

自然言語の解析による知識獲得と拡張

— 四文字漢字列をもちいて —

田 中 康 仁
(姫路短期大学)

吉 田 将
(九州工業大学)

自然言語の理解や文脈解析には知識データが必要である。この知識データの獲得と拡張方法について述べる。日本語の中には漢字列がしばしば使われる。この中で四文字漢字列に注目する。この四文字漢字列が二文字・二文字の語基に分割されるものを抽出する。このデータを多量に集め、さらに助詞、助動詞等のパターンにより知識データの拡張を考えている。

日本科学技術情報センターの抄録中より四文字漢字列を抽出し、整理して約16万5千件の知識データを得た。この知識はさらに拡張することができる。

The Acquisition and Expansion of Knowledge Data By Analyzing Natural Language

YASUHIITO TANAKA

SHO YOSHIDA

Himeji College

Kyushu Institute Technology

1-1-12 Shinzaike Honmachi
Himeji
670 JAPAN

680-4 Kawatu
IIZUKA
820 JAPAN

Knowledge data is indispensable in the comprehension of natural language or its context analysis. This paper describes the method for acquiring and expanding such knowledge data.

Among Kanji (Chinese character) strings which are frequently used in the Japanese language, the authors focused on the four-character Kanji strings. Those which could be segmented into word bases of 2 Kanji characters were selected. The authors considered increasing knowledge data by gathering a large extent of such character strings and analyzing them by using particles and auxiliary verb patterns. By this method, 165,000 cases of knowledge data were acquired from selecting pertinent 4 character Kanji strings from the abstracts published by the Japan Information Center of Science and Technology.

1. はじめに

自然言語の理解や文脈解析には知識データが必要である。この知識データを多量に準備するにはどのような方法があるのだろうか？

その方法を紹介すると同時に収集された知識をさらに拡大する方法を考えた。この方法について述べる。知識データの収集については四文字漢字列を利用した。

対象文献は日本科学技術情報センターの抄録データをもちいている。

2. なぜ四文字漢字列を対象とするか？

四文字漢字列を研究対象としたか、その理由を述べてみる。

(1) 機械的抽出ができ、しかも四文字漢字列は種類が多い。

延べ漢字列は二文字漢字列が多い。

(2) 四文字漢字列は二文字⊕二文字に分割できる。つまり2つの語基に分割できる。

(3) 助詞等を使って語の結合を表現することができる。

例

会長辞任 { (i) 会長が辞任する
(ii) 会長を辞任する
(iii) 会長の辞任

(4) 同一語基の中でも、いろいろな訳語を取る。それ故専門用語として、また機械翻訳の訳語選択の面からも研究できる。

例 限界利益 → marginal gains

限界圧力 → limiting pressure

(5) 四文字漢字列から収集した二文字⊕二文字の結合を利用し、知識の拡大が容易にできる。

この方法については、この中で詳細に述べる。

3. 四文字漢字列の分割による知識データの獲得

3.1 四文字漢字列の収集

四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。四文字漢字列は二文字⊕二文字の2つの基礎的概念語(語基)に分割できる。およそ80%~90%のデータが二文字⊕二文字に分割でき、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。

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のようなになる。日本科学技術情報センターの各分野の抄録文から延べ96.5万件の四文字漢字列を抽出し、分析した。この結果16万4千件の知識データが得られた。4回の分析結果を磁気テープ化した。4回の分析毎にデータ量が増大したのであ

るが種類はその割合には増加しなかった。

分野	データ量		種類	知識データ	
	延漢字列	種類		知識データ	種類
(1) 管理システム 技術	78,304	31,932	27,635	0.408	
(2) 電気・金属 機械・土木	159,717	58,591	48,217	0.365	
(3) 物理・電気 機械・化学	512,143	90,840	71,250	0.291	
(4) 管理システム, 原子力, 物理, 金属, 機械, 化学	415,048	114,977	88,630	0.277	
総 合	965,212	-	164,538	-	

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。*はカテゴリを付加中)

