自然言語処理 65－2
（1988. 3. 25）

自然言語の知識獲得

― 語と語の関係について，朝日新聞記事データの分析 ―

田中康仁
（姫路短期大学）

吉田将
（九州工業大学）

自然言語の分析による知識データの獲得を行っている。前回「知識データ（語と語の関係）による多義性の解消」昭和62年3月の研究発表の発展した内容である。前回のものは日本科学技術情報センターのデータを利用したが，今回は朝日新聞のデータを用いた。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，多義性のための幾つかの方法と問題の検討を行い，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'中心とした新聞データのKWICを使い，その中から手作業で知識データを集めた。

約20万行のKWICを解析し，10万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期に来ている。

Acquisition of Knowledge Data for Natural Language

— From Asahi News paper —

YASUHITO TANAKA
Himeji College
1-1-12 Shinzaike Honmachi
Himeji
670 JAPAN

SHO YOSHIDA
Kyushu Institute Technology
680-4 Kawatu
IIAUKA
820 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 200,000 lines of the KWIC list 100,000 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 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ま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1　引く（動詞）

(i) 引っ張る・曳く・挽く・索く
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
substract from, deduct, reduce, abate,
cut down, take off, allow discount

(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

このように「引く」という語を調べると14以上の意味がある。

例2　解く（動詞）

(i) ほどく　untie, undo, unbind, loosen
unloosen, unfasten, unravel
disentangle, unpack

(ii) 縫ったものをほどく　unsew

(iii) 分解する　disjoint, take to pieces

(iv) 解答する　solve, work out, answer
resolve, dispel, remove
clear away

(v) 解除する　dissolve, cancel, riscinel
remove, lift, absolve
release, disengage

(vi) やめさせる　relieve

• その他

このように「解く」という語を調べると6個以上の意味がある。

「新和英大辞典」研究社より引用

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」や「解く」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。

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であろうか？

(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
loft 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) 格文法と意味マーカ

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ーのようなものを付けなければならず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。これと他のものとを利用し多義性の解消に役立てている。

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) 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 packge

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 一般的方針

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いう疑問が起る。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こ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こ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 電話を磨く
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 電話をかけなおす
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 電話を開設する	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
	⋮

一つの語彙に関係する語彙は限られている

表 1. 一つの語に関係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的文法と＂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。

```
823 14MPG5T:  構。私のように出来の悪い子供 を かかえ，経済的に余裕のない親に
821 08MPG1T:  ある。五，六人で尾をもち，頭 を かかえ，傷つけないように運んだ。
822 12MPG5T:  （主婦 46歳） 家のローン を かかえ，娘の嫁入り支度は血がに
820106MKA2T:  子育て，仕事といろいろな問題 を かかえて，エネルギッシュに生き
821 08MPG2T:  ，ポーランドという複雑な問題 を かかえている。軍事力増強を図っ
821 09MKA1T:  では，どこでも同じような問題 を かかえている。授産施設と更生施
821 08EPG1T:  も世話しなければならない家族 を かかえている。母娘の生活や健康
820319ESK2T:  は険しい。赤字財政という難問 を かかえているからだ。政府は利権
820323MKM-T:  るこの新会社は，自らも研究陣 を かかえているが，最大の特徴は，
823 11ELG-T:  ラヤ山脈を中心に数多くの高山 を かかえているが，世界最高峰のチ
822 16MPG1T:  も職場に相談相手がいない悩み を かかえているという。心の問題に
820323MPG5T:  である。 国鉄は，多くの問題 を かかえているとはいえ，日本列島
822 16ELG-T:  それだけに頭 を かかえているのが海上保安部。イ
822 21MGH2T:  調整していくのか，難しい問題 を かかえているのも見逃せない。
823 13MKA1T:  ンティア）は，さまざまな課題 を かかえている地域に，同協会が資
821 18MPG1T:  わらず，なお相当数の余剰人員 を かかえており，このまま政府直営
821 12MBKJT:  力のつよさには，しばしば感動 を おぼえた。たとえば，昨日，十七
820315MPG5T:  子供は，こうしてひとつのこと を おぼえていくのですが，新しいこ
821 07ENVAT:  た時のこの母のあたたかい愛情 を おぼえていた。昼まで仕事をして
823 12ELG-T:  どからだがあたたかかったこと を おぼえている。
822 11MSP1T:  楽しさを教え，先生方に指導法 を おぼえてもらうためだ。三年間に
820319MSP1T:  んだ。琴風も「横綱には取り口 を おぼえられたが，下位に取りこぼ
823 11MPG5T:  納税者の一人として激しい憤り を おぼえる。
822 16MPG1T:  たちは新しい戦慄（せんりつ） を おぼえる。戦慄というややおおげ
821 12MBKJT:  とをきけば，世界の国々は恐怖 を おぼえるであろうし，とりわけ文
821 09ELG-T:  の阿Qに，人々が覚りない感動 を おぼえるのは，阿Qの時代と同様，
823 10MNVUT:  る，にひとつひとつ新鮮な感動 を おぼえるらしかった。 そういう
822 22EGR-T:  プリンス演出とは全く違う舞台 を おみせしたい」という。
823 11ESK1T:  情を察知，社長や常務理事の顔 を おりたが，畠山のあとの社長に，
823 13MPG5T:  師の中で，政治，物価等の批判 を おりまぜて笑わす芸は，おそらく
822 22MSK1T:  出した約六万円を奪い，ズボン を おろさせ，すぐに追いかけられな
821 12MSK1T:  美哉子さんは郵便局で五十万円 を おろし，学校などに確認しないま
821 12MSP1T:  佳田の海は「まだ膝 を おろしかけていたので待ったした。
```

