

テキスト・データベースから採取した言語表現資料 に基づいてリレーショナル・データベース・ソフト ウェアを利用して言語記述作業を行なうためのマ ニュアル的ノート

竹内，義晴
九州大学言語文化部

<https://doi.org/10.15017/6796297>

出版情報：言語科学. 27, pp. 36-55, 1992-03-02. 九州大学言語文化部言語研究会
バージョン：
権利関係：



テキスト・データベースから採取した言語表現資料に基づいて
リレーショナル・データベース・ソフトウェアを利用して
言語記述作業を行なうためのマニュアル的ノート

竹内 義晴

目次

- 0. 初めに
- 1. 議論の前提
- 2. 言語表現資料の採取
 - 2.1. パーソナルコンピュータ端末を立ち上げ、大型計算機センターと接続
 - 2.2. テキスト・データベースにアクセス
 - 2.3. テキスト・データベースから資料を検索・採取
 - 2.4. 大型計算機からパーソナルコンピュータへのデータ転送
 - 2.5. 「SIGMA」システム、大型計算機接続、端末エミュレータの終了
- 3. リレーショナルデータベースを利用した言語記述のための、
パーソナルコンピュータ上でのデータ変換・加工
 - 3.1. パーソナルコンピュータのエディタとワープロを使った
手書き作業カードの作製
 - 3.2. テキストデータのリレーショナル・データベース・データへの変換
 - 3.2.1. 「dBASE III PLUS」の起動とデータ構造の仮決定、
「dBASE III PLUS」の終了
 - 3.2.2. リレーショナル・データベース・データ変換のための
テキストデータの加工
 - 3.2.3. リレーショナル・データベース・データへの変換
- 4. 言語記述作業
 - 4.1. 手書き作業カードを用いた暫定的記述
 - 4.2. 記述項目のとりあえずの設定（データ構造の仮設計）と
リレーショナル・データベースへの入力、記述途中でのデータ構造の変更
- 5. 記述作業の諸問題と若干の考察
 - 5.1. 記述と抽象化について
 - 5.2. 記述作業の諸問題
 - 5.3. まとめ

0. 初めに

テキスト・データベース、またはコーパスと呼ばれる、計算機処理が可能な言語テキストが様々な研究機関で制作され、言語研究などに利用できるようになっている。ドイツ語関係では、私の同僚、樋口忠治の制作したトーマス・マン全集・ゲーテ全集のテキスト・データベースが、それぞれ「トーマス・マン・ファイル」・「ゲーテ・ファイル」として九州大学大型計算機センターで公開されている。また、マンハイムのドイツ語研究所の「マンハイム・コルプス」、フライブルク大学の「フライブルク・コルプス」などが、知られている。光学文字読み取り装置（OCR）の普及により、活字化された言語

この論文は1991年度文部省科学研究費の補助を受けてなされた、一般研究C「言語表現の意味の揺れに基づく言語学意味論の認知科学的位置づけのための研究」（課題番号：02610225、代表者：竹内義晴）の成果の一部である

資料のテキスト・データベース化の作業は、これまでの人力に頼る部分が大幅に減少されるだろう。種類・話者（作家・ジャーナリスト・政治家・茶飲み話をする人、等々）・話題・内容など、必要に応じて研究者の目的にかなうテキストを、比較的短時間に計算機可読化することが可能になってきている。私たちの利用可能なテキスト・データベースは、様々な言語において、今後、質・量・種類ともに充実してゆくに違いない。

（テキスト・データベース利用のような知的処理能力の量的拡大の可能性は、科学技術の発展が切り開いたものである。しかし、テキスト・データベースを制作する作業が光学文字読み取り装置の普及などにより、容易になればなるほどに、その作業の遂行はアルバイト作業員などの労働力に依存することになるだろう。その場合、大学の研究費など公的な学術研究予算の充実によって、テキスト・データベース制作・維持・公開・管理に対する支援が得られることは、非常に大切なポイントである。科学技術の切り開いた利便さは、その利便供与にかかわる予算的平等の配慮がなされない限り、資力を備えた少数者に独占される危険を備えてもいる。このことについては、辞書作製の目的で作成されたテキスト・データベースが、出版社の知的資産として囲い込まれ、手軽には公開されていないイギリスの事情に言及するまでもない。）

テキスト・データベースの蓄積が増大してゆくことは確実なことである。しかし、その利用方法について、私たちの人間は、どのような知恵を蓄積してきたのだろう。私の知る限り、研究者同士が情報交換し、テキスト・データベース利用のためのノウ・ハウを公開し、より能率的に研究を進めることの可能な研究環境の確立のために、互いに協力し合っているということをおそらく聞かない（ドイツ語研究者の間に、「日本ドイツ語情報処理研究会」という集まりがあるが、私は、現在のところ参加していないので、どのような成果を挙げているのか、残念ながらよく知らない）。

私は現在、文部省科学研究費の補助を受けて、テキスト・データベースから言語表現の資料を収集し、それを整理・分析することにより、言語表現の意味の核心を抽出する作業を行っている（竹内 1991、1992a、b を参照していただきたい）。全くの手探りにも似た作業を続けているといつてよいのだが、その手探りの実際を、記録に留め、ここに公開することは決して無駄なことではないだろう。もちろん、よりすぐれた作業の段取りを確立している研究者の方々に、私の稚拙さを笑っていただくだけでなく、御教示・御指導いただければ、大きな収穫である。さらに、私の作業手順の内でもわずかも、参考にしていただけることがあれば、また喜ばしい。

私自身の経験から言えば、計算機を利用するような、人間にとって非常に不自然な作業においては、作業の手順をマニュアル化し、繰り返し類似の作業がとどこおりなく行えるような作業基盤を整備しておくことはかなり重要なことである。各々のステップは単純なことであっても、研究行為全体としての作業構成はある程度複雑なものにならざるを得ないからである。

例えば、大型計算機センターに自分の端末を接続するとか、あるアプリケーション・ソフトウェアのためのデータを、他のソフトウェアのために変換するというような、なんでもない作業のことを考えてみる。それらの作業を行なうためには、計算機やソフトウェアの製作者がほとんど「場当りの」に定義した記号の羅列を、間違いなくキーボードから入力しなければならない。現在の計算機は、いくらかの改良はあるものの、わずかな誤りも許容しない、融通の効かないものであるから、これらの作業は単純であっても、決して気楽な作業ではない。しかも、これらの作業は、自然言語で会話をしたり、または、紙に鉛筆で文字を書いたり、ケシゴムでそれを消すというような、自然な、人間に馴染んだ作業ではない。単純ではあっても、記号の並びをキーボードから正確に入力するような、機械的なことがらを正確に覚えておくようには、私たち人間の頭脳はできて

いない。そのような作業を、毎日繰り返すのならまだしも、「忘却に手ごろな間合い」をおいて繰り返すのであるから、私たち人間にとっての負担は決して軽くはない。

このような、何回も繰り返さなくてはならない単純な作業に対して、その度に、それぞれの分厚いマニュアルを本棚から取り出してきて、ページを繰り、読み直すのでは、時間や手間・労力の問題としても、大変な損失である。私は、各々の研究者が、自分の研究スタイルにあった、研究における計算機作業のマニュアルを持っていて当然であると思う。そして、おそらく、そのようなマニュアル作りの基本的な骨組みや、あるいは少なくとも発想などは、多くの似たような研究方法やテーマに携わる研究者間で（例えば、言語表現の意味の記述に関心を持つ研究者の間で）共有できるのではないだろうか。ともあれ、そのような、研究方法のノウハウ公開の試みとして、本論文を執筆する。目標は、情報交換の促進と、取るに足らないものであるにしても、議論の積み重ねである。御批判・御教授をよろしく願いたい。