延漢字列/種類は減少し0.2程度まで減少することがわかる。このことは再現性が高いことを表わしている。

この四文字漢字列から二文字⊕二文字に分割できる知識データは近く公開する予定である。

1ページ400個のデータを表示するように印刷し、約410ページの資料が得られた。この資料は前接語、後接後順に分類したものの2種類を用意している。

4. 知識データの拡大方法

4.1 文中の語と語の関係について

日本語文の中に四文字漢字列の語と語の関係がどのように含まれているかを調べてみた。二文字・二文字の2つの基礎的概念語の中にどのような助詞や助動詞等が含まれているか調べてみた。

○ 1文字 ○ , ○ 2文字 ○
○ 3文字 ○ , ○ 4文字 ○
○ 5文字 ○ ,

この関係からKWICを作り新しいデータを集中的に収集することができる。

例

～的な～, ～的に～

たとえば「的な」、「的に」, を中心にしたKWICを作ると新しい展開がみられる。

四文字漢字列の分割による語と語の結合関係を明らかにした東京女子大 水谷 静夫教授 の資料からも助詞、助動詞等を中心とした発展を考えていた。

このような関係を実際の文中から抽出したものを(表2)に示す。

この表は日本電気 大山 裕, 首藤正道らが分析した資料をさらに加工したものである。

Seq	結合子	頻度	例
1-1	～ の ～	59	産業の発展
1-2	～ を ～	9	問題を解決
1-3	～ な ～	9	重要な課題
1-4	～ に ～	11	大量に消費
1-5	～ が ～	10	供給が不足
1-6	～ 的 ～	4	世界的規模
1-7	～ 性 ～	1	安定性確保
1-8	～ て ～	2	価格で安定
1-9	～ や ～	1	科学や技術
1-10	～ は ～	2	石油は不足
1-11	～ も ～	1	開発も要請
1-12	～ ・ ～	2	政治・経済
1-13	～ 、 ～	1	有償、無償
2-1	～ する ～	4	消費する生活
2-2	～ した ～	1	安定した供給
2-3	～ 的に ～	4	安定的に確保
2-4	～ 的な ～	2	長期的な課題
2-5	～ 性の ～	1	安定性の確保
2-6	～ 性を ～	2	安全性を確保
2-7	～ 上の ～	2	経済上の問題

表2 語と語の関係と結合子

この文中の語と語の関係の中に含まれる助詞、助動詞、動詞等を集めてみた。さらに助詞相当語などの機能語も収集し、文字数別に整理した。詳細は資料(1)を参照していただきたい。

4.2 知識データの拡大

4.1で収集した助詞、助動詞等の表をもとにして知識データの拡大を試してみる。

例	一様分布	関係有り/無し
	↓	
	一様の分布	×
	一様が分布	×
	一様を分布	×
	一様に分布	○
	一様や分布	×
	一様な分布	○
	⋮	

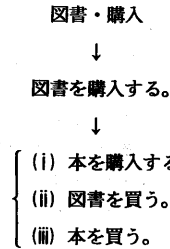
正しい結合だけを印を付けて抽出する。

このようにすれば網羅的に収集でき抜けがなく、確実に知識データを増やすことができる。ディスプレイを用いて対話形式で行うこともできる。また用紙に印刷しておいて、多くの人々に配布し、処理を分散させることもできる。このようにすることで短期間に知識データを得て、拡大することができる。

