

推移律に基づく共起グラフのクラスタ分け

田中久美子
東京大学工学部

岩崎英哉
東京農工大学工学部

コーパスから得られた共起情報は、共起関係を枝、語をノードとするグラフと見なすことができる。同じ語が複数の意味を持つことがあるという曖昧性によって、この共起グラフは全体でひとつの大きなグラフを形成している。本稿では、曖昧性を「共起関係の推移律」という観点から捉え、共起グラフから同じ意味・話題を成す部分グラフを推移律が成立するように抽出して全体をクラスタ分けするための手法を提案する。クラスタ分けされたグラフの性質などについても考察する。さらに、実際に20Mのコーパスから作成した共起グラフに対して本手法を適用した結果、および、それを用いた主成分分析による記事の分類を行なった結果も合わせて報告する。

Clustering Co-occurrence Graph by Transitivity

Kumiko TANAKA

Hideya IWASAKI

Faculty of Engineering

Faculty of Technology

The University of Tokyo

Tokyo University of Agriculture and Technology

Co-occurrence information, obtained from a corpus, is a graph, regarding words as nodes and co-occurrences as branches. Subgraphs of different themes stay connected because of ambiguous words. This paper assumes the ambiguity as the non-transitivity in co-occurrence relationship and tries to cluster a graph into subgraphs in which transitivity holds. Several properties of such subgraphs are also discussed. The clustering method is applied to a co-occurrence graph obtained from 20M Japanese newspaper articles and its results are analyzed. They are also used to classify the articles.

1 はじめに

意味の曖昧性の解消は自然言語の根本的な問題のひとつであるが、それを補助するためにクラスタリングの研究がさかんである。クラスタリングとは、何らかの基準によって語をいくつかのグループに分けることをいう。たとえば、既存の類語辞典やシソーラスは人間が手で行なったクラスタリングの結果である。

共起関係や辞書から得た語群の機械的なクラスタリングは、巨大なシソーラスの自動作成や既存の類語辞典の見直しを目的として研究されている。またより応用に近い目的としては、文書の自動分類 [Iwayama, 1995]、また機械翻訳時の訳語選択の語の曖昧性の解消の補助 [Fukumoto, 1994] が挙げられるであろう。本研究ではグラフの観点からのクラスタリングの一手法を提案する。

クラスタリングの方法にはすでに様々な提案 [Tokunaga, 1995][Brown, 1992][Hindle, 1990] がなされている。Li と阿部 [Li, 1996] は各クラスタの情報理論における Description Length が最小となるようにグラフを分割していく方法をとっており、結果が pp-attachment などに有効であることを示した。また、Pereira [Pereira, 1993] はエントロピーが小さくなるようにクラスタ分けする。このようにクラスタリングでは、いづれの研究もある数理的な基準をもって稠密な語群を抽出しているが、結果として得られたクラスタの言語的意味が必ずしも明確とは言いがたい。

本研究の目的は、共起関係から意味まとまりを抽出することである。コーパスから得られた共起情報は、共起関係を枝、語をノードとするひとつの大きなグラフ（以後“共起グラフ”と呼ぶ）とみなすことができる。共起グラフにおいては、異なる話題に属す語群は、意味の曖昧な語を介して共起関係でつながっている。しかし、その曖昧な語をはさんで、異なる話題の語の間には共起関係はない。すなわち、語が曖昧である時には共起関係に推移律が成り立たないという特徴がある。この特徴を利用して、推移律の成り立つ部分グラフを抽出することによって、共起グラフをクラスタ分割を試みる。

共起グラフをつくるもととなる共起情報には共起の度合（“共起度”と呼ぶ）があるので、共起度を考慮することによって、様々な共起グラフが考えられる。本研究においては、共起度は共起情報から共起グラフを作成する際にのみ参照し、共起グラフの構造のみから曖昧性を排除してクラスタ分けすることを試みている。

