行するとシステムのモードが「推敲モード」から「HELPモード」に変わる。「推敲モード」に戻るには、PF3キーを押す。

◆ D S N A M E コマンド ◆

(1) 入力形式

<u>DSNAME</u>	日本語テキストファイル名
---------------	--------------

(2) 機能

- ・ 以後のコマンドの処理対象となる日本語テキストファイルのデータセット名の設定/変更を行う。

(3) オペランドの説明

a. 日本語テキストファイル名

処理したい日本語テキストファイルのデータセット名を指定する。

◆ M A C R O コマンド ◆

(1) 入力形式

<u>MACRO</u>	マクロ・コマンド名
--------------	-----------

(2) 機能

- ・ マクロ・コマンドを定義する。

(3) オペランドの説明

a. マクロ・コマンド名

新しく定義したい (または内容を参照/変更したい) マクロ・コマンド名を指定する。

(4) 注意

- ・ マクロ・コマンドは、「SUIKOU.MACRO」という名前の区分データセットにメンバとして作成する (フル・スクリーンエディタを起動する) ので、マクロ・コマンド名の指定は8文字以内。また、作成したマクロ・コマンドを実行するには、マクロ・コマンドの先頭に「/」(スラッシュ)を付加すること。消去する時にはTSSモード (または「推敲」を終了してREADY状態) で

DELETE SUIKOU.MACRO(マクロ・コマンド名)

のように行う。

◆ M L I S T コマンド ◆

(1) 入力形式

<u>M L I S T</u>	なし
------------------	----

(2) 機能

・登録されているマクロ・コマンド名の一覧を表示する。

(3) オペランドの説明

(4) 注意

・このコマンドは登録されているマクロ・コマンド名の一覧を表示するだけなので、その内容が見たければ MACRO コマンドを使用すること。

◆ E X I T コマンド ◆

(1) 入力形式

<u>E X I T</u>	なし
----------------	----

(2) 機能

・SUIKOU コマンドを終了する。

◆ N P R コマンド ◆

(1) 入力形式

<u>N P R</u>	出力先
--------------	-----

(2) 機能

・各コマンドの L I S T オペランドを指定した時に出力するプリンタ名の設定/変更を行う。

(3) オペランドの説明

a. 出力先

出力したいプリンタ名 (「SY (・)」, 「I・・・」, 「M・・・」のいずれか) を指定する。

(4) 注意

・出力先を省略した場合、現在設定されているプリンタ名を表示したあと出力先の入力を促す。

初期値は SY (0) となっている。

◆ N E D I T コ マ ン ド ◆

(1) 入力形式

<u>NEDIT</u>	[EDIT FORMAT] [DICTION (辞書データ セット名)] [MACLIB (マクロライブラリ名)]
--------------	--

(2) 機能

- ・FDMS和文エディタを起動する。

(3) オペランドの説明

a. EDIT|FORMAT

NEDITセッション開始時にEDITモードとするか、FORMATモードとするかを指定する。省略するとEDITを指定したと見なす。

b. DICTION (辞書データセット名)

使用するカナ漢字変換辞書のデータセット名を指定する。省略するとマスタ辞書を指定したと見なす。

c. MACLIB (マクロライブラリ名)

マクロライブラリを使用するときマクロライブラリ名を指定する。

◆ T S S コ マ ン ド ◆

(1) 入力形式

<u>TSS</u>	なし
------------	----

(2) 機能

- ・TSSコマンド実行モードに移る。(プロンプト「TSS:」が出力されるので、これに続けて実行したいTSSコマンドを入力する。)