語基と語基の結合は2つの語の間に独特の動詞等を含んだ組合も考えられる。それらは、その語基と語基の固有のものとして入力すればよい。

5. その他の知識データの拡大方法

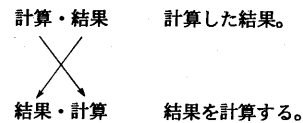
1. 連想による知識の拡大



シソーラスや類語を利用することで容易に拡大できる。

2. 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



16万件の知識データを利用して知識データを拡大することができる。全てに逆が成立するわけではない。

3. 複合語の生成

二つの語の結合関係を一つの語を中心として集めてみると次のようになる。



図1 語の構成図

このような構成図を作ることにより複合語の生成状況がわかる。

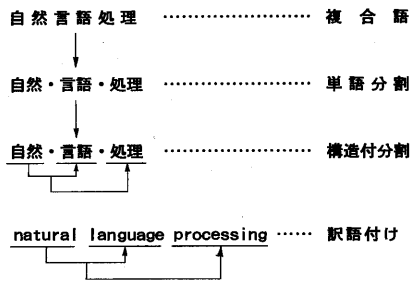


図2 訳語付けのプロセス

このようなシステムを完成させるためには各語基に対する対応訳語を収集、選択合成できるようにしなければならない。

7. シソーラス

語を類似した概念ごとに集め、整理し、上位概念へと発展させ大系化したものである。

類似した概念を集めているため語の持っている特性を大系的につかむことができる。これと他のものを利用して多義性の解消に役立っている。

シソーラスとこの研究の語と語の関係の照合を行い、さらに詳細な概念分類を行わなければならない。そのためにもシソーラスは重要な役割を持っている。

7.1 シソーラスとの照合

この数十万種類の知識データと照合することによって不足している知識データを補充するとか、シソーラスの概念分類をさらに詳しく意味分類し、機械翻訳の多義語の判別、その他に役立つ。

機械可読の大規模なシソーラスが提供されることを期待する。

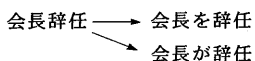
四文字漢字列の分割と助詞、助動詞等の附加により語と語の関係が得られる。この語と語の関係は膨大なものとなり収集作業、整理作業が大変であると考えられる。

確かに n 個の単語を考えると n^2 個の組合せが考えられ、それに助詞、助動詞等の附加により kn^2 個にふくらむ。このなかで意味のあるものは非常に少ないが kn^2 個について調べなければならない。

このような作業は大変であるのでシソーラス体系と照合し、意味がありそうなグループを重点的に調べることを考える。

このようにして無限に近い関係を有限の関係に押えることができる。

① シソーラスとの照合の意味(I)



この語と語の関係から会長、社長、役員等が同一の意味マーカ上にあるか？

シソーラスとの照合による知識データの拡張や拡張意味マーカの確認を行う。

② シソーラスとの照合の意味(II)

四文字漢字列からの語と語の共起関係とシソーラスとの照合は次のような意味がある。

- シソーラスの正しさの検証に役立つ
- 意味マーカの細分化、統合化に役立つ
- 訳し分けの判断と例外の抽出に役立つ

③ シソーラスとの照合の意味(III)

四文字漢字列からの語と語の関係の知識データとシソーラスを組合せることにより、どの概念と動詞が結ばれるかを知ることができる。



A1とBという動詞の間に結合関係があるとわかればA1に属するA2グループ内全部の語にBという動詞が結合するか否かを検討することができる。

もしA1個有のものであればA1とBとの結合とみることができる。

もしA2のグループ内の語とBが結合することがわかればそれは同一の訳語を取るか否かを調べる。

さらにA3まで拡大し、Bと結合を調べる。同様の方法を取りシソーラスの上位概念へと発展させて考える。

語と語の結合を全ての語について調べることは出来ないので、該当する動詞との結合を語と語の関係とシソーラスとの照合により機械的に知り、その後、該当箇所を局所的に詳細に調べる。このようにすると大巾に労力の削減をはかることができる。

このためからも語と語の関係の知識データでは前接語は基礎的概念語になるようにしている。このためにも四文字漢字列の分割は大変意味がある。

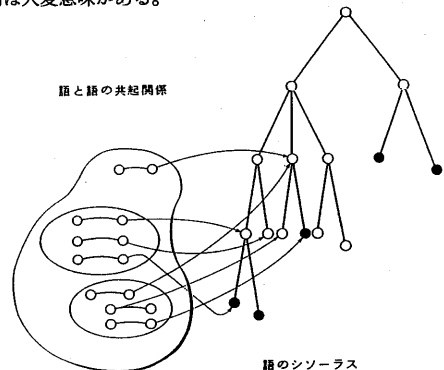


図3 語と語の共起関係と語のシソーラスとの関係

7.2 シソーラスと長単位用語

文の中に使われる語には複合語や長単位用語が多くもちいられている。これら長単位用語から基礎語を導く方法があるか、長単位用語の多くの語がシソーラスの体系に組込まれていないかならなければならない。

