

知識データ（語と語の関係）による多義性の解消

田中 康 仁
(姫路短期大学)

吉 田 将
(九州工業大学)

知識データによる多義性の解消方法について、多義性の問題点、多義性の解消のための幾つかの方法と問題点の検討を行い、この中で特に語と語の関係による知識データが多義性の解消のために有効であることがわかった。

知識データの収集方法としては、格助詞「を」中心とした一般文のKWICを使い、その中から局所的解析で知識データを集めた。

約79万行のKWICを解析し、28万種類の語と語の関係の知識データを得た。

この知識データを翻訳し、整理することにより機械翻訳の多義性の解消ははかれる。翻訳等に少し、費用はかかるが解決の第1歩がつかめた。知識データをさらに収集し、整理し、新しい観点から文法規則の体系化を進めるべき時が来は始めている。

A Method for Appropriately Selecting the Multivocal Words By Utilizing
Knowledge Data (Relationship Between Words)

YASUHITO TANAKA

SHO YOSHIDA

Himeji College

Kyushu Institute Technology

1-1-12 Shinzaike Honmachi
Himeji-shi Hyogo-ken
670 JAPAN

1-1 Sensui-cho Tobata-ku
Kitakyushu-shi Hukuoka-ken
840 JAPAN

This paper describes the results of considering the problems and some methods for solving the multivocal problems in words by using knowledge data. As a result, it was found that the knowledge data based on the relationship of words was especially effective in solving the multivocal word problems.

The knowledge data was gathered from partially analyzing general sentences by using a KWIC list with the kakujoshi (postpositional case auxiliary) "wo (を)" as its base.

From analyzing approximately 790,000 lines of the KWIC list 280,000 types of knowledge data (relationships between words) were obtained.

By translating these knowledge data, and re-arranging them, the problem of multivocal words in machine translation can be solved. Though a small cost may be required to translate the data, the knowledge data obtained through this study has shown some possibility to act as a method for solving the problem of multivocal words.

The time has come to gather more knowledge data, re-arrange it and systemize the grammatical rules from a new aspect.

1. はじめに

機械翻訳をはじめとする自然言語の研究における重要な課題は多義性の解消である。この問題については幾つかの提案がなされているが、まだまだ完全な解決方法はない。ここでは今までの研究と問題点を分析し一つの解決方法である“語と語の関係による知識”を用いる方法と、この知識データの収集方法について、具体的に述べる。

2. 多義性について

どのように多義性が発生するかを具体的に考える。

例を用いて説明する。

例 引く(動詞)

(i) 引っ張る・曳く・索く・挽く

draw, pull, haul

(ii) 引きずる・曳く

drag, draggle, trail (すそなどを)

blend [引きまげる]; [引きまげる] bend

(iii) 引きつける

attract, draw, catch, arrest, win

(iv) 導く

lead

(v) (線・地図を)描く

draw, let fall [垂線など]

(vi) 引き入れる, 導く

lead, admit, [敷設する] lay on

install

(vii) 引用する

cite, quote, refer to

(viii) (字を)捜し出す

look up, see = (consult turn,
refer to)

(ix) 減ずる

subtract from, deduct, reduce

abate, cut down, take off, allow

discomat

(x) 塗る lay on, apply, daub

(xi) すり減らす blunt

(xii) ひいきする patronize

(xiii) 抜きとる → 引き抜く

(xiv) こっそり盗む → 盗む

• その他

「新和英大辞典」研究社より引用

このように「引く」という言を調べると14以上の意味がある。

このように幾つもの意味を我々は文章中又は音声の中から適切なものを判断している。

この作業を計算機で行うとすれば「引く」という一語を操作しても判別することはできない。何らかの他の要素と組合せなければならない。それではこの他の要素としてはどのようなものであればよいのであろうか。1つの語の多義性を平均30~40とすればこれを判別するには最低7Bit以上の要素の語があればよいことになる。語には冗長度が必要なので漢字1字(16Bit)程度以上の情報があれば判別できることがわかる。

3. 多義性の解消方法

多義性の解消方法にはどのようなものがあるのだろうか?