本稿の構成は以下の通り。次節で本稿で抽出する推移律が成り立つグラフとはいかなるものかを定義して性質を述べ、3節で抽出アルゴリズムを提唱する。4節では実際に実験を行なってその有効性を確かめる。これを用いて実際に新聞記事の自動分類を試みた結果をも報告する。

2 語義の曖昧性と推移律

二つの語がコーパス中に隣接して高い頻度で出現する時、その二つの語が共起関係にあると定義する。語をノード、枝を共起関係と見なすとコーパスからは巨大な共起グラフを抽出することができる。このグラフを適当な基準で部分グラフに分けることをクラスタリング、各部分グラフをクラスタと定義する。さらに、語すなわちノードの集合を V 、共起関係すなわち枝の集合を E としてグラフを $G = \langle V, E \rangle$ 、ノードの数を $|V|$ と記述する。

2.1 共起関係における推移律

本節では共起関係と意味の曖昧性とのつながりを論じる。集合 V 上に定義されたある関係 R と $a, b, c \in V$ に対して、つぎの反射律、対称律、推移律を考える。

- 反射律 aRb .
- 対称律 $aRb \rightarrow bRa$.
- 推移律 $aRb, bRc \rightarrow aRc$.

R が共起関係である時、上のそれぞれが成立することを以下の様に解釈することができる。

- 反射律 ある語が同じ語と共起することを指す。
- 対称律 共起の順序における対称性を示す。たとえば、医者と看護婦の共起は順序に依存しないので、対称律は成り立つ。一方、国会議事堂は国会の後に議事堂は起こりやすいが、その逆が成り立つことは稀なので、対称律は成り立たない。
- 推移律 語 b の意味が曖昧でないことを指す。たとえば、 $a =$ 台所、 $b =$ 勝手、 $c =$ 我儘 の3語では a と c の間に共起関係がないため、推移律は成り立たない。

曖昧性が排除された連結な部分グラフは、ひとつの話題に関連する語から成ると考えられる。そのため推移律が成り立たない部分を排除するような部分グラフを抽出することによってクラスタリングを行なう。

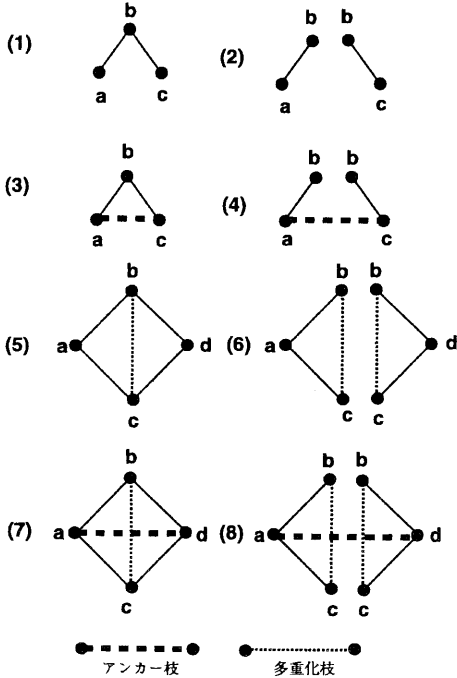


図 1: グラフの分割

3 推移律とグラフの分割

3.1 分割, 多重化, 完全グラフ

推移律を曖昧性という観点から捉えると、以下のようになる。

最も単純な場合として、3語から成る共起グラフを考える。図1-(1)は、推移律の成立しないグラフであり、先にあげた例にあてはめると、*a*が台所、*b*が勝手、*c*が我儘となる。ここで*ac*間に枝(共起関係)がないのは、*b*(勝手)という語に曖昧性(二面性)、すなわち、「台所」的な意味(「勝手口」の勝手)と「我儘」的な意味(「自分勝手」の勝手)があるためである。そこで、*b*を“多重化”して、それぞれの意味に対して異なるノードを割りあてると、図1-(2)のように*ab*と*bc*のふたつの部分グラフに分割される。つまり、推移律が成立しない場合には、曖昧性を持つノードを多重化することにより曖昧性が解消され、グラフが分割される。

