

TR-I-0111

文節候補選択処理のための係り受けデータ

Dependency Relationship Database  
for Selecting the Phrase Candidates

柿ヶ原 康二\*

泉 昌明\*\*

Koji KAKIGAHARA

Masaaki IZUMI

1989.9

概要

ATR自動翻訳電話研究所では日本語の言語現象を明らかにする上で必要となる各種の基礎データを収集した大規模な言語データベースの構築を目指しており、その1つとして言語データベースに格、係り受け関係を記述している。この格、係り受け関係のデータは係り受けの構文関係、意味関係等の詳細な記述がなされており、さまざまな言語処理の基礎データとして利用できるものである。

現在この係り受け関係データを用いて、音声認識結果の文節ラティスの文節候補から最適な候補を選択する手法の検討を行っている。特に、そのままでは適用できない部分に対してデータ構造の変換と、個々の単語に意味素性の付加を行った。本稿では変換後の係り受け辞書データの内容とその作成手順について説明する。

\*ATR 自動翻訳電話研究所

ATR Interpreting Telephony Research Laboratories

\*\*日進ソフトウェア(株)

NISSIN Software Company

©ATR 自動翻訳電話研究所 1989

©ATR Interpreting Telephony Research Laboratories

## 目次

1. はじめに .....	1
2. 係り受け辞書データ作成手順 .....	1
3. dBASE上のデータ構造変換 .....	7
3.1 処理内容説明 .....	7
3.2 プログラム説明 .....	8
4. UNIX上のデータ構造変換 .....	10
4.1 処理内容説明 .....	10
4.2 プログラム説明 .....	10
5. 意味素性の付与 .....	12
5.1 意味素性 .....	12
5.2 係り受け辞書データ .....	12
6. おわりに .....	16
参考文献 .....	17
<付録1 係り受け構造変換プログラム (dBASE) > .....	18
<付録2 係り受けデータ構造変換処理 関数一覧 (UNIX) > .....	24

## 1. はじめに

日本語の意味構造の解析には単語間の係り受け関係が重要な役割を担っている。ATR自動翻訳電話研究所では日本語の言語現象を明らかにする上で必要となる各種の基礎データを収集した大規模な言語データベースの構築を目指しており、その1つとして言語データベースに格、係り受け関係を記述している。この格、係り受け関係のデータは係り受けの構文関係、意味関係等の詳細な記述がなされており、さまざまな言語処理の基礎データとして利用できるものである。

現在この係り受け関係データを用いて、音声認識結果の文節ラティスの文節候補から最適な候補を選択する手法の検討を行っている。特に、そのままでは適用できない部分に対してデータ構造の変換と、個々の単語に意味素性の付加を行った。本稿では変換後の係り受け辞書データの内容とその作成手順について説明する。

## 2. 係り受け辞書データ作成手順

図1に係り受け辞書データの作成手順を示す。図中の各種データの内容は以下のようである。

### (1) テキストデータ

電話会話、キーボード会話の対話文のデータであり、文章番号、形態素分けされた単語、個々の単語の文中の位置、正規表現、品詞番号等をdBASEのデータとして記述している。

### (2) 係り受けデータ

対話文から抽出した係り受け関係のデータであり、dBASEのデータベースとして作成されている。係り受けの構文関係、意味関係、テキストデータの単語の位置等が記述されている。

### (3) 付属語等の付いた係り受けデータファイル

テキストデータと係り受けデータから、付属語、接頭語、複合名詞中の先行する名詞等を付加して係り受け関係を記述したdBASEのデータファイル。

### (4) MS-DOS上のテキストファイル

付属語等の付いた係り受けデータファイルをMS-DOS上のファイルに変換したもの。dBASEのコマンド(copy)を用いて変換する。

### (5) UNIX上のテキストファイル

MS-DOS上のテキストファイルをUNIX上のテキストファイルに変換したもの。UNIX上のプログラム(rfilter)を用いて変換する。

### (6) 係り受け辞書データ

UNIX上のテキストファイルに対して、係り受け関係を用いたプログラム等での利用が容易な形態にデータ構造の変換をしたもの。追加情報として単語の意味素性を付加している。