第一章では、個々での議論の前提となっている、研究上の制約・諸条件について、第二章では言語表現資料をテキスト・データベースから採取する作業について述べる。第三章ではテキスト・データベースから採取した言語表現資料を大型計算機からパーソナルコンピュータに転送し、リレーショナル・データベース・ソフトウェアを用いて記述するためにデータを変換・加工する作業について述べる。第四章では実際に言語記述をする作業について述べ、さらに第五章では、テキスト・データベースから言語表現資料を採取し、リレーショナル・データベース・ソフトウェアを用いて記述する研究方法の本質的問題について若干の考察を加えるとともに、記述作業についての具体的な、しかし、周辺の問題点についても触れる。

1. 議論の前提

ここで参考にした作業は、ある言語表現の使用例を、テキスト・データベースから採取し、その使用例について、リレーショナル・データベース・ソフトウェアを用いて意味にかかわる様々な記述・分析・整理の作業をするというものである。

筆者のおかれた研究条件の制約から、テキスト・データベースは、前述の九州大学大型計算機センターで公開されている「トーマス・マン・ファイル」を使用した。データ検索に利用したのは、やはり、同センターで利用されている「テキストデータベース管理システム SIGMA」である。「トーマス・マン・ファイル」その物が、この「SIGMA」によって使用されることを前提にしたテキスト・データベースでもある。リレーショナル・データベース・ソフトウェアは、日本アシュトン・テイト社の「dBASE-III PLUS」を使用した。エディタはメガソフト社の「MIFES」を使用、また、ワードプロセッサは、パーソナルメディア社の「ビーナス」を使用した。大型計算機センター端末、そして、リレーショナル・データベースソフトウェアとワードプロセッサ使用のためのパーソナルコンピュータとして、「MS-DOS vers. 3.3」をオペレーション・システムする日本電機の「PC-9801vm」とエプソン社の「PC-386VR」を使用した。更に、パーソナルコンピュータを電話回線を通じた大型計算機センター端末として利用するために、オムロン社のモデムと、佐賀大学の渡辺・武政両氏が開発し九大センターで公開されている「日本語 PFDエミュレータ（高速多機能 TSSPFD エミュレータ）」を使用している。

これらの前提条件は、本当に取るに足らないことであるので、相当部分省略したが、実際には、それぞれの研究者がそれぞれの研究条件において、必要に応じて、独自のマニュアル化・手続きの確立を考えなくてはならない。（この場合、マニュアルを個人の使用と限定するか、あるいは、共同研究や学生の教育のための基礎資料として活用するか、

などの要因もまた、マニュアル全体の構成にかかわってくるはずである)

2. 言語表現資料の採取

2.1. パーソナルコンピュータ端末を立ち上げ、大型計算機センターと接続

計算機センターと接続するための通信ソフトウェア(ターミナル・エミュレータ)を、それぞれの計算機センターが数種類用意しているはずである。すでに述べたように、私は、渡辺・武政両氏が開発し九大センターで公開している「日本語 PFD エミュレータ(高速多機能 TSSPFD エミュレータ)」を利用しているが、パソコンの初期設定などについては、一度行なえば、機械を買い替えたり、転勤するなどのことがない限り、必要ないので省略する。また、この程度の所までは、パソコンを納入する業者や計算機センターに相談すれば、面倒をみてくれるべきことなので、計算機に素人の言語研究者が心配しなくてもよいことのようにだ。

パーソナルコンピュータを立ち上げ、端末エミュレータ・ソフトを使い、メニューに従って端末モードにする。次に、問題なのは、計算機センターと接続するための手続きであるが、しばらく、大型計算機を使わないとすぐ忘れてしまう(実際私の場合、大型計算機のテキスト・データベースから、言語使用例を採取してしまうと、パソコンだけで作業することが多く、次に計算機センターと接続するのは、数か月から半年後ということもある。そうすると、課題番号から、パスワードまですっかり忘れてしまうので、このことを好い加減にして置くと本当に困ってしまう)。だから、この手続きは、パスワードの秘密保持の問題は別として、マニュアルにしっかりと組み込んでおくほうがよい。

(1)のように、(1a)の操作によってモデムを通じて計算センターを呼び出し、(1b)の操作によって自分の課題番号とキーワードを使って、計算機の利用を開始する(以下、下線部がキーボードからの入力、「[RETURN]」はリターンキーを押す操作とする)。

「日本語 PFD エミュレータ」の場合、この手続きがファンクション・キーによって可能になっているので、実際の入力(2)ですむ。

- ```

1) ATDP 462 [RETURN] -----(1a)
 CONNECT ----- (センターからの表示)
 JCET005 SYSTEM READY ---- (センターからの表示)
 LOGON TSS A72014K/PASSWARD [RETURN] ----- (1b)
 (省略)
 READY ----- (センターからの表示)

2) SHIFT+F1
 CONNECT ----- (センターからの表示)
 JCET005 SYSTEM READY ---- (センターからの表示)
 SHIFT+F2
 (省略)
 READY ----- (センターからの表示)

```

### 2.2. テキスト・データベースにアクセス

前述したように、テキスト・データベース「トーマス・マン・ファイル」はテキストデータベース管理システムである「SIGMA」によって使用されることを前提に作成されて

いる。それでこのファイルを九州大学の大型計算機センターで使用するのには、まず、「SIGMA」システムに入らなくてはならない。大型計算機が「READY」の表示をしている状態で(3a)のように「SIGMA」と入力すると(3a)の表示がセンターから返され、「SIGMA」のシステムに入る。

```
3) READY ----- (センターからの表示)
 SIGMA[RETURN] ----- (3a)
 SIGMA)----- (「SIGMA」からの表示) -- (3b)
```

「SIGMA」を初めて使う場合は、(3a)と(3b)との間に(3')のやり取りが入り、ユーザーがテキストデータベース管理システムのために使用する領域が大型計算機の記憶領域に設定されてから、「SIGMA」のシステムに入る。(操作(3'a)はリターンキーを押すだけの空入力である、以下同様)

```
3') NUMBER OF BLOCKS:=50[RETURN]
 . . .
 MASTER KEY:=[RETURN] ----- (3'a)
```

ユーザーが使用する記憶領域の大きさの目安は、50 ブロックで1メガバイト程度(5インチの2HD フロッピー・ディスク約一枚分)のものである。私は現在 50 を指定しているが、検索結果としての出力ファイルが大きい場合には、もっと大きなファイルを指定しなければならない。ユーザーは 10から4776までの数値が指定できるが、あまり大きな数値を指定すると、センターから高額の記憶領域使用料を請求され、私は一度胆がつぶれる思いをした。

更に、「トーマス・マン・ファイル」を利用する場合には、サブ・ファイルを指定するキーボードからの入力を簡略化するためのプロファイルの設定をする作業を、(4)のやり取りでもって行くと便利である。

```
4) SIGMA)PROF[RETURN]
 PROF:P[RETURN]
 MEMO OR SIGMA(M OR S)?S[RETURN]
 PREFIX:=A70152B[RETURN]
 PROF:END[RETURN]
```

