自然言語処理　54－3
（1986. 3. 28）

# 自然言語の分析による知識データ

田中康仁
（姫路短期大学）

吉田将
（九州大学）

## 1. はじめに

自然言語の研究の流れを振り返り，将来を見通してみると次のような4つの段階に区分けすることができる。

第一段階　自然言語研究の始り

語とその簡単な属性と単純な文法規則によって自然言語の処理を行った時代

第二段階　自然言語処理の応用が開花した時代

語に意味属性を付け，ある程度のカテゴリ化を行い，語にはシソーラス化も行って，文法も単純なものから格構造を使用するまでになった。

自然言語処理の応用が盛んになった。

第三段階　“語の文法”の時代

語の属性や意味マーク等のほかに語の使われ方を細かく調べた情報を基にした“語の文法”が完成し文法規則，格構造も詳細になり，階層構造が導入され，より体系的になっている時代。

第四段階　“文脈解析”の時代

文と文の関係，動詞の因果関係や動詞の状態が詳細に分析され，体系的にまとめられる時代。

第五段階　“文章解析”の時代。

文章と文章の解析や研究が進み，文章の理解が人間と同じようになる時代。

現在は第二段階が終り，第三段階に入った状態である。一部の研究者は第四段階の研究を行っている。ここでは“語の文法”を充実させるため，知識データを膨大な日本語文の中から得る方法について述べている。一つの方法として「4文字漢字列の分割による知識データの獲得」を提案し，実際に収集し，カテゴリー付を進めてきた。この対象データとしては日本科学技術情報センターの抄録テープを使用した。

## 2. 4文字漢字列の分割による知識データの獲得

4文字漢字列は日本語文の中から機械的に抽出することが可能である。4文字漢字列は2文字・2文字の2つの基礎的概念語に分割できる。およそ，80％～90％のデータが2文字2文字に分割でき，2つの語は助詞，助動詞，動詞等を使って文や句にすることができる。

例

情報・伝達　→　情報ヲ伝達スル

進学・指導　→　進学ヲ指導スル

このような分析を4回行った。この分析結果をまとめると，表1のようになる。日本科学技術情報センターの各分野の抄録文から延96.5万件の4文字漢字列を抽出し，分析した。この結果約17万件程度の知識データが得られた。（詳細については分析中である。）

| | データ量 / 分野 | 延漢字列 | 種類 | 知識データ | 種類／延漢字列 |
|---|---|---|---|---|---|
| (1) | 管理システム 技術編 | 78,304 | 31,932 | 28,033 | 0.408 |
| (2) | 電気・金属 機械・土木 | 159,717 | 58,391 | 48,000 | 0.365 |
| (3) | 物理・電気 機械・化学 | 312,143 | 90,840 | 71,200 | 0.291 |
| (4) | 管理システム，原子力，物理，金属，機械，化学 | 415,048 | 114,977 | 89,000 | 0.277 |
| | 総合 | 965,212 | － | （約17万） | － |

表1　分析データ量と収集した知識データ量

（－は分析中を示す。＊はカテゴリーを付加中）

この4回の分析で延漢字列が増加するほどには種類は増えない。種類／延漢字列の数値が0.408から0.277へ移っていることでわかる。これは語の接続が全ての語と関係するのではなく，ある程度限られた範囲にしぼられることがわかる。この4回の分析で実用に供する程度の語と語の関係の知識データを得ることができた。

## 3. 新しい知識データの収集方法

① 集められた知識データを利用する。

4文字漢字列の分割により知識データが得られている。この知識データを利用し，語と語の2項関係の各語を逆転させ，逆転した語の間に関係があるか調べる。

例1


計算した結果

結果を計算する

例2

輸入・製品　　輸入した製品

製品・輸入　　製品を輸入する。

十数万件の知識データが整理されて集められているが，このデータを利用することにより知識データを増加させることが出来る。但し例1，例2のように全ての場合に逆転が成立するわけではない。

例3

現行・設備　　現行の設備

設備・現行　　関係は成り立たない

逆転したデータが知識データ・ファイルに有るか否か事前に調べ，有るものは省き，残りを分析すればよい。しかし調べるのは人間であり，調査員の語彙能力，分野の知識に大きく依存する。調査員の選択が重要である。約2割程度の増加が見込める。