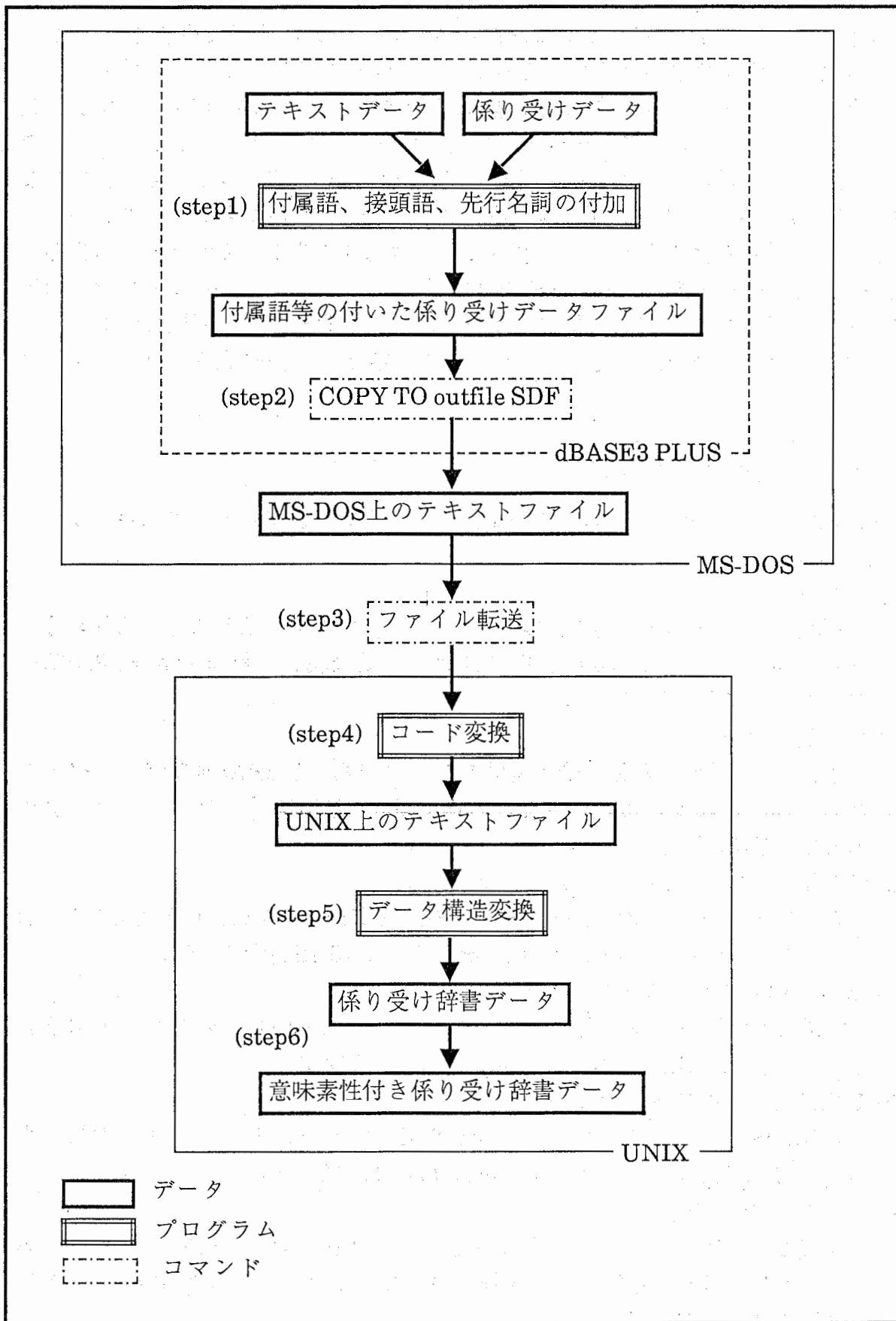


図1 係り受け辞書データ作成フロー

以下に順を追って作成方法を説明する。

(step1)

対話文のテキストデータと係り受けデータは図2と図3に示されるようなdBASE上のデータファイルである。この係り受けデータは自立語間の係り受け関係を記述しており、その自立語に後接している付属語はテキストデータを参照しなければならない。接頭語や複合名詞などの前接する単語の情報を得る場合も同様である。そこでdBASE上でプログラムを作成し、テキストデータと係り受けデータから図4に示される付属語等の付いたデータファイルを作る。このdBASE上のプログラムについては後で説明する。

(step2)

dBASE上のデータファイルはそのままでは利用できないので、dBASEのCOPYコマンド(COPY TO 出力ファイル名 SDF)を用いてMS-DOS上のテキストファイルに変換する。

(step3)

MS-DOS上のテキストファイルをエミュレータ(VT98)を用いてUNIX上に転送する。

(step4)

UNIX上でテキストファイルを利用できるようにするため、コード変換を行う。(rfilter <入力ファイル名>出力ファイル名)

(step5)

UNIX上で作成したプログラムを用いてデータ構造の変換を行う。プログラムについては後で説明する。

(step6)

意味素性の付与を行う。これは手作業で行っている。

レコード	文章番号	位置番号	単語	正規	品詞
1	D11601	999.00	通訳者		0
2	D11601	1.00	はい		11
3	D11601	2.00	。		0
4	D11602	999.00	質問者		0
5	D11602	1.00	会議事務局	会議事務局	30
6	D11602	2.00	でしょ	です	12
7	D11602	3.00	う	う	12
8	D11602	4.00	か	か	16
9	D11602	5.00	。		0
10	D11603	999.00	事務局		0
11	D11603	1.00	はい	はい	11
12	D11603	2.00	そう	そう	8
13	D11603	3.00	で	で	15
14	D11603	4.00	ございます	御座います	32
15	D11603	5.00	が	が	16
16	D11603	6.00	。		0
17	D11604	999.00	質問者		0
18	D11604	1.00	[		0
19	D11604	2.00	えーと	え	33
20	D11604	3.00	]		0
21	D11604	4.00	先日	先日	4
22	D11604	5.00	そちら	其方	6
23	D11604	6.00	の	の	15
24	D11604	7.00	国際会議	国際会議	30
25	D11604	8.00	の	の	15
26	D11604	9.00	資料	資料	4
27	D11604	10.00	を	を	15
28	D11604	11.00	送っ	送る	32
29	D11604	12.00	て	て	14
30	D11604	13.00	もらっ	貰う	19

図2 テキストデータ

レコード	文章番号	ID	コード	LOC1F	LOC1T	LOC2F	LOC2T
1	D11601	01	PRD	1.00		2.00	
2	D11601	01	TMA	4.00		11.00	
3	D11601	02	POS	5.00		7.00	
4	D11602	02	RNG	7.00		9.00	
5	D11602	01	OBJ	9.00		11.00	
6	D11602	01	CON	11.00		13.00	
7	D11602	05	CAU	13.00		27.00	
8	D11602	01	OBJ	21.00		23.00	
9	D11602	05	CAU	23.00		27.00	
10	D11603	01	TMA	3.00		6.00	
11	D11603	01	OBJ	5.00		6.00	
12	D11603	02	POS	6.00		8.00	
13	D11603	01	OBJ	8.00		14.00	
14	D11603	04	DGR	13.00		14.00	
15	D11603	01	CON	14.00		16.00	
16	D11603	01	CON	16.00		18.00	
17	D11604	05	PRL	18.00		59.00	
18	D11604	01	OBJ	28.00		30.00	
19	D11604	05	CAU	30.00		54.00	
20	D11604	01	SPA	36.00		38.00	
21	D11604	05	PRL	38.00		43.00	
22	D11604	01	REC	51.00		54.00	
23	D11604	04	CND	54.00		57.00	
24	D11604	03	GAI	58.00		57.00	
25	D11604	01	PRD	58.00		59.00	
26	D11604	02	APP	8.00		11.00	
27	D11604	01	OBJ	11.00		13.00	
28	D11604	01	CON	13.00		16.00	
29	D11604	02	DLM	6.00		8.00	
30	D11604	01	RNG	8.00		16.00	

