

自然言語の知識獲得【その1】

— 語と語の関係について、朝日新聞記事データの分析 —

— 'が' について —

田中 康 仁
(姫路短期大学)吉 田 将
(九州工業大学)

自然言語の分析による知識データの獲得を行っている。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 'が' 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14.4万行のKWICを解析し、5.4万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がはかれる。翻訳等に少し費用はかかるが解決の第一歩がつかめた。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 (NO. 1)

— From Asahi News paper —

YASUHIRO TANAKA

Himeji College

1-1-12 Shinzaike Honmachi
Himeji
670 JAPAN

SHO YOSHIDA

Kyushu Institute Technology

680-4 Kawatu
Iizuka
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 144,000 lines of the KWIC list 54,000 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1 上がる(動詞)

- (i) 登る(昇る) go up, come up, rise, ascend, mount, climb up
- (ii) 騰貴する rise, go up, advance, appreciate
急に上がる jump 非常に上がる soar, skyrocket
- (iii) 昇給する, 昇級する rise, be rised, be advanced, be promoted, get promoted
- (iv) 進歩する progress, make progress, advance improve
- (v) 入る enter (the parlor), go (=come or step), (into the house)
- (vi) 訪問する call (on a person, at a person's house)
- (vii) 供えられる be offered
- (viii) 生じる issue (from), accrue (from), be derived, be obtained
- (ix) 済む be completed, be finished, be ready to serve, be through with (work), get off duty
- (x) 足りる be sufficient
- (xi) 死ぬ, 枯れる die, be dead
- (xii) 商売がだめになる fall off, be ruined, go to the dogs
- (xiii) 上簇(じょうそく)する begin spinning
- (xiv) 捕われる, 獲得される be arrested, be caught, be rounded up, be nabbed be found, be recovered
- (xv) 飲食する take, have, eat, drink
- (xvi) 興奮する lose self-control, lose one's presence of mind, get excited, get hot, get nervous

(xvii) (すごろくなどで)上がりになる come out, win

例2 出る(動詞)

- (i) 現われる appear, come out, emerge (from, out of, on), make one's appearance, turn up, come forward, present itself
- (ii) 出没する haunt, infest
- (iii) 見つかる be found, be restored to, get (a thing) back
- (iv) (料理などが)出される be brought, be served
- (v) (道が…に)通ずる lead to, strike up on, come upon, enter
- (vi) (結果として)生ずる come out
- (vii) 外へ出る go (=come) out, get out, stir out, turn out, take one's way out, step out
- (viii) 出席(出勤)する be present at (a ceremony), attend (a meeting), appear, report (at)
- (ix) 勤務する work (in), serve (in), be in the service (of), hold an office (in) (役所に)
- (x) 出場(参加)する join, participate in, take part in, enter for
- (xi) (候補に)立つ stand (for), run (for)
- (xii) 乗り出す go (=sally) forth into, go upon launch into, enter upon
- (xiii) 掲載される appear (=come out) in, be published, go into

その他出版される, 売れる, 出発する, 去る, 卒業する, 発生する, 産出する, ~から生ずる, 突き出る, 脱出する, 差し出る, 範囲外に出る, 出費がある, 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「上がる」や「出る」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, 止ぶす, 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〕	群衆〔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。

次のような問に対して充分答えられるほどのデータはない。

- (i) 格フレームと意味マーカだけでどの程度日本語処理はできると考えるか?
- (ii) 意味マーカと格だけで訳しわけが行なえるか?
- (iii) 意味マーカを全部の単語に付けなければならないか?
- (iv) 意味マーカを細かくすれば労力が多くなり、あいまい性も多くなるか?
- (v) 意味マーカの区分と全ての動詞の対象語の区分とが一致するか?
- (vi) Vを検証した資料はあるか?一致しない場合の例外リストはあるか?
- (vii) Top-downアプローチはどこまで言語活動を解析するために有効か?

この意味マーカを詳細に研究するためにはシソーラスとの照合が重要である。

(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。

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 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。

025	電話が鳴る	030	電話の声
026	電話で伝える	031	電話の部品
027	電話で話す	032	電話の金
028	電話で連絡する	033	電話の料金
029	電話に出る	034	電話のベル

⋮

一つの語彙に關係する語彙は限られている

4. 語と語の關係データの抽出

4-1 一般の方針

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この疑問が生ずる。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關係する語は限られている。

人はある状況を見てそれにふさわしいある一定の表現をする。決して別の表現をしない。



のような状況を見れば“雷が鳴っている”という表現をする。

決して“雪が降っている”とは言わない。又簡単な表現形式を望む。複雑で長い表現はもちいられない。

我々は物事や状況から意味の最小単位を表現することを知っている。この意味の最小単位を集めなければならない。

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關係を数えあげることが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關係を数えあげることが大変困難である。

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關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	013	電話を撤去する
002	電話をきる	014	電話を売る
003	電話を持ち上げる	015	電話を販売する
004	電話をこわす	016	電話を買う
005	電話を握る	017	電話を購入する
006	電話を持つ	018	電話を磨く
007	電話を改良する	019	電話を受ける
008	電話を作る	020	電話を盗聴する
009	電話を製作する	021	電話をかけなおす
010	電話を組立てる	022	電話を待つ
011	電話を開設する	023	電話を持たせる
012	電話を引く	024	電話を聞く