この設定によって、例えば(5)の入力で、「トーマス・マン・ファイル」の中のそれぞれのサブ・ファイルを簡略化した形で指定できる。この設定をしない場合の入力は、例えば「ブッデنبロック家の人びと」を指定する場合、(5b)のように少しだけ面倒になる。

|                        |                                                              |
|------------------------|--------------------------------------------------------------|
| 5a) <u>S. MANN. BB</u> | Die Buddenbrooks (ブッデنبロック家の人びと)                             |
| <u>S. MANN. DF</u>     | Doktor Faustus (ファウスト博士)                                     |
| <u>S. MANN. ERZ</u>    | Erzählungen (短編)                                             |
| <u>S. MANN. EW</u>     | Der Erwählte (選ばれし人々)                                        |
| <u>S. MANN. FK</u>     | Bekenntnisse des Hochstaplers Felix Krull (詐欺師フェリックス・クルルの告白) |
| <u>S. MANN. JS1</u>    | Joseph und seine Brüder 1 (ヨセフとその兄弟 1)                       |
| <u>S. MANN. JS2</u>    | Joseph und seine Brüder 2 (ヨセフとその兄弟 2)                       |
| <u>S. MANN. KH</u>     | Königliche Hoheit (大公殿下)                                     |
| <u>S. MANN. LT</u>     | Lotte in Weimar (ワイマールのロッテ)                                  |

|                       |                                |
|-----------------------|--------------------------------|
| <u>S. MANN. NT</u>    | Nachträge (補遺)                 |
| <u>S. MANN. RA1</u>   | Reden und Aufsätze 1 (演説と論文 1) |
| <u>S. MANN. RA2</u>   | Reden und Aufsätze 2 (演説と論文 2) |
| <u>S. MANN. RA3</u>   | Reden und Aufsätze 3 (演説と論文 3) |
| <u>S. MANN. RA4</u>   | Reden und Aufsätze 4 (演説と論文 4) |
| <u>S. MANN. ZB</u>    | Der Zauberberg (魔の山)           |
| <u>.S. MANN. ALLE</u> | 全ファイルの指定                       |

5b) S. 'A70152B. MANN. BB'

### 2.3. テキスト・データベースから資料を検索・採取

例えば、私は、前述のように、竹内 1992a、bにおいて、ドイツ語の形容詞「empfindlich (日本語の『敏感な』、『感じやすい』などに相当する)」の意味の揺れを記述し、その意味の核心に迫る試みを行った。「トーマス・マン・ファイル」全体からこの形容詞の使用例を検索・採取するためには、(6a)または(6b)のように、「SIGMA」システムの SIGMA 状態、または DO 状態から「SEARCH」コマンドを入力する。そうすると、システムはレコード区切り記号の入力を促すプロンプトを表示して検索作業に入る(入力部分の下線が前半で終わっているのは、そのような省略形で入力できるということ)。このプロンプト表示以降のやり取りは、(7)の通りである。

```

6a) SIGMA>SEARCH[RETURN]
6b) DO>SEARCH[RETURN]

7) RECORD DELIMITERS
DO1:=#[RETURN] ----- (7a)
DO2:=[RETURN]
 ITEM DELIMITERS
DO2:=[RETURN]
 KEYWORDS
A01:= EMPFINDLICH ----- (7b)
A02:=[RETURN]
 LOGICAL FORMULAS
Z01:=A1
Z02:=[RETURN]
REPORT(Y/N)?[RETURN]
FILE:=.S. MANN. ALLE[RETURN]
PASSNUMBER:=[RETURN]
 SEARCHING S. MANN. BB ...
 RETRIEVED TEXTS
QUESTION 01 (Z01) = 8 8 ----- (7c)
 (省略)
 SEARCHING S. MANN. ZB ...
 RETRIEVED TEXTS
QUESTION 01 (Z01) = 20 91 ----- (7d)
TOTAL = 20 91
CPU (SEC/1000) = 928 10132
FILE:=[RETURN] ----- (7e)
DO:REFILE[RETURN] ----- (7f)

```

```

REPROT(Y/N)?[RETURN]
 RETRIEVED TEXTS
QUESTION 01 (Z01) = 91
TOTAL = 91
QUESTION:=1[RETURN] ----- (7g)
NEW RECORD DELIMITER:=#[RETURN] ----- (7h)
NUMBERING(N/Y)?[RETURN]
OUTPUT-FILE:=EMPF. MANN[RETURN] ----- (7i)
QUESTION 01 (Z01) = 91
QUESTION:=[RETURN] ----- (7j)

```

(7)のやり取りについて最低限の説明を付け加える。(7a):「トーマス・マン・ファイル」の文は「#」で区切られているので、レコード区切り記号は「#」を入力する。(7b):「トーマス・マン・ファイル」のテキストはすべて大文字である。検索文字列「EMPFLINDLICH」の直前に空白を置くことによって「unempfindlich」などの合成語が採取されることが防止される。(7c): 検索第一番目のファイル「ブッデンプロック家の人々」では、検索例は 8例、最初の検索であるから、合計も 8例である。(7d): 検索最後のファイル「魔の山」では、検索例は20例、検索例の合計は 91例という結果になった。(7e): これ以上の検索はしないので空入力。(7f): 検索結果をコピーヤリスト、検索などの操作可能なファイルに変換。(7g): ファイルに変換する検索項目を入力。ここでは、一項目のみの検索だったので、検索番号「1」を入力。(7h): 新しいレコード区切り記号も「#」。(7i): 変換したデータのファイル名は「EMPF. MANN」。(7j): 他には変換を行わないので、空入力。ここまでの作業で、「トーマス・マン・ファイル」中の「\_EMPFLINDLICH」という文字列の含まれる例がすべて、「EMPF. MANN」というファイルに集められた(注釈:「\_」は空白)。

「トーマス・マン・ファイル」は大文字で構成されていて、自然言語としての大文字を「〈」記号を前置するなど、融通の効かない計算機による処理を可能にするために様々な工夫をこらした、しかし、私たち人間には読みにくいデータである。検索結果のデータの始めの数行を(8)に示すが、計算機による検索作業を終了した「EMPF. MANN」は、もはや、このような、読みにくい形式を必要とするものではない。むしろ、私たち人間が読みやすい、大文字・小文字の混在する、自然な形式で構成されていることが好ましい。「SIGMA」システムにおいて「トーマス・マン・ファイル」を利用する場合に、テキスト・データベースから抽出したファイルを大文字・小文字混在の形式に変換するためには、(9)のやり取りをすればよい。変換後のファイル名を変換前と同じにすると、変換前の内容が消去されて、新しい内容が書き込まれる(上書きされる)。大文字・小文字混在の形式に変換した後のデータの始まりの部分のが(10)に示されたものである。

```

8) #0115535 DANN LEHNTE SIE SICH ZUR=UCK UND SCHLO$ DIE <AUGEN , DIE M=UD
 E UND EMPFLINDLICH WAREN .
 #0125932 WAR DER VERSTORBENE <KONSUL , MIT SEINER SCHW=ARMERISCHEN <LI

```

```

9) DO:=. S. SMALL[RETURN]
 INPUT-FILE:=EMPF. MANN[RETURN]
 OUTPUT-FILE:=EMPF. MANN[RETURN]
 REPORT(N/Y)?[RETURN]
 INPUT-FILE:=[RETURN]

```

- 10) #0115535 dann lehnte sie sich zur=uck und schlo\$ die Augen, die m=uide und empfindlich waren.  
#0125932 war der verstorbene Konsul, mit seiner schw=armerischen Liebe

## 2.4. 大型計算機からパーソナルコンピュータへのデータ転送

大型計算機は高速・大容量であるから、「トーマス・マン・ファイル」のような大きなデータを公開し、検索作業をするのには適しているかもしれない。しかし、小回りの効く、使い勝手の良さ、という点では、パーソナルコンピュータに軍配があがる。例えば、数ヶ月来、私のキャンパスと大型計算機を結ぶ電話回線のセンター側のモデムの調子が悪いということで（センターではいろいろと回復の努力をしてくださっているようではあるが）、「PFD」というシステムを使って大型計算機上でデータの画面編集をすることができないでいる。パーソナルコンピュータだけの仕事であれば、故障した機械を修理に出せば、数週間で返ってくるだろう。または、回線のスピードが遅く（この問題は、私の研究予算の乏しさに由来する問題で、高速のモデムを購入し、料金を支払って NTT の回線を利用できれば、解決するのだから、八つ当たりの的ではある）、データが端末と大型計算機の間を行き来する間、表示がディスプレイ画面をたらたらと流れるのをいらいらしながら眺めているのは精神衛生上非常によくはない。