図3 係り受けデータ

レコード	文章番号	ID	コード	係り元	付属1	付属2	付属3	付属4	正規元	係り先	付属5	付属6	付属7	付属8	正規先						
1	D11602	01	PRD	事務局	で	し	よ	う	か	事務局	で	し	よ	う	か	です					
2	D11604	01	TMA	先日						先日	送	っ	て			送る					
3	D11604	02	POS	会議	の					会議	資	料				資料					
4	D11604	02	OBJ	資料	を					資料	送	っ	て			送る					
5	D11604	05	CAU	送	っ	て				送る	も	ら	っ	た	ん	で	貰う				
6	D11604	01	OBJ	興味	を					興味	持	ち	ま	し	て		持つ				
7	D11605	01	TMA	本日	は					本日	ご	さ	い	ま	す	で	し	よ	う	か	御座います
8	D11606	01	OBJ	会議	に					会議	参	加					参加				
9	D11606	04	DGR	せ	ひ	と	も			せ	ひ	と	も				参加				
10	D11606	01	OBJ	カード	が					カード	あ	り	ま	す	の	で	有る				
11	D11606	01	SPA	これ	に					此	書	い	て				書く				
12	D11606	05	PRL	書	い	て				書く	登	録					登録				
13	D11606	01	REC	そ	ち	ら	に			其	方	送	り				送る				

図4 付属語付き係り受け辞書データ



### 3. dBASE上のデータ構造変換

#### 3.1 処理内容説明

dBASE上の係り受けデータの係り受け関係から、その係り受け関係が存在する文章の番号、係り受け関係の構文構造、意味構造、係り元の単語の文章中の位置、係り先の単語の文章中の位置を知ることができる。この係り元と係り先の単語(自立語)に対して付属語、接頭語、複合名詞の先行する名詞列(a)を求めるには、文章番号と位置番号をもとに、テキストデータを参照すればよい。テキストデータには品詞番号が記述されているので、求める単語(付属語など)の品詞番号をサーチし単語の表記を得る。以下にステップを追って説明する。

- (step 1) 各ファイル名[テキスト,係り受け,付属語付き係り受けデータ]を入力(INPUT)し、オープン(USE)する。
- (step 2) 作業領域(SELECT)を係り受けデータとする。
- (step 3) start変数に1を代入する。(STORE)
- (step 4) 係り受けデータのレコードをstartとし、文章番号,ID,コードなどを変数にセットする。(STORE)
- (step 5) テキストデータの処理を済んだデータについて文章番号のレコードを仮に削除する。(DELETE) 但し、文章番号が昇順に更新した場合だけで、もし小さい文章番号が係り受けデータにあれば、その文章番号を復活させる。
- (step 6) 作業領域(SELECT)をテキストデータとする。
- (step 7) 位置番号が、0でないか?  
[YES] : テキストデータから文章番号と位置番号を検索(LOCATE)して、レコードを見つけると、正規と係りの自立語を変数にセットし、次にもし品詞の4,5,6,7,30,31,18が(3回)あれば、変数にセットし、次にもし品詞の12,13,14,15,16,34,35,36が(4回)あれば、変数にセットする。(STORE)  
[NO] : (step 10)に進む。
- (step 8) 作業領域(SELECT)を付属語付き係り受けデータにし、空白レコードを追加する。
- (step 9) 各セットした変数を置き換えて(REPLACE)、付属語付き係り受けデータファイルに出力する。
- (step 10) start変数に1を加算する。
- (step 11) 作業領域(SELECT)を係り受けデータに戻す。
- (step 12) (step 4)から(step 11) までを係り受けデータファイルが終わるまで、繰り返す。

