

日本語聖書コンコーダンス・システム J A C O B の開発

図書館情報大学 石塚英弘、井上薫孝*

JACOB(Japanese Automated Concordance of Bible)は、聖書のように複雑な構造を持ったテキスト・データベースのために開発された対話型システムである。聖書は、本文のほか、脚注、引照、チェーン式引照、重要語句の解説などの関連情報、梗概、体系的索引など、多様な情報が含まれた複雑な構造をしている。そこで、次に述べるような機能を設定して、このように複雑なテキスト・データベースでも扱うことができるシステムを開発した。まず、画面の左側に本文を、右側に関連情報を表示して、この2つを同時に見られるようにした。また、チェーン式引照に示された関連箇所画面、梗概画面、体系的索引画面などを簡単なコマンドで次々に渡り歩くことができるようにした。

Development of JACOB(Japanese Automated Concordance of Bible)

Hidehiro Ishizuka and Yoshitaka Inoue
University of Library and Information Science
1-2, Kasuga, Yatabe, Tukuba, Ibaraki, 305 Japan

Present authors have developed a system named JACOB(Japanese Automated Concordance of Bible). JACOB is an on-line interactive system for complex text database. Bible consists of text, related information, an abstract of volume and a systematic index. Related information contains a footnote, a cross reference, a chain of reference and explanation of an important word/phrase.

Text is displayed in the lefthand part of the screen, and related information is also displayed in the righthand part of the same screen. In this system, there are four types of screens: for text and related information, for corresponding text and related information pointed by a chain of reference, for an abstract, and for a systematic index. A user of the system can navigate these screens using a simple command.

*現在(株)松下電器産業勤務

1.はじめに

最近、種々の分野で文章のデータベース（以下、テキスト・データベースという）が作られ、それに対応した検索機能を持つシステムが作られるようになってきた。しかし、今のところ作られているテキスト・データベースの構造は簡単で、人文科学で広く使われている本に見られるような、注や引照などを含む複雑な構造ではない。また、その複雑な構造を操作できるシステムも見当たらない。

たとえば情報検索の分野では、米国化学会の雑誌論文を始めとする種々の雑誌論文、法律情報、ニュース記事、辞典などのテキスト・データベース（情報検索では full-text database と言われている）がある。¹⁾また、これに対応して、BRS, DIALOG などの商用情報検索システムでは、N語以内に指定した2つの単語があるものを捜すなど従来のキーワード検索とは異なる機能を用意するようになっている。また、テキスト中にある単語を通常的手法で最初から最後まで検索すると処理に時間が掛かるため、文字列の高速マッチング機能を用意したシステムも報告されている。

また文章研究の分野では、英国や米国などで小説や詩などのテキスト・データベースが作られ、ワードリスト、語の頻度、索引、コンコーダンス（厳密に言えば KWIC）などが可能なソフトウェア・パッケージ OCP²⁾(Oxford Concordance Program) が利用されている。ただし、OCP のKWICで言う context とはあらかじめ指定された単語数であって、文やパラグラフといった本来の意味での文脈ではないことに注意する必要がある。日本でもテキスト・データベース用のシステム SIGMA³⁾を用いて、トーマス・マンのテキスト・データベースを作り、文章研究を行なった例⁴⁾が報告されている。

一方、オフィスで扱う文書のテキスト・データベース（ドキュメント・データベースといわれる）の研究がある。ここでは章・節といった論理構造が考えられているが、注などを含めた構造にはなっていない。

本稿では、本文以外に注、引照、重要語句の説明などを含む複雑な構造のテキスト・データベースと、それを操作できるシステムを開発したので、その概念と機能について報告する。このシステムのデータベースは日本語訳の聖書なので、システム名を JACOB (Japanese Automated Concordance of Bible) と名づけた。

2. システムの特徴

本システムは次に述べるように、読書、コンコーダンス、重要語辞典の3つの機能を合わせ持っており、これらを組み合わせて使うことができる。

1) 読書の機能

本文とそれに対応した注、引照などの関連情報とを同じ画面で見ることができる。また、本文を任意の箇所から読むことができる。

2) コンコーダンスの機能

重要語（たとえば復活、愛、死など）がどこで、どのような文脈で使われているかを見ることができる。この場合の『文脈』とは意味的に一纏まりの単位（文）であって、KWICの場合のような予め決められた文字数の機械的文脈ではない。