そういう訳で、検索・採取済のデータは、なるべく早く大型計算機から転送してパーソナルコンピュータで扱った方が仕事の能率が上がる。「日本語 PFD エミュレータ」には、大型計算機のファイルと端末のディスク上のファイルをやり取りする機能があるので、その機能を利用する（前述の通り、回線のスピードが遅いので、私の場合は、かなり時間がかかる）。転送は「SIGMA」システムの中に入ったまま行なうが、そのやり取りは以下（11）の通り。

```

11) DO: ----- (11a)
 F2 ----- (11b)
 (省略) ----- (11c)
 DO: ----- (11a)
 F3 ----- (11d)
 表示データをファイルに書き込みます ----- (11e)
 DO:LIST EMPF.MANN[RETURN] ----- (11f)
 #0115535 dann lehnte sie sich zur=uck und schlo$ die Auge
 es besteht kraft seiner Nat=urlichkeit und seinerintakte
 n Menschlichkeit.
 #
 DO: ----- (11a)
 F3 ----- (11g)
 表示データのファイルへの書き込みを一時中止します -- (11h)
 DO: ----- (11a)
 F2 ----- (11g)
 表示データのファイルへの書き込みを終わります ----- (11i)
 DO: ----- (11a)

```

例によって、最低限の説明を付け加える。(11a): 「SIGMA」の DO 状態における入力待ちのプロンプトであるが、次の操作が、ファンクション・キーを用いたパーソナルコンピュータへの指示なので、なにもしない。(11b): ファンクション・キー F2=「HOST」を押すと、エミュレータが指示を出すので、それに従って、フロッピー・ディスク上に

ファイルを指定する・指定が終ると、ファンクション・キー「F2」の表示が黄色に変わる。その後、大型計算機の端末としては「SIGMA」システムに戻る。(11d): 「F3=DISKW」を押すと(11e)の表示を出して、ファンクション・キー F3 の表示が黄色に変わる。端末としての状態は「SIGMA」システムの入力待ちのままであるが、パーソナル・コンピュータは、画面に出力された情報をすべてディスクに書き込む状態になる。

(11f): 「SIGMA」システム内に作製されたファイル「EMPF.MANN」を画面に表示してから、「SIGMA」の入力待ち状態に戻る。パーソナル・コンピュータはこの画面に出力された情報をすべてディスクに書き込む。(11g): ファンクション・キー「F2」または「F3」を押すと、それぞれ、(11h)または(11i)の表示を出して、そのファンクション・キーの表示が元の色に戻る。パーソナル・コンピュータは、それぞれ、表示データのファイルへの書き込みの一時中止、または書き込みの終了の状態になる。端末としては「SIGMA」システムの入力待ち状態のまま。

## 2.5. 「SIGMA」システム、大型計算機接続、端末エミュレータの終了

「SIGMA」システムの終了と、大型計算機接続の終了についてのやり取りは(12)の通り。これで、計算機は、利用負担金などについての情報を表示して「SYSTEM READY」のメッセージを出す。ファンクション・キー「F10=MENU」を押して、指示に従って、エミュレーター・プログラムを終了する。モデムの「PHONE」ボタンを押して、電話回線の接続を切る。

```
12) DO:END END[RETURN]
 READY
 LOGOFF[RETURN]
 (省略)
 JCETO10 SYSTEM READY
```

## 3. リレーショナルデータベースを利用した言語記述のための、パーソナルコンピュータ上でのデータ変換・加工

### 3.1. パーソナルコンピュータのエディタとワープロを使った手書き作業カードの作製

パーソナルコンピュータを利用して言語記述作業を行なうといっても、私の経験から言って、言語事例は、やはり一例ごとに、図(13)のような形式の用紙に出力して、まず最初は、紙と鉛筆でもって記述作業をすることが必要であるし、能率もよい。

#### 13) 手書き用の記録用紙

テキスト

自由記入部分

なぜなら、パーソナルコンピュータは、いくら便利であるといっても、やはり、紙に鉛筆で書き込む作業の自在さ、そして、ペラペラとめくり、シートを較べてみるという紙の扱いやすさには勝てないからである。この、作業を行なうために、パーソナルコンピュータのエディタやワープロを使って、分析・記述をする人間にとって読みやすく、使いやすい記録用紙を制作する。

大型計算機からパーソナルコンピュータに転送したデータは、(10)と同形式のものなので、ドイツ語特有の文字「ä」、「ö」、「ü」、「Ä」、「Ö」、「Ü」、「ß」などを「=a」、「=o」、「=u」、「=A」、「=O」、「=U」、「\$」で代替していて、まだまだ読みやすいとは言えない。また、自然な文章ならば、適切に行われているはずの分綴も行われていない。これらの、文字の変換、分綴によって、文章としての様式を整えることで、言語資料に立ち向かう場合の、視覚インターフェイスの障壁を普通の読書なみに低くすることができる。

この作業については、パーソナルコンピュータやワープロ・エディタの機種・ソフトウェアの種類が多様であり、その操作も比較的日常的に馴れ親しんでいるものなので、詳述しない。これらの些細な作業は、つまらないゼイタク（紙や時間・労力の無駄使い）のように思われるかもしれない。しかし、計算機がパーソナルコンピュータとして普及し、手軽に扱える時代に生きる私たちにとって、マン・マシン・インターフェイスの問題をできるだけ少なくする気配りが必要なのであり、それによって、便利なのは機械を利用して、能率を悪くしたり、疲労をため込んだりすることが避けられるのである。

### 3.2. テキストデータのリレーショナル・データベース・データへの変換

市販のリレーショナル・データベース・ソフトウェアの種類も増えており、筆者は、どの製品が優れているとか判断できる立場にはない。むしろ、どの製品が優れている、などの情報に振り回されて、時間とエネルギーを無駄にしまったりした経験だけが残っている。現在使用しているものと仕様・操作感覚の違うソフトウェアの扱いになれるのに時間を取られ、それにもかかわらず、結局の所、これまでのものと機能や扱いの点で大きな違いがないのにつかりさせられたりするのである。

それで私は、機能の点では不満タラタラなのであるが、前述の「dBASE III PLUS」を使い続けている。最初にこのソフトウェアを使うことになったという「くされ縁」からである（このソフトウェアは、結局のところ、顧客管理や売上処理のために作られているので、例えば、半角 254 文字以上のデータは、メモファイルという特殊な処理をしないで扱えず、しかも、そうすると、参照しにくい・他のファイルとの互換性がないなど、文科系の研究データを処理するのには、本当に使いにくい）。しかし、いずれにしても、限られた研究資源の制約のもとで作業をしなくてはならないのだから、この「dBASE III PLUS」を用いた作業手順についての覚え書きを書き記しておくことにする。

ここでは、リレーショナル・データベース・ソフトウェアを用いて、テキスト・データベースから収集した言語資料についての記述作業をしようというのだから、最低限、テキスト・データベースから収集した言語資料がデータの一部として組み込まれなくてはならない。そのためのデータ構造を定義し、テキスト・データベースから収集した言語資料を、このデータ構造に読み込む手順について述べる。

### 3.2.1. 「dBASE III PLUS」の起動とデータ構造の仮決定、「dBASE III PLUS」の終了

すでに述べたように、このソフトウェアでは、通常モードでは反角で 254 文字以上のデータは扱えない。言語資料というのは、その程度の文字数でおさまる性格のものではないのだが仕方がないので、最大の文字幅を取ったデータ項目（以下では、「dBASE III PLUS」の語法にならって「フィールド」と呼ぶ）を一項目、言語資料そのものために用意する。更に、この言語資料についての、出典を表わすインデックスのための 9 文字分のフィールドも必要である。それ以上、どのような記述項目を設定したらよいのか、という点については、それぞれの記述対象の性格、記述目的などの条件にあわせて、考慮しなくてはならないことであり、章を改めて説明する。