(1) 語と品詞

left は名詞, 形容詞, 副詞, 動詞で少しづつ使われ方が異なる。

left (n) 左, 左方, 左側, 左翼, 左党

left (a) 左の, 左方の, 左側の, 左翼の

left (ad) 左に, 左側に, 左方に

left (v) leave の過去・過去分詞

leave 去る, 止(よ)す, 退校, 放置する

残す, 遺贈する, ゆだねる, 託す,

渡す, させる, 行き過ぎる, ……

品詞によってはあまり多義性に有効であるとは思われない。

(2) 専門用語

専門用語はある特定の分野で使われるもので用語

の中には多義性はあまりみられない。しかし用語によっては多義性がある場合もある。

文部省の専門用語を調べた中では数%以下である。一般用語と専門用語の間に起きる多義について特に注意しなければならない。

専門用語は専門区分を付けることによって利用時に判別を助けることが出来る。

多義性の多い例

association

会合〔化学〕	群叢〔植物〕
関連〔動物〕	集落〔星の〕〔天文〕
協会〔図書館〕	対合〔染色体の〕〔遺伝〕
群集〔植物〕	連合〔動物〕

(3) 複合語

専門用語とまではいえないが複数の語や語基が結合して複合語を作っている。これは語長が長く、多義性は発生しにくい。

leap year	うるう年
race cup	優勝杯

(4) 慣用表現

自然言語の中には慣用的な表現がある。これを集め辞書にすることによって多義性をうまく解消することができる。

例	as soon as	～するや否や
	so help me God.	誓って申します。

これは今後研究しなければならないテーマである。

(5) 格文法と格

格文法により文を解析し、その格のとりうる意味にもとづき多義性を判別するという方法が一般的に用いられている。しかし格の意味によって各動詞のもつ多義性が全て判別できるものではない。又、各語彙に意味マーカのようなものを付けなければならない。この作業は大変な労力が必要である。

この方法を用いても多くの例外が発生し、個々の事例を分析し、何が規則に適用でき、何が例外かを調べなければならない。

(6) 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多義性の解消に役立てている。

(7) 語と語の関係

語は色々な語と結合するが、よく調べてみると特定の語との共起関係が強いものが見うけられる。この共起関係の強いものを多量に集め、利用すれば、語の多義性が解消できる。

例	問題を解く	→ solve a problem
	包を解く	→ untie a package

ここではこの語と語の関係についてのデータ収集方法を述べる。

(8) その他

文と文の解析によって語の多義性を解消する方法等が最近研究されている。

今後の研究に期待したい分野である。

省略文、代名詞の指示物、文脈等の研究がある。

4. 語と語の関係データの抽出

4-1 一般的方法

一つの語は無限に多くの語と結合することができるので、語の活動範囲や条件を明確にすることができないのではないかとこの疑問が起る。また語自身も無限にあり、これらを全て調べあげることも大変な労力と時間がかかる。しかし、実際の語を調べてみると一つの語に関係する語は限られている。

例えば、電話という語を考えてみると、電話の特性は通信の手段、物体、場所、……というように限られる。通信の手段としての機能、電話独特の特徴は電話独特のものである。これについては語と語の関係を数えあげるとは簡単であり有限である。一般的な物体、場所としての語と語の関係を数えあげるとは大変困難である。

但し、これらのうち主要なものは簡単にまとめることができる。語に特有な語や使用頻度の高い語と語の関係をテーブルにまとめ、その他のものはシステムに

プリセットされたデフォルト値を用いるようにする以外に方法はないであろう。

001 電話をかける	018 電話を聞く
002 電話をきる	019 電話を受ける
003 電話を持ち上げる	020 電話を盗聴する
004 電話をこわす	021 電話をけなす
005 電話を握る	022 電話を待つ
006 電話を持つ	023 電話を持たせる
007 電話を改良する	024 電話を聞く
008 電話を作る	025 電話が鳴る
009 電話を製作する	026 電話で伝える
010 電話を組立てる	027 電話で話す
011 電話を閉路する	028 電話で連絡する
012 電話を引く	029 電話に出る
013 電話を撤去する	030 電話の声
014 電話を売る	031 電話の部品
015 電話を販売する	032 電話の金
016 電話を買う	033 電話の料金
017 電話を購入する	034 電話のベル