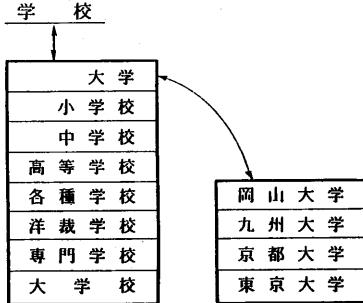


図4 語と語の関係をさらに発展させるために小さい意味の集合を考える。小さい意味の集合に上位、下位の意味の集合を作る。

7.3 シソーラスと知識の体系化

高品質の機械翻訳システムや文章理解システムを実現してゆくためには機械に知識をうめこまなければならない。この知識体系がシソーラス体系であり、概念体系である。

これらを図に示すと次のようになる。

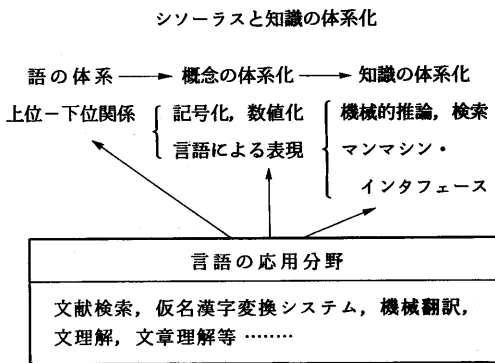


図5 シソーラスと知識の体系化

我々は機械にうめこむ知識体系のためにシソーラスを利用していると言える。

8. 知識データの評価

知識データの収集方法が確立し、知識データが大量に収集できるようになってきた。今後は知識データの評価を行い、何が不足しているか、収集する知識データの重複はどの程度発生しているか、どのような分野の知識データが不足しているか等を検討しなければならない。

また、集められた知識データの追加、修正が簡単に行えるような環境を作ってゆかねばならない。

知識データ抽出作業は第一歩を進めた段階である。今後このデータを機械翻訳システムまで組込むとすると次のような段階を通らなければならない。

シソーラスについても評価を行う方法の確立が必要である。このための条件は何かをあげておかなければならない。

- ① シソーラスの語数
- ② シソーラスと応用分野
- ③ シソーラスの内容の公開、機械可読性
- ④ シソーラスと各種利用ユーティリティの整備
- ⑤ 知識の継承と推論システムの関係

等が考えられる。

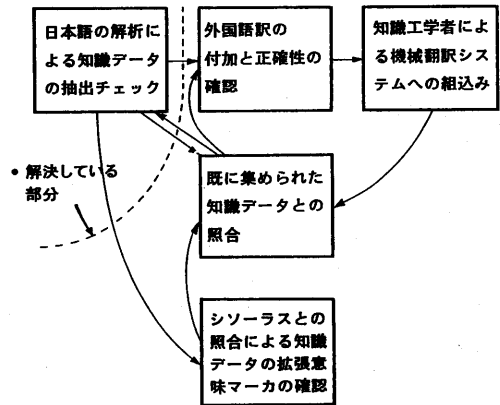


図6 語と語の関係の知識データが機械翻訳システムに組込まれるまでの作業プロセス

9. 今後の課題

① 四文字漢字列の知識データを全部翻訳する。

- 1件当りの翻訳・チェック費用 500円
- 16万件の翻訳費用 8,000万円
- 1日当りの作業量(1人) 100件/日
- 延人日(4人で約2年間) 1,600人日

費用が限られた場合としては頻度の高いものから翻訳する方法と、ある語から順次翻訳する方法が考えられる。

翻訳した内容の一部を最後に示す。

もし、これだけの費用を投入するとすれば優秀な人材への投入費用、データの検査や照合のための費用としてどの程度割りふるかも考えなければならない。

さらに、翻訳速度の向上、正確さを保つための方法を考えなければならない。

② この実験では集められなかったデータ等について、対象とする分野が異っていたためか、使われることが無くなってしまったか、等を検討する必要がある。

③ 機械翻訳システムや仮名漢字変換システム、音声や文字認識システムへ応用し、実用化する。

④ 四文字漢字列からの語と語の関係の知識データが多量に安価に入手可能になると自然言語の研究も新しい方向に進まねばならない。哲学の言葉に「量的拡大は質的变化をもたらす」とあるように次の発展が必要である。