---

(a) 係り受けデータには意味の中心となる単語のみを記述している。

(step 13) 各ファイルをクローズする。

図5に係り受けデータ構造変換プログラムのチャート図を示す。

### 3.2 プログラム説明

このデータ構造変換プログラムはdBASEのプログラムとして作成している。プログラム名はkakari.prgである。付録1にプログラムリストを示す。

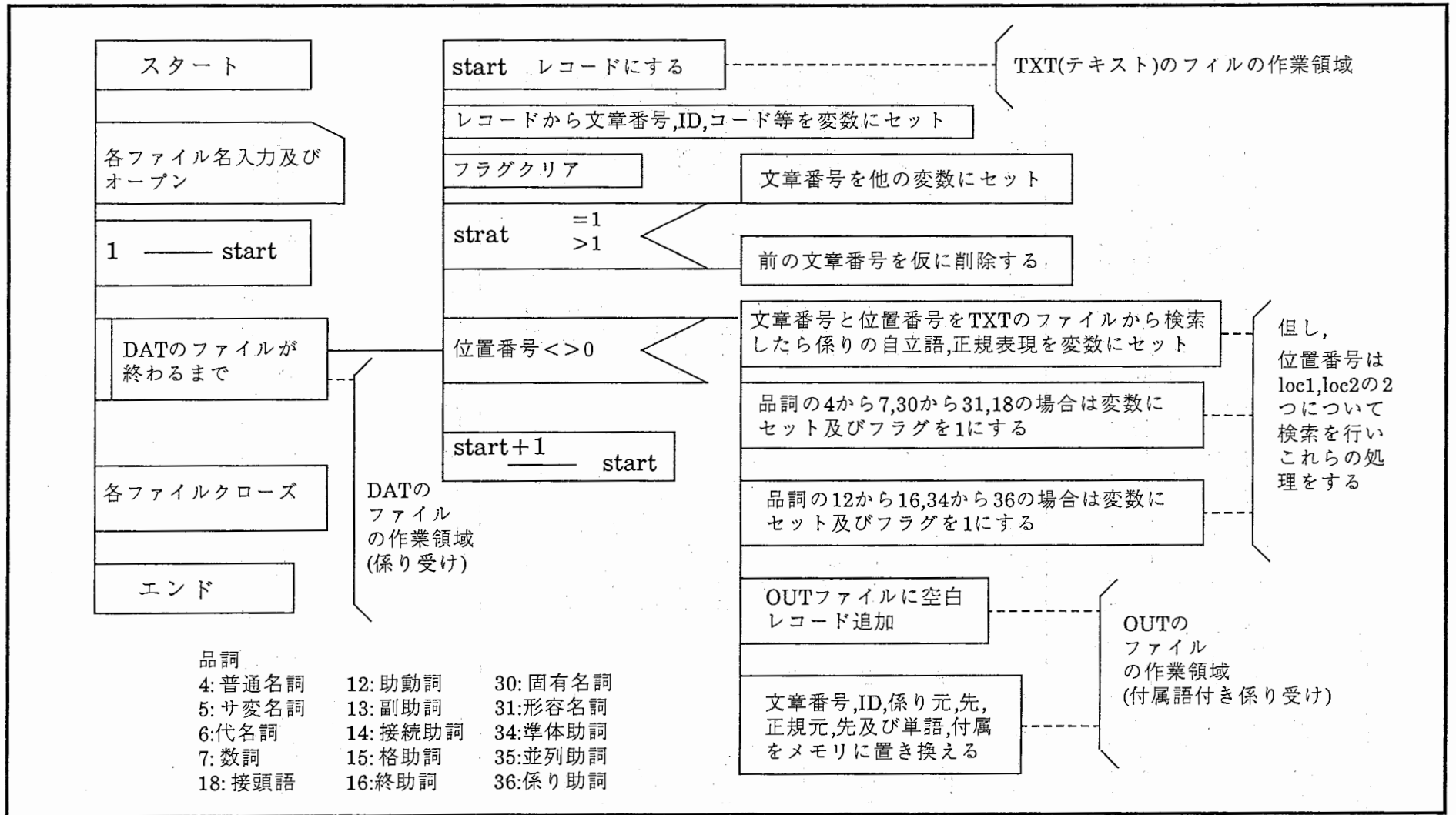


図5 係り受けデータ構造変換プログラムのチャート図 (kakari)

## 4. UNIX上のデータ構造変換

### 4.1 処理内容説明

UNIX上でのデータ構造変換は、dBASE上で行ったデータ構造変換後のテキストファイルを他のプログラムでの利用を容易にするためや人手による加工を容易にするために行ったものである。すなわち、図4に示されるデータを図6に示されるデータ構造に変換する。図中のXXXXXXXXは意味素性を付与するための領域である。図7にデータ構造変換プログラムのチャート図を示す。

ID	コード	頻度	表層表現	正規表現	意味素性
01	PRD	000	何-でしょうか : でしょうか	: 何 : です	: XXXXXXXX ; : XXXXXXXX
01	RNG	000	事-で : 聞き	: 事 : 聞く	: XXXXXXXX ; : XXXXXXXX
01	CON	000	聞き : たいのですが	: 聞く : たい	: XXXXXXXX ; : XXXXXXXX
02	RNG	000	場-の : 事で	: 場 : 事	: XXXXXXXX ; : XXXXXXXX
01	SPF	000	駅-からは : 行けば	: 駅 : 行く	: XXXXXXXX ; : XXXXXXXX
04	CND	000	行け-ば : 宜しいですか	: 行く : 宜しい	: XXXXXXXX ; : XXXXXXXX