表1 一つの語に關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關係で規定しなければならない。使用頻度の低い語と語の關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。

自然言語の一般的知識をあげてみると次のようになる。

これらを利用して総合的に自然言語解析、理解、生成を進めたいかなければならない。

語の持っている知識をあげてみると次のようになる。

1. 語に関する属性
 - 1.1 語の属性
 - 語、品詞、発音、アクセント、仮名表記
 - 1.2 長単位用語、専用用語
2. 語と語に関する属性 (成立する前提条件)
 - 2.1 上位、下位の關係 (シソーラス構造)
 - 2.2 反対語、否定語
 - 2.3 部分、全体の關係
 - 2.4 順序關係
 - 2.5 比較關係 (大小、高低)
3. 語と語に関する属性 (2) (成立する前提条件)
 - 3.1 格による關係
 - 3.2 文構成要素の共起による關係
 - 3.3 慣用表現
4. 語と語に関する属性 (3) (成立する前提条件)
 - 4.1 連想による關係
5. 文の關係
 - 5.1 因果關係 (動詞と動詞の)
 - 5.2 場の設定条件
 - 5.3 文と語の關係
 - 5.4 文と文の連なりの關係

表2 語の持っている知識

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。

19180 ものそめて、いやなものには目をつぶる性癖がある。しかし、米ソ対立の激化
 19181 かけなくてすむのに……」と非難されることがある。しかし、老人ホームは入
 19182 と、おまじないを唱えるといふという笑話がある。しかし「当たり前」た
 19183 ンリー・フォントの命はもつと延びる可能性がある。しかし「威厳を保って死
 19184 ……またケガをしなれば良いか、と思うことがある。しかし自分とはたえ…と
 19185 存続の領域では、裁判所の役割に大きな限界がある。しかし少なくとも平等原
 19186 界の議論は労働戦線統一で塗りこめられた感がある。しかし新情勢に際れ、マ
 19187 ジアの指導者の不安を呼ぶような歴史的記憶がある。しかし多くの指導者は、
 19188 河本氏に対して善理解く、切り崩しにくい面がある。しかし半面、「その会社
 19189 の関係をすっきりさせぬとその努力にも限界がある。しかも、改革の方向はず
 19190 上げせ」とてもなったら大変、といった懸念がある。しかも、公務員給与抑制
 19191 のように解することには専門家の有力な議論がある。しかも、事故が二以上重
 19192 まえより原子力船の母港としての安全性に疑問がある。しかも私としては、この
 19193 課税対象の拡大で税収入を増やす利点がある。しかも専当局の人員削減
 19194 どうしてもその背景を深く理解しておく必要がある。したがって次のような三
 19195 きをつくっている一農民にまでそういう意識がある。じゃ、それに基づくと中国
 19196 き聖明歌謡から一言。これにはもう一節続きがある。すなわち、一呼吸おいて
 19197 ーンを運めるにあたっては、いくつかの原則がある。すなわち、客観的に、バ

19119 ケストラ・コンダクター」と表示された部分がある。この機械を操作すれば、
 19120 子規派の重鎮となる。句はやや古風だが風格がある。この句、説明は要るまい
 19121 たりまえだという、非常にはっきりした原則がある。この原則が無視されると
 19122 土砂崩れのため、一時通行止めになったことある。この事故で、同国運は上
 19123 一ゲームとなり、翌日に410で勝った実績がある。この時の監督が高木部長
 19124 会場のPA(拡声装置)の方法は改善の余地がある。この次はカールマンやミ
 19125 行政の政治的中立性に、改めて強い疑念の声がある。この正念場にどう立ち向
 19126 の自由」に、いっそう深い配慮をせしめる必要がある。この代表団の一員として
 19127 本来のものであることを納得させる強烈な力がある。この地域の音楽レコード
 19128 地。そして、大洲市の秋の行事に「芋炊き」がある。この地方を流れる、ふだ
 19129 人の文化・歴史・慣習と深くかかわる価値観がある。この点、アメリカとて同
 19130 液はぬめり。ヤマノイモのトロロに似た風味がある。この結實物の正体は、ベ
 19131 企業との間には、最高裁が認めたとように相違がある。この判決後、米国の弁護
 19132 村役場の近くに村立園保沢内病院がある。この病院の良いすの並ん
 19133 ならず、何らかの形で、実用性を求める傾向がある。この変化に対応するには
 19134 ばロサンゼルスに KROQ というFM放送局がある。この放送局は、三年前に
 19135 他の重要議題には、文化と民主主義の問題がある。この議論の出発点は世界
 19136 よせられるため、日中も気温が下がることがある。これから一週間の気圧の
 19137 のだ。上層と下層で雲の動きが異なることがある。これが「閃電雲」
 19138 各乳期層の中には、平均直径一の炭化銀粒子がある。これが光に反応して、乳