語と語の関係による知識データが安価に大量に入手可能になる時代を迎えた。

-
- ① 文法の体系化(単純化, 詳細化)をすることができる
 - ② 構文解析における構文木の多発防止
 - ③ 機械翻訳の多義性の解消
訳の向上がはかられる
 - ④ 文字認識, 音声認識の精度向上をはかる
 - ⑤ 同音, 同形異義語の判別を簡便化する
 - ⑥ 自然言語処理の意味解析の発展をうながす

10. おわりに

自然言語の知識獲得は自動的に抽出するのが望ましいが、構文解析の多義性を解消するためにも知識データが必要であるとか、困難な面が多い。

ここでは単純な方法で得られる四文字漢字列とその整理、拡張の方法について述べた。又、機械翻訳に必要な訳語付けの方法についても考察した。

この研究の一部は文部省科学研究費によって行った。

文部省科学研究費特定研究「言語情報処理の高度化」課題番号 62110006 A04

総括者 長尾 眞 分担代表者 野村 雅 昭

最後に四文字漢字列の分割について指導して下さった東京女子大学 水谷 静夫 教授 に感謝の意を表します。

参 考 文 献

- (1) 稲永, 小西: カナ文字の機械処理のための辞について AL 76-39 電子通信学会技術研究会資料 1976
- (2) 野村雅昭: 四字漢語の構造 国立国語研究所報告 54 「電子計算機による国語研究Ⅶ」 1875.3
- (3) 野村雅昭: 現代漢語の語構成について 情報管理 VOL18 No.11 1976.2
- (4) 水谷静夫: 語構成概説 岩波国語辞典第三版 岩波書店 1980.3
- (5) 小西友七編: 英語基本動詞辞典 研究社出版 1980.9
- (6) 内田裕士・増山顕成: 機械翻訳における概念の変換について 情報処理学会自然言語処理研究会 39-3 1983
- (7) 鈴木重幸・鈴木康之: 日本語文法・連語論(資料編) 言語学研究会編 むぎ書房 1983
 (この資料は国語学の研究者が連語として取り扱い動詞) の分類を行っている。
- (8) 勝俣銓吉郎編 新英語大辞典 研究社
- (9) 金田一京助 他 新明解国語辞典 三省堂
- (10) 森田良行 基礎日本語 角川書店
- (11) 鶴丸弘昭・内田 彰・日高 達・吉田 将: 意味辞書開発のための国語辞書の活用について 情報処理学会第28回(昭和59年前期)全国大会 3M-5 1984
- (12) 鶴丸弘昭・内田 彰・日高 達・吉田 将: 国語辞書を利用した単語間の階層付けシステムの開発について 情報処理学会第28回(昭和59年前期)全国大会 3M-6 1984
- (13) 牧野武則・石川淳子: Venus, 翻訳のための知識の学習について 情報処理学会第28回(昭和59年前期)全国大会 6M-4
- (14) 吉田 将: 機械辞書について 日本語学 3(7) 明治書院 1984
- (15) 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
- (16) 溝口文雄 他 大特集: 機械翻訳 情報処理 Vol 26 No. 1985.10
- (17) 新田義彦 他 計算言語学 情報処理 Vol 27 No. 8 1986.8
- (18) 水谷静夫・石綿敏雄 他 文法と意味 I 朝倉日本語講座
- (19) 水谷静夫 他 文字・表記と語構成 朝倉日本語新講座
 (この中の水谷静夫の論文と付録の表が四文字漢字列) と意味の研究をまとめたものである。
- (20) 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 第27回全国大会論文集 情報処理学会 1983.10
- (21) 田中康仁・水谷静夫・吉田 将: 語と語の関係について — 知識データ収集方法についての試み — 第20回情報科学技術研究会論文集 日本科学情報センター 1983.10
- (22) 田中康仁・吉田 将: 自然言語の分析とデータ量 — 異った分野のデータを基にした分析 — 情報処理学会第29回(昭和59年後期)全国大会 6N-4 1984
- (23) 田中康仁・西川仙之・吉田 将: 語と語の関係について — 知識データの自動カテゴリー化について — 情報処理学会第29回(昭和59年後期)全国大会 6N-5 1984
- (24) 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研究会 41-4 1984
- (25) 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 「自然言語処理技術」 シンポジウム 1984.11
- (26) 田中康仁・吉田 将: 自然言語の分析による知識データ 情報処理学会自然言語処理研究会 54-3 1986.3
- (27) 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会論文集 1986.3
- (28) 田中康仁・吉田 将: Acquisition of Knowledge Data by analyzing Natural Language 11th International Conference on Computational Linguistics COLING '86 1986.8
- (29) 田中康仁・吉田 将: 知識データ(語と語の関係)に多義性の解消 情報処理学会自然言語処理 60-3 1987.3
- (30) 田中康仁: 語と語の関係解析用資料 — “を”を中心とした解説編・資料編(I), (II) 文部省科学研究費特定研究「言語情報処理の高度化」総括班 1987.3
 (これは前回作成した語と語の関係の資料である。) (JICSTデータの分析によって作成した。)
- (31) 田中康仁 語と語の関係による知識データについて計量国語学論集 秋山書店 1987.3