KWICの例

表 2 助詞，助動詞を中心としたKWIC

格助詞 '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は機械的に容易に作成することができる。

このKWICを基にして姫路短期大学の学生達にデータの抽出を行わせた延べ20数人の学生が約1ケ月の期間をかけて作業を行った。データの抽出内容は図書カードに記入した。このカードを集め計算機の入力データとした。入力は2回にわけて行った。延べデータが増える割合ほど種類は増えていない。このことから知識データの重複があることがわかる。
これはこのようなデータが再現性があることを示している。手作業による抽出であるが，これは和語が多いためとデータ量（KWIC）があまり多くないため丁寧に抽出を行った。

手作業による大量の知識データの収集は単純作業の繰返しであまり学問的価値が無いが，一旦集められ整理された大量の知識データは多くの人々に利用され，それから作り出される新しい知的な生産物は広く社会に利用され大きな意味を持ってくる。また，大量の知識データの多くの分野にわたる利用方法の研究も多くの人々に喜ばれ有意義である。

	種　類（A）	延データ（B）	A/B
第1回入力データ	80,442	126,334	0.636
第2回入力データ	29,150	40,957	0.712
総　合	101,676	167,291	0.608

表3　朝日新聞 語と語の関係（'を'）データ

収集した知識データの一部「とく，解く」と「ひく，引く」を調べJICSTのデータと比較すると，朝日新聞記事データと科学技術文献抄録とではかなり異っていることがわかる。これは新聞と科学技術文献抄録とでは対象語彙も異るので当然のことかもしれない。

次にその結果を示す。

	朝日新聞		JICST抄録	
語	種　類	延べ数	種　類	延べ数
とく，解く	29 (3)	35 (8)	89 (3)	523 (241)
ひく，引く	107 (12)	199 (64)	40 (12)	80 (37)

（ ）内は朝日新聞とJICST抄録の共通数

表4　語と語の関係分析表（とく，ひく）

又，図に示すと次のようにもなる。

新聞と科学技術の分野では語と語の関係が全く異っている。共通部分は少ない。

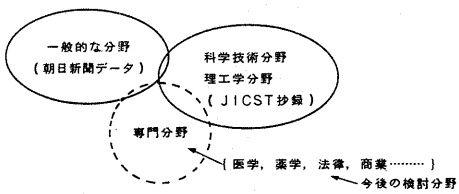


図1　分析データの関係

朝日新聞記事データの分析結果をまとめると次のようになる。

朝日新聞（84日分）の分析結果
（'を'を中心とした語と語の関係データ抽出）

(1) 84日分の新聞データより'を'のKWIC抽出　約20万行
(2) 語と語の関係の抽出できたデータ　167,291件
(3) 同一内容を集約して得られたデータ　101,676件
(4) 作業に関係した学生数（1ケ月×25人＝25人月）　25人月
(5) 入力の費用（1件10円）　約160万円