KWICの例

表3 助詞、助動詞を中心としたKWIC

格助詞「が」を選んだ理由は次のようなことがらである。

(1) 名詞と動詞の関係がつかみやすい。

自動詞の関係や、形容詞、形容動詞の関係がつかみやすい。

(2) 「が」は格助詞として使われているので使用頻度が高い。

(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」に次いで「が」が
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10数人の学生が約1ヶ月の期間をかけて
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り返し
であ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。

(4) 朝日新聞記事データの特徴

- ① 新聞は広い分野を対象範囲としている。
- ② 和語の語と語の関係が握抱できる。
- ③ データ数を多くすることができる。
- ④ 語と語の関係を網羅的に抽出できる。
- ⑤ 頻度情報も得られる。

(5) 手作業を主体とした方式の特徴

- ① 抽出作業でこまかい配慮ができる。
- ② 再入力の費用がかかる。1件10円程度。
- ③ 14.4万件程度のKWICならばこの手法がよい。
- ④ 作業員の質が不均一であるため作られたデータの再
分析、チェックを機械的に処理する方法を考えなければ
ならない。

このためには語の長さを適当な長さ以上は調査対象デ
ータとすることにより解決できる。

しかし、個々のデータの検討もしなければならない。

- ⑤ データ収集の費用がかかる。しかし、大学では学生に
勤労奉仕させる場合や授業の演習の一部として行わせ
ると安く行える。

抽出されたデータの一部を示す。

SeqNo		頻度
1	あて	1
2	コントロール	1
3	ドライバー	1
4	リズム	1
5	音程	1
6	感覚	1
7	気	2
8	気象	1
9	計画	1
10	歯車	1
11	手もと	1

12	制球	が	狂う	2
13	精神	が	狂う	1
14	打線	が	狂う	1
15	投球	が	狂う	1
16	答え	が	狂う	1
17	予想	が	狂う	1
18	予定	が	狂う	1

語と語の関係(が)

表4 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)

SeqNo	語と語の関係	頻度
1	合意 がある	3
2	合意 ができる	10
3	合意 がない	3
4	合意 がみられる	13
5	合意 が固まる	1
6	合意 が困難だ	1
7	合意 が出来る	3
8	合意 が成る	1
9	合意 が成立する	6
10	合意 が生まれる	10
11	合意 が絶対に必要である	1
12	合意 が続く	1
13	合意 が通る	1
14	合意 が必要だ	2
15	合意 が不可欠だ	2
1	高校生 が加わる	1
2	高校生 が救出する	1
3	高校生 が考える	1
4	高校生 が座る	1
5	高校生 が最も多い	1
6	高校生 が増える	1
7	高校生 が対象である	1
8	高校生 が提案する	1
9	高校生 が入隊する	1
10	高校生 が発見する	1
11	高校生 が優勝	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)

(6) 「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た。

「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5.4万件の知識データが得られた。

分析結果(が)

KWICの行数	14.4万件
収集したデータ(延データ)	76,371件
重複を取り除いたデータ	54,745件

(7) 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、(作成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。

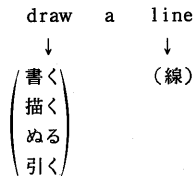
③ 多くの語と語の関係が得られるため動詞の辞書が作りやすい。多くの例文を思い付きやすくなるため辞書が充実する。

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
- 5.4万件の翻訳費用 2,700万円

- 1日当りの作業量(1人) 50件/日
- 延人日(3人で約2年間) 1100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。一部翻訳した内容を最後に示す。

- ② 'の', 'から', 'へ'...等, 'が'以外の助詞について'語と語の関係'の抽出を試みる。
- 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。
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

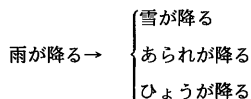
この5.4万種類の知識データと照合することによって不足している知識データを補充するとか, シソーラスの概念分類をさらに詳しく意味分類し, 機械翻訳の多義語の判別, その他に役立つ。

この資料は5.4万種類ある。これは膨大なデータであるから何かカテゴライズすべきであるという考えがある。しかし, 何かをカテゴライズするとその例外が発生する。又, 動詞をカテゴライズするとそれに対応する名詞もカテゴライズしたくなる。

名詞に何か区分を付けようとするとなん十万の名詞があらわれる。労力の最適化も全体的作業の中から考えなければならない。ここではシソーラスとの照合による解決を提案する。

そのため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, 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

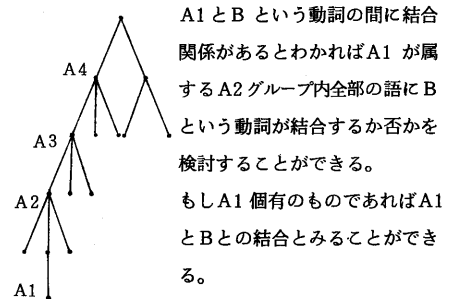
② シソーラスとの照合の意味(II)