一つの語彙に関する語彙は限られている。

表1. 一つの語に関する語は限られている

高いとか美しい……という語は使用頻度も高く、個別に語の活動範囲や条件を決めにくいものもある。これらについては一般の文法と“高い”とか“美しい”で最も多く使われる語の意味を含ませ、それ以外の場面で使用する特別の場合の高いとか、美しいという意味の使用条件を語と語の関係で規定しなければならない。使用頻度の低い語と語の関係については個別規則を使い、さらに一般文法を適用することになる。

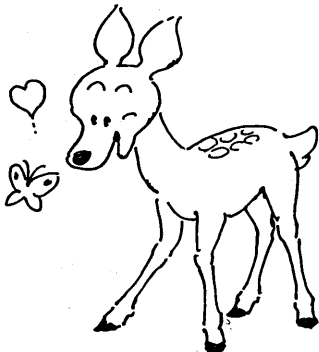
4-2 知識データの収集方法

一般文の中から助詞、助動詞を利用し、KWIC を用いて知識データを抽出する方法を利用した。

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、……

KWICの例を次にあげてみる。



Seq.No	KWIC
244509	期、角育期、開花期、結実期というサイクルを 繰り返す。
244510	不完全なら完全な葉まで上記を 繰り返す。
244511	セトンを加えて溶解したあと水を加える操作を 繰り返す。
244512	1.1を混ぜて清水を注ぎ換えます。抽出操作を 繰り返す。
244513	抽出液を1.1に抽出液を加え、以下抽出液の抽出 繰り返す。
244514	抽出液を1.1より再び抽出液を加える操作を 繰り返す。
244515	をグリブジョーでつまんで抽出液を抽出する。 繰り返す。
244516	を筒の中には上流管をとりつてクロマトグラフ法を 繰り返す。
244517	はラジが動かないことによるが、抽出液の抽出を 繰り返す。
244518	抽出液を一定比率まで変換抽出するサイクルを 繰り返す。
244519	で完全に抽出、非抽出液もまた干渉分析法を 繰り返す。
244520	またカッピングとホルムディングを 繰り返す。
244521	はラジが動かないことによるが、抽出液の抽出を 繰り返す。
244522	期から抽出液へ、可動相を異なる極性の抽出液を 繰り返す。
244523	の比重量の異なる抽出液の抽出を 繰り返す。
244524	しばしば使われるが、底層と抽出液のサイクルを 繰り返す。
244525	タンクにビュースに接続してスイッチングを 繰り返す。
244526	同じ物の領域に加える抽出液の抽出を 繰り返す。
244527	第2、第3層は操作を 繰り返す。
244528	抽出液を濃縮して抽出液を抽出する。 繰り返す。
244529	これら抽出液をむろとろに抽出液を加え、抽出液を 繰り返す。
244530	抽出液を抽出し、抽出液を抽出する。 繰り返す。
244531	抽出液を抽出し、抽出液を抽出する。 繰り返す。
244532	抽出液を抽出し、抽出液を抽出する。 繰り返す。
244533	抽出液を抽出し、抽出液を抽出する。 繰り返す。
244534	抽出液を抽出し、抽出液を抽出する。 繰り返す。
244535	抽出液を抽出し、抽出液を抽出する。 繰り返す。
244536	抽出液を抽出し、抽出液を抽出する。 繰り返す。
244537	抽出液を抽出し、抽出液を抽出する。 繰り返す。
244538	抽出液を抽出し、抽出液を抽出する。 繰り返す。
244539	抽出液を抽出し、抽出液を抽出する。 繰り返す。

('を' を中心としたKWICの例、'を' の左側を分類したものを)