② 助詞，助動詞を利用し，KWICを用いて知識データを抽出する。

助詞，助動詞としては次のものを考えている。

が，を，に，へ，と，から，より，により，の，

する，した，に対する，に関する，……

例

安全を守る。

安全を確保する。

方法を拡張する。

機能を拡張する。

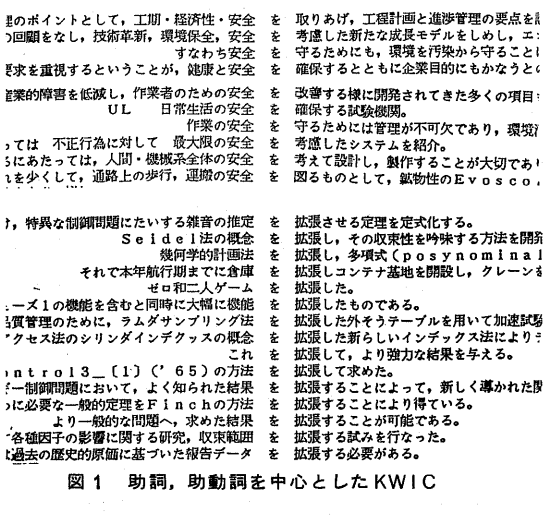



図1　助詞，助動詞を中心としたKWIC

KWICは機械的に作成することができる。このKWICの例で説明すると“を”の前の“安全”“を”の後の“拡張”という語を別々に抽出しておき，プログラムで前接語，後接語のテーブルにより機械的に知識データを抽出することができる。また，この抽出されたデータを集約することにより頻度情報も得られる。

この方法による語と語の関係の抽出は全て機械的に出来るのではなく，次の点に注意しなければならない。

- 長単位用語から基本概念語を抽出しなければならない。

  例

  希塩酸溶液　を　そそぎます

  溶　液　を　そそぐ

- 動詞を終止形にしなければならない。

- 助詞，助動詞等の直前，直後の語が必ずしも語と語の関係を持つものではない。

  例

  ～方法をつぎつぎに実行した。

  ごみ以外を集めたもの。

- 文中に現われた格助詞は表層的なもので語と語の関係を適切に表わしていない場合がある。

  例

  積乱雲が雨を降らせた。

  雨が降る。

格助詞を変えなければならない場合がある。

使役，受身（自発，尊敬）は注意しなければならない。

- 語の係り受け関係が少し複雑な場合

  Ａを求めた解Ｄで割る。

“Ａを求めた”ではなく“Ａを割る。”を抽出しなければならない。

- 並列関係

  例

  りんご　と　みかん　を　食べる。

直前直後の関係だけでは“りんごを食べる”は抽出できない。

- 機械的に処理できなかったデータをリストし，手作業で分析しなければならない。

幾つかの問題点，例外はあるがこれらは発生する割合が少ないので，手作業による修正を行うとしている。この方法によって“を”を中心としたKWICを作り分析中である。対象としたデータは日本科学技術情報センターの抄録文であり，データ量は約79万件（KWICの行数）である。

このデータの中から重複をまとめ，約20万件程度の知識データが得られる予定である。“を”以外の助詞，助動詞へも拡大する予定である。

79万語の“を”を中心としたKWICから前接語，後接語を抽出する方法を考えなければならない。

前接語の抽出方法

(i) KWICの前接語が含まれている部分の文字列から漢字，片仮名，記号(-)の部分文字列を抽出する。

(ii) (i)で抽出した文字列を最後の文字から分類し，同一のものは集計し，まとめ，頻度情報を付ける。

(iii) (ii)で集約したDataのうち頻度2以上のものをリストする。（5万行程度になる）

(iv) (iii)でリストした中から基礎的概念語を手作業で抽出する。（約2万語程度になる予定）

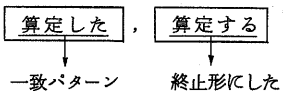
後接語の抽出方法

(i) KWICの後接語が含まれている部分の文字列から8文字を抽出する。

(ii) 最初の文字から漢字コード順に分類し，同一のものはまとめ1件にする。約8万行程度になる。

(iii) 集約したDataからKWICのパターンと一致させる文字列と終止形に変形する文字列を抽出する。

例




前接語，後接語をテーブルに納めプログラムで語と語の関係を抽出する。抽出できないものは2つのテーブルを順次増やすことにより処理することができる。また処理できないものはリストし手作業によってテーブルを改良しなければならない。