とりあえず、以上の言語資料のインデックスと、言語資料本体のための二項目を持つデータ構造をリレーショナル・データベース・ソフトウェア「dBASE III PLUS」上に定義する。ハードウェアに合わせたアプリケーション・ソフトウェアの初期設定などの問題は省略すると「dBASE III PLUS」を起動するには「dbase」というコマンドを「MD-DOS」のシステムから入力する。

「dBASE III PLUS」では、データのそれぞれの項目が格納されるスペースをフィールドと呼んでいる。ここで定義されるフィールドはそれぞれ「NUMBER」、「TEXT」と名づけることにする。データ形式は文字型、データ幅は 9 と 254 である。また、これらの定義をしたデータの格納されるファイルを例えば「B:EMPF.DBF」としよう。以上のことを定義するために「dBASE III PLUS」では、「CREATE STRUCTURE」コマンドを用いる。このコマンドを用いると、メニュー方式でデータ構造が設計できるので、詳細は述べない。また、このコマンドを用いるのにも、「ASSIST」コマンド、またはそれと同等のファンクション・キー「F2」から、メニュー方式の操作が可能である。

また、データ構造を定義するにあたって、すでに類似の記述作業を行っていて、その時に使用したデータ構造が応用できる場合、「COPY STRUCTURE TO」コマンドを用いて、データ構造のみを新しいデータベース・ファイルに移し換えることができる。この作業を行なうには、古いデータベース・ファイルを使用する状態から新しいデータベース・ファイルを指定してコマンドを実行することになる。例えば、データベース・ファイル「B:EMPF.DBF」を使用しているとすると、(14)の入力を行なうと、このファイルのデータ構造が新しいデータベース・ファイル「B: SCHWER.DBF」に転写される。

#### 14) COPY STRUCTURE TO B: SCHWER.DBF [ RETURN ]

以上でデータ構造が定義できたら、メニューを終了し、さらに「quit」コマンド（ファンクション・キー「SHIFT+F8」）で「dBASE III PLUS」を一応終了しておく。

### 3.2.2. リレーショナル・データベース・データ変換のためのテキストデータの加工

「dBASE III PLUS」でテキスト・データベースから収集した言語資料を扱うためには、言語資料を 254 文字以下に圧縮しなくてはならず、分綴について考慮することもできない。また、このソフトウェアでは、ドイツ語特有の文字は扱えない。私は、このソフトウェアで言語資料を扱うために、英語タイプライターを用いてドイツ語を表記する場合の慣用に従い、「ä=ae」、「ö=oe」、「ü=ue」、「ß=ss」の代替を採用し、大文字・小文字混じりの文字面を選択することにする（それが自分の視覚にとっていくらかでも負担の少ないからであるが、手間を省くことを優先するのだったらトーマス・マン・ファイルの表記法をそのまま引き継ぐという選択もあるだろう）。この作業は、前章で行つ

た、手書き作業カード作製とは全く独立に、または、そのカード作製に用いたファイルから、行なうことになる。

言語資料を圧縮するためには、実際には、記述の目的を手書き作業カードによる言語資料の内容に依存した整理の作業が前提となるが、そのことについては章を改めて述べる（第四章）。その、整理作業とは別に、単純・機械的な圧縮手段としては、ドイツ語を扱う場合、辞書や事典などで慣習化している省略表記法の利用、再現が可能な限りにおいての母音の省略法、固有名詞の頭文字による代替などの工夫が考えられるが、この処理も、やはり記述目的・条件により扱いが異ならざるを得ない。例えば、形態論的な問題に感心がある場合は、線的な言語資料について、いかにその前後の切断位置を工夫するかが大切であって、省略などは問題外である。

ここの作業についても、やはり、ワープロ、またはエディタを用いた作業なので、詳細は述べないが、ワープロやエディタの検索・置換機能などを用いた効率的な手順の工夫が必要である（このパラグラフの作業は、前項で述べた、マン・マシン・インターフェイスをできるだけ人間サイドにやさしいものにならなければならないという原則とに全く対立する。機械側の条件に合わせて、人間が不便を忍ばなくてはならないということは、本当にいらだたしい）。

テキスト・データをリレーショナル・データベース・データに変換するための前提として、さらに、テキスト・データの構造をどのようにリレーショナル・データベース・データの構造に対応させるのかという問題がある。いくつかの対応の取り方があるが、ここでは、最も単純な手順を説明しておく。

テキスト・データが改行で区切られることを、NUMBER、TEXT の二つのフィールドから成り立つデータの単位（レコード）と対応させ、最初の 9 文字分を NUMBER フィールドのためのデータ、残りの改行までのデータを TEXT フィールドのためのデータとする（インデックスそのものの幅は 8 文字であるが、テキストとの間の空白をインデックスのデータとして読み込むことにする。この空白をテキスト側のデータとすると、テキストの文の記入スペースが 1 文字減ってしまうからである）。この仕方で加工したテキスト・データを、2 レコード分以下 (15) に示す。

- 15) #0115535 dann lehnte sie sich zurueck und schloss die Augen, die muede und empfindlich waren  
#0125932 war d. verstrb. K....d.l.s.Geschlechtes gewesen, d.unalltgl., unbrgrl.u.differenz.Gefuehle gekannt u.gepflegt hatte, so schienen s.bd.Sohne d.erste BBs zu sein, d.vor d.freien u.naiven Hervortreten solcher Gefuehl empf zurueckschreckten.

このように加工したテキスト・データを、例えば「B:EMPF.TDB(TO DATA BASE)」という名のファイルに格納しておく。

### 3.2.3. リレーショナル・データベース・データへの変換

前項で加工したテキスト・データをリレーショナル・データベース・データに変換するために再び dBASE III PLUS を起動する。前々項で定義したデータ構造のデータ・ファイルを指定し、テキスト・データをリレーショナル・データベース・データに呼び込むコマンドは以下 (16) である。

16) USE B:EMPF.DBF [ RETURN ]  
APPEND FROM B:EMPF.TDB SDF [ RETURN ]

これで、テキスト・データがリレーショナル・データベース・データに変換される。リレーショナル・データベース・データの第一番目のレコードは(17)のようになっている。

| 17) フィールド名 | データ                                                                                  |
|------------|--------------------------------------------------------------------------------------|
| NUMBER     | #0115535                                                                             |
| TEXT       | dann lehnte sie sich zurueck und schloss die Augen, die muede und empfindlich waren. |

#### 4. 言語記述作業

リレーショナル・データベースを用いて、言語記述を行なうといっても、すぐさま、キーボードに向かい、データ構造を決定し、記述を機械的に行ってゆくというような、インスタントなトコロテン式作業が可能なのはならない。それは、言語研究の奥行の深さ、対象領域の複雑さからいって当然のことである。だからこそ、3.1.項において手書き作業カードの作製を行った。手書き作業カードを用いて、暫定的なデータ記述を行なうから、さらに進んだ段階において、リレーショナル・データベースなどの機械の助けを借りた作業が有効になってくるのである。一般的に、手書き作業を、計算機の助けを借りた記述、そして、分析結果をまとめる、論文執筆などの作業という研究完成への手順が考えられるのではないだろうか（もちろん、研究の動機づけや記述対象の選択・または、問題点に遭遇した場合の研究文献を用いた思考作業を別にすることである）。

##### 4.1. 手書き作業カードを用いた暫定的記述

研究・記述目的によるが、大抵の場合、第一ラウンドの暫定的記述においては、「とりあえず」の、「おおまか」で、「直観的」な記述作業が行なわれれば、それでよいのではないだろうか。記述を行なうからには、研究者はその記述対象に対して、自分の見識に基づいた何らかの予測、直観をもっていて、その予測の正しさを経験的に検証しようとするものである。しかし、どのような研究対象においても同じことなのだろうが、言語現象についても、実際記述をしてみると、意外な、不思議な、予測に合わない、不可解なことがでてくるものである。（そうでなかったら、データにあたり、記述・検証などを行なう必要がない！）