一方、図1-(3)のように*a*、*b*、*c*の間に推移律が成立する場合には、先と同じように*b*を多重化してもグラフは分割されない(図1-(4))。これは、多重化された部分(ノード*b*)に曖昧性がないためと

説明することができる。

以上ではノードひとつにある曖昧性に注目して説明したが、この考えを拡張すると、「ふたつのノードとその間の枝」の曖昧性を考えることもできる。図1-(5)は*bc*に曖昧性がある例で、*bc*を多重化することにより、曖昧性が解消されてふたつの部分グラフに分割される(図1-(6))。図1-(7)は*ad*間に共起関係があるために、*bc*の多重化ではグラフが分割しない(図1-(8))。すなわち*bc*に曖昧性はないのである。なお、図1-(7)はノード数が4の完全グラフであることに留意されたい。*bc*のように多重化される枝を“多重化枝”，*ad*のように多重化による分割を阻止する枝を“アンカー枝”と呼ぶことにする。

一般には、注目するグラフ中の一部を多重化することによってグラフが分割される場合、多重化された部分グラフには曖昧性があると解釈することが可能である。したがって、推移律という観点からは、

分割された各々の部分グラフは、そのいかなる一部を多重化してもそれ以上分割されることはない

ようにグラフを分割するのが理想である。ところが実際にはこの制限を満すグラフは完全グラフしかあり得ず、クラスタ分けに用いるためには制限が厳しすぎる。アルゴリズム的に考えても、あるグラフから完全グラフを抽出する問題はNP完全なので、完全グラフを考えるのは現実的ではない。

3.2 推移律グラフ

そこで制限を緩めることを考える。制限の緩め方には二通り考えられる。

5ノード以上の完全グラフでは、多重化枝を一本定めると、図2-(1)のようにアンカー枝が複数存在する。しかし、実は分割を阻止するためだけならば、1本のアンカー枝で十分である(図2-(2))。アンカー枝の数が少ないほど制限は緩い。

もうひとつの制限の緩め方は、推移律そのものの緩和である。2.1節で述べた通常の推移律は3ノード間(図2-(3)のように*acd*間や*abd*間)で成り立つが、以下のような緩い推移律を定義することができる。

$$v_1 R v_2, \dots, v_{n-1} R v_n \rightarrow v_1 R v_n$$

この緩い推移律を*abd*と*acd*に適用すると、図2-(4)となる。*a-b-d*を通る経路の最小値と*a-c-d*を通る経路の最小値の大きい方を“アンカー距離”と

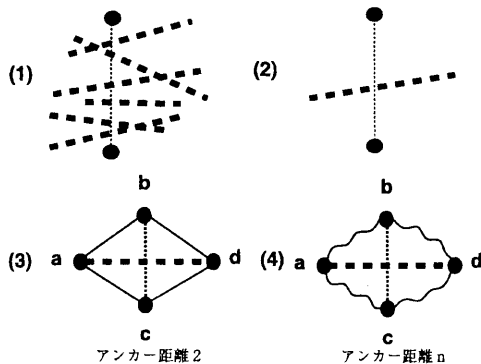


図 2: 分割のための緩い条件

いうこととする。たとえば図 2-(4) の $a-b-d$ の長さの最小値が 4, $a-c-d$ の長さの最小値が 6 の時, アンカー距離は 6 となる。アンカー距離は $n = 2$ の時が図 2-(3) で, もっとも制限がきつい。

このような緩めた条件を満たすグラフを推移律グラフと呼ぶ。 $m, n < |V| - 1$ に対してあるグラフ G が (m, n) 推移律グラフであるとは,

あらゆる多重化枝 $e \in E$ に対し, アンカー枝 $e' \in E$ でアンカー距離が n 以下のものが m 本存在する。

と定義する。

図 3 は, さまざまな制限による推移律グラフの集合の関係を示している。縦軸はアンカー枝の本数 m を示し, m が小さいほど条件が緩い。 m が大きい推移律グラフは m が小さいグラフの部分グラフとなっている。横軸はアンカー距離 n を示し, n が大きいほど制限が緩い。 n が小さい推移律グラフは n が大きいグラフの部分グラフとなっている。