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。



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。

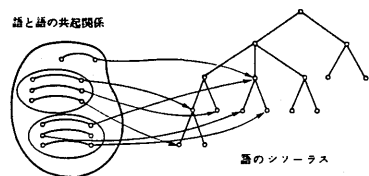


図1 語と語の共起関係と語のシソーラスとの関係

④ 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ければならない。

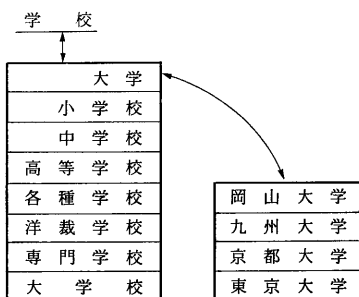


図2 語と語の関係をさらに発展させるために小さい意味の集合を考える。小さい意味の集合に上位，下位の意味の集合を作る。

⑦ シソーラス体系と概念

シソーラスは同じ様な語を集め階層構造をなしたものである，と考えられてきたが，もって深く考えてみるべきものである。語は概念に付けられたラベルであり，語の背景にある概念を明確にし，定義してゆかなければならない。

概念の詳細化には2つの方法が考えられる。

- 1) 既存の知識，辞書から集める例えば百科辞典から集める方法である。

この方法は特定の語に対しては有効であるが全ての概念について網羅的に集めることはできない。

- 2) 日本語の場合“の”による名詞の結合関係より集める。

この方法はあらゆる語（概念）に対して有効である。集め易いし，集めながら，体系化し，不足データを補うことができる。

似た概念は同じような属性構造をなしており，このことからシソーラスの検証に役立つし，似た概念を照合させることにより不足属性を調べあげることができる。

- 1) の方法，2) の方法等を組合せ概念の定義を明確にすることができる。

次に上記2)の方法を具体化した例を述べる。

東大工学部藤崎教授，亀田助手によって作成された朝日新聞84日分KWICから“コート”に関する使用例を調べてみると次のようになる。

黒テン	の	コート	
フードつき	の	コート	
ミンク	の	コート	2

毛皮	の	コート	2
キルティング	の	コート	
冬物	の	コート	
四分の三丈	の	コート	
隣	の	コート	3 (スポーツ)
茶色	の	コート	
ピンク	の	コート	
ローズ色	の	コート	
中国側	の	コート	(政治)
アルゼンチン側	の	コート	(政治)
苦情	の	コート	
神宮	の	コート	(テニス)
スエード	の	コート	
紙	の	コート	
婦人	の	コート	
十二面	の	コート	(テニス)
グレー	の	コート	3
白熱	の	コート	(テニス)
コート	の	状態	(テニス)
コート	の	外	(テニス)
フード付きコート	の	ような姿	
コート	の	感じ	
コート	の	上	(テニス)
ミンクのコート	の	淑女	
神宮のコート	の	常連	(テニス)
隣のコート	の	バスケット部	
コート	の	ポケット	
コート	の	襟	

表6 コートに関する語と語の関係

このコートを詳細に分析すると英語のCoat (上着)とCourt (庭)の2つに大別できる。さらにCourtはスポーツと政治の場で使われていることが判る。さらにこれらを詳細に体系化すると次のようになる。

コート(1) (Coat)	色	ピンク
		ローズ色
素 材	材	グレー
		茶色
		：
		毛皮
		ミンク
		黒テン
		スエード
		：
		キルティング
		紙
部 分 丈	丈	ポケット
		襟
		フード
		フード付，フードつき
対 象 者	者	四分の三丈
		婦人
		淑女
		：
時 期	期	冬物
		：
		：
着 ぐ ち	ち	感じ
		苦情
		：

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、修正が簡単にでき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。

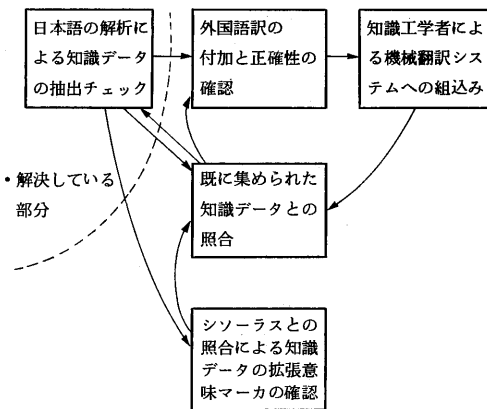


図3 語と語の関係の知識データが機械翻訳システムに組込まれるまでの作業プロセス

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。

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。

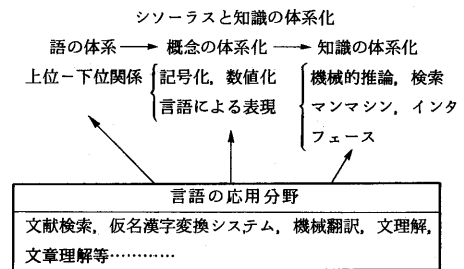


図4 シソーラスと知識の体系化

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。

最後に、この研究のために朝日新聞KWIC Tape を利用させて下さった東京大学工学部 藤崎教授、亀田氏に深く感謝の意を表します。

さらに、この研究の一部は文部省の科学研究費によって行った。

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
- (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
(これは前々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
- (24) 田中康仁 語と語の関係解析用資料—朝日新聞記事デ
ータ分析——“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 1987.11

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, No31, 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回全国大会(1987)

以 上

自然言語の知識獲得 (その2) 資料編

— 語と語の関係について, 朝日新聞記事データの分析 —

— 'が' について —

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 — — 'が' について — の続き (その 2) 資料編

Acquisition of Knowledge Data for Natural Language (NO. 2)

— From Asahi News paper —

YASUHITO TANAKA

SHO YOSHIDA

Himeji College

Kyushu Institute Technology

1-1-12 Shinzaike Honmachi
Himeji
670 JAPAN

680-4 Kawatu
Iizuka
820 JAPAN

It is reference documents of "Aquisition of Knowledge Data for Natural Language (No. 1)".