また朝日新聞記事データの分析で得られた知識データを分析すると次のような特徴がある。

新聞データの特徴

① 新聞データは一般的な用語の語と語の関係が得られる。
（一部には新聞特有の語もある。）

② 和語の分析に向いた資料が得られる。

③ 科学技術分野の語と語の関係と比べ重複が少ない。

④ 延16万件程度の資料分析でも，重複が多く，良く使われる関係には頻度情報が得られた。

抽出されたデータの一部を示す。

業法	を	設定する	1
業法	を	大幅手直しす	2
業務	を	する	2
業務	を	委託する	2
業務	を	引き受ける	1
業務	を	営む	1
業務	を	開始する	1
業務	を	拡大する	1
業務	を	含む	1
業務	を	記載する	1
業務	を	禁じる	1
業務	を	経験する	1
業務	を	兼務する	1
業務	を	限定する	2
業務	を	行う	5
業務	を	合理化する	1
業務	を	再開する	3
業務	を	始める	1
業務	を	執行する	1
業務	を	実施する	1
業務	を	手がける	1
業務	を	集中する	1
業務	を	正常に行う	2
業務	を	通じる	1
業務	を	停止する	3
業務	を	統括する	1
業務	を	分ける	1
業務	を	目的とする	1

表5　朝日新聞記事データより抽出した知識データ（先頭より分類）

CD	を	運用する	1
F16	を	運用する	1
これ	を	運用する	2
ドル	を	運用する	1
一億円	を	運用する	3
規定	を	運用する	1
技術	を	運用する	1
財投資金	を	運用する	1
手形	を	運用する	1
昇格措置	を	運用する	1
制度	を	運用する	1
代金	を	運用する	1
中期国債	を	運用する	1
法	を	運用する	1
法律	を	運用する	1

表6　朝日新聞記事データより抽出した知識データ（後接語より分類）

このほか朝日新聞記事データKWICを利用し'を'以外にも"が"の知識データを抽出している。

知識データの収集（'が'を中心とした）

'を'以外'が'を中心とした語と語の関係を基にした知識データの抽出を試みている。'が'については次のような

「前の語」　が　「後の語」

関係の共起データを集めている。

「後の語」としては次のようなものがある。

① 形容詞　② 形容動詞　③ 自動詞

④ 他動詞＋（受身又は使役の助動詞）

'が'→'を'にすべきもの。

⑤ 他動詞で'を'格が省略されたか別のところに移ったもの

⑥ そ の 他

等が考えられる。

朝日新聞データについてKWIC LISTを作成し，抽出は手作業によって行っている。約14.4万件のKWIC LISTから約6万件の知識データが得られ，'87年末には完了する予定である。

- この方法の特徴

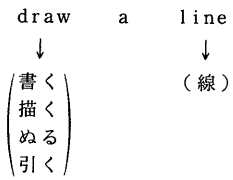
① 文を分析することによって得られた知識データであり，（作為的なデータではない）。
機械翻訳の訳しわけ等に適用すると知識データのヒット率が高くなる。

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 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



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。

⑥ 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
- 10万件の翻訳費用　5,000万円
- 1日当りの作業量（1人）　50件／日
- 延人日（4人で約2年間）　2,000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），と（解く）を最後に示す。

② '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

③ 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。

④ 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。

⑤ 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。

⑥ シソーラスとの照合

この10万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。

① シソーラスとの照合の意味(I)

雨が降る → { 雪が降る / あられが降る / ひょうが降る }

雨，雪，あられ，ひょう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意味マーカの確認

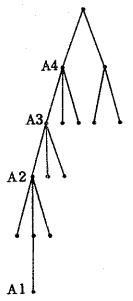
② シソーラスとの照合の意味(II)