前接後，後接語のための作業テーブルを次に示す。

| | |
|---|---|
| 両理論 | 3 |
| 改良理論 | 2 |
| 乱子乱流理論 | 2 |
| 計量理論 | 2 |
| 運動量理論 | 2 |
| 混合距離理論 | 2 |
| 行列理論 | 14 |
| 反応行列理論 | 2 |
| 核分裂理論 | 2 |
| 配向緩和理論 | 2 |
| 確率論 | 7 |
| 化学量論 | 16 |
| 記録 | 17 |
| メンテナンス記録 | 2 |

表2　前接語作業表の一部

| | |
|---|---|
| 与えるか。 | 7 |
| 与えるかという | 3 |
| 与えるかどうか | 5 |
| 与えるかについ | 17 |
| 与えるかもしれ | 4 |
| 与えるかを確立 | 3 |
| 与えるかを検討 | 7 |
| 与えるかを考察 | 6 |
| 与えるかを調べ | 13 |
| 与えるが，5－ | 3 |
| 与えるが，この | 11 |
| 与えるが，これ | 11 |

表3　後接語作業表の一部

この方法は機械的方法が多いため試行錯誤が行える。費用が安く，処理速度が早いという特徴がある。全体的処理図は図2に示す。

③　3文字漢字列の分割による知識データの獲得

4文字漢字列の分割と同様に3文字漢字列を分析することにより知識データを得る。3文字漢字列を分析すると次のようになる。

(1) 3文字で1つの語となり分割が不可能なもの。

例　　不可決，　不思議

(2) A（B1＋B2）のように分解され，AB1，AB2が意味のある語となるもの

例　　輸・出入，　国・内外

(3) （A1＋A2）のように分解されA1B，A2Bが意味のある語となるもの。

例　　入出・力　　上下・院

(4) SA，AS　Aが主要な語で，それ等に接頭語(S)，接尾語(S)が付いたもの。

例　　未・解決　，　近代・化

(5) AB　A，Bともに主要な語でそれが結合したもの。

例　　金・鉱山　，　障害・児

(6) A・B・C　A，B，Cが主要な語で，それらが並列的に結合したもの。

例　　年・月・日，市・町・村，英・独・仏

大きくわけると上記6つのようになる。これに各語が持つ品詞成分を付け加えることにより，さらに細分類することができる。この考え方は東京女子大学水谷静夫の分析方法を用いている。このような方法を基礎にして延169,163件，種類33,155件のデータを分析中である。機械的に収集した3文字漢字列であるため，この中から有効なデータを抽出することができる。

④ 慣用句による知識データの収集

日本語の中には慣用的に用いられる句が多い。これらを構成している一語，一語を取りあげて，その意味を考えても全体の句の意味には結び付かない場合が多い。意味の深い部分に立ち入れば，又，語の連想から各語の意味を結合すれば全体が合成できるものもある。しかし，これを機械的に行うには，かなり無理な面がある。

例　　反吐(へど)が出る ⟶ 嫌悪感を感じる。

　　鼻が高い ⟶ 自慢する。

慣用表現をおきかえ可能な語に変えることを考える。このような方法を取れば機械翻訳等に便利である。

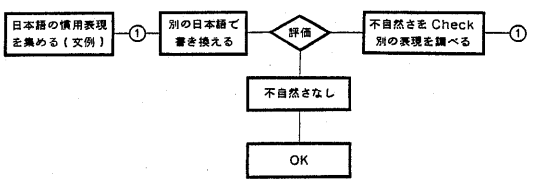



図3　慣用表現の表現変え

慣用表現を機械的に収集するにはどのようにすればよいかという問題がある。そこで語に定義された数個の意味だけを与え（あまり深い意味はもちいない。），これを基準にして個々の語の意味から合成されるものと慣用表現の持つ意味を照合すれば判断することができる。又，慣用表現は語と語の結び付きが強いものであるから語と語の共起と考えられる。このことを利用し機械的に抽出することができる。

慣用表現は既に多くの人々が慣用句の研究として出版物にまとめている。ここでは出版物を利用し研究の材料とする。

慣用表現を分析するために，次の二つのファイルを作成する。

(1) 慣用表現ファイル

このファイルは慣用表現と，その属性を入力する。

(2) 慣用表現例文ファイル

この例文ファイルは例文と慣用表現に対応するNOを持っている。

これら2つのファイルの項目は次のようなものを予定している。

- 慣用表現ファイルの項目

1. 登　録　NO　（SEQ．NO）
2. 慣用表現
3. 読　み
4. 分かち書き
5. 自立語の見出し
6. 文，句の区別
7. 品詞，又は品詞相当のもの
8. 活　用
9. 語　尾
10. 置換可能語
11. 置換にあたっての文字列操作1（前）
12. 置換にあたっての文字列操作2（後）
13. 訳　語
14. 共起語
15. その他区分