図6 係り受け辞書データ構造

### 4.2 プログラム説明

このプログラムはC言語で作成されており、ファイル名はbuncheng.cである。includeファイルはkakar.hである。付録2に個々の関数の説明を示す。

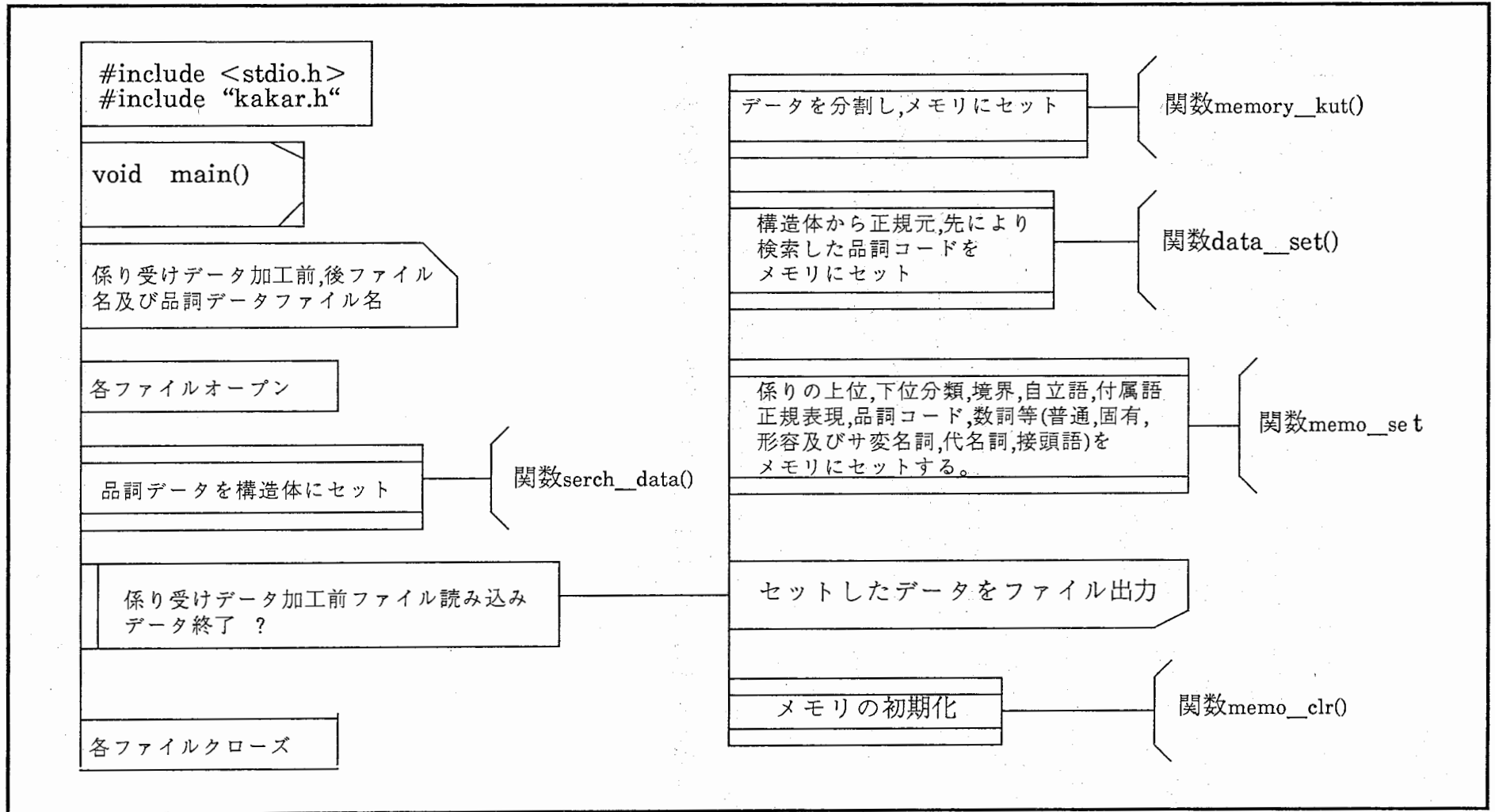


図7 係り受けデータ構造変換プログラムのチャート図 (buncheng )

## 5. 意味素性の付与

### 5.1 意味素性

係り受けデータに付与した意味素性の一覧を表1に示す。基本的な意味素性の構造は上位素性/下位素性の2階層となっている。上位の素性はタスクに無関係な一般的なものであり、下位の素性はタスクに依存したものになっている。ただし、サ変名詞、動詞、形容詞については個別に素性を付与しており、今後これらの意味素性による分類の作業が必要である。また副詞は1つの素性しか付与していないが、これらについても素性による分類が必要であろう。

### 5.2 係り受け辞書データ

意味素性を付与した係り受け辞書データの一部を図8に示す(b)。図中の頻度の項目は係り受け関係の出現頻度等を記述するために設けたものである。現在作成している係り受け辞書データは言語データベースの中の表2に示される係り受けデータから作成したものである。現在、44142語の対話文から9029項目の係り受け辞書データを作成している。この係り受け関係数はのべ数であり、重複するものが多数存在する。

係り受け辞書データファイルは次の2種類がある。

- (1) kakou-all.jpn : 付属語を含む係り受け関係を記述。
- (2) kakou-all-2.jpn : 接頭語、付属語、前接する名詞を含む係り受け関係を記述。

---

(b) 現状は一部意味素性の付与されていない自立語が残っている。それらの自立語の素性はXXXXXXXXと記述されている。

表1 意味素性一覧

名詞	抽象概念 (abstract,abs)	論理(logic,log); 事、理論、方法 状態(state,sta); 様子、場合 言語(language,lan); 日本語、何語 学術(learning,lea); 分野、専門 意向(intention,int); 興味、希望 価値(value,val); 必要 記号(sign,sig); 名前 労役(labor,lab); 用件
	具象物 (concrete,con)	文書(document,doc); 要約、原稿 交通(transport,tra); バス、地下鉄 物品(article,art); 写真 貨財(money,mon); お金、カード 任意(various,var); これ、どれ
	人間 (human,hum)	人間(human,hum); 発表者、学生 任意(various,var); だれ、どなた
	場所 (location,loc)	施設(institution,ins); ホテル、会場 地域(region,reg); 京都、海外 位置(position,pos); 間、方、近く 任意(various,var); どこ
	時間 (time,tim)	時間(time,tim); 間、後、今、時間 任意(various,var); 何時、いつ
	数 (number,num)	数量(amount,amo); 字数、何名 単位(unit,uni); 円、ドル 費用(cost,cos); 登録費、参加料 価格(price,pri); 無料、いくら
	行為 (action,act)	予定 act/yti、制限 act/sgn、 参加 act/san、発表 act/hap (総数42)
	任意 (diverse,div)	ほか、なに
動詞 (predicate,pre)	申し込む pre/msk、来る pre/kur、 書く pre/kak、持つ pre/mot (総数48)	
形容詞 (adjective,adj)	近い adj/tik、多い adj/ooi、 遠い adj/toi (総数16)	
副詞 (adverb,adv)	既に adv/adv、ちょうど adv/adv、 もう adv/adv (総数41)	