3) 重要語辞典の機能

重要語の説明と関連する本文とを同時に見ることができる。

本システムではこれらの機能を実現するために、次のような画面型インターフェースを採ることとした。端末画面を左右に分割し、左側に本文を表示し、右側にはその本文に関連する注、重要語の説明、引照などを表示する。また、コマンド欄を画面下部に設定し、上記の3つの機能を自在に指定できるようにした。

3. 聖書

本システムの機能の背景には、聖書の持つ特徴や独特の言葉使いがあるので、ここで説明しておきたい。

3.1 聖書の構成

聖書は旧約聖書39巻、新約聖書27巻、合計66巻から構成されている。各巻は巻号でなく、たとえばマタイ伝による福音書など書名で呼ばれることが多い。各巻は章に分かれ、章は節に分かれている。節は1ないし2つの文から構成されており、一纏まりの意味を表わす単位である。そのため、本システムでは節を本文の最小単位として取り扱っており、文脈は節を単位としている。

次に本データベースの源とした冊子体の聖書のデータ項目を示す。

1) 本文

日本聖書刊行会訳を使用。引照のある場合は、その語句の肩に(1),(2)のマークが付けられている。欄外注は*、チェーン式引照は、重要語句の説明には が同

様に付けられている。

2) 引照

同じような語句、概念、出来事などが出てくる箇所、相互に関連のある箇所、参照箇所、比較箇所を、ロマ書6章23節といったように書名・章・節の形式で欄外に付けたもの。本文中のマークに対応している。

3) 欄外注

異本での訳、原語、直訳などを欄外に付けたもの。

4) 脚注

本文中より語句、文などが抜き出されていて、それに関する簡単な注解が付けられている。

5) チェーン式引照

重要な語句、概念、人名、地名などが出てくる箇所を書名・章・節で示したものの。本文中の印が付いた該当箇所、その言葉が最初に言及される箇所、当該箇所のすぐ前の箇所、すぐ次ぎの箇所、最後に言及される箇所、その言葉の説明がある箇所の合計6つが並んで示されている。鎖のように結んであるのでチェーン式というが、これにより言葉の流れ、相互関係を理解し、聖書全体から正しい意味を把握することができる。

6) 重要語句の説明

重要な語句、概念、人名、地名などに関する説明。

7) 索引

体系的な索引で、コンコーダンスの機能に加えて、辞典的な機能も持つ。言葉を組織だった幾つかの項目に分けており、その区別を見出しを使って示している。それぞれの項目には、聖書の箇所だけでなく、その内容が抜き出されており、説明文のような役割も果たしている。さらに、重要語句の説明箇所が書かれていることもある。また、『愛する、愛されている』のように見出しが2つある場合や、『愛、キリストの』のように修飾語が後ろについている場合もある。

8) 梗概

各書の目次。その書の概略や、ある事柄について述べている所を知ることができる。

9) 結論

各書の著者、書かれた年代、場所、背景、特色、主題などについて述べている。

3.2 本文以外の情報の必要性

聖書は難解な書物である。しかし、聖書には同じような語や概念、出来事が複数の箇所に出てくることが多く、それらを総合的に検討すれば、その意味するところを理解できることも多い。そのため、本文に加えて前述のような種々の情報がつけ加えられている。たとえば、コンコーダンスといわれる一種の索引は昔から作られている。また、関連のある箇所を引照やチェーン式引照で示すことも行なわれる。さらに、最近では本文中の重要な語や句に対して注解や説明を欄外に書いたものも出版されている。

これらの工夫は有効であるが、冊子体であるために次に示す欠点も持っている。

- 1) 引照箇所を見るのに手間が掛かる。
- 2) ある本文からどの語が索引に出されているかを調べることは著しく困難である。
- 3) 索引の見出しで修飾語が後置されていることは度々あるが、その場合にその修飾語で検索することは困難である。

これらの欠点を本システムでは解決した。

3.3 コンコーダンス

コンコーダンス(concordance)は語句索引、特に聖書の語句索引という狭い意味にとられていることが多い。また、1958年 H.P.Luhn がKWIC(keyword in context)索引を実現してみせた後には、計算機処理によるコンコーダンス即、KWICと思う人も多い。しかし、桜井によって簡潔明解に述べられたように⁶⁾、コンコーダンスとはその由来から言えばより広く深い意味を持った言葉である。すなわち、『特定の用語のもとに内容の調和一致(concordantia)する文脈を集めたもの』がコンコーダンスである。文献⁷⁾によれば、コンコーダンスという用語は聖書、文法、音楽、図書館学など分野により異なった意味に使われていると言うが、聖書の場合のコンコーダンスとは既に述べたように聖書に書かれていることを統一的に理解するためのものである。