- 慣用表現例文ファイルの項目

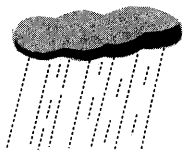
1. 登　録　NO
2. 例　文
3. 慣用表現ファイルの対応NO
4. その他

置換可能語は慣用表現をできるだけ簡単な表現に書き換えるためのものである。

慣用表現を置換可能な表現に変えた場合，追加したり，削除する文字が有るか否か，有ればその手順は常に一定か，例文で検証する。不自然であれば置換可能な語に変える。

このほか，自然言語の表現の中にはある特定の現象や事象については特定の言葉を用いるということがある。語には共起性の強いものがある。

例えば



「雨が降る。」を考えてみると"雨"のかわりに「水，水滴，$H_2O$，雨水」に変える表現は適切でない，また"降る"のかわりに「落ちる，落下する」と変えても良い表現ではない。「雨が降る。」は一種の慣用表現とみなすことができる。

計算機に言語を処理させるにはこのような語と語の共起についても多量のデータを集めておかなければならない。

以上4つの新しい知識データの獲得方法について述べた。この4つの中で②のKWICを用いて知識データを得る方法が有効な手段と考えられる。

## 4. 収集した知識データの分析と利用方法

日本科学技術情報センターの抄録文を分析し，表1の知識データを得た。知識データの収集過程における各種資料から判断し，4文字漢字列の延 80 %に当る漢字列を収集しているとみなすことができる。（抄録文に関して）この知識データは応用研究に充分対応することができる。

① 4文字漢字列の分析内容

4文字漢字列として使用頻度の高いものを示すと表4のようになる。また，前接漢字列の種類は約4,700種類，後接漢字列の種類は約3,200種類であり，その使用頻度の高いものをあげると表5，表6のようになる。この収集した知識データを見易くするために1つの語についての接続状態を調べてみると表7のような結果を得た。これにより各語の接続の関係が明らかになる。

② 概念の結合関係の分析

概念の結合関係，概念と動詞の結合関係について分析することはこの研究の一つの課題である。前接語，後接語を幾つかの概念（時間，場所，空間，道具，……等）に分類し，これらの結合関係の強弱を見つけ出すことである。

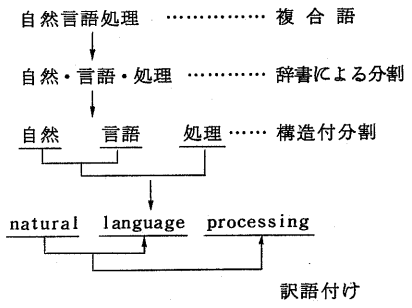
このためには次のような方法を考えている。前接語，後接語に類語新辞典（角川）のコードを割り付ける。同一コードの組を集めることにより，語のグループができる。これによりカテゴリー化ができる。類語新辞典の分類は10進分類を基本とし，1,100のカテゴリーに分類されている。それゆえ，語と語の関係は $1,000 \times 1,000 = 100$ 万のカテゴリー空間のどこかに入る。約28,000件のデータでは2％程度の部分空間に入る。17万件のデータでは約10数％程度の部分空間を満す。（重複があるため。）

これにより概念の結合関係がわかるし，今後新しく出現する語と語の結合を予想することもできる。

③ 複合語の構造付分割と訳語への利用

複合語の翻訳にあたっては，翻訳の基礎となる複合語の分割が重要である。ただ単純に分割するのではなく構造を持った状態に分割することが必要である。

例1




この構造を持った漢字列の分割のためには次の5つの方法が考えられる。

1) 最小語数に分割する。
2) 分割された各語の係り受けは非交差である。
3) 語と語の知識データを使う。
4) 語の持つ特殊性を利用する。
（以上，以前，等，中，… ）
5) 複合語の分割パターンの頻度を利用する。