以下に、私がドイツ語の形容詞「empfindlich」の記述を行なった場合に手書きで行なった記述項目を参考までに挙げておく。繰り返すが、これは、その時の研究・記述目的に応じて、臨機応変に設定することがらである。

- 18) ドイツ語テキストへの日本語の対応（翻訳）  
 表層的な統語構造についての記述  
 直観的な意味記述  
 抽象的な修飾関係の記述  
 意味を成立させる文脈・常識のはたらき、などの要因についての記述  
 その他の覚え書き・コメントなど

ここでの記述は、リレーショナル・データベースを用いた記述への予備的作業であるから、できるだけ、柔軟な姿勢で臨むことが必要である。なぜなら、いくら、パーソナルコンピュータが大型計算機よりも融通が効くといっても、やはり計算機は計算機、なかなか融通が効かない。次項で述べるが、リレーショナル・データベースを用いた記述においても、定義された記述事項についての記録しかできないのである。それに較べて、紙の上の鉛筆の作業は、書いたものを消したり、復元させたりすることにおいては、ある種の能率は悪いものの、空間の使い方という点では、変幻自在・融通無碍・勝手気儘である。記述対象が、自然言語という、自由自在な奥行の深い、柔かな対象である以上、この手書き作業の有利な側面は最大限に活かされなくてはならず、またそのことが、後に続く抽象化という作業をより良質のものとするはずなのである。

#### 4.2. 記述項目のとりあえずの設定（データ構造の仮設計）と リレーショナル・データベースへの入力 記述途中でのデータ構造の変更

前項の手書き作業を行なうことによつて、または、その作業を行なうという研究者の行為そのものにおいて、対象を記述するために、そして、その記述から対象についての何らかの知見を得るために、どのような記述項目が必要とされるのかが次第に明らかになってくるだろう。一応の手書き作業が終了した段階で、この記述項目を整理し、リレーショナル・データベースのデータ構造に組み込む作業が必要になる。

私が、今回の記述において設定した記述項目の一部は、以下(19)のようなものであった。パーソナルコンピュータを用いた作業の利点は、データ量や記述項目数を気にしないですむこと、データの変更・取り消しが可能であることである（ただし、取り消したものの復活は難しい）。だから、これらの項目は、ある程度思いつくままに、無駄と思われることがらについても、設定しておいて構わないだろう。（ハード・ディスクの普及している現在、記録スペースについてはあまり気にする必要がなくなってきた）

- 19) 表層の統語構造
  - 修飾表現（係りの対象）
  - 修飾の仕方（付加語的か述語的か）
  - 修飾表現（の指示対象）の性質
  - 被修飾表現
  - 被修飾の内容
  - 直観的な意味記述
  - 日本語への対応
  - 平行する表現の記述
  - 意味的な修飾構造
  - 感受する主体
  - 感受する主体を表わしている言語表現
  - 感受する主体を表わしている言語表現の文法的記述
  - 感受する主体の性質
  - 感受する主体を特定させる要素
  - 感受の枠組
  - 感受の対象
  - 感受する主体を表わしている言語表現の文法的記述
  - 感受の対象の性質
  - 感受の対象を特定させる要素
  - 否定的・肯定的ニュアンスについての記述

実際には、項目の数は 50 近くになったが、それらの項目についての記述は、あまり意味のないものから、当面の研究目的にすぐ役立つもの、将来的な研究対象として興味深いものまで、結果として多岐に及んだ。これらの項目について、リレーショナル・データベースのデータ構造に組み込むためには、「dBASE III PLUS」を起動し、以下 (20) の手順で、記述を行なうデータ・ファイルを指定し、「MODIFY STRUCTURE」コマンドでデータ構造の変更をおこなう作業に入る。データ構造の変更作業については、メニュー方式で行われるので、説明は省略する。

```
20) USE B:EMPF.DBF [RETURN]
 MODIFY STRUCTURE [RETURN]
```

以後「EDIT」コマンドを使って、手書き作業の記録を参照しながら、それぞれの項目について入力を行なっていく。これは、「dBASE III PLUS」の上では、データ構造の変更作業によって、実際は、すべてのレコードの新しく追加されたフィールドに空白のデータが入っているの、レコードを変更するという作業になる。「EDIT」コマンドを使ったレコード変更作業に入るための入力と、このモードにおいてよく使う操作手順を (21) に列挙する。

```
21) EDIT [RETURN] -- レコード変更作業に入る
 ←→ ----- カーソルの左右の移動
 CTRL+A ----- カーソルの単語単位の左移動
 CTRL+F ----- カーソルの単語単位の右移動
 ↑ ↓ ----- フィールドの上下の移動
 CTRL+C ----- レコードのページ (画面) を一枚めくる
 CTRL+R ----- レコードのページ (画面) を一枚戻す
 CTRL+W ----- レコード変更作業の終了
 ESC ----- レコード変更作業の中断 (ポインタのあるレコード
 の変更は実行されない)
```

以上の作業を行ないながら、データベースの記述を行っていくが、この作業において、フィールドの追加・変更が当然必要になってくる。繰り返しになるが、そのような、予測した方針ではどうしても記述できないことがらに会えるからこそ、研究は面白い。困難にぶちあたり、頭を悩ませて、方針を変更し、考え方・ものの理解を新たにすることにこそ記述・分析作業の意義がある。考え方、つまりものの理解の仕方が更新されるから、記述項目の変更が必要になるのであって、その場合にはもう一度 (20) の手順に従って、データ構造の変更を行なう作業に入る。この場合に気をつけるべきことがいくつかある。

まず、新しいフィールドをこれまでのフィールドの間に追加する場合には、その位置にポインタを移動し「CTRL+N」の操作を行ない、新しいフィールドを設定する。これまでのフィールドの位置をデータ構造の中で同じフィールド名で移し換える場合には、手順が少し面倒である。新しいフィールドを同名のフィールドで設定すると、フィールド名が重複してしまうので、これまでであった位置のフィールドを削除 (「CTRL+U」) してから、正確に同名のフィールドを新しく設定する。この場合、誤って、一文字でも異なるフィールド名を設定して、データ構造の変更を実行してしまうと、そのフィールド名のデータがなくなってしまうので、本当に気をつけなくてはならない。

またフィールド名を変更する場合には、データ構造変更によって、新しいフィールドを設定してから、「dBASE III PLUS」のコマンド・モードにいったん戻る。それから、「REPLACE」コマンドを用いて、古いフィールドを新しいフィールドに移し、その後で、

また、データ構造変更によって古いフィールドを削除する。例えば、表層の統語構造を記述するフィールドの名前を「SYNT」から「SYNT\_STR」に変更する場合、以下(22)の通り「REPLACE」コマンドを入力すると、すべてのレコードが新しい「SYNT\_STR」フィールドに移される。その後にも「MODIFY ASTRUCTURE」コマンドによる作業によって、古いフィールド「SYNT」を削除する。

22) REPLACE ALL SYNT\_STR WITH SYNT [ RETURN ]

一つの特定のフィールド、例えば統語構造を記述するフィールド「SYNT\_STR」について、一連のレコードの変更を行なう必要の生じることある。例えば、このフィールドの記述作業の途中で、記述の方法を(23a)のような言語を用いたものから(23b)のような形式的なものに変更することにしたとする。そうすると、これまで記述したレコードのこのフィールドについて、すべてデータを書き換えなくてはならない。そのような場合には「BROWSE」コマンドを用いて特定フィールドのみの変更作業に入ると、「EDIT」コマンドを用いる場合のように、いちいちレコードのページ(画面)をめくらなくてはむので効率的である。特定フィールドのみの変更作業に入る「BROWSE」コマンドの入力と、このモードにおいてよく用いる操作手順を(24)に示す。

23a) empfindlich は sein と結びつき、述語として、関係代名詞 die によって指示される名詞句 die Augen に係る

b) ((sein empfindlich)(die\_rel=(die Augen)))