この観点から言えば、単なる索引やKWICでは不十分で、体系的索引であると共にその言葉が使われている文脈を容易に見られる必要がある。ここで文脈とは意味的

に一纏まりの単位であるべきで、聖書の場合で言えば節であったり、またその前後の節を含めた文章ということになる。演者らはコンコーダンスをこのように考え、本システム上に実現した。

4. システムの機能と実行例

本システムでは、2.で述べた読書、コンコーダンス、重要語辞典の機能を組み合わせて使えるようになっている。本システムはコマンドを入力することにより、図1に示すような状態遷移を行なう。

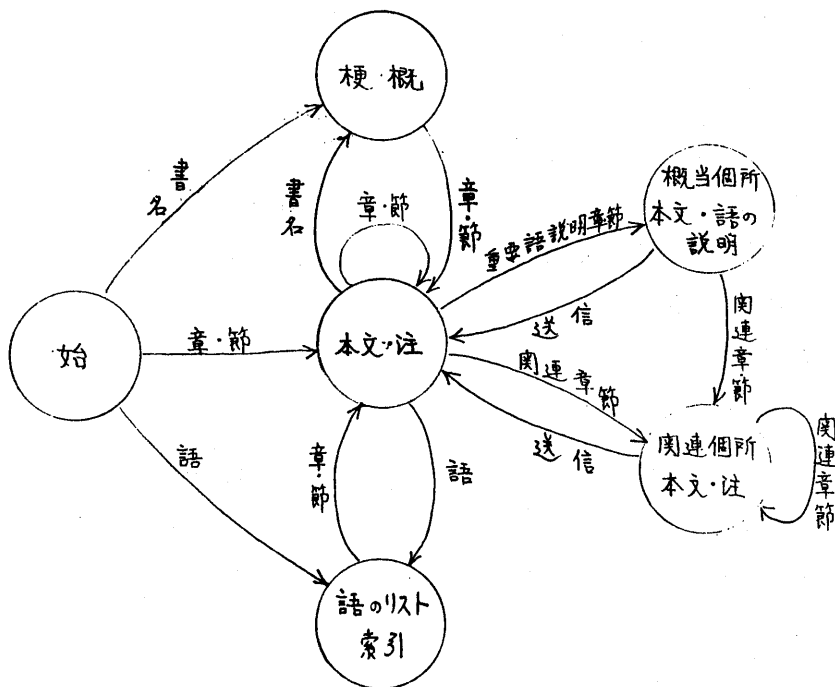


図1. 状態遷移図

図2に本文および注ほかの表示画面の例を示す。この画面で『死』に関するチェーン式引照に示されている箇所、ロマ書6章11節を表示するコマンドを入力すると、図3に示す関連箇所表示画面が出力される。そこで送信キーのみを押せば、元の図2の画面に戻る。また、図4に『罪』に関する体系的索引を表示した例を示す。

本システムはHITAC M-260H VOS3 上にVSAMを用いてPL/Iでプログラムを書いて実現した。データベースには今のところロマ書全16章が格納されている。

聖書...読書と検索システム

本文表示 ローマ人への手紙6章23節 SCROLL [HALF]

6:23 (1)†罪からくる報酬は†死です。 チェーン式引照
 しかし、神の下さる賜物は、(2)私たちの主 罪 ロマ6:23 創3:8 ロマ5:12
 キリスト・イエスにある永遠のいのちです。 ロマ14:23 1ヨハ5:16 ヨハ8:3
 7:1 それとも、(1)兄弟たち。あなたがた 4
 は、律法が人に対して権限を持つのは、そ 死 ロマ6:23 創2:17 ロマ6:11
 の人の生きている期間だけだ、と言うこと 1コリ15:55 黙21:8 ロマ6:23
 を知らないのですか。――私は律法を知っ 重要語句
 ている人々に言っているのです。―― 死―生命が失われること。死は罪に
 (1)7:2 夫のある女は、夫が生きている間は、 対する罰として人間に与えられたも
 律法によって夫に結ばれています。しかし、 ので、生命の根柢である神との関係
 夫が死ねば、夫に関する律法から解放され が断たれることを意味する。死は肉
 ます。 体的死と霊的死に分けられ、肉体的
 7:3 ですから、夫が生きている間に他の男 死はすべての人が一度は経験するも
 に行けば、姦淫の女と呼ばれるのですが、 のであるが、キリストの贖罪を信じ
 夫が死ねば、律法から解放されており、た る者は霊的死から救われ、永遠の生
 とい他の男に行っても、姦淫の女ではあり 命が与えられる。
 ません。 脚注
 7:4 私の兄弟たちよ。それと同じように、 [罪から来る報酬は死です]―罪の
 あなたがたも、(1)キリストのからだによっ] 検索範囲 []
 15 コマンド [R ROMA 6:11]