ID	コード	頻度	表層表現	正規表現	意味素性
01	PRD	000	何-でしょうか : でしょうか	: 何 : です	: div/div ; : pre/des
01	RNG	000	事-で : 聞き	: 事 : 聞く	: abs/log ; : pre/kik
01	CON	000	聞き : たい の です が	: 聞く : たい	: pre/kik ; : pre/tai
02	RNG	000	場-の : 事 で	: 場 : 事	: loc/ins ; : abs/log
01	SPF	000	駅-からは : 行け ば	: 駅 : 行く	: loc/ins ; : pre/yuk
04	CND	000	行け-ば : 宜しい です か	: 行く : 宜しい	: pre/yuk ; : adj/yrs
04	MAN	000	どう : 行け ば	: どう : 行く	: div/div ; : pre/yuk
01	TOO	000	地下鉄-で : 行っ て	: 地下鉄 : 行く	: con/tra ; : pre/yuk
01	SPF	000	駅-から : 行っ て	: 駅 : 行く	: loc/ins ; : pre/yuk
01	SPT	000	駅-まで : 行っ て	: 駅 : 行く	: loc/ins ; : pre/yuk
01	SPF	000	そこ-から : 乗り	: 其処 : 乗る	: loc/pos ; : pre/nru
01	OBJ	000	番-か : 乗り	: 番 : 乗る	: abs/sig ; : pre/nru
01	OBJ	000	タクシー-に : 乗り	: タクシー : 乗る	: con/tra ; : pre/nru
02	SEL	000	バス-の : 番 か	: バス : 番	: con/tra ; : abs/sig
01	OBJ	000	時間-が : かかり ます か	: 時間 : 掛る	: tim/tim ; : pre/kkr
01	PRD	000	分-位です : です	: 分 : です	: tim/tim ; : pre/des
01	TOO	000	バス-で : 行く より も	: バス : 行く	: con/tra ; : pre/yuk

図8 係り受け辞書データの例



表2 係り受け元データ

会話	ファイル名		
電話会話	tel-022-03	tel-032-04	
	tel-022-04	tel-032-05	
	tel-022-05	tel-032-06	
	tel-022-06	tel-032-07	
	tel-022-07	tel-032-08	
	tel-022-08	tel-032-09	
	tel-022-09	tel-032-10	
	tel-022-10	tel-032-11	
	tel-022-11	tel-032-12	
	tel-022-12	tel-032-13	
	tel-022-13	tel-032-14	
	tel-022-14	tel-032-15	
	tel-032-01	tel-032-16	
	tel-032-02	tel-032-17	
	tel-032-03	tel-032-18	
	キーボード会話	key-022-04	key-052-03
		key-032-01	key-052-04
		key-032-02	key-062-01
key-032-03		key-062-02	
key-032-04		key-062-03	
key-042-01		key-072-01	
key-042-02		key-072-02	
key-042-03		key-082-01	
key-042-04		key-082-02	
key-042-05		key-092-01	
key-052-01		key-092-02	
key-052-02			

## 6. おわりに

音声認識結果の文節ラティスに対する文節候補選択処理で用いている係り受け辞書データについて、その内容と作成方法について説明した。現在作成している係り受け辞書データは言語データベースの係り受けデータの一部から作成したものであり、加工されていない係り受けデータが多く残されている。今後は多くのデータから係り受け頻度等を抽出し、言語処理への統計量の適用などに役立てる必要がある。

## 参考文献

- [1] 荻野：「文法と意味I」朝倉書店 (1983)
- [2] 南：「文法と意味II」朝倉書店 (1985)
- [3] 大野：「類語国語辞典」角川書店 (1985)
- [4] 井上、小倉、森元：「言語データベース用格・係り受け意味体系」ATRテクニカルレポート TR-I-0029 (1988)
- [5] 柿ヶ原、桶谷、泉：「SL-TRANSにおける文節候補選択処理、問合せ処理」ATRテクニカルレポート TR-I-0110 (1989)
- [6] 柿ヶ原、森元：「係り受け関係を用いた文節候補選択手法」ATRテクニカルレポート TR-I-0109 (1989)
- [7] 柿ヶ原、森元：「係り受け関係を用いた文節候補選択についての一考察」情報処理学会第38回全国大会 (1989)
- [8] 柿ヶ原、森元：「SL-TRANSにおける文節候補の削減」情報処理学会第39回全国大会 (1989)
- [9] dBASE III PLUS マニュアル「基本編、プログラミング編、リファレンス編」

## &lt;付録1 係り受け構造変換プログラム (dBASE) &gt;

```

*****
* Kakari-uke data sakusei program      *
* 係り受けデータ加工プログラム        *
* '88.11.18 by kakigahara             *
* '89.7.3 by izumi                    *
* 7.14 VER 3.0 *
*****

CLEAR ALL
SET TALK OFF
SET DELETED ON

STORE 0 TO start

CLEAR
@ 1,1 GET start PICTURE "9999"
READ
INPUT "TXTデータのファイル名を入力してください。" TO txtfname
INPUT "DATデータのファイル名を入力してください。" TO datfname
INPUT "OUTデータのファイル名を入力してください。" TO outfname

SELECT 1
USE &txtfname ALIAS TXT
SELECT 2
USE &datfname ALIAS DAT
SELECT 3
USE &outfname ALIAS OUTFILE

SELECT DAT
GOTO TOP
STORE 0 TO FLG
DO WHILE .NOT. EOF()
  GO start
  STORE SPACE(6) TO bunnum
  STORE SPACE(2) TO idnum
  STORE SPACE(7) TO code
  STORE SPACE(30) TO ftango1
  STORE SPACE(30) TO ftango2
  STORE SPACE(30) TO ftango3
  STORE SPACE(30) TO ftango4
  STORE SPACE(30) TO ftango5
  STORE SPACE(30) TO ftango6
  STORE SPACE(30) TO tango1