24) BROWSE FIELDS SYNT\_STR [ RETURN ] -- 特定フィールドのみの変更作業に入る  
 CTRL+C ----- 画面を一枚めくる  
 CTRL+R ----- 画面を一枚戻す  
 CTRL+W ----- レコード変更作業の終了  
 ESC ----- レコード変更作業の中断 (ポインタのあるレコードの変更は実行されない)

以上のような手順の繰り返しによって、言語記述作業が進められ、言語研究のための分析を加えられた資料がリレーショナル・データベースのデータとして作製される。

## 5. 記述作業の諸問題と若干の考察

### 5.1. 記述と抽象化について

前項で、言語資料の記述・分析しリレーショナル・データベースのデータ化する作業について述べた。

すでに述べたように、私は実際、このような記述作業を行って、ドイツ語の形容詞「empfindlich」の意味の核心に迫ろうとした。詳しくは、竹内 1992a,b の論文を見ていただくしかないが、記述作業の結果、形容詞「empfindlich」の意味を絞り込みすことができたと思っている。一つの言語表現についての様々な意味のヴァリエーションは、その言語表現の骨組みをなす非常に抽象的な「意味の核心部分」と、人間の豊かな知識システムのインタラクションによって成り立つ。そして、その場合、言語表現の意味構造は、知識プロセッシングの中で、情報の結びつきを限定的な形で制御し、方向づけている。そのような明快な結論に至ることができた。

この、私の研究の方法的な特徴は、以下のようにまとめることができる。言語現象の資料としては、できるだけ間口を広く、様々な例外をふくむ事例を集めること。（といっても、このような仕事の場合、記述対象の性格上、一つ一つのレコードの記述には、かなり時間を要するものである。個人で作業をして、研究結果を出そうという場合、100例前後のデータについてまず、作業をしてみる。そしてさらに必要な場合は、より多くの資料に当るとというのが、研究の実際条件から考えての目安になるだろう。）そして、その間口を広くして集めた資料について綿密な検討を加え、経験的な手続きを経てことからの本質的・核心に迫ってゆくことを目指したのである。

手書き作業カードを用いて、まず、言語による直観的な記述を行ない、さらに、リレーショナル・データベースのデータとして、分析を加えてゆく研究方法の一つの利点は「繰り返し」にある。作業手順の性格上、この方法においては「繰り返し」が幾重にも必要とされるのである。

人間の知性の働きというものは、カッチリとした計画・方針を立てて、それに従って、決まった作業を積み重ねれば、豊かな成果が得られるというものでは必ずしもない。むしろ、無駄に思われるかもしれない、同じようなことの繰り返しのなかで試行錯誤を重ねながら、そのことからの思いがけない点に気づき、ことからの本質に迫ることが、ひょつとしたらできるかもしれないのである。（私は、九年間福岡市に暮らして、今年初めて、「福岡市の木」であるクロガネモチが冬になると、とてもきれいな赤い実を濃い緑の葉陰からのぞかせることに気がついた。とても感動している。福岡市には、このクロガネモチが街路樹として至る所に植えられているのである）

もちろん、研究者はなんらかの予測があるからこそ、その記述対象を選ぶのだが、間口を広くして集めた資料について、その予測を実際に検討を進めてみると、やはり、予測はそのままでは通らない。当初の予測を修正するか、または廃棄し新しい予測をたてるか、ということが必要になってくる。この場合、自己の思考過程を点検し直す、または研究文献に当り、ことがらへの切り込みの視点の正当性について点検を加えるなどの作業が介在しなくてはならないのはもちろんである。科学の仕事の基本が「仮説・検証」の繰り返しであることを考えれば、ここで取られたような、繰り返しと試行錯誤を内在的に必要とする研究方法は理にかなったものであるといえるのではないだろうか。

私の行なった研究においては、言語資料に繰り返し繰り返し分析を加え、幾度も予測を変更してゆく過程において、言語表現の様々な意味のヴァリエーションの基底を成している意味の骨組みが抽象化されていった。言語資料の記述・分析を試行錯誤的に繰り返す過程そのものが、言語表現の基底にある何らかの抽象的な意味構造への、洞察を得る作業なのである。それは、言語資料そのものの背景を成す、言語テキスト生成の過程に含まれる様々な指向性のベクトルを探り当て、整理し、さらに、その言語表現の意味構造に固有な複合的ベクトル結合の可能性を抽出する作業であるともいえる。

言語表現の意味構造の抽象化の過程とは、個別の言語資料において、その言語資料の特殊な条件に依存する意味の要素を切り分けてゆく作業に他ならない。しかし、この作業は、言語資料における個別的な意味の要素を切り捨てるのでは決していない。個別的な意味の要素を「ゴミバコ」に捨ててしまうのではないのである。個別的な意味の要素は、個別的な条件と関連づけて整理される。コンテキストやコミュニケーション状況などの個別的な条件があるからこそ、言語表現は個別・特殊な意味を伝え・個別・特殊な解釈が成り立つのである。この作業のために、リレーショナル・データベースにより記述項目を変更しながら記述してゆく方法は以下の点で有効である。この方法によれば、意味の核心的部分についての記述が残るだけではない。核心部分以外の意味要素であると判断されたことがらについて、その意味要素がどのような条件のもとにその言語資料に立

ち現われるのかを、新たに記録項目を設定して、記録に残すことができる。さらに、その記録項目について、すでに記録済みのレコードを対象に繰り返し検討し直すことが可能となる。

より具体的には、個別的な意味の要素を個別的な条件と関連づけて整理し、その記録をデータとして集めることにより、個別的な事例において言語表現の意味を揺らす諸要因について、まとまりをもった記述ができあがる。私の研究では、このような諸要因が人間の知識システムとして捉えられた。このような人間の知識システムについての知見が得られるということは、言語表現の意味の核心を探るという研究の目的からは外れていることではある。しかし、言語の意味の核心を追求する研究の副産物として、そのような人間の知識システムの存在が確かめられ、また、その構造の一端がうかがえることは決して無意味なことではない。人間の言語行動、テキスト生成活動は、いずれにしても、人間の包括的な認知行動の一部である。であるなら、言語表現の意味の問題も、最終的には、やはりそのような科学の大枠において理解されるべきことからであるからだ。

## 5.2: 記述作業の諸問題

このような作業によって構築されたデータベースは、その作製の過程において、対象となった言語資料についての洞察を与えてくれた。また、データベースに蓄えられた記録を参照しながら、言語の問題についての新たな考察を進めることも可能である。しかしまた、このデータベースを、今後の可能性として、例えば、他の研究で作製された同様のデータベースと結合し、利用することも可能であろう。今回の仕事では、ドイツ語の形容詞「empfindlich」の記述を行なったが、この形容詞とパラディグマをなす「traurig」、「grücklich」などの形容詞群の記述を同様の手法で行ない、例えば、その「感受の主体」についてのデータを取り出し・まとめることができる。そのようにまとめられたデータベースを利用して、この形容詞群の「感受の主体」になりうる対象の掘りについて総合的、または比較・対照的考察を行なうことも可能である。リレーショナル・データベースの用い方とすれば、むしろ、そのような、データベース間の結合・相関関係を利用した積極的な利用が、発想としては本筋なのだろう（そのようなリレーショナルな作業のために、ここで用いた「dBASE III PLUS」というアプリケーション・ソフトウェアがよくできているとは、あまり思えないのだが）。

そのような、複数データベースを結合させ、相関された利用を、将来的な課題として見すえておくならば、少なくとも、以下の点は考慮に入れなくてはならない。一つには、記述項目のある程度の共有化である。今後、リレーショナル・データベースを用いた、同様の研究が増えるのならば、データの相互利用のためにも、記述項目について、ある程度の共通の枠組を設定できないと（意味研究なら、その分野に限ってでも）せっかく作り上げたデータベースが、その場限りにおいて利用されるだけで、あとは死蔵さるということになってしまう。さらに、その記述項目についての、記録の様式についても、同様の配慮がなされないと有効・有機的な活用ができないことになってしまう。