001	さびしさ	が 必要だ	1	046	教師	が 必要だ	1
002	アプローチ	が 必要だ	1	047	金	が 必要だ	2
003	スケープゴート	が 必要だ	1	048	空港	が 必要だ	1
004	チームワーク	が 必要だ	1	049	訓練地	が 必要だ	1
005	チャンネル	が 必要だ	1	050	形成	が 必要だ	1
006	データ	が 必要だ	1	051	計画	が 必要だ	1
007	パイプ	が 必要だ	1	052	警戒	が 必要だ	1
008	バルブ	が 必要だ	1	053	決断	が 必要だ	1
009	ホルモン	が 必要だ	1	054	検査	が 必要だ	1
010	圧力	が 必要だ	1	055	検討	が 必要だ	1
011	安静	が 必要だ	1	056	研究	が 必要だ	1
012	安定	が 必要だ	1	057	見直し	が 必要だ	1
013	案	が 必要だ	1	058	見直すこと	が 必要だ	1
014	意識	が 必要だ	3	059	原則	が 必要だ	1
015	意欲	が 必要だ	3	060	減税	が 必要だ	1
016	一致	が 必要だ	1	061	交流	が 必要だ	1
017	延長	が 必要だ	1	062	公表	が 必要だ	1
018	加工	が 必要だ	1	063	向上	が 必要だ	1
019	加熱	が 必要だ	1	064	工作費	が 必要だ	1
020	介入	が 必要だ	1	065	工夫	が 必要だ	1
021	解決	が 必要だ	1	066	考え方	が 必要だ	1
022	解明	が 必要だ	1	067	行動	が 必要だ	1
023	改革	が 必要だ	1	068	高温	が 必要だ	1
024	改正	が 必要だ	2	069	合意	が 必要だ	1
025	開発	が 必要だ	1	070	歳月	が 必要だ	2
026	開放	が 必要だ	1	071	賛成	が 必要だ	1
027	外圧	が 必要だ	1	072	姿勢	が 必要だ	1
028	外貨	が 必要だ	1	073	思いやり	が 必要だ	1
029	外交	が 必要だ	1	074	指導力	が 必要だ	1
030	拡張	が 必要だ	1	075	指標	が 必要だ	1
031	確保	が 必要だ	1	076	支え	が 必要だ	1
032	覚書	が 必要だ	1	077	施策	が 必要だ	1
033	革新	が 必要だ	1	078	視点	が 必要だ	1
034	環境	が 必要だ	2	079	資金	が 必要だ	2
035	監視	が 必要だ	1	080	歯止め	が 必要だ	1
036	緩和	が 必要だ	4	081	時間	が 必要だ	2
037	観客	が 必要だ	1	082	自制	が 必要だ	1
038	観点	が 必要だ	1	083	実験	が 必要だ	1
039	期間	が 必要だ	1	084	実行	が 必要だ	1
040	機関	が 必要だ	1	085	実施すること	が 必要だ	1
041	議論	が 必要だ	1	086	取り決め	が 必要だ	1
042	許可	が 必要だ	1	087	取り組み	が 必要だ	1
043	協力	が 必要だ	1	088	手術	が 必要だ	1
044	強化	が 必要だ	1	089	手段	が 必要だ	2
045	教育	が 必要だ	1	090	手当て	が 必要だ	1