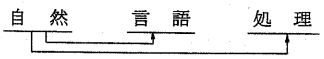
等が考えられる。

例１の場合には知識データとして次の２つの知識データが有ることを仮定している。

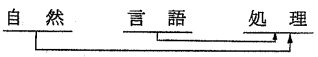
知識データ ： 言語（ヲ）処理（スル）
自然（ナ）言語

もし，これらの知識データがなければ次のような色々な構造が考えられる。

例２




例３




複合語の翻訳にあたっては正しい構造に分割することが第一であるが，次に分割された各語がどのような訳語を取るかということが重要である。（この考え方の前提は分割された語が１対１の訳語を取るという場合である。）多量の複合語に訳語を付けるということについては研究の第一歩を踏み出した状況である。

④ 格構造の部分構造を抽出する。

格構造の分析のために知識データを利用することを考える。日本語の機械処理では格構造を結合価文法で表わしたものが多く用いられている。これは文の構造を把握するために重要な役割をはたしている。しかし，現段階では用語に対する文型パターンが少なすぎる。これを充実するためには，知識データをこれら文型パターンにあてはめ，文型パターンを増やさなければならない。

例 吸う ： N〔hum〕が N〔Con〕をV
{hum →人間，Con →具象物}

しかし，“脱脂綿が水を吸った。”という文はこの文型にはあてはまらない。我々は実際の抄録文から使われているサンプルを集め格構造の部分構造を抽出し，より強固な結合価文法の構築をめざしている。

⑤ ２文字漢字列の分析用基礎資料となる。

２文字漢字で構成されている語を調べると，延べ語数は非常に多いが，種類はそれほどでもない。このため２文字漢字で構成されている語の共起関係を調べることはKWICを使うにしても膨大になり大変である。そこで“語と語の関係”で得られた関係が文中に有るか否か調べる。

これら文を抽出し知識データの有効性の判定資料として使うことができる。

分析の手順を示すと次のようになる。

- 文中に知識データによる共起関係があると予想できるもの。
  - 共起関係が認められる文である。
  - 共起関係が認められない文である。
- 文中の共起関係を知識データでは抽出できなかったもの。

このような利用方法が考えられる。

⑥ 既に考えてきた応用分野

そのほか，これまでに知識データを使っての応用分野としては次のようなものを考えている。

文字認識，音声認識の後処理に利用し精度向上をはかる。仮名漢字変換システムの同音異義語の選択，機械翻訳の多義語の選択等に役立つ。これらへの応用のためにも着実な知識データが準備されなければならない。