(3) オペランドの説明

(4) 注意

- ・「TSSモード」から「推敲モード」に戻るには、空行を入力すればよい。

◆LIST コマンド◆

(1) 入力形式

LIST	コマンド名	[字種]	出力先
------	-------	------	-----

(2) 機能

・各コマンドの結果をプリントアウトする。

(3) オペランドの説明

a. コマンド名

結果をプリントアウトしたいコマンド名を指定する。

b. 字種

a. のコマンド名がKWIC、R-KWIC、COUNT、LENGTH、XREF、INDEX、FREQUENCYの場合に字種（A：英字，C：漢字，H：ひらがな，K：カタカナ，N：数字のいずれか）を指定する。

c. 出力先

出力先のプリンタ名（「SY（・）」，「I・・・」，「M・・・」のいずれか）を指定する。

3. 4 備 考

各コマンド（その他に含まれるコマンドを除く）は、実行時に結果をデータセットに保存する為、データセットの数が増えるが、これらのデータセットは翌日には消えるようになっている。

※ ここで述べたコマンド・オペランド等はまた暫定的なものであり、今後もさらにコマンドの追加・変更を行う予定である。その際には利用者の意見も参考にしたいと思っている。このためにコマンドMESSAGEを追加した。

◆MESSAGE コマンド◆

MESSAGE	なし
---------	----

・プロンプト「推敲：」に続けてMESSAGEと入力するとプロンプト「MSG：」が出力される。この後に使ってみて気づいたことや、改善した方がよい所、あった方がよい機能などのコメントを入力する。一行に収まらなければ適当な所でENTERキーを押せば、再び「MSG：」が出力される。推敲モードに戻る

研究開発

には空行を入力する。

※ 追加（または変更）したコマンドについては、時々HELPコマンドを実行して確認しておくとい。

4. 結 び

このツールのプログラムの最初の版は、我々の研究室で作った日本語 SNOBOL 4^[7, 8] で書いた。日本語 SNOBOL 4 処理系は解釈実行型なので処理速度は本来非常に遅い。しかし FACOM M382 を利用しているかぎり、例に用いた程度の大きさの文章についての反応速度は、そのまま満足できるものもかなりある。仕様の固まったものから少しずつ FORTRAN 77 で書き直して処理速度の改善を行っている。表 4.1 にコマンドの処理時間（CPU 時間）をいくつか示す。

しかし、この推敲支援ツールは、未だ試作段階で、これからさらにいろいろな機能を追加して試して見る予定である。例えば辞書^[9]の使用も試みつつあるが、処理時間の増大が気に掛かる。

文章を作るにはまず（1）何を書くかを定め、次に（2）構想を固めて、（3）書きおろす。それを（4）読みかえて、（5）書き改める、といった段階を経ることになる。ここで述べたツールは（4）と（5）の過程の一部を援助するにすぎない。はじめに述べたように、推敲を行うのは人間であることをもう一度指摘して結びとする。

表 4.1 各コマンドの処理時間（単位 mS）

昭和 59 年 10 月 18 日現在

文章（文字数）	SENT ENCE*	R-SENT ENCE**	PASS IVE**	COUNT ひらがな**	KWIC 漢字列*	INDEX 漢字列*
A (5007)	87	351	518	368	193	55
B (10367)	150	553	1326	651	294	67
C (23583)	298	1291	3980	1762	655	118
D (36605)	536	3462	7461	2222	770	128

* FORTRAN77

** 日本語 SNOBOL4

参考文献

- [1] Cherry, L., Writing Tools, IEEE Trans. on COM, Vol. COM-30, NO. 1, pp. 100-104, 1982
- [2] Heidorn, G. E., et al., The EPISTLE text-critiqu-

- ing system, IBM System J., Vol. 21, No. 3, pp. 305-326, 1982
- [3] 吉村賢治ほか, 文節数最小法を用いたべた書き日本語文の形態素解析, 情報処理学会論文誌, Vol. 24, No. 1, pp. 40-46, 1983
- [4] 木下是雄, 理科系の作文技術, 中公新書, 1981
- [5] 田中康仁ほか, 科学技術文献抄録における片仮名の解析, 情報処理学会第26回全国大会, 7H-6, 1983
- [6] 坂本義行, 日本語の文節分かち書きの自動化, 情報処理学会第19回全国大会, pp. 447-448, 1978
- [7] 牛島和夫ほか, JEF日本語処理機能を追加したSNOBOL4について, 九州大学大型計算機センター広報, Vol. 16, No. 2, pp. 155-179, 1983
- [8] Ushijima, K., et al., SNOBOL4 with Japanese text processing facility, Proc. ICTP' 83, pp. 235-240, 1983
- [9] 吉田将ほか, 公用データベース日本語単語辞書の使用について, 九州大学大型計算機センター広報, Vol. 16, No. 4, pp. 335-361, 1983