KWIC の例

Seq.No	KWIC
555605	組合せによる条件付き冷却過程不安定に対する計算法を 繰り返す。
555606	の方法をさらに一般化した冷却過程不安定に対する計算法を 繰り返す。
555607	抽出液に対する全体の抽出液の抽出液を利用する計算法を 繰り返す。
555608	しおいに置きかけた抽出液を抽出する計算法を 繰り返す。
555609	抽出液を抽出し、抽出液を抽出する計算法を 繰り返す。
555610	抽出液を抽出し、抽出液を抽出する計算法を 繰り返す。
555611	に基づいて抽出液を抽出し、抽出液を抽出する計算法を 繰り返す。
555612	に基づいて抽出液を抽出し、抽出液を抽出する計算法を 繰り返す。
555613	抽出液を抽出し、抽出液を抽出する計算法を 繰り返す。
555614	抽出液を抽出し、抽出液を抽出する計算法を 繰り返す。
555615	抽出液を抽出し、抽出液を抽出する計算法を 繰り返す。
555616	新しい方法によってのPCAによる計算法を 繰り返す。
555617	抽出液を抽出し、抽出液を抽出する計算法を 繰り返す。
555618	抽出液を抽出し、抽出液を抽出する計算法を 繰り返す。
555619	よい、計算法を抽出し、抽出液を抽出する計算法を 繰り返す。
555620	抽出液を抽出し、抽出液を抽出する計算法を 繰り返す。
555621	抽出液を抽出し、抽出液を抽出する計算法を 繰り返す。
555622	抽出液を抽出し、抽出液を抽出する計算法を 繰り返す。
555623	抽出液を抽出し、抽出液を抽出する計算法を 繰り返す。
555624	抽出液を抽出し、抽出液を抽出する計算法を 繰り返す。
555625	抽出液を抽出し、抽出液を抽出する計算法を 繰り返す。
555626	抽出液を抽出し、抽出液を抽出する計算法を 繰り返す。
555627	抽出液を抽出し、抽出液を抽出する計算法を 繰り返す。
555628	抽出液を抽出し、抽出液を抽出する計算法を 繰り返す。
555629	抽出液を抽出し、抽出液を抽出する計算法を 繰り返す。

('を' を中心としたKWICの例、'を' の左側を分類したものを)

KWIC の例

表2 助詞、助動詞を中心としたKWIC

格助詞 'を' を 'を' を選んだ理由は次のようなことからである。

- ① 名詞と動詞の関係がつかみやすい。
他動詞は必ず目的語を必要とする。
他動詞は自動詞よりも数が多く、使われる頻度も高い。
- ② 文字 'を' は容易に判別でき、KWIC が作りやすい。
- ③ を、が、へ、から、より、の……等を用い少量の File で実験的にKWICを作ってみた。'を' が一番語と語の関係をつかみやすい。

このような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“計算法” “を”の後の“繰返す”という語を別々に抽出しておく。'を'の前の語を前接語、'を'の後の語を後接語と名付ける。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, 可能)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を作り分析した。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79万行(KWICの行数)である。

このデータの中から重複をまとめ、約28万件程度の知識データが得られた。順次“を”以外の助詞、助動詞へも拡大する予定である。

79万行の“を”を中心としたKWICから前接語、後接語を抽出する方法を述べる。

● 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)で集約したデータは154,402件である。
- (iv) この前接語の中で1~3文字の漢字列とカタカナと漢字1文字で構成されているものを抽出すると約4万件のデータを得る。
 この中をさらに分析し、不用なものを省き整理する。
- (v) (iii)で集めた前接語と照合し、前接語として含まれていないものは追加する。
 一部分は作業中であるが、ほぼ問題なく作業ができる見通しがついている。
- (vi) 最終的には約3万5千語程度に整理される。

● 後接語の抽出方法

- (i) KWICの後接語が含まれている文字列から7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て1件とする。
 このデータは70,552件になった。(頻度2以上)
- (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

列挙し、

 →

列挙する

 ↓ ↓
 一致パターン 終止形にしたもの

(iv). 後接語のテーブルは20,483件になった。

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する。

前接語，後接語のための作業テーブルを次に示す。

散乱理論	6
光子散乱理論	2
多重散乱理論	8
交換力理論	2
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3 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
与えるかについて	17
与えるかもしれ	6
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
与えるが，その	7
与えるが，それ	3
与えるが，メチ	3
与えることから	16
与えることがあ	4
与えることがて	67
与えることがわ	67
与えることが可	5
与えることが示	15
与えることが知	8
与えることが判	27

表4 後接語作業表の一部

● 分析結果

- ① 実験に使用したKWIC総行数 788,477行
- ② 語と語の関係による知識データとして有効なデータ 533,730行
- ③ 語と語の関係による知識データの抽出(種類) 287,354件