添付資料(1)

知識獲得の方法(枠組の中での知識拡張方法: 検証による知識獲得)

1-1	～ の ～	3-15	～ などの ～
1-2	～ を ～	3-16	～ なりを ～
1-3	～ が ～	3-17	～ にして ～
1-4	～ な ～	3-18	～ による ～
1-5	～ に ～	3-19	～ になる ～
1-6	～ で ～	3-20	～ の方を ～
1-7	～ や ～	3-21	～ のみを ～
1-8	～ も ～	3-22	～ のもつ ～
1-9	～ と ～	3-23	～ のない ～
1-10	～ へ ～	3-24	～ のため ～
1-11	～ 的 ～	3-25	～ までの ～
1-12	～ 性 ～	3-26	～ をより ～
1-13	～ 化 ～	3-27	～ をどう ～
1-14	～ . ～	3-28	～ 的にも ～
1-15	～ , ～	3-29	～ 的には ～
2-1	～ から ～	4-1	～ あるいは ～
2-2	～ され ～	4-2	～ していく ～
2-3	～ した ～	4-3	～ していた ～
2-4	～ する ～	4-4	～ といった ～
2-5	～ との ～	4-5	～ としての ～
2-6	～ とか ～	4-6	～ てきない ～
2-7	～ ない ～	4-7	～ てさえも ～
2-8	～ なり ～	4-8	～ において ～
2-9	～ にも ～	4-9	～ における ～
2-10	～ には ～	4-10	～ に必要な ～
2-11	～ まで ～	4-11	～ に基づく ～
2-12	～ への ～	4-12	～ に対する ～
2-13	～ より ～	4-13	～ に対して ～
2-14	～ での ～	4-14	～ に向けて ～
2-15	～ では ～	4-15	～ に関して ～
2-16	～ 的に ～	4-16	～ に関する ～
2-17	～ 的な ～	4-17	～ に於いて ～
2-18	～ 性の ～	4-18	～ について ～
2-19	～ 性を ～	4-19	～ になった ～
2-20	～ 上の ～	4-20	～ に属する ～
2-21	～ 側に ～	4-21	～ によって ～
2-22	～ 化の ～	4-22	～ のための ～
2-23	～ 内の ～	4-23	～ のために ～
3-1	～ および ～	4-24	～ のなかの ～
3-2	～ された ～	4-25	～ のなかで ～
3-3	～ される ～	4-26	～ のように ～
3-4	～ しまう ～	4-27	～ を用いて ～
3-5	～ すべき ～	5-1	～ に基づいて ～
3-6	～ すると ～	5-2	～ に於いての ～
3-7	～ だけを ～	5-3	～ についての ～
3-8	～ ている ～	5-4	～ に対しての ～
3-9	～ できる ～	5-5	～ においての ～
3-10	～ である ～	5-6	～ にも拘らず ～
3-11	～ という ～	5-7	～ に関しての ～
3-12	～ として ～		
3-13	～ と同じ ～		
3-14	～ とかの ～		

添付資料(2)

限界

bound ; ある境界, 囲い内に限る。～を設ける。

limit ; これ以上越えてゆくことの出来ない空間・時間・範囲の制限

margin(marginal) ; 限界(に近い状態)