◇ プログラムライブラリ開発課題の募集について

センターではプログラムライブラリ開発課題を随時募集しています。

開発計画をお持ちの方は「プログラムライブラリ開発課題申請書」に必要事項を記入し、センター図書室に提出して下さい。

なお、対象となる課題は広く各分野で使われる可能性のあるプログラム、またはある専門分野に限られるが、その分野では広く使われる可能性のあるプログラムで、新規開発および書換えを含みます。その他詳細については利用の手引ライブラリ編を参照し、不明な点はセンター図書室にお問合せ下さい。

付録A 日本語文章推敲支援ツール「推敲」・コマンド一覧表 (3. 3節における説明順)

機能	コマンド	オペランド
文章の統計	<u>STATISTIC</u>	なし
文の切り出し	<u>SENTENCE</u>	<u>NODISP</u> <u>LIST</u>
文末からソート	<u>R-SENTENCE</u>	<u>NODISP</u> <u>LIST</u>
受動態抽出	<u>PASSIVE</u>	<u>NODISP</u> <u>LIST</u>
指示詞に下線を引いて NLPに出力	<u>KOSOA</u>	SY (O K S U) →注1
段落の最初の文と最後の文 だけを出力	<u>STRUCT</u>	SY (O K S U) →注1
見出し抽出	<u>HEADING</u>	<u>NODISP</u> <u>LIST</u>
括弧の対応 (数)	<u>PAREN</u>	<u>LIST</u>
1文/1レコードに変換	<u>RENUMBER</u>	OUT (日本語テキストファイル名) →注2
KWIC	<u>KWIC</u>	字種 OUT (データセット名) →注3 <u>NODISP</u> <u>LIST</u>
逆順KWIC	<u>R-KWIC</u>	字種 <u>NODISP</u> <u>LIST</u>
文字列長・数・割合の一覧	<u>COUNT</u>	字種
指定した範囲の長さの 文字列を表示	<u>LENGTH</u>	字種 SIZE (i1-i2) <u>LIST</u> →注4
XREF	<u>XREF</u>	字種 <u>NODISP</u> <u>LIST</u>
文字列一頻度一覧	<u>INDEX</u>	字種 <u>NODISP</u> <u>LIST</u>
頻度順一覧	<u>FREQUENCY</u>	字種 <u>NODISP</u> <u>LIST</u>
コマンドの説明	<u>HELP</u>	[コマンド名]
処理対象テキストの 設定/変更	<u>DSNAME</u>	日本語テキストファイル名
マクロ・コマンドの登録	<u>MACRO</u>	マクロ・コマンド名
マクロ・コマンド名の一覧	<u>MLIST</u>	なし
コマンド終了	<u>EXIT</u>	なし

機 能	コ マ ン ド	オ ペ ラ ン ド
プリンタ名指定	<u>NPR</u>	出力先→注5
FDMS和文エディタ起動	<u>NEDIT</u>	[EDIT FORMAT] [DICTION (辞書データセット名)] [MACLIB、(マクロライブラリ名)]
TSSモード	<u>TSS</u>	な し
結果の出力	<u>L I S T</u>	コマンド名 [字種] 出力先→注5

⊕NODISP : このオペランドを指定すると結果をディスプレイに表示しない。

⊕L I S T : このオペランドを指定するとコマンド実行後、結果をプリンタ（あらかじめNPRコマンドで指定しておく）に出力する。

⊕字種：A-英字，C-漢字，H-ひらがな，K-カタカナ，N-数字のいずれかを指定する。

《注1》結果を出力するNLP（日本語ラインプリンタ）の出力クラスを指定する。

省略するとSY(O)に出力する。

《注2》日本語テキストファイル名には、変換結果を出力するデータセット名を指定する。

省略すると「RENUMBER.JEF」という名前のデータセットに出力する。

《注3》KWICリストを保存したい時に、保存するデータセット名を指定する。

《注4》i1-i2には表示したい文字列の長さの範囲を整数で指定する。

省略すると「10文字以上」を出力する。

《注5》結果の出力先（「SY(・)」「I・・・」「M・・・」のいずれか）を指定する。

付録B 日本語文章推敲支援ツール「推敲」・コマンド一覧表 (アルファベット順)