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)

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が属する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との結合を調べる。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，該当個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。

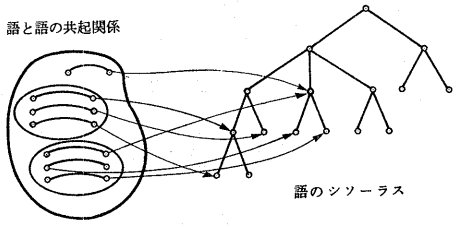


図2　語と語の共起関係と語のシソーラスとの関係

⑦ 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「量的拡大は質的変化をもたらす」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

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。

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。

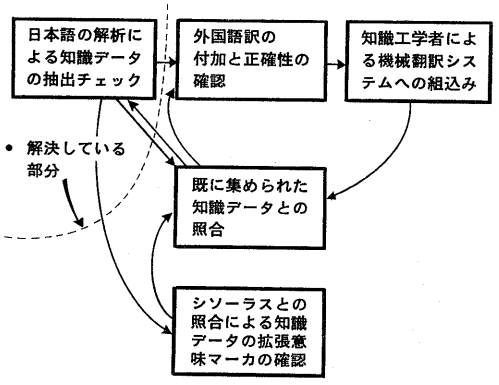


図3 語と語の関係の知識データが機械翻訳システムに組込まれるまでの作業プロセス

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。

この研究の一部は文部省の科学研究費によって行った。

また，朝日新聞データ 84日分は東大（工）藤崎教授，亀田氏（現・東京工科大学助手）によって整備されたものを使用した。朝日新聞，東大・藤崎教授，亀田氏に感謝の意を表す。

文部省科研費課題番号
（代表者 田中康仁）（61580033）
文部省科研費特定研究言語
（代表者 長尾 真） A04
文部省科研費課題番号
（代表者 藤崎博也）（61880005）
文部省科研費課題番号
（代表者 吉田 将）（60302090）

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集会発表論文集　　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 情報科学技術研究集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　№ 1985. 10

(16) 新田義彦　他　　計算言語学　　情報処理 Vol 27 № 8　1986. 8

(17) 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。）

(18) 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－ 秋山書店　　1987. 3

(19) 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

(20) 田中康仁，吉田　将　　慣用表現について－収集と整理－　　情報処理学会情報学基礎　5－1　　1987. 6

(21) 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。）

(22) 小西友七編　　英語基本動詞辞典　　研究社出版 1980. 9

(23) Morton Benson，他　The BBI Combinatory dictionary of English John Benjamins／丸善 1986. 11

藤崎・亀田関連論文リスト

〔1〕 藤崎博也・亀田弘之・荻野綱男：　"新聞記事の分かち書き処理とそれに基づく語彙調査"　情報処理学会第30回全国大会　5G-2，pp. 1679－1680(1985)

〔2〕 亀田弘之・藤崎博也：　"大量の新聞記事データを対象とした語彙調査"　情報処理学会第31回全国大会　3H－8，pp. 1375－1376（1985）

〔3〕 藤崎博也・亀田弘之：　"新聞記事データを対象とする自単位切りとそれに基づく語彙調査"　情報処理学会研究報告　Vol. 85，No. 31，NL－51－2（1985）

〔4〕 亀田弘之・藤崎博也・明石孝祐：　"新聞記事データを対象とする語彙調査結果"　情報処理学会第32回全国大会　1S－1，pp. 1563－1564（1986）

〔5〕 亀田弘之・藤崎博也：　"高機能な検索のできる大規模日本語データベースの構成"　情報処理学会第33回全国大会　4K－7，pp. 1831－1832（1986）

〔6〕 亀田弘之・藤崎博也：　"新聞記事を対象とする用字調査"　情報処理学会第33回全国大会　4K－8，pp. 1833－1834（1986）

〔7〕 藤崎博也・亀田弘之：　"自動単位切りによる新聞記事の語彙調査"，特定研究「情報化社会における言語の標準化」　研究成果報告書，木下是雄(編)，pp. 661－675（1986）