001	樹立	が 必要だ	1	043	代表者	が 必要だ	1
002	修正	が 必要だ	1	044	知識	が 必要だ	1
003	準備	が 必要だ	1	045	注意	が 必要だ	1
004	処置	が 必要だ	1	046	注目	が 必要だ	1
005	署名	が 必要だ	1	047	調査	が 必要だ	1
006	助成	が 必要だ	1	048	調整	が 必要だ	7
007	承認	が 必要だ	1	049	追い上げ	が 必要だ	1
008	触れ合い	が 必要だ	1	050	通訳	が 必要だ	1
009	審議	が 必要だ	1	051	撤退	が 必要だ	1
010	新設	が 必要だ	1	052	転換	が 必要だ	1
011	森林	が 必要だ	1	053	電灯	が 必要だ	1
012	人材	が 必要だ	1	054	電力	が 必要だ	1
013	水	が 必要だ	1	055	努力	が 必要だ	3
014	制度	が 必要だ	4	056	統治	が 必要だ	1
015	政策	が 必要だ	1	057	同意	が 必要だ	4
016	整備	が 必要だ	1	058	道義	が 必要だ	1
017	製品	が 必要だ	1	059	内張り	が 必要だ	3
018	積み重ね	が 必要だ	1	060	入院	が 必要だ	1
019	設備	が 必要だ	1	061	年月	が 必要だ	1
020	節度	が 必要だ	1	062	能力	が 必要だ	1
021	絶縁	が 必要だ	1	063	把握	が 必要だ	1
022	宣言	が 必要だ	1	064	廃棄	が 必要だ	1
023	専門家	が 必要だ	1	065	配慮	が 必要だ	1
024	戦争	が 必要だ	1	066	費用	が 必要だ	1
025	選択	が 必要だ	1	067	避難	が 必要だ	1
026	措置	が 必要だ	1	068	付き添い	が 必要だ	1
027	組織	が 必要だ	1	069	福祉	が 必要だ	1
028	操作	が 必要だ	1	070	文言	が 必要だ	1
029	相互性	が 必要だ	1	071	兵器	が 必要だ	1
030	増加	が 必要だ	2	072	平和	が 必要だ	1
031	増強	が 必要だ	4	073	壁	が 必要だ	1
032	増税	が 必要だ	1	074	方法	が 必要だ	1
033	多数	が 必要だ	1	075	法	が 必要だ	2
034	妥協	が 必要だ	5	076	防衛費	が 必要だ	1
035	体力	が 必要だ	4	077	魔性	が 必要だ	1
036	対応	が 必要だ	1	078	猛省	が 必要だ	1
037	対策	が 必要だ	1	079	輸血	が 必要だ	1
038	対処	が 必要だ	1	080	理解	が 必要だ	1
039	対話	が 必要だ	1	081	力	が 必要だ	2
040	待遇	が 必要だ	1	082	連帯	が 必要だ	1
041	態勢づくり	が 必要だ	1	083	論議	が 必要だ	1
042	態度	が 必要だ	1				

語と語の関係(が)

001	おそれ	が 強い	4	046	拒否反応	が 強い	1
002	とばく性	が 強い	1	047	競争力	が 強い	1
003	アラブ世界	が 強い	1	048	恐れ	が 強い	36
004	イメージ	が 強い	1	049	空気	が 強い	1
005	トップスポーツ	が 強い	1	050	傾向	が 強い	1
006	ドル	が 強い	1	051	警戒感	が 強い	1
007	ニーズ	が 強い	1	052	警戒気分	が 強い	1
008	ニュアンス	が 強い	1	053	結びつき	が 強い	1
009	ムード	が 強い	1	054	懸念	が 強い	1
010	メーカー	が 強い	1	055	権限	が 強い	1
011	悪くなっている面	が 強い	1	056	見込み	が 強い	1
012	悪臭	が 強い	1	057	見通し	が 強い	1
013	圧力	が 強い	1	058	見方	が 強い	1
014	依存	が 強い	1	059	限界	が 強い	1
015	意見	が 強い	1	060	個性	が 強い	1
016	意向	が 強い	2	061	後半の方	が 強い	1
017	意識	が 強い	2	062	公算	が 強い	1
018	意味合い	が 強い	1	063	向き	が 強い	1
019	異論	が 強い	1	064	抗議の声	が 強い	1
020	印象	が 強い	1	065	考え	が 強い	1
021	引き	が 強い	1	066	考え方	が 強い	2
022	営利性	が 強い	1	067	高感	が 強い	1
023	影響力	が 強い	2	068	国の力	が 強い	1
024	英側	が 強い	1	069	再生力	が 強い	1
025	円	が 強い	1	070	志向	が 強い	1
026	塩分	が 強い	1	071	自閉傾向	が 強い	1
027	塩味	が 強い	1	072	執念	が 強い	1
028	押し	が 強い	1	073	実験性	が 強い	1
029	可能性	が 強い	1	074	実態	が 強い	1
030	家父長権	が 強い	1	075	邪魔意識	が 強い	1
031	海運	が 強い	1	076	若島津の方	が 強い	1
032	革新	が 強い	1	077	主体性	が 強い	1
033	感じ方	が 強い	1	078	上手段	が 強い	1
034	観念	が 強い	1	079	色あい	が 強い	2
035	関心	が 強い	1	080	色彩	が 強い	1
036	危機感	が 強い	2	081	心臓	が 強い	1
037	希望	が 強い	1	082	親のほう	が 強い	1
038	期待	が 強い	2	083	勢い	が 強い	2
039	気持	が 強い	1	084	勢力	が 強い	1
040	気持ち	が 強い	1	085	性質	が 強い	1
041	気持のほう	が 強い	1	086	政治色	が 強い	1
042	気分	が 強い	4	087	政府	が 強い	1
043	季節風	が 強い	1	088	精神症状	が 強い	1
044	規則色	が 強い	1	089	声	が 強い	1
045	疑い	が 強い	1	090	責任感	が 強い	2