```

```

STORE SPACE(10) TO tango12
STORE SPACE(10) TO tango13
STORE SPACE(10) TO tango14
STORE SPACE(10) TO tango15
STORE SPACE(30) TO tango2
STORE SPACE(10) TO tango22
STORE SPACE(10) TO tango23
STORE SPACE(10) TO tango24
STORE SPACE(10) TO tango25
STORE 文章番号 TO bunnum
STORE ID TO idnum
STORE コード TO code
STORE LOC2F TO loc2
STORE LOC1F TO loc1
STORE 0 TO fu1,fu2
STORE 0 TO fu3,fu4
STORE 0 TO fu5,fu6
STORE 0 TO fu7,fu8
STORE 0 TO fu9,fu10
STORE 0 TO fu11,fu12
STORE 0 TO fu13,fu14
STORE 0 TO setrec1,setrec2
IF (bunnum = SPACE(7))
  EXIT
ENDIF
IF (start = 1 .OR. FLG = 0)
  STORE bunnum TO bunbk
ENDIF
IF (start > 1 .AND. FLG <> 0)
  IF ( bunnum <> bunbk )
    IF ( VAL(RIGHT(bunnum,5)) < VAL(RIGHT(bunbk,5)) )
      SELECT TXT
      GOTO TOP
      LOCATE FOR 文章番号 = bunnum
      IF FOUND()
        SET DELETED OFF
        RECALL WHILE 文章番号 = bunnum
        SET DELETED ON
      ENDIF
    ELSE
      SELECT TXT
      GOTO TOP
      LOCATE FOR 文章番号 = bunbk
      IF FOUND()
        DELETE WHILE 文章番号 = bunbk
      ENDIF
    ENDIF
  ENDIF

```

```

STORE bunnum TO bunbk
CLEAR
@ 1,1 SAY bunnum
ENDIF
ENDIF
ENDIF
SELECT TXT
GOTO TOP
IF (loc1 <> 0 .AND. loc2 <> 0)
LOCATE FOR 文章番号 = bunnum .AND. 位置番号 = loc1
IF FOUND()
STORE 単語 TO tango1
STORE 正規 TO seikimoto
setrec1 = (RECNO() + 1)
GOTO (setrec1 - 2)

IF (品詞 > 3 .AND. 品詞 < 8).OR.(品詞 > 29 .AND. 品詞 < 32).OR.(品詞 = 18)
STORE 単語 TO ftango1
STORE 1 TO fu9
GOTO (setrec1 - 3)
IF (品詞 > 3 .AND. 品詞 < 8).OR.(品詞 > 29 .AND. 品詞 < 32).OR.(品詞 = 18)
STORE 単語 TO ftango2
STORE 1 TO fu10
GOTO (setrec1 - 4)
IF (品詞 > 3 .AND. 品詞 < 8).OR.(品詞 > 29 .AND. 品詞 < 32).OR.(品詞 = 18)
STORE 単語 TO ftango3
STORE 1 TO fu11
ENDIF
ENDIF
ENDIF
GOTO (setrec1)
IF (品詞 > 11 .AND. 品詞 < 17) .OR. (品詞 > 33 .AND. 品詞 < 37)
STORE 単語 TO tango12
STORE 1 TO fu1
GOTO (RECNO() + 1)
IF (品詞 > 11 .AND. 品詞 < 17) .OR. (品詞 > 33 .AND. 品詞 < 37)
STORE 単語 TO tango13
STORE 1 TO fu2
GOTO (RECNO() + 1)
IF (品詞 > 11 .AND. 品詞 < 17) .OR. (品詞 > 33 .AND. 品詞 < 37)
STORE 単語 TO tango14
STORE 1 TO fu3
GOTO (RECNO() + 1)
IF (品詞 > 11 .AND. 品詞 < 17) .OR. (品詞 > 33 .AND. 品詞 < 37)
STORE 単語 TO tango15
STORE 1 TO fu4

```

```

        ENDIF
    ENDIF
ENDIF
ENDIF
GOTO TOP
LOCATE FOR 文章番号 = bnum .AND. 位置番号 = loc2
IF FOUND()
    STORE 単語 TO tango2
    STORE 正規 TO seikisaki
    setrec2 = (RECNO() + 1)
    GOTO (setrec2 - 2)
    IF (品詞 > 3 .AND. 品詞 < 8).OR.(品詞 > 29 .AND. 品詞 < 32).OR.(品詞 = 18)
        STORE 単語 TO ftango4
        STORE 1 TO fu12
        GOTO (setrec2 - 3)
        IF (品詞 > 3 .AND. 品詞 < 8).OR.(品詞 > 29 .AND. 品詞 < 32).OR.(品詞 = 18)
            STORE 単語 TO ftango5
            STORE 1 TO fu13
            GOTO (setrec2 - 4)
            IF (品詞 > 3.AND.品詞 < 8).OR.(品詞 > 29.AND.品詞 < 32).OR.(品詞 = 18)
                STORE 単語 TO ftango6
                STORE 1 TO fu14
            ENDIF
        ENDIF
    ENDIF
    GOTO (setrec2)
    IF (品詞 > 11 .AND. 品詞 < 17) .OR. (品詞 > 33 .AND. 品詞 < 37)
        STORE 単語 TO tango22
        STORE 1 TO fu5
        GOTO (RECNO() + 1)
        IF (品詞 > 11 .AND. 品詞 < 17) .OR. (品詞 > 33 .AND. 品詞 < 37)
            STORE 単語 TO tango23
            STORE 1 TO fu6
            GOTO (RECNO() + 1)
            IF (品詞 > 11.AND.品詞 < 17).OR.(品詞 > 33.AND.品詞 < 37)
                STORE 単語 TO tango24
                STORE 1 TO fu7
                GOTO (RECNO() + 1)
                IF (品詞 > 11.AND.品詞 < 17).OR.(品詞 > 33.AND.品詞 < 37)
                    STORE 単語 TO tango25
                    STORE 1 TO fu8
                ENDIF
            ENDIF
        ENDIF
    ENDIF
ENDIF
ENDIF
ENDIF