約28万件の知識データの中で一部分は整理されていない部分があるが，これらは今後手直しをしてゆく方針である。

抽出されたデータの一部を示す。

Seq No	語と語の関係	頻度
1	過程 を しらべる	9
2	過程 を とる	6
3	過程 を 解析する	10
4	過程 を 含む	7
5	過程 を 記述する	12
6	過程 を 経る	18
7	過程 を 検討する	17
8	過程 を 研究する	31
9	過程 を 考える	10
10	過程 を 考察する	19
11	過程 を 考慮する	6
12	過程 を 示す	33
13	過程 を 述べる	22
14	過程 を 紹介する	6
15	過程 を 説明する	11
16	過程 を 調べる	27
17	過程 を 追跡する	5
18	過程 を 通す	5
19	過程 を 明らかにする	12
20	崩壊過程 を 調べる	8
21	成長過程 を 観察する	5

(' を ' の前の語を後から分類し，頻度5以上のリスト)

表5 語と語の関係

この処理プロセスを図1に示す。

● この方法の特徴

- ① 文を分析することによって得られた知識データであり，(作為的なデータではない)。機械翻訳の訳しわけ等に適用すると知識データのヒット率が高くなる。
- ② 頻度情報が付いている。
- ③ 多くの語と語の関係が得られるため動詞の辞書が作りやすい。多くの例文を思い付きやすくなるため辞書が充実する。
- ④ 文の構文解析を行わずに得られる。知識データを得るために文の構文解析をする方法が考えられる。さらに，構文解析の構文木を減らすために知識データを必要とする，という矛盾から抜け出せる。
- ⑤ 処理時間があまりかからない。

- ⑥ 機械翻訳において訳文の生成がより適切に行うことができる。

例

draw	a	line
↓		↓
(書く)		(線)
(描く)		
(ぬる)		
(引く)		

知識データに「線を引く」があるため、この訳語を優先させる。

- ⑦ ボトム・アップのアプローチである。

4-3 今後の課題

- ① 「語と語の関係」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
- 25万件の翻訳費用 1.25億円
- 1日当りの作業量(1人) 50件/日
- 延人日(10人で約2年間) 5,000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、ある動詞から順次翻訳する方法が考えられる。一部翻訳した内容(引く)を最後に示す。

- ② 新聞データより「語と語の関係」を抽出する。東大(工)藤崎教授を中心に整備している、新聞データにこの考えを適用し、一般文からの知識データの抽出を行う。

- ③ 「の」、「から」、「へ」…等、「を」以外の助詞について「語と語の関係」の抽出を試みる。

例

夜	が	ふける	毒	が	ある
音	が	する	時間	が	ない
虹	が	出る	人形	が	動く

- ④ この実験では集められなかったデータ

例えば

車	を	引く
糸	を	引く
		⋮

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。

- ⑤ 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。

- ⑥ 語と語の関係でも多義性が判別できない場合が少し発生する。

例

手を引く

(i) 人の手を引く

(ii) 危険な仕事等から抜け出す

これについては今後さらに検討しなければならない。

この方法による知識データの収集は成功したが、今後、各種の方法で知識データが増えると思われる。これについては次のことを考えなければならない。

5. 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。

おわりに

機械翻訳の一つの大きな問題点である多義性の解消について知識データを利用することで明るい見通しを与えることができた。自然言語の分析は大変な仕事なので、なるべく規則による解決をはかろうとするが、規則にはある限界があり、細かい部分には効果がない。細かい部分を考えるにあたってはBottom upによる自然言語の解析と知識データの収集により、規則の大系化、再構築が必要である。この作業は大変根気のいる作業である。一面では理論的でない面があるが、これは次のStepへの発展のためには通らなければならない道程であると信じている。