実際には、このような研究の方法が、ある程度浸透し、様々な経験を踏まえられない限り、このような問題についての生産的な方向の議論は為され得ないだろう。しかし、この、記述様式の問題は、データベースの結合や相互利用ということを考えるまでもなく、単一のデータベースの制作においても重要である。とすれば、私たちはやはり以下に述べるような工夫を、たとえその一つ一つが他愛のないものであっても、意識的な経験の積み重ねとして実践してゆく必要がある。

現状の計算機は、何の工夫もしない場合、「形容詞」という記述と「adjective」とい

う記述、または「adj」のような省略記述が同等のものであるとは判断しないから、ばらばらの記述をされた記述項目について、種類分け・数え上げ、など、計算機の能力を活かした作業が成り立たないだろう。しかし、人間は、記録作業を行なうに当って、その時の気紛れや、記憶違いで、ある時は「adjective」と書くところを、ある時は「adj」のような省略した書き方をしてしまう。あるいはキーボードの操作ミスということだってあるだろう。このようなことを避ける工夫にはいろいろあるだろうが、記録項目ごとに記入様式のリストなどを作って、気をつけることが基本的な対策だろう。しかし、より積極的には、ファンクション・キーなどに、よく使う入力を登録して、記入の誤りをなくす工夫も有効である。しかし、代用できるキーの種類は限られているから、登録の書き換えが頻繁に必要なことになる。いっそのこと、ソフトウェアのデータベース機能をそのまま用いて、入力のための辞書を制作し、「形容詞」、「adjective」または「adj」のいずれの書き方をしても、「形容詞」とデータに記入されるような工夫もできる。

また、周辺的なことなので簡単に触れておくだけにするが、言語資料を記述する作業について、次のような配慮が役立つかもしれない。テキスト・データベースから採取したデータは、コマ切れで、言語記述のために、特に意味の問題などを扱う場合には、適していない。文脈や、話題・情報の流れを追うためには、時には数十ページにわたって、内容を読み込まなければ、適切な記述ができないことすらある。この点で、大型計算機は使い勝手の悪いから、テキスト・データベースに入って、資料に当るよりも、図書そのものの頁を繰った方が、効率は良い。このような作業のために、例えば「トーマス・マン・ファイル」を利用して、意味記述作業を行う場合には、オリジナルの全集や、翻訳などをできるなら手許に置けた方が良く、また、全集の目次のリストなどを用意して、文脈についての把握が容易になる工夫が必要である。この点でまた、翻訳などの文献リストの作製も、作業の効率化のために有効である。

さらに、このような計算機を使って、言語資料の記述作業を効率良く行なうためには、「MS-DOS」のバッチ・ファイルのレベルからプログラムを組んで、大型計算機の端末エミュレータ、リレーショナル・データベース・ソフトウェア、エディタなどを有機的に結びつけると便利である。例えば、「ファンクション・キーへの登録の書き換え」は「CONFIG.DB」という「MS-DOS」のファイルの書き換えによって行うことになっている。バッチ・ファイルによって「dBASE III PLUS」からエディタに移行して、「CONFIG.DB」ファイルの書き換えが簡単に行なえるようにしておくとも便利である。また、リレーショナル・データベース・ソフトウェアそのものの内部でプログラムを組み、「データ入力」、「データ構造の変更」などのモードへの移行や、レコードのポインタの移動が、メニュー方式によって簡単に実行できるように工夫することもできる。

いずれにしても、計算機というのは、基本的に融通の効かないしろものであるから、それを耐え忍ぶためには、まず人間が、ありとあらゆる工夫をしなければならない。（私の行っている工夫は、例えばプログラムなどにしても、とても人に見せられるものではなく、まだまだ、時々刻々手を加えているような段階である。従って、公開はしないけれど、個人的に興味のある方は、情報提供の相互主義の原則のもとで、工夫をし合つてよいと思っている。）

### 5.3. まとめ

以上、コンピュータを用いて言語記述を行うについての私の自己流の方法のマニュアル的説明を行ない、さらに若干の本質的問題と、いくつかの周辺的問題について議論してきた。つきつめて言えば、私達は結局、幻想の計算機というユートピアに目をくらまされて、現実の計算機の愛想の悪さに振り回されているだけなのかもしれない。もつと、

扱いやすい、人間にやさしい計算機が登場し、また、融通無碍のデータベース・ソフトウェアの登場すれば、ここで述べてきたことなどの大半は反古と化すだろう。しかし、実は、私は、竹内 1989 において論じたのだが、計算機は現状ではただか、人間の思考の補助をする程度の役割しか果せないものであって、主役はあくまでも人間なのだと考えている。結局、言語研究において、計算機を用いたとしても、結論を出す作業は結局は人間の仕事なのである。そのことをさえわきまえていれば、機械に振り回されるといっても、最後のイニシアティブは人間の側が握っていることができる。計算機を用いる場合には、人間は現実を踏まえた目的意識を持たなくてはならないのである。そもそも私は、そのようなことを念頭において、計算機を用いて言語記述を行なう可能性について、ここに論じてきたつもりである。

これまで述べてきたように、計算機を扱うことは煩わしい。しかし、そうであるとしても、本論で述べたような、テキスト・データベースから検索により言語資料を抽出・採取し、また、リレーショナル・データベース・ソフトウェアを用いて、試行錯誤的に記述作業を行うことのメリット、積極的に評価すべき点は、これまで検討してきた通りである。計算機の不便についてはまず第一に工学の方々に工夫していただかなければならない。それに較べると、我々ユーザーができる工夫などは、ここで議論したように、取るに足らないものである。しかし逆に、計算機の積極的な面を評価して、その学術研究への応用を試行錯誤するのは、今度は、言語研究の分野に限っていえば、言語研究者が前面に立つて為すべきことである。そして、その点で、これまで為されてきたことは、まだまだ貧弱であると思う。これは、本論の稚拙さの言い訳のようであるが、私はそのつもりではない。たとえ、稚拙であっても、破れかぶれの試みをお互いに公開し合い、研究の前進を目指すことが、今、1990年代前半のこの分野についての状況に見合ったことなのではないだろうか。そのような研究者間の交流の活性化を今後に期待したい。

## 文献

- 有川節夫他、1987、「テキストデータベース管理システム SIGMA 第 2 版について」、九州大学大型計算機センター広報、vol. 20, No6
- 竹内義晴 1989、「言語研究のための補助手段・思考実験手段としてのドイツ語処理ミニパーザの試み」、言語科学24号、pp 71-96、九州大学言語文化部言語研究会
- 1990、「言語固有の意味領域の確定のための考察」、言語科学25号、pp 68-97、九州大学言語文化部言語研究会
- 1991、「自然言語の意味論の基本問題：世界についての知識と自然言語固有の意味との関係について」、言語科学26号、pp 35-62、九州大学言語文化部言語研究会
- 1992a、「自然言語における言語固有の意味の追求 — テキスト・データベースによって検索した ドイツ語の形容詞「empfindlich」の使用例についての意味の揺れの記述から」、ドイツ文学89号、日本独文学会（投稿中）
- 1992b、「意味解釈に対する統語関係の未決定性と、意味理解・知識処理における言語表現の意味関係の役割について — テキスト・データベースによるドイツ語の形容詞『empfindlich』の記述から」、ドイツ文学論集25号、日本独文学会中国四国支部（投稿中）
- 樋口忠治・篠原武、1987、「テキストデータベース『トーマス・マン・ファイル』の完成と再編成について」、九州大学大型計算機センター広報、vol. 20, No6
- 渡辺健次・武政尹士、1988、「日本語 PFD エミュレータ — PC-9801シリーズ対応」、九州大学大型計算機センター広報、vol. 21, No5
- dBASE III PLUS マニュアル、日本アシュトン・テイト社、1987