001	絶望論	が強い	1	031	彼	が強い	1
002	先生たち	が強い	1	032	批判	が強い	2
003	宣伝員	が強い	1	033	漂白力	が強い	1
004	早慶	が強い	1	034	表象性	が強い	1
005	側面	が強い	1	035	不安	が強い	1
006	足腰	が強い	1	036	不安の方	が強い	1
007	尊敬心	が強い	1	037	不満	が強い	3
008	体質	が強い	1	038	負担感	が強い	1
009	対策の声	が強い	1	039	風	が強い	1
010	耐震性	が強い	1	040	風雨	が強い	1
011	地域性	が強い	1	041	風当たり	が強い	1
012	地肩	が強い	1	042	米欧	が強い	1
013	痛み	が強い	1	043	米国	が強い	1
014	抵抗	が強い	2	044	癖	が強い	1
015	抵抗力	が強い	1	045	法大	が強い	1
016	締めつけ	が強い	1	046	防御機能	が強い	1
017	投機性	が強い	1	047	北の風	が強い	1
018	東欧勢	が強い	1	048	面	が強い	2
019	毒性	が強い	2	049	野党	が強い	1
020	突き上げ	が強い	1	050	予想	が強い	1
021	日本経済	が強い	1	051	与野党	が強い	1
022	認識	が強い	2	052	要因	が強い	1
023	廃止論	が強い	1	053	要求	が強い	1
024	発言力	が強い	1	054	要請	が強い	6
025	発想	が強い	1	055	要素	が強い	1
026	判断	が強い	1	056	要望	が強い	1
027	反対	が強い	1	057	抑える力	が強い	1
028	反対意見	が強い	1	058	留任の線	が強い	1
029	反動性	が強い	1	059	力	が強い	1
030	反発	が強い	2				

語と語の関係(が)

SeqNo.			頻度
1	チョウ	が 飛ぶ	1
2	灰皿	が 飛ぶ	1
3	外相	が 飛ぶ	1
4	機	が 飛ぶ	1
5	酷評	が 飛ぶ	1
6	魂	が 飛ぶ	1
7	質問	が 飛ぶ	2
8	実弾	が 飛ぶ	1
9	小鳥	が 飛ぶ	1
10	笑い	が 飛ぶ	1
11	声	が 飛ぶ	1
12	声援	が 飛ぶ	1
13	石	が 飛ぶ	1
14	値	が 飛ぶ	1
15	蝶	が 飛ぶ	1
16	怒声	が 飛ぶ	1
17	当たり	が 飛ぶ	1
18	白鷺	が 飛ぶ	1

語と語の関係 (が)

1.	うだつ	が 上がる	1	① rise in the world ② succers in the world
2.	コスト	が 上がる	1	costs rise
3.	トイレットペーパー	が 上がる	1	prices of toilet rolls rise
4.	ドル	が 上がる	1	① Dollars appreciate ② Dollars get stronger
5.	ペース	が 上がる	1	① speed up ② The pase is accelated
6.	意気	が 上がる	1	① be elated ② be in high spirits
7.	右手	が 上がる	1	a right hand is raised
8.	営業成績	が 上がる	1	The results of operations improves
9.	温度	が 上がる	2	① The temperature rises ② The temperature goes up
10.	価格	が 上がる	2	① Prices rise ② Prices go up
11.	火	が 上がる	1	① burst into flame ② The fire starts ③ The fire is froken out
12.	火の手	が 上がる	1	furst into flame
13.	火柱	が 上がる	1	a pillar of flame shots up
14.	花火	が 上がる	1	① Fireworks are displayed ② Fireworks are sot off
15.	学年	が 上がる	1	be promoted to (the second year of ~)
16.	気温	が 上がる	1	① The atmospheric temperature rises ② The atmospheric temperature goes up
17.	氣勢	が 上がる	1	be in high spirits
18.	業績	が 上がる	5	The results improves
19.	金利	が 上がる	1	① The interest rate rises ② The interest rate goes up
20.	効果	が 上がる	2	① be effective ② work out effective
21.	効率	が 上がる	1	The efficiency improves

22.	工場	が 上がる	2	—
23.	高度	が 上がる	2	go up higher
24.	支持率	が 上がる	1	① The support rate rises ② The support rate goes up
25.	手	が 上がる	1	a hand is raised
26.	蒸気	が 上がる	1	① steam is raised ② steam is generated
27.	人気	が 上がる	1	① rise in popularity ② win more popularity
28.	水温	が 上がる	2	① The water temperature rises ② The water temperature goes up
29.	水準	が 上がる	1	The standard goes up
30.	性能	が 上がる	1	The performance improves
31.	成績	が 上がる	2	get higher marks
32.	生産効率	が 上がる	1	The production efficiency improves
33.	生産性	が 上がる	1	The productivity improves
34.	声	が 上がる	3	① a voice is made ② a voice arises
35.	息子	が 上がる	1	—
36.	足	が 上がる	1	① a foot is raised ② a leg is raised
37.	打率	が 上がる	1	The batting average improves
38.	大歓声	が 上がる	1	① a big shout of joy arises ② a big shout of joy is heard
39.	値打ち	が 上がる	1	rise in value
40.	地位	が 上がる	1	① get a higher position ② promote to a higher position
41.	調子	が 上がる	1	① go well ② The pitch is raised
42.	賃金	が 上がる	1	The wage increases The wage rises
43.	鎧五郎	が 上がる	2	—