範囲 1:見出し語 2:類義語 3:意味・説明 4:索引 5:チェーン式引照
 6:重要語句の説明 7:脚注 8:引照 9:欄外注 0:梗概

図2. 本文・注ほか表示画面

聖書...読書と検索システム

関連情報表示: コマンド=R SCROLL [HALF]

ロマ6:11 このように、あなたがたも、自 脚注
 分は(1)罪に対しては†死んだ者であり、神 キリストと一体化された者は、キリ
 に対しては、キリスト・イエスにあって生 ストが十字架で死んだように、自分
 きた者だと、思いなさい。 も罪に対しては死に、キリストが復
 活して今も生きておられるように、
 自分もキリストにあって新しいいの
 ちに生きていると思うべきである。:
 11 @死@ 6 : 11 創 2 : 17 ロマ
 5 : 12 ロマ 6 : 23 黙 21 : 8 ロ
 マ 6 : 23
 引照
 (1)ロマ6:2 ; (1)絶対にそんなことはあ
 りません。(2)罪に対して死んだわた
 したちが、どうして、なおもその中
 に生きていられるでしょう。
 (1)ロマ7:4 ; 私の兄弟たちよ。それと
 同じように、あなたがたも、(1)キリ
 ストのからだによって、(2)律法に対
 しては(3)死んでいるのです。それは、
] 検索範囲 []

16 コマンド []

範囲 1:見出し語 2:類義語 3:意味・説明 4:索引 5:チェーン式引照
 6:重要語句の説明 7:脚注 8:引照 9:欄外注 0:梗概

図3. 関連箇所表示画面

聖書…読書と検索システム

関連情報表示：コマンド=R 07 6:23

SCROLL [PAGE]

漢字：罪

読み：ツミ

類義語

000そむき

000違反

索引

定義

信仰から出ていないこと

ロマ14:23

源

アダムの違反

ロマ5:12

ロマ5:16

新生していない人にとっての結果

死

ロマ6:23

新生した人の務め

罪の奴隷にならない

ロマ6:6

罪に従わない

ロマ6:6

ロマ6:12

罪を征服する

ロマ6:14 - 6:22

対策

御霊に従って歩む

ロマ8:1 - 8:14

3 コマンド [

] 検索範囲 [

]

範囲 1:見出し語 2:類義語 3:意味・説明 4:索引 5:チェーン式引照

6:重要語句の説明 7:脚注 8:引照 9:欄外注 0:梗概

図4. 体系的索引表示画面

文献

1) Carol Tenopir: "full-text databases", Annual Rev. Info. Sci. & Tech., Vol.19, p.216-246(1984).

2) 長瀬真理、西村弘之: "コンピュータによる文章解析入門— OCPへの招待—", pp.210、オーム社、1986年10月。

3a) 有川節夫、篠原 武、白石修二、玉越可: "研究者向き情報システムSIGMA について"、九州大学大型計算機センター広報、Vol.14, No.4, p.550-573(1981)。

b) 有川節夫ほか: "テキストデータベース管理システムSIGMA について"、情報処理学会情報学基礎研究会報告、No.6-4 (1987)。

4) 樋口忠治、篠原 武: "公用データベース トーマス・マン・ファイル/SIGMA の公開"、九州大学大型計算機センター広報、Vol.16, No.4, p.389-393(1983)。

5) いのちのことば社、聖書 注解・索引・チェーン式引照付き、1981年刊、2115p。

6) 桜井宜隆: "コンコーダンス"、図書館情報大学附属図書館報、Vol.2, No.4, p.5(1986)。

7) Allen Kent and Harold Lancour ed. "Encyclopedia of Library and Information Science", Vol.5, p.592-605. 1971