ただ単純な知識データの収集と、問題の解決ではない。知識データを十分に集めれば体系化しやすく、何が主体か例外か判りやすくなる。

数十万件程度の知識データが簡単に収集できるようになった現在では、次に発展の方向を探らねばならない。

参 考 文 献

- (1) 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
- (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 「自然言語処理技術」シンポジウム 1984.11
- (3) 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
- (4) 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
- (5) 勝俣銓吉郎編 新和英大辞典 研究社
- (6) 金田一京助 他 新明解国語辞典 三省堂
- (7) 西尾 実 他 岩波国語辞典 岩波書店
- (8) 森田良行 基礎日本語 角川書店
- (9) 田中康仁 専門用語の自動抽出 第17回情報科学技術研究会発表論文集 日本科学技術情報センター 1980.10
- (10) 長田孝治, 田中康仁 他 専門用語の造語成分 第18回 情報科学技術研究会発表論文集日本科学技術情報センター 1982.3
- (11) 吉村賢治, 山下明男, 日高 達, 吉田 将 専門用語の自動収集システムについて 自然言語処理研究会 42-1 情報処理学会
- (12) 田中康仁, 吉田 将 専門用語の自動収集について 1987年情報学シンポジウム 情報処理学会 1987.1
- (13) 花田岳美, 佐々木 肇 日本語における学術用語の特色と問題点 1987年情報学シンポジウム 情報処理学会 1987.1
- (14) 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
- (15) 溝口文雄 他 大特集:機械翻訳 情報処理 Vol 26 No.1985.10
- (16) 新田義彦 他 計算言語学 情報処理 Vol 27 No.8 1986.8