最も制限を緩くして共起グラフを分割して得られた推移律グラフの集合が図 3 の $GS2$ であり, 「単独のノード」および「ふたつのノードとその間の枝」の曖昧性が, 緩い推移律の成立するアンカー枝 1 本によって解消されている。また, 最も条件をきつくして得られる推移律グラフの集合は完全グラフで, 図では $GS0$ を含む底辺にあたる。

4 推移律グラフの抽出法とその考察

推移律グラフの抽出法には, ひとつ大きなグラフをある基準で分割して小さい部分グラフを得るトップダウンの方法と, ある枝から始めてある性質を満

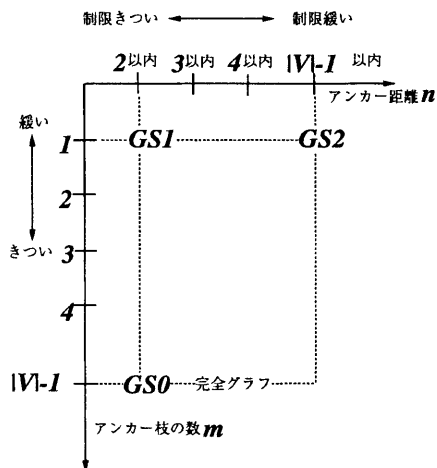


図 3: (m, n) 推移律グラフ

たす部分グラフを膠着的に大きくしていくボトムアップの方法の二種類が考えられる。推移律グラフでは局所的にどのようにノードがつながっているかという観察から出発しているの, 後者が適している。

本節では, 分割すべき共起グラフは対称律が成り立つように作られているものとする。

4.1 抽出アルゴリズム

ここでは図 3 における $GS1$, すなわち $(1, 2)$ 推移律グラフを抽出する方法を示す。

0. $G_0 = \langle \emptyset, \emptyset \rangle, i = 0$
1. G_0, \dots, G_i に含まれないノード v を G からひとつ選び, それを含む任意の三角形を G_{i+1} に入れる。
2. G_{i+1} の一辺 e を底辺, G_0, \dots, G_{i+1} のいずれにも含まれない v を頂点とする三角形に対し, 条件
 - e の両端点から距離 1 のノード $v' \in G_{i+1}$ と, v との間に枝がある (図 4)
 が成り立つ時, v を G_{i+1} に入れる。また, G_{i+1} 中のノード v'' と v との間の枝をすべて G_{i+1} に入れる。
3. 2 を G_{i+1} が拡張できなくなるまでくり返す。
4. $i = i + 1$ として 1 に戻る。