1	限界利益	marginal gain(s)
2	限界費用	marginal cost
3	限界基準	boundary standard
4	限界決定	limiting determination
5	限界収率	marginal profit rate
6	限界濃度	limiting concentration
7	限界分析	marginal analysis
8	限界圧縮	limiting compression
9	限界圧力	limiting pressure
10	限界価値	marginal value
11	限界距離	limiting distance
12	限界原価	marginal cost
13	限界効率	marginal efficiency
14	限界条件	boundary condition
15	限界単価	marginal unit price
16	限界定理	limit theorem
17	限界部分	marginal part
18	限界容量	limiting capacity
19	限界利点	marginal profit point
20	限界計算	limit calculation
21	限界意見	marginal opinion
22	限界効用	marginal utility
23	限界思考	marginal thinking
	その他 ; 限界検査	marginal checking
	限界曲線	limiting curve

添付資料(3)

限界

bound ; ある境界, 囲い内に限る。～を設ける。
 limit ; これ以上越えてゆくことの出来ない空間・時間・範囲の制限
 margin(marginal) ; 限界(に近い状態)

1	信 頼 限 界	confidence limit confidence bound
2	許 容 限 界	admissible limit allowable limit
3	適 用 限 界	applicable limit applicable bound
4	妨 害 限 界	limit of interference bound of interference
5	安 全 限 界	safety limit
6	管 理 限 界	limit of control
7	供 給 限 界	limit of supply
8	成 長 限 界	limit of growth
9	価 格 限 界	price limit
10	機 能 限 界	limit of function bound of function
11	効 率 限 界	limit of efficiency bound of efficiency
12	危 険 限 界	risk limit risk bound
13	図 的 限 界	graphic limit
14	量 的 限 界	quantitative limit
15	低 減 限 界	limit of decrease bound of decrease
16	爆 発 限 界	explosion limit explosion bound
17	装 置 限 界	device limit
18	騒 音 限 界	loudness limit
19	能 力 限 界	limit of ability
20	規 格 限 界	bound of standard
21	採 算 限 界	accounting limit accounting bound
22	効 力 限 界	limit of effect bound of effect
23	最 低 限 界	minimum limit
24	対 象 限 界	bound of object
25	仕 様 限 界	bound of specification
26	下 方 限 界	lower limit
	その他 ; 処 置 限 界 ; 処 置 限 界 ; 誤 差 限 界 ; 検 査 限 界	action bound limit of error check bound

添付資料(4)

中心

1	中心課題	main theme (subject, problem, exercise)
2	中心転送	{ • central transfer • transfer of the center
3	中心程度	central degree (standard, grade)
4	中心装置	main equipment (device, unit)
5	中心記述	main description
6	中心命題	central proposition
7	中心付近	{ • near the central • center's neighborhood
8	中心概念	central concept (notion)
9	中心観念	main idea
10	中心人物	{ • leader • central figure
11	中心方向	{ • principal direction • direction to center
12	中心欲求	main request
13	中心定理	central theorem
14	中心温度	central temperature
その他	中心示度	(a typhoon with) a central atmospheric pressure
	中心円	center circle
	中心曲面	center surface
	中心距離	distance between centers
	中心角	central angle
	中心線	central line
	中心点	central point, center, omphalos
	中心拡大	central extension
	中心対称	central symmetry
	中心射影(投影)	central projection
	中心列	central series
	中心差分	central difference
	中心積率	central moment
	中心運動	central motion
	中心傾向	central tendency
	中心原子	neutral atom
	中心価標	centric bond
	中心式	centric formula

添付資料(5)