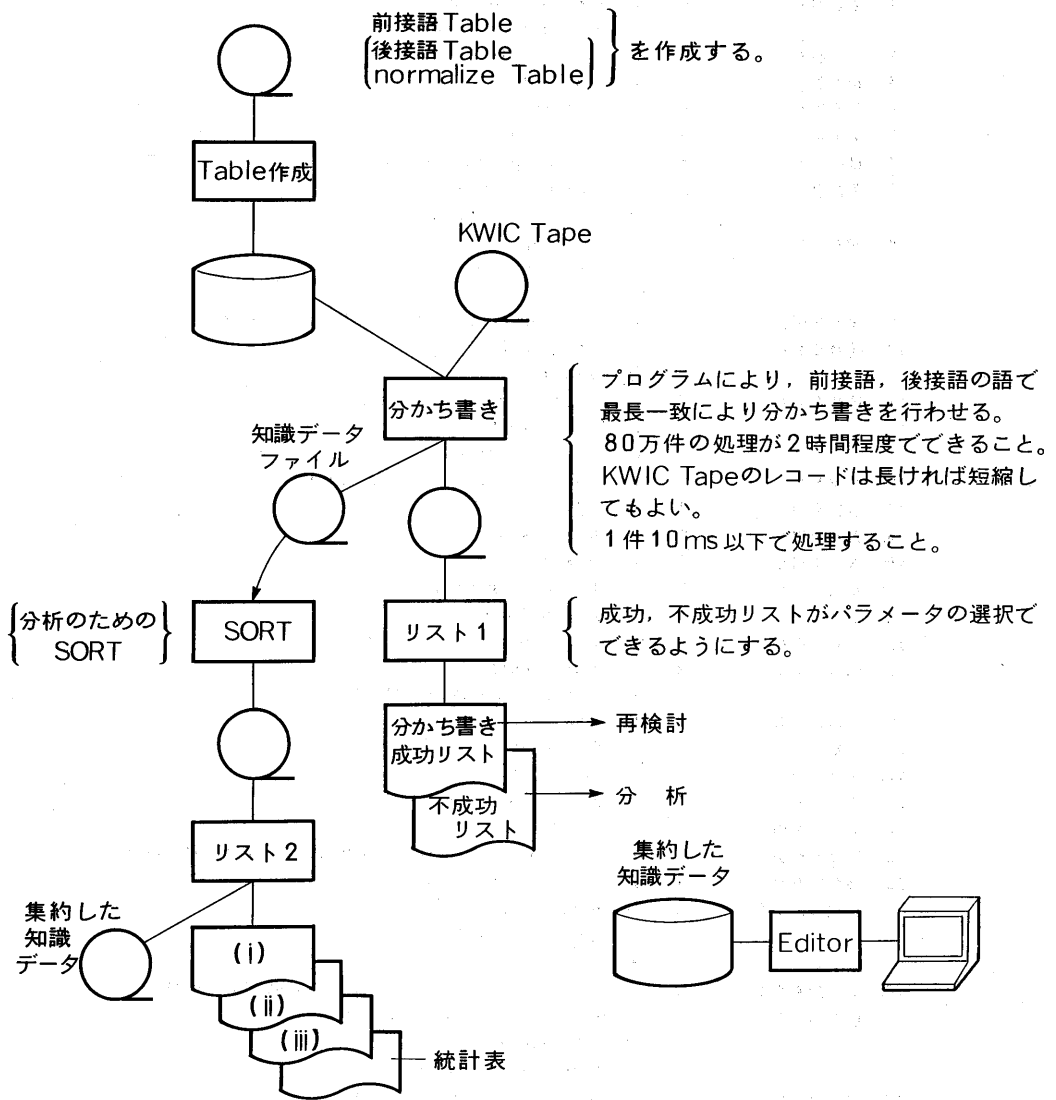
謝 辞

この研究の一部は文部省科研費課題番号(代表者 吉田 将)(60302090)と文部省科研費課題番号(代表者 藤崎博也)(61880005)と文部省科研費特定研究言語(代表者 長尾 真, 野村雅昭)(A04)と文部省科研費課題番号(代表者 田中康仁)(61580033)によって行った。

No.	語と語の関係	頻度	訳	語
1.	手を引く	2	① take someone by the hand ② pull out of a risky business	
2.	水を引く	1	① lay on water	
3.	線を引く	5	① draw a line	
4.	尾を引く	5	① left its mark on	
5.	名を引く	1	① look up a name	
6.	目を引く	3	① draw a person's attention	
7.	例を引く	4	① draw on examples ② quote an examples	
8.	和を引く	1	① minus sum ② subtract sum ③ take sum	
9.	関心を引く	6	① win someone's affections (or interests)	
10.	関数を引く	1	① subtract function ② take function	
11.	興味を引く	9	① win someone's interests ② be interested in ③ take an interest	
12.	曲線を引く	3	① draw a curve (or curved line)	
13.	空気を引く	1	① install the air	
14.	形状を引く	1	① draw a shape	
15.	人目を引く	1	① draw attention	
16.	接線を引く	2	① draw a tangent line	
17.	繊維を引く	1	① pull a fiber	
18.	注意を引く	1	① draw a person's attention	
19.	注目を引く	2	① draw a person's attention ② attract a person's attention	
20.	直線を引く	2	① draw a straight line	
21.	白線を引く	1	① draw a white line	
22.	用途を引く	1	① look up an usage	
23.	アームを引く	1	① draw an arm	
24.	カーブを引く	2	① draw a curve	
25.	商品名を引く	1	① look up a commercial name (or a name of goods)	
26.	戦争論を引く	1	① look a study of war	
27.	抵抗分を引く	2	① minus a resistance ② subtract a resistance ③ take a resistance	
28.	同時線を引く	2	① draw a simultaneous line (or a synchronous)	
29.	販売高を引く	1	① minus a sales value (or a price) ② subtract a sales value ③ take a sales value	
30.	理論値を引く	1	① minus a theoretical value ② subtract a theoretical value ③ take a theoretical value	
31.	コア推力を引く	1	① minus a core drift ② subtract a core drift ③ take a core drift	
32.	トレーラーを引く	1	① drag a trailer	
33.	残存窒素量を引く	1	① minus a survival nitrogen quantity ② subtract a survival nitrogen quantity ③ take a survival nitrogen quantity	
34.	非常ボタンを引く	1	① pull an emergency switch	
35.	くじを引く	0	① draw a lots	
36.	弓を引く	0	① draw a bow	
37.	垂線を引く	0	① draw perpendicular lines	

表 6 語と語の関係と訳語

図1 K W I Cデータから知識データの抽出



分析のための SORT キーについて

	前接語	'を'	後接語	頻度
(i)	①	②	④	③
(ii)	④	②	①	③
(iii)	②	③	④	①

(高頻度のものは慣用表現である。)

このシステムのメリット

- ① やりなおしが簡単にできること。
- ② 経費がかからずまちがいが少ない。
- ③ 統計処理ができる。
- ④ 人手が最小限におさえられる。
- ⑤ Tape Operationが中心であること。