グラフは有限であるので, この操作は必ず終了し, 計算量は高々 $O(|V|^3)$ で済む。

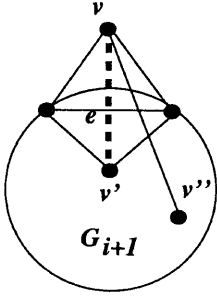


図 4: (1, 2) 推移律グラフ $GS1$ の抽出法

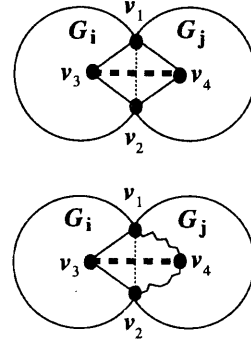


図 5: アルゴリズムに関する考察

4.2 アルゴリズムに関する考察

上のアルゴリズムが正しいことを示すには、

- 本来は別々になるべき部分グラフがひとつにまとまってしまうこと；
- 本来はひとつにまとまるべき部分グラフが別々になってしまうこと；

が共に起こり得ないことを示す必要がある。

前者に関しては、部分グラフのボトムアップに求めるという方法からして起こり得ないのは明らかである。以下、後者を示す。

抽出された部分グラフを、 G_0, G_1, G_2, \dots とし、これが後者の意味で正しい分割ではなかったと仮定する。すると、ある枝 v_1v_2 で接しあう部分グラフ G_i と G_j が存在して、図 5 の上段のようになっているはずである。しかし、これは G_i の方が先に作られたとしても G_j の方が先に作られたとしても、このアルゴリズムのグラフ拡大の方法からして起こり得ない。つまり、 G_i が先に作られた場合には、 v_4 というノードは必ず G_i に取りこまれるはずである。逆もまた同様である (証明終わり)。

この証明をみるとわかる通り、本アルゴリズムでは v_1v_2 (多重化枝) に対して v_3 と v_4 (アンカー枝の両端点) とが対称な位置関係にあることが重要である。仮に図 5 の下段のように、抽出アルゴリズムの第 2 ステップにおいて v と e の両端点との距離は 1、 v' と e の両端点との距離は 2 以上で可、というような非対称な条件で推移律グラフを拡大していくと、拡大の順序によっては正しい分割結果を得ることができない。なぜならば、 G_j の方が先に作られた場合には v_3 に関しては拡大の条件は満たされて G_j に取りこまれるので G_i と G_j と分かれることはないが、運悪く G_i が先に作られたとすると、

$v_1v_2v_4$ が拡大の条件を満たしていない時に G_i のグラフの拡大が不可能になってしまい、本来はひとつにまとまるはずの部分グラフが別々になってしまうことが起こりうるのである。

$GS2$ という推移律グラフの最大の集合を抽出するにも上のアルゴリズムを部分的に変更するだけでよい。具体的にはステップ 2 において、 v と v' と枝 e の両端点との距離を 1 でなく、 $|V|-1$ 以下とする。

5 実験

朝日新聞の教育記事、環境記事それぞれ 10M (語数は計およそ 1120 万語) を形態素解析器 JUMAN により名詞、動詞、形容詞、副詞を抽出した。ある語から 3 語先までの共起語を数え、相互情報量によって共起の度合 (共起度) を計算した。この時、共起度がある閾値より大きいものだけを考慮し、双方向となっていない枝は捨てることによって対称律が成立するように共起グラフを作成した。分析に用いた語数は 40049 語であった。

5.1 推移律グラフの抽出

前の節のアルゴリズムにより推移律グラフの集合 $GS1$ を抽出した。共起度の閾値によってアルゴリズムを適用する共起グラフが異なり、結果も異なるので、まずはどの閾値で抽出したグラフを用いるかを決めなくてはならない。実験の結果、得られた $GS1$ は、閾値によって

- 1 大きなひとつのグラフ
- 2 ある程度の語数 (5 ~ 100 程度) の複数のグラフ
- 3 多数の小さなグラフ

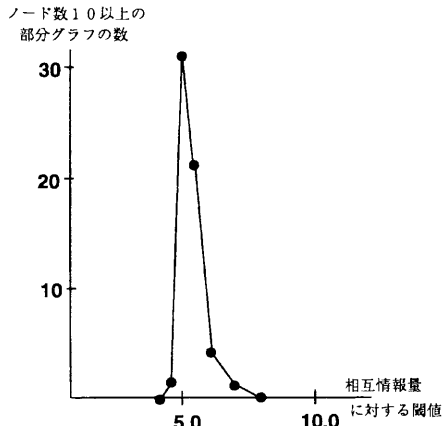


図 6: 閾値と推移律グラフの関係

のいずれかになる。閾値が高いと3となり、閾値が低いと1となる。また、高い閾値によって得られる推移律グラフは低い閾値によって得られる推移律グラフの部分グラフとなる。1の場合は、ひとつの話題にまとまってしまう、意味処理には用いることができない。逆に3の場合は大域的な話題が掴めないので、2の傾向を示すものを抽出したい。そのためには、結果がある一定以上のノード数をもつ推移律グラフの数が最大となる閾値を最適と定めれば良い。すなわち、最適なクラスタ分割を与える共起グラフを共起情報から抽出するように、一連の処理を自動化することが可能である。