## 5. おわりに

知識データの収集，整備は日本語処理システムが一層発展するために必要である。約１００万件程度のデータが集まれば，新しい研究段階になるであろう。

この知識データについては解決しなければならない多くの問題があるが少し先が見通せる状態になってきた。

知識データを利用することは高品質で，より高度の利用技術が提供されることになるが，これを実現するためには着実なデータ分析の作業が必要である。

この研究の一部は文部省科研費課題番号（60302090）（代表者 吉田 将）によって行われた。

## 参考文献

(1) 田中康仁，水谷静夫，吉田 将 語と語の関係について 情報処理学会 自然言語処理 41－4 1984－1

(2) 田中康仁，吉田 将 自然言語の分析による知識データの収集 情報処理学会「自然言語処理技術」シンポジウム 1984－11

(3) 田中康仁，水谷静夫，吉田 将 語と語の関係による

知識データの収集 — 自動カテゴリー化について日本科学技術情報センター第21回情報科学技術研究集会論文集　1985. 3

(4) 田中康仁，吉田　将　自然言語における知識データについて情報処理学会第31回（昭和60年度後期）全国大会　3H−9

(5) 村木　　，慣用句・機能動詞結合・自由な語結合
日本語学第4巻第1号　1985. 1　明治書院

(6) 宮地　裕　"慣用句の意味と用法"　明治書院
1982. 10

(7) 倉持，阪田　"必携慣用句辞典"　三省堂　1985. 1

(8) 水谷静夫，石綿敏雄　他，　「文法と意味Ⅰ」
朝倉日本語新講座3　朝倉書店

(9) 大野　普　「類語新辞典」　角川書店

(10) 田中康仁，吉田　将　慣用表現について　情報処理学会第32回（昭和61年前期）全国大会　1S−3
1986. 3

(11) 田中康仁，水谷静夫，吉田　将
語と語の関係による知識データの収集について
日本科学技術情報センター 第22回情報科学技術研究集会論文集　1986. 3

(12) 小西友七編　英語基本動詞辞典
研究社出版　昭和55年9月

(13) 本間　茂，山階正樹，小橋史彦
連語解析を用いたべた書きかな漢字変換日本語文書処理21−2　情報処理学会研究報告　85-JDP-21　1985.5

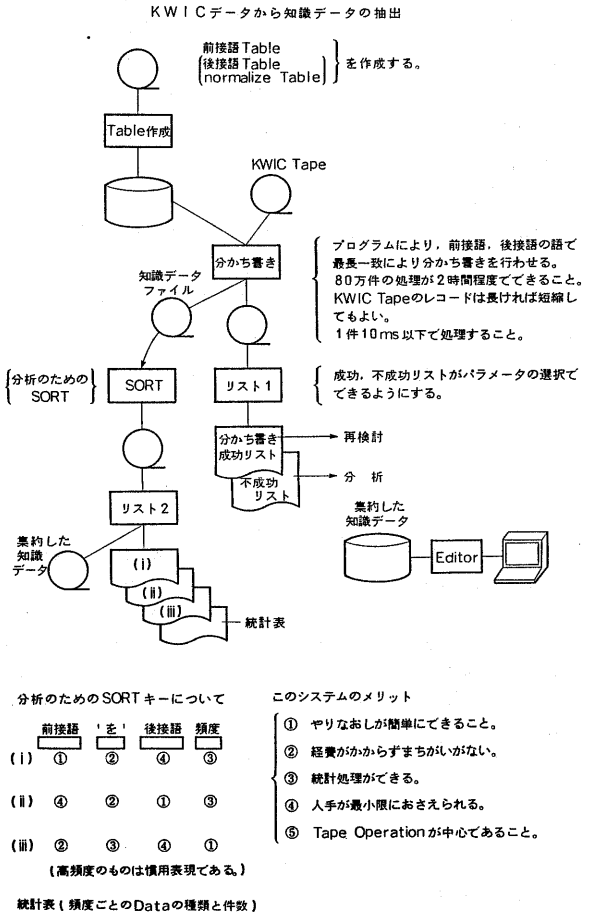



図2　語と語の関係抽出プロセス

| seq. | 漢字列 | 頻度 |
|---|---|---|
| 1 | 意思決定 | 412 |
| 2 | 品質管理 | 356 |
| 3 | 化学工業 | 341 |
| 4 | 研究開発 | 273 |
| 5 | 中小企業 | 222 |
| 6 | 情報処理 | 221 |
| 7 | 在庫管理 | 218 |
| 8 | 生産管理 | 175 |
| 9 | 科学技術 | 158 |
| 10 | 問題解決 | 156 |
| 11 | 大気汚染 | 152 |
| 12 | 技術革新 | 152 |
| 13 | 情報検索 | 149 |
| 14 | 品質保証 | 143 |
| 15 | 経営管理 | 135 |
| 16 | 相互関係 | 135 |
| 17 | 調査結果 | 132 |
| 18 | 生産計画 | 131 |
| 19 | 教育訓練 | 125 |
| 20 | 目的関数 | 117 |
| 21 | 技術開発 | 112 |
| 22 | 経営情報 | 106 |
| 23 | 環境問題 | 104 |
| 24 | 比較検討 | 97 |
| 25 | 相互作用 | 95 |
| 26 | 産業構造 | 93 |
| 27 | 経営計画 | 92 |
| 28 | 経済成長 | 91 |
| 29 | 組織開発 | 90 |
| 30 | 企業経営 | 89 |
| 31 | 情報科学 | 87 |
| 32 | 構成要素 | 86 |
| 33 | 原価計算 | 84 |
| 34 | 必要条件 | 83 |
| 35 | 経営科学 | 81 |
| 36 | 工作機械 | 80 |
| 37 | 技術予測 | 79 |
| 38 | 実験結果 | 74 |
| 39 | 作業研究 | 73 |
| 40 | 事例研究 | 72 |
| 41 | 制約条件 | 71 |
| 42 | 労働時間 | 71 |
| 43 | 市場調査 | 69 |
| 44 | 長期計画 | 69 |
| 45 | 労働組合 | 69 |
| 46 | 評価基準 | 68 |
| 47 | 作業時間 | 68 |
| 48 | 工程管理 | 67 |
| 49 | 実態調査 | 67 |
| 50 | 評価方法 | 65 |