```

```

ENDIF
SELECT OUTFILE
APPEND BLANK
REPLACE 文章番号 WITH bunnum
REPLACE ID with idnum
REPLACE コード with code
IF (fu9 = 1 .AND. fu10 = 1 .AND. fu11 = 1)
  REPLACE 単語3 WITH ftango1
  REPLACE 単語2 WITH ftango2
  REPLACE 単語1 WITH ftango3
ENDIF
IF (fu9 = 1 .AND. fu10 = 1 .AND. fu11 = 0)
  REPLACE 単語2 WITH ftango1
  REPLACE 単語1 WITH ftango2
ENDIF
IF (fu9 = 1 .AND. fu10 = 0 .AND. fu11 = 0)
  REPLACE 単語1 WITH ftango1
ENDIF
REPLACE 係り元 WITH tango1
IF fu1 = 1
  REPLACE 付属1 WITH tango12
ENDIF
IF fu2 = 1
  REPLACE 付属2 WITH tango13
ENDIF
IF fu3 = 1
  REPLACE 付属3 WITH tango14
ENDIF
IF fu4 = 1
  REPLACE 付属4 WITH tango15
ENDIF
REPLACE 正規元 WITH seikimoto
IF (fu12 = 1 .AND. fu13 = 1 .AND. fu14 = 1)
  REPLACE 単語6 WITH ftango4
  REPLACE 単語5 WITH ftango5
  REPLACE 単語4 WITH ftango6
ENDIF
IF (fu12 = 1 .AND. fu13 = 1 .AND. fu14 = 0)
  REPLACE 単語5 WITH ftango4
  REPLACE 単語4 WITH ftango5
ENDIF
IF (fu12 = 1 .AND. fu13 = 0 .AND. fu14 = 0)
  REPLACE 単語4 WITH ftango4
ENDIF
REPLACE 係り先 WITH tango2
IF fu5 = 1

```



```
REPLACE 付属5 WITH tango22
ENDIF
IF fu6 = 1
  REPLACE 付属6 WITH tango23
ENDIF
IF fu7 = 1
  REPLACE 付属7 WITH tango24
ENDIF
IF fu8 = 1
  REPLACE 付属8 WITH tango25
ENDIF
REPLACE 正規先 WITH seikisaki
ENDIF
FLG = 1
start = start + 1
SELECT DAT
ENDDO
CLOSE DATABASES
QUIT
```

<付録2 係り受けデータ構造変換処理 関数一覧 (UNIX) >

名前	係り受け辞書データ加工(メインプログラム)	
機能	本プログラムは各入出力ファイルを入力,オープンし、品詞データを構造体にセットしデータベース化して、加工前データを読み込み,(正規表現からデータベースを検索して品詞コードを)メモリにセットし、それらのメモリを配置,結合し、ファイル出力する。付録3 dBASE上のデータ構造変換	
呼び出し形式	void main()	
リターン値	無し	
引数名	タイプ	説明
備考	本プログラムのファイル名はbuncheng.cである。 また加工前データとはdbase3 plus(pc-9801)で作成し、unixに転送したものである。	

名前	データを検索及びセット(data_set)	
機能	本プログラムは構造体から一致する正規表現を検索し、その品詞コードのデータをセットする。	
呼び出し形式	void data_set(s,v)	
リターン値	無し	
引数名	タイプ	説明
s	char *	検索するためのデータ(正規表現)のポインタ
v	char *	品詞コードをセットするポインタ
備考		

名前	文字列分割(memory__kut)	
機能	本プログラムは文字列の中から文字列を分割し、その文字列から空白を取り除いて、メモリにセットする。	
呼び出し形式	void memory__kut(rd__p)	
リターン値	無し	
引数名	タイプ	説明
rd__p	char *	取り出される文字列のポインタ
備考		

名前	メモリーセット(memo_set)	
機能	本プログラムは取り出してきたデータの上位,下位分類,境界,自立語,付属語,正規表現,品詞コード,数詞等(普通,固有,形容及びサ変名詞,代名詞,接頭語)をメモリーに配置し、ファイルに出力する。	
呼び出し形式	void memo_set(OUT_FP)	
リターン値		
引数名	タイプ	説明
OUT_FP	FILE *	出力ファイルのファイルポインタ
備考		

名前	構造体にデータセット(serch_data)	
機能	本プログラムは以前作成したデータファイルから正規表現と品詞コードのデータを構造体にセットする。	
呼び出し形式	void serch_data(HINSI_FP)	
リターン値		
引数名	タイプ	説明
HINSI_FP	FILE *	以前作成したデータファイルのポインタ
備考		

名前	コロンで文字列の位置を検索(data_hinsi)	
機能	本プログラムは文字列データから'!'で文字列の位置を検索して、正規表現と品詞コードを取り出す。	
呼び出し形式	void data_hinsi(data)	
リターン値	無し	
引数名	タイプ	説明
data	char *	取り出される文字列のポインタ
備考		



名前	文字列を取り出す(moji_o_tol)	
機能	本プログラムは文字列中の文字列のあった位置('!又は' 'でない)から空白までの文字列を取り出す。	
呼び出し形式	char *moji_o_tol(s)	
リターン値	取り出した文字列のポインタ	
引数名	タイプ	説明
s	char *	取り出される文字列のポインタ
備考		

## ●その他の関数の説明

関数名	機能
memo_clr()	各メモリの領域をクリアする。
dat_set1()	正規表現を構造体にセットする。
dat_set2()	品詞コードを構造体にセットし、構造体のポインタを更新する。
kuhaku_del()	文字列の後ろの空白を取り除く。
buff_clear()	文字列の内容をクリアする。
mid()	文字列の中からある範囲の文字列を取り出す。
whtec()	文字列の前にある範囲の空白を結合する。