図6は閾値を0.5きざみで共起グラフを作成し、前節のアルゴリズムを適用し、ノード数が10以上であった推移律グラフの数を調べている。2に相当する構造をもつ閾値は非常に狭い範囲に限定される。閾値5.0では最も多くの(ノード数10以上の)推移律グラフが抽出され、32グラフの中に異なる937語が含まれた。

出力結果の一部を付録に示す。グループ1は原子力発電の事故問題(環境)、グループ2は政治・安全保障・核問題(環境と教育両方の記事に含まれる)、グループ3は医大における移植手術(教育)、グループ4は住宅(環境)、グループ5は受験(教育)などと話題が限定されていることが観察されるであろう。

5.2 新聞記事の分類

環境と教育記事それぞれから1000記事をランダムに抽出し、計2000記事を分類の対象とする。前節

表 1: 正しく教育記事が分類された割合

上位記事数	正しい分類の割合
100	98.0%
200	97.5%
300	93.3%
400	89.3%
500	86.2%
600	82.2%
700	80.0%
800	77.0%
900	74.7%
1000	72.5%

で求めた32個の推移律グラフを用いて、これらを以下の順序で主成分分析した。

1. 各記事ごとに、各推移律グラフと共通する語数を計算し、32次元ベクトルを求める。
2. 2000記事の32次元ベクトルから主成分分析を行なう。
3. 各記事のベクトルの第一主成分を大きい順にソートする。

結果を表1に示す。第一主成分の大きい上位1000記事には教育記事が72.5%含まれた。この値は低いように思えるが、第一主成分の最も大きかった100記事を見るとその中には98.0%の精度で教育記事が含まれている。上位200記事では精度は97.5%であった。第一主成分が中央の値に近づくにつれ、環境記事か教育記事か曖昧になっていくことを示している。実際、教育、環境記事という分類は非常に大まかであり、特に双方に政治、訴訟、医学に関する記事が混在し、このようなものの第一主成分は中央に近い値を取ることが観察された。以下の記事は教育に分類された環境記事の例である。いずれも環境の意味がほとんどないため、第一主成分は中央に位置する。

- 例1 日本の反アパルトヘイト市民運動グループが南ア紙に意見広告—日本で反アパルトヘイト(人種隔離)の運動を続けている市民グループ、アフリカ行動委員会(牛嶋保夫代表)が南アフリカの五つの新聞に意見広告を出すことになり、二百万円の募金と賛同者の募集を行っている。ネルソン・マンデラ氏の釈放でア

バルトヘイト完全撤廃へ動き出した南アフリカの民衆を勇気づけ、日本での反バルトヘイト運動の輪を広げるのがねらい。三月末ごろ、英語の週刊新聞「ニュー・ネーション」や日刊紙「ザ・スター」など五紙に掲載する。...

- 例2 「詐欺的商取引」などへの関与禁止を新設 — 日本弁護士連合会（藤井英男会長、会員一万三千九百人）は二日、臨時総会を開き、弁護士の行動規範となる「弁護士倫理」規定を三十五年ぶりに全面改定した。前文で弁護士の社会的責任を明記し、豊田商事や地上げ業者、暴力団の顧問弁護士が不正に手を貸していた反省を踏まえ「詐欺的商取引、暴力などの違法・不当行為」を助長、利用することを禁じた条文を新設したのが特徴。弁護士の不祥事が後を絶たない中で、日弁連としても内部規律の強化を迫られた。（3面に解説）新倫理規定は全部で八章六十一条。詳細に規律を定めたため、昭和三十年三月施行の旧規定が七章三十五条だったのに比べ、倍以上に膨らんだ。旧規定が理事会決議なのに対し、新規定は日弁連最高意思決定機関である総会で議決する手続きを取った。最大の特色は一四条（違法行為の助長）の新設。...