表4　4文字漢字（高頻度順1〜50位）

| seq. | 前接漢字列 | 種類 | 頻度 | seq. | 前接漢字列 | 種類 | 頻度 |
|---|---|---|---|---|---|---|---|
| 1 | 問題 | 290 | 1152 | 51 | 環境 | 89 | 341 |
| 2 | 方法 | 282 | 1139 | 52 | 理論 | 88 | 328 |
| 3 | 管理 | 266 | 2270 | 53 | 報告 | 88 | 306 |
| 4 | 計画 | 262 | 1479 | 54 | 開発 | 86 | 819 |
| 5 | 関係 | 229 | 776 | 55 | 結果 | 86 | 532 |
| 6 | 作業 | 220 | 597 | 56 | 制度 | 85 | 290 |
| 7 | 情報 | 212 | 873 | 57 | 設計 | 85 | 260 |
| 8 | 技術 | 202 | 822 | 58 | 制御 | 85 | 221 |
| 9 | 装置 | 193 | 684 | 59 | 資料 | 82 | 254 |
| 10 | 時間 | 190 | 739 | 60 | 利用 | 82 | 220 |
| 11 | 方式 | 189 | 380 | 61 | 製品 | 81 | 195 |
| 12 | 条件 | 173 | 779 | 62 | 構成 | 79 | 154 |
| 13 | 処理 | 167 | 661 | 63 | 関数 | 77 | 431 |
| 14 | 活動 | 166 | 644 | 64 | 段階 | 77 | 235 |
| 15 | 分析 | 164 | 658 | 65 | 速度 | 77 | 150 |
| 16 | 状況 | 159 | 461 | 66 | 項目 | 77 | 137 |
| 17 | 構造 | 157 | 501 | 67 | 目標 | 76 | 225 |
| 18 | 機能 | 149 | 293 | 68 | 市場 | 76 | 223 |
| 19 | 業務 | 148 | 389 | 69 | 検査 | 76 | 197 |
| 20 | 過程 | 139 | 371 | 70 | 試験 | 75 | 125 |
| 21 | 部門 | 138 | 515 | 71 | 価格 | 73 | 149 |
| 22 | 能力 | 134 | 309 | 72 | 範囲 | 72 | 163 |
| 23 | 企業 | 133 | 691 | 73 | 分布 | 71 | 319 |
| 24 | 会社 | 133 | 465 | 74 | 機構 | 71 | 135 |
| 25 | 内容 | 133 | 387 | 75 | 測定 | 70 | 173 |
| 26 | 調査 | 131 | 404 | 76 | 形態 | 69 | 143 |
| 27 | 工場 | 130 | 392 | 77 | 経済 | 67 | 298 |
| 28 | 状態 | 127 | 253 | 78 | 体制 | 67 | 186 |
| 29 | 研究 | 125 | 67 | 79 | 手段 | 67 | 159 |
| 30 | 組織 | 124 | 328 | 80 | 形式 | 67 | 98 |
| 31 | 計算 | 117 | 365 | 81 | 変化 | 66 | 186 |
| 32 | 設備 | 117 | 315 | 82 | 手順 | 65 | 140 |
| 33 | 可能 | 117 | 305 | 83 | 訓練 | 64 | 256 |
| 34 | 教育 | 116 | 496 | 84 | 工程 | 64 | 242 |
| 35 | 政策 | 114 | 358 | 85 | 機器 | 64 | 220 |
| 36 | 効果 | 111 | 307 | 86 | 変数 | 64 | 205 |
| 37 | 基準 | 108 | 392 | 87 | 行動 | 64 | 174 |
| 38 | 分野 | 108 | 338 | 88 | 要因 | 63 | 132 |
| 39 | 費用 | 106 | 237 | 89 | 部品 | 62 | 150 |
| 40 | 機関 | 102 | 348 | 90 | 目的 | 61 | 143 |
| 41 | 産業 | 102 | 259 | 91 | 予測 | 60 | 258 |
| 42 | 機械 | 101 | 301 | 92 | 実験 | 60 | 110 |
| 43 | 評価 | 99 | 329 | 93 | 効率 | 59 | 217 |
| 44 | 対策 | 98 | 310 | 94 | 水準 | 59 | 213 |
| 45 | 工業 | 95 | 712 | 95 | 部分 | 58 | 78 |
| 46 | 期間 | 93 | 180 | 96 | 原価 | 57 | 150 |
| 47 | 決定 | 92 | 680 | 97 | 領域 | 57 | 118 |
| 48 | 要素 | 92 | 292 | 98 | 手続 | 57 | 93 |
| 49 | 生産 | 92 | 228 | 99 | 単位 | 56 | 84 |
| 50 | 特性 | 92 | 156 | 100 | 社会 | 55 | 240 |