44.	土煙	が 上がる	1	a cloud of dust rises
45.	2つ	が 上がる	1	—
46.	日	が 上がる	1	The sun rises
47.	年代	が 上がる	1	① Ages go up ② Ages rise ③ Ages are higher
48.	年齢	が 上がる	1	get older
49.	燃料代	が 上がる	1	① cost of fuel goes up ② cost of fuel rises
50.	能率	が 上がる	1	① be efficient ② The efficiency improves
51.	悲鳴	が 上がる	1	a scream arises
52.	批判	が 上がる	1	a criticism arises
53.	評価	が 上がる	1	① be estimated higher appreciate ② rise in evaluation
54.	物価	が 上がる	1	Prices rise
55.	物価だけ	が 上がる	1	only prices rise
56.	包装料	が 上がる	1	① Wrapping fees rise. ② wrapping fees go up
57.	防御率	が 上がる	1	The defense rate goes up
58.	雷雨	が 上がる	1	a thunderstorm stops
59.	利回り	が 上がる	1	① an investment yield ② an investment rises ③ an investment improves
60.	隆の里	が 上がる	1	Takanosato goes up (on the ring)
61.	累進税率	が 上がる	1	Rates of progressive taxes rise

1.	PLO	が 出る	1	PLO attends (a meeting)
2.	イタリア	が 出る	1	Italy attends (a meeting)
3.	カーペット	が 出る	1	—
4.	ビザ	が 出る	1	a visa is issued
5.	ブレ	が 出る	1	① be out of focus ② be unsteady given ③ be unsteady expressed
6.	意見	が 出る	4	—
7.	一光側	が 出る	2	—
8.	影響	が 出る	2	be influenced by ~
9.	音	が 出る	1	① a sound is made ② a sound is produced ③ a sound arises
10.	感情	が 出る	1	① The emotion is revealed ② The emotion is betrayed
11.	犠牲者	が 出る	1	be victimiyed
12.	議員	が 出る	1	① a member of parliament attends a meeting ② a member of parliament represente a meeting
13.	給料	が 出る	1	① Salary is given ② Salary is paid
14.	許可	が 出る	3	① an allowance is given ② an allowance is issued
15.	傾向	が 出る	1	The tread occurs
16.	見方	が 出る	13	① The view is revealed ② The view is presented
17.	原簿	が 出る	1	① The original register is presented ② The original register is shown
18.	個人的なもの	が 出る	1	① The personal thing is presented ② The personal thing is shown
19.	考え方	が 出る	2	The way of thinking is revealed
20.	高望山	が 出る	1	Koboyama participates in the match
21.	差	が 出る	1	① The difference arises ② The difference occurs

22.	斉藤	が 出る	1	Saito attends (a meeting)
23.	3人	が 出る	1	① Three people attend (a meeting) ② Three people gathers
24.	死者	が 出る	1	Lives are lost
25.	資質	が 出る	1	The quality is revealed
26.	主戦論	が 出る	2	① an aggressive opinion is presented ② an aggressive opinion is expressed
27.	柔軟論	が 出る	1	① a flexible opinion is presented ② a flexible opinion is expressed
28.	焼き魚	が 出る	1	a grilled fish is served
29.	人質たち	が 出る	1	Hastages are released
30.	性格	が 出る	1	a character is revealed
31.	政党	が 出る	1	(a representative of) the political party attends (a meeting)
32.	声	が 出る	1	① a voice arises ② a voice is made
33.	線	が 出る	1	a line appears
34.	捜査員	が 出る	1	an investigator appears
35.	走者	が 出る	1	① a runner runs ② gains a lead
36.	大阪行き	が 出る	1	a train for asaka leaves
37.	店員	が 出る	1	a store clerk appears
38.	答申	が 出る	1	a report is submitted
39.	動き	が 出る	1	① The trend occurs ② The movement occurs
40.	特急	が 出る	1	an express leaves
41.	発言	が 出る	1	① an opinion is presented ② an opinion is expressed
42.	判断	が 出る	1	① a decision is made ② a judgement is made
43.	悲観	が 出る	1	a pessimistic view is expressed
44.	批判	が 出る	1	① a critical opinion is expressed ② a criticism arises

45.	票	が 出る	1	① The votes are won ② The votes are gathered
46.	品	が 出る	1	Dignity is revealed
47.	不満	が 出る	1	① Dissatisfaction is revealed ② Dissatisfaction is expressed
48.	復権	が 出る	1	—
49.	母親	が 出る	2	a mother appears
50.	本	が 出る	1	a book is published
51.	命令	が 出る	2	an order is issued
52.	役所	が 出る	1	a representative of the authority attends (a meeting)
53.	訳本	が 出る	1	a translation of the book is published
54.	予想	が 出る	1	① an expectation is made ② an expectation is released ③ an expectation is announced
55.	利用者	が 出る	1	~ is begun to use
56.	600人	が 出る	1	① 600 people attend (a meeting) ② 600 people gather