分類に主成分分析を用いること自体は、自動処理の観点からは問題がある。主成分分析の結果、主成分を用いた分類がどのようなものになるかという点は、人間が観察する以外にない。たとえば、本実験でも第一主成分だけでなく、第二主成分までを用いて4つに分類することなども可能である。

6 まとめと今後の展望

コーパスから得られた語の共起関係をクラスタリングする手法に関して述べた。曖昧性を「語の共起関係の推移律」と見なし、共起グラフから同じ意味・話題を成す部分グラフを推移律が成立するように抽出して全体をクラスタ分けするための手法を提案した。クラスタ分けされたグラフの性質などについても考察した。さらに、実際に20Mのコーパスから作成した共起グラフに対して本手法を適用した結果を示した。また、それを用いて主成分分析法で2000記事の分類を行ない、全体で72.5%の精度（適用率を落とすと98%）の精度で正しく分類できることを示した。

新聞記事の分類の手法として、主成分分析は完全なる自動化を目指せないで、今後はその点を改善

することが重要な課題である。同時に、より細かく分類されている新聞記事を用いて同じ実験を行ない、クラスタリングの精度を測る必要もある。

また、現在のところはクラスタリングの結果は意味の階層構造などは考えられていないが、集合の包含関係から階層構造に関しても考えていきたい。また、より高次の曖昧性の解消を行なったグラフの抽出アルゴリズムも考えたい。さらに、この枠組を共起度を反映した重み付きグラフに対応するように拡張することも課題である。

参考文献

- [Li, 1996] H. Li and N. Abe. 1996. Clustering Words with the MDL Principle. In *Proceedings of the 15th International Conference on Computational Linguistics*, vol.1, pp.4-10.
- [Iwayama, 1995] M. Iwayama and T. Tokunaga. 1995. Hierarchical Bayesian Clustering for Automatic Text Classification. In *Proceedings of the International Joint Conference on Artificial Intelligence '95*.
- [Tokunaga, 1995] T. Tokunaga, M. Iwayama and H. Tanaka. 1995. Automatic Thesaurus Construction based on Grammatical Relations. In *Proceedings of the International Joint Conference on Artificial Intelligence '95*.
- [Fukumoto, 1994] F. Fukumoto and J. Tsujii. 1994. Automatic Recognition of Verbal Polysyny. In *Proceedings of the 14th International Conference on Computational Linguistics*, vol.2, pp.762-768.
- [Pereira, 1993] F. Pereira, N. Tishby and L. Lee. 1993. Distributional Clustering of English Words. In *Proceedings of the 31st ACL*, pp. 183-190.
- [Brown, 1992] P. Brown, V. Pietra, et al. 1992. Class-based n-gram Models of Natural Language. *Computational Linguistics*, 18 (4), pp. 283-298.
- [Hindle, 1990] D. Hindle. 1990. Noun Classification from Predicate — Argument Structures. In *Proceedings of the 28th ACL*, pp.168-175.

付録

グループ例 1

圧力異常自動水位原子炉漏れ安全装置停止手
動作制御運転低下冷却原型教習高圧温度炉
心出力循環系統臨界下がる故障一次蒸気注入
配管沸騰下げる放射能隔離損傷破損気圧事故
増殖福井県高温細管軽水新型二次実証高速出
口発生器給水振動流れる改良液体日本海腐食
関西電力定期美浜想定原発美浜町福井隣接高
浜取り付ける下部大飯止め振れる見つかる原
因施工起きる稼働燃やす燃料金具投棄廃棄物
放射性海洋割れる三方郡大飯郡交換高浜町設
計届く検査二本低レベル極東固体処理専用海
軍原潜貯蔵点検六ヶ所村施設廃液潜水艦青森
県操業青森燃す濃縮本原