表5　前接漢字列種類頻度の高いもの順（1～100位）

| seq. | 後接漢字列 | 種類 | 頻度 |
|---|---|---|---|
| 1 | 生産 | 189 | 1231 |
| 2 | 企業 | 160 | 865 |
| 3 | 作業 | 156 | 848 |
| 4 | 技術 | 147 | 939 |
| 5 | 経営 | 142 | 1062 |
| 6 | 情報 | 140 | 1236 |
| 7 | 経済 | 135 | 698 |
| 8 | 管理 | 127 | 560 |
| 9 | 各種 | 122 | 186 |
| 10 | 労働 | 114 | 618 |
| 11 | 教育 | 114 | 524 |
| 12 | 自動 | 106 | 366 |
| 13 | 基本 | 105 | 396 |
| 14 | 販売 | 105 | 357 |
| 15 | 社会 | 101 | 435 |
| 16 | 研究 | 100 | 731 |
| 17 | 最適 | 99 | 315 |
| 18 | 計画 | 99 | 269 |
| 19 | 設計 | 94 | 287 |
| 20 | 開発 | 93 | 264 |
| 21 | 製品 | 92 | 279 |
| 22 | 処理 | 88 | 364 |
| 23 | 産業 | 87 | 290 |
| 24 | 一般 | 85 | 190 |
| 25 | 組織 | 84 | 397 |
| 26 | 主要 | 83 | 169 |
| 27 | 環境 | 82 | 579 |
| 28 | 安全 | 82 | 331 |
| 29 | 品質 | 81 | 762 |
| 30 | 機械 | 80 | 389 |
| 31 | 市場 | 79 | 282 |
| 32 | 製造 | 76 | 400 |
| 33 | 工場 | 73 | 177 |
| 34 | 利用 | 70 | 310 |
| 35 | 輸送 | 69 | 201 |
| 36 | 会計 | 68 | 342 |
| 37 | 投資 | 68 | 215 |
| 38 | 関連 | 68 | 152 |
| 39 | 使用 | 67 | 207 |
| 40 | 工業 | 66 | 249 |
| 41 | 特許 | 65 | 281 |
| 42 | 計算 | 65 | 255 |
| 43 | 標準 | 65 | 199 |
| 44 | 国際 | 61 | 271 |
| 45 | 制御 | 61 | 241 |
| 46 | 基礎 | 59 | 184 |
| 47 | 評価 | 58 | 205 |
| 48 | 専門 | 58 | 204 |
| 49 | 時間 | 58 | 200 |
| 50 | 設備 | 58 | 141 |

表6　後接漢字列種類頻度の高いもの順（1～50位）

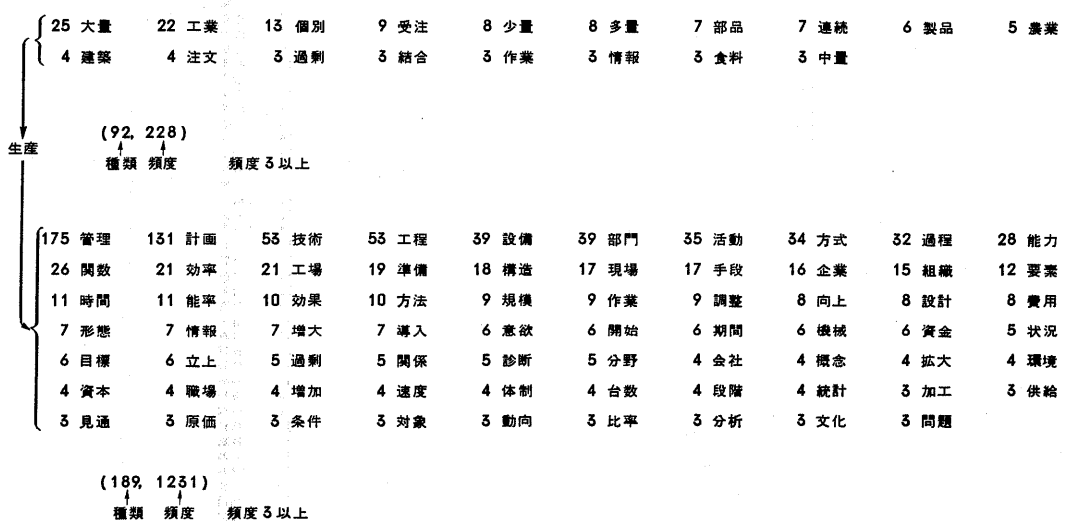



表7　語と語の接続関係（生産）