〔8〕 藤崎博也・亀田弘之・森田敏生・田口　茂：　"高機能・大規模な日本語用字・用語データベースのための品詞解析"，情報処理学会第35会全国大会（1987）

以　上

朝日新聞データ分析結果（ひく，引く）

Seg.No.	語と語の関係	頻度	訳語
1	かぜをひく	9	catch a cold
2	くじをひく	2	draw lots
3	そでをひく	1	pull 〈 a person 〉 by the sleeve
4	ひき金をひく	1	pull (squeeze) a trigger
5	カゼをひく	6	catch a cold
6	一線をひく	1	draw a line between
7	陰影をひく	2	have a shadow
8	右上手をひく	1	take upper right
9	何かをひく	1	draw something
10	荷車をひく	1	draw a cart
11	楽器をひく	1	play a musical instrument
12	関係者の注目をひく	1	draw attention of persons concerned
13	関心をひく	4	win his or her affections
14	眼をひく	1	draw attention
15	気をひく	2	win his or her interests
16	興味をひく	5	win his or her interests , be interested in , take an interest
17	曲をひく	2	play a tune
18	糸譜をひく	1	① look up a genealogy (pedigree) ② have a genealogy
19	血をひく	2	be descended from
20	言葉をひく	1	look up a word
21	腰をひく	1	draw one's waist
22	字引をひく	1	① look up 〈 a word 〉 in a dictionary ② consult a dictionary
23	持ち手をひく	1	draw a holding hand
24	耳をひく	2	pull an ear
25	辞典をひく	1	① look up 〈 a word 〉 in a dictionary ② consult a dictionary
26	手をひく	3	① take her child by the hand ② pull out of a risky business
27	心をひく	2	attract
28	親の血をひく	1	be descended from one's parents
29	人目をひく	1	draw attention
30	世間の目をひく	1	make a noise in the world
31	税金をひく	2	subtract tax
32	線をひく	1	draw a line
33	村営水道をひく	1	install a village water service
34	段ボール紙をひく	1	lay carton
35	注意をひく	1	draw a person's attention
36	注目をひく	4	draw a person's attention

37	長女をひく	1	take the eldest sister
38	伝統をひく	2	have a tradition
39	尾をひく	6	left it's mark on
40	鼻かぜをひく	2	have a cold in the nose , have a nose cold
41	風邪をひく	2	catch a cold
42	目をひく	6	draw a person's attention
43	油をひく	2	oil
44	流れをひく	4	① be descended from ② belong to the school of
45	例をひく	4	quote an examples
46	0.14を引く	1	subtract 0.14
47	DEを引く	1	subtract DE
48	そでを引く	2	pull 〈a person〉 by the sleeve
49	ひもを引く	1	draw a string
50	まわしを引く	1	take the belt
51	アゴを引く	1	draw the jaw (a chin)
52	クジを引く	1	draw lots
53	サイドラインを引く	1	draw a side line
54	マークを引く	1	draw a mark
55	マルクスを引く	1	quote Marx
56	ミツを引く	1	drop honey
57	ラインを引く	1	draw a line
58	ラブレーを引く	1	quote Rabelais
59	リボンを引く	1	pull a ribbon
60	レバーを引く	1	pull a lever
61	ロープを引く	1	pull a rope
62	右下手を引く	2	take lower right
63	右上手を引く	1	take upper right
64	気を引く	1	attract
65	給与所得控除を引く	1	subtract earned income deducation
66	興味を引く	4	win her interests , be interested in , take an interest
67	偶数番号を引く	1	take an even number
68	経費を引く	1	subtract expenditure
69	言葉を引く	2	look up a word
70	裁判を引く	1	put ～ on trial
71	山車を引く	1	pull a float
72	仕掛けを引く	1	make a contrirance
73	仕事を引く	1	retire
74	使用料を引く	1	subtract a use rate
75	四%を引く	1	subtract 4%
76	糸を引く	2	① pull a string ② pull the wires
77	字引を引く	1	① look up 〈a word〉 in a dictionary ② consult a dictionary
78	辞書を引く	2	〃