機能	コマンド	オペランド
文字列長・数・割合の一覧	<u>COUNT</u>	字種
処理対象テキストの設定/変更	<u>DSNAME</u>	日本語テキストファイル名
コマンド終了	<u>EXIT</u>	なし
頻度順一覧	<u>FREQUENCY</u>	字種 <u>NODISP</u> <u>LIST</u>
見出し抽出	<u>HEADING</u>	<u>NODISP</u> <u>LIST</u>
コマンドの説明	<u>HELP</u>	[コマンド名]
文字列-頻度一覧	<u>INDEX</u>	字種 <u>NODISP</u> <u>LIST</u>
指示詞に下線を引いて NLPに出力	<u>KOSOA</u>	SY (O K S U) →注1
KWIC	<u>KWIC</u>	字種 <u>OUT</u> (データセット名) →注2 <u>NODISP</u> <u>LIST</u>
指定した範囲の長さの 文字列を表示	<u>LENGTH</u>	字種 <u>SIZE</u> (i1-i2) <u>LIST</u> →注3
結果の出力	<u>LIST</u>	コマンド名 [字種] 出力先→注4
マクロ・コマンドの登録	<u>MACRO</u>	マクロ・コマンド名
マクロ・コマンド名の一覧	<u>MLIST</u>	なし
FDMS和文エディタ起動	<u>NEDIT</u>	[<u>EDIT</u> <u>FORMAT</u>] [<u>DICTION</u> (辞書データセット名)] [<u>MACLIB</u> (マクロライブラリ名)]
プリンタ名指定	<u>NPR</u>	出力先→注4
括弧の対応 (数)	<u>PAREN</u>	<u>LIST</u>
受動態抽出	<u>PASSIVE</u>	<u>NODISP</u> <u>LIST</u>
逆順KWIC	<u>R-KWIC</u>	字種 <u>NODISP</u> <u>LIST</u>
文末からソート	<u>R-SENTENCE</u>	<u>NODISP</u> <u>LIST</u>
1文/1レコードに変換	<u>RENUMBER</u>	<u>OUT</u> (日本語テキストファイル名) →注5
文の切り出し	<u>SENTENCE</u>	<u>NODISP</u> <u>LIST</u>

機 能	コ マ ン ド	オ ペ ラ ン ド
文章の統計	<u>STATISTIC</u>	なし
段落の最初の文と最後の文 だけを出力	<u>STRUCT</u>	SY (O K S U) →注1
TSSモード	<u>TSS</u>	なし
XREF	<u>XREF</u>	字種 <u>NODISP</u> <u>LIST</u>

⊕NODISP：このオペランドを指定すると結果をディスプレイに表示しない。

⊕LIST：このオペランドを指定するとコマンド実行後、結果をプリンタ（あらかじめNPRコマンドで指定しておく）に出力する。

⊕字種：A－英字，C－漢字，H－ひらがな，K－カタカナ，N－数字のいずれかを指定する。

〈注1〉結果を出力するNLP（日本語ラインプリンタ）の出力クラスを指定する。

省略するとSY (O) に出力する。

〈注2〉KWICリストを保存したい時に、保存するデータセット名を指定する。

〈注3〉i1－i2には表示したい文字列の長さの範囲を整数で指定する。

省略すると「10文字以上」を出力する。

〈注4〉結果の出力先（「SY (・)」 「I・・・」 「M・・・」のいずれか）を指定する。

〈注5〉日本語テキストファイル名には、変換結果を出力するデータセット名を指定する。

省略すると「RENUMBER.JEF」という名前のデータセットに出力する。