グループ例 2

安保国連査察決議北朝鮮制裁撤回首脳脱退外
交高官再開条件朝鮮民主主義伝える特別南北
非核米朝履行会談核拡散韓国期限疑惑協定脅
威重村表明両国議題通告交渉十二日声明宣言
対話中止朝鮮復帰加盟外務省拒否協議実務政
府接触応じる米国合意次官受け入れる正常大
使訪問解決外務核兵器軍縮相互問題解決話し
合う保有国義務進展北京三回十九日二回国務
省主席首席聯合外相細川首相大統領日露
官邸宮沢国防慰安中国日韓下川総理演説海部
理由連行代表蔵相渡辺来日調印香港時事訪日
薄木法相訪米国務長官閣僚強制台湾日報南京
羽田従軍条約発効閣議批准

グループ例 3

医師産婦人科小児科病院医療外科教授治療小
児注射薬品患者通院内科皮膚手術感染院長肝
炎療法心臓男児肝臓移植女児小沢承認倫理血
液副作用臓器血管摘出脳死一歳閉鎖生体提供
信州信州大学医科カ月申請京都大学判定停止
医学二歳医大四歳八歳六歳鳥根三歳死亡男女
五歳七歳年齢男性未満長男少女少年十歳九歳
女性未婚既婚

グループ例 4

延べ住宅二階平方メートル建物木造建てる土
地敷地民家建つ校舎駐車場宅地農地所有

グループ例 5

英語言語中国語日本語外国語学ぶ学科国語算
数習う会話覚える教え数理解科科目論文教科
先生学力授業教える宿題後期受験選択出題物
理試験必修履修勉強単位学年学部工学部理工
学部経済学部前期日程倍率分割方式工学科農学

部理学部経営選抜段階定員二次分離募集業学
部連続出願2月公立国公立除く一倍3月試験
日1月前半志願国際関係来春確定面接実技門
前払五大前年志望合格私大受験生集計上回る
今春最多選考偏差値不合格入試替え玉顔写真
得点平均点成績満点浪人短大万三千人内申書
評価進路推薦可否私立採点進学生徒高卒万人
指導開示高校中学中退業者事前相談昨春増設
中高答案用紙遅刻大卒採用要録請求非公開本
人所見原簿追放打ち出す中学校鳩山文相就職
教師高槻公文書条例森本文書公開埼玉答申審
査異議

グループ例 6

伊方住民上告訴訟敗訴二審判決水害最高裁無
罪一審勝訴確定棄却取り消す地裁差し戻す多
摩川逆転原告側大東下級法廷結審言い渡す行
政訴訟弁論改修適格控訴賠償却下和解抗告高
裁請求戸塚差し止める懲役福岡熊本三次審理
新潟水保提訴併合勧告申し立てる命ずる河川
弁護原告裁判被告公判堤防支払う認容民事訴
訟損害細川熊本県裁判所破棄所見提示求刑騒
音執行猶予無期量刑檢察一時金一陣四日市救
済認定工事団長宮崎北川殺害起訴当事者禁固
三年罰金冒頭陳述患者北野症状

グループ例 7

一冊高校生小学生中学生読む雑誌大人読書百
円感動千円大学生男子毎日活字忙しい図書
感想出版新聞表彰指定楽しい入賞

グループ例 8

一酸化炭素水素窒素化物硫黄自動車酸性大気
汚染二酸化二酸化炭素排出物質硫酸濃度測定
石炭火力

グループ例 9

効率実用化太陽電池発電利用電気風力水力
地熱電力燃料自家発電出力

グループ例 10

異議公開公文書審査開示条例請求非公開文書
内申書要録高槻指導進路成績答申本人所見
原簿森本偏差値教師生徒選抜進学入試学力高
校中学中退業者推薦可否私立受験生評価公
立追放相談打ち出す中学校鳩山文相事前中高
合格受験不合格満点埼玉国公立志望浪人私大
替え玉顔写真採点試験教科

グループ例 11

育児仕事子育て家事母親父親働く保育園保母
職場保育保育所預ける出産園児赤ちゃん産む
妊娠女性結婚出生