中心

1	人 間 中 心	{	• ~と考えること	humanism to attach the greatest importance to mankind
2	在 庫 中 心		• ~に	stocks being the primary portion
3	事 例 中 心		• ~で	mainly by examples
4	委 託 中 心		• ~て	mainly on commission
5	欧 州 中 心	{	• ~と考えること	Europeanism to attach the greatest importance to Europe
6	会 社 中 心		• ~と考えること	to attach the greatest importance to the company
7	規 制 中 心		• ~で	mainly by regulation
8	経 験 中 心		• ~で	mainly by experience
9	高 校 中 心		• ~に	high-schools being the primary portion
10	作 業 中 心		• ~に	work being the primary portion
11	仕 事 中 心		• ~の	business oriented
12	市 場 中 心		• ~の	market oriented
13	資 本 中 心		• ~と考えること	to attach the greatest importance to capital
14	生 産 中 心		• ~と考えること	to attach the greatest importance to production
15	石 炭 中 心	{	• ~と考えること	to attach the greatest importance to coal
			• ~に	coal being the primary portion
16	鉄 鋼 中 心	{	• ~と考えること	to attach the greatest importance to iron and steel
			• ~に	iron and steel being the primary portion
17	電 話 中 心	{	• ~に	telephone being the primary portion
			• ~で	mainly by phone
18	読 者 中 心		• ~の	readers oriented
19	品 質 中 心		• ~の	quality oriented
20	物 流 中 心		• ~に	material transportation being the primary portion
21	福 祉 中 心		• ~と考えること	to attach the greatest importance to welfare
22	名 詞 中 心		• ~に	nouns being the primary portion
23	目 的 中 心	{	• ~と考えること	to attach the greatest importance to purpose
			• ~の	purpose oriented
24	問 題 中 心	{	• ~で	mainly by exercises
			• ~に	exercises being the primary portion
25	利 益 中 心	{	• ~と考えること	to attach the greatest importance to profit
			• ~の	profit oriented
26	学 習 中 心	{	• ~と考えること	to attach the greatest importance to study
			• ~で	mainly by study
27	企 業 中 心	{	• ~と考えること	to attach the greatest importance to enterprise
			• ~の	enterprise oriented
28	技 術 中 心	{	• ~と考えること	to attach the greatest importance to technology
			• ~の	technology oriented
29	機 能 中 心	{	• ~と考えること	functionalism to attach the greatest importance to functionalism
30	一 定 中 心		• ~に	definiteness being the primary portion

四文字漢字列とそれに関連する内容について既に発表した内容一覧

1. 出 版 物

- (1) 田 中 康 仁
語と語の関係による知識データについて
「計量国語学と日本語処理」 — 理論と応用 —
水谷静夫教授還暦記念会編 秋山書店 1987. 3

2. 日本科学技術情報センター(JICST)研究集会論文集

- (1) 田 中 康 仁, 水 谷 静 夫, 吉 田 将
語と語の関係について — 知識データ収集についての試み
第 20 回情報科学技術研究集会発表論文集 1984. 3
- (2) 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
第 21 回情報科学技術研究集会発表論文集 1985. 3
- (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について
— 大量データによる分析 —
第 22 回情報科学技術研究集会発表論文集 1986. 3

3. 情報処理学会自然言語研究会

- (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, 吉 田 将
自然言語の分析による知識データ
情報処理学会自然言語処理 54-3 1986. 3
- (4) 田 中 康 仁, 吉 田 将
知識データ(語と語の関係)による多義性の解消
情報処理学会自然言語処理 60-3 1987. 3

4. 情報処理学会全国大会

- (1) 田 中 康 仁, 水 谷 静 夫, 吉 田 将
語と語の関係について
情報処理学会第 27 回(昭和 58 年後期)全国大会 5F-7
1983. 10
- (2) 田 中 康 仁, 長 田 孝 治, 吉 田 将
自然言語の分析とデータ量
情報処理学会第 28 回(昭和 59 年前期)全国大会 6L-9
1984. 3
- (3) 田 中 康 仁, 吉 田 将
自然言語の分析とデータ量
— 異った分野のデータを基にした分析 —
情報処理学会第 29 回(昭和 59 年後期)全国大会 6N-4
1984. 10
- (4) 田 中 康 仁, 西 川 仙 之, 吉 田 将
語と語の関係について
— 知識データの自動カテゴリー化について —
情報処理学会第 29 回(昭和 59 年後期)全国大会 6N-5
1984. 10
- (5) 田 中 康 仁, 吉 田 将
語と語の関係について — 大量データによる分析 —
情報処理学会第 30 回(昭和 60 年前期)全国大会 5G-1
1985. 10
- (6) 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 年後期)全国大会 3H-9
1985. 10
- (7) 田 中 康 仁, 吉 田 将
語と語の関係による知識データの獲得
— 'が' と 'を' を中心とした —
情報処理学会第 35 回(昭和 62 年後期)全国大会
1987. 10