

係り受け関係と相互情報量を用いた単語の意味獲得

前川 篤志 伊藤 毅志 古郡 延治

電気通信大学情報工学科
東京都調布市調布ヶ丘 1-5-1

計算機によって語彙の曖昧性を解消するには、単語の意味を表現、あるいは記述する必要がある。本稿では、単語の意味を複数の類義語で記述する手法を提案する。まず、既存のシソーラスでは捉えにくい単語間の関係を、コーパスから得られた情報をもとに類義語の集合として抽出する。次に、文脈を考慮して優先度を計算し、その順序集合づくり、これによって所与の単語の意味を記述する。ここでは、動詞「開く」を用例として実験を行い、その結果をクラスタリングと人手によって評価する。クラスタリングによる評価では、無作為に分割したときの平均よりも結果が得られ、人手による評価では、所与の単語の意味記述に使われた類義語の上位5語までのものが、82%の確率で人間の評価と一致した。

キーワード：コーパスベース、係り受け関係、相互情報量、類義語、クラスタリング

Determining Word Senses using Syntactical Relation and Mutual Information

Atsushi MAEKAWA Takeshi ITO Teiji FURUGORI

Department of Computer Science, The University of Electro-Communications
1-5-1 Choufugaoka, Chofu, Tokyo, Japan

Determining word sense(s) for any word used in text is a necessary process in natural language processing. In this paper, we propose a method for describing the meaning of the word in text with a set of synonyms that appear in similar contexts. First, using syntactical information, we get synonyms of the given word from a corpus. Then, using mutual information, we select a few of them as the ones that describe the meaning of the given word. We have got a success rate of 82%, when compared with an human judgment over an experiment with the verb *hiraku*.

Keywords: corpus-based NLP, syntactical relation, mutual information, synonym, clustering

1. はじめに

語彙の曖昧性解消は言語処理上の重要問題の一つである。人間が言語を理解する際にも、文中の曖昧性を解消し単語のもつ特定の意思を決定していると考えられている。たとえば、英語の“bank”には、少なくとも「銀行」と「土手」という2つの意味があり、“I go to the bank to get money.”という文では、“bank”のもつ曖昧性は何らかの方法で解消され、「銀行」という意味が推測される。

計算機上によって語彙の曖昧性を解消するには、まず、単語の意思を表現、あるいは記述する必要がある。従来、意思記述には、①辞書の語義定義文を用いる方法[1,2,3]、②文の用例を用いる方法[4,5]、③係り受けの関係をj用いる方法[6,7,8]の3つの方法がある。

①は、辞書の語義定義文からベクトルを作成する方法である。この方法は、辞書に掲載されている全ての語義を対象としてベクトル化できるといふ利点があるが、新語や新語義は人手で定義しなければならず、自動的に単語の意思を獲得できないという欠点がある。

②は、語義が既知である文の用例を集め、これを基にベクトルを作成する方法である。この方法は、多くの用例からベクトルを作成するため、多くの観点から単語間の関係を捉えることができるが、①同様、単語の意思を自動的に獲得できないという欠点がある。

③は、動詞と係り受け関係にある名詞のクラスタリングによって得られたクラスターで、その動詞がもつ意思を表現している。この方法は、事例を基に単語の意思を定義するため、新語、新語義を獲得しようとするうえで有効である。しかし、クラスタリングの際に、既存のソーラスを用いて名詞間の類似性を計っている点に問題があると思われる。川村ら[9]によれば、現在のソーラスには「分類観点によって変化する語と語の関係を捉えることができない」という問題があり、この問題がクラスタリングに影響を与えていると考えられる。また、得られたクラスターは、仮想動詞として、その類義語1語でしか表されておらず、クラスターの表す意味が複数の類義語にまたがる場合を考慮していない。

本研究では、単語の意思を複数の類義語を用いて記述する。たとえば、「会議を開く」といったときの「開く」の意思を「開催する」と特定してしまうのではなく、この文で「開く」と文脈上同様の意思をもつだろうと推測される別の単語の集合によって記述する。この背景には、「同様の意思で用いられる単語は、同様の文脈上に現れる」[10,11,12]という仮説がある。

以下、本稿では、まず、2章で、本研究で使う意思の記述法を提案する。また、その際、必要となる類義語の抽出方法を述べる。3章では、動詞「開く」が使われている文を対象に行った単語の意思記述（または推定）の実験結果を示し、4章でクラスタリングと人手で行ったその評価実験の結果をもとに言語処理にもつ意義を考察する。

2. 類義語による単語の意思記述

我々は、単語の意思はその類義語を用いて表すことができると仮定する。しかし、単語がもつ意思を表す類義語が唯一であるとは限らない。したがって我々は、文脈に沿った複数の類義語を用いて単語の意思を記述することを提案する。たとえば、「会議を開く」における「開く」の意思は{開催する, 催す, …}、「本を開く」の意思は{開ける, 広げる, …}のように表す。これは、次の方針で実現させることを試みる。

- i 文脈に依存しない類義語を複数抽出する。
- ii 文脈に合うように、抽出した類義語に優先度を与え、順序集合を構成する。

上記iの類義語抽出方法には、ソーラスとコーパスを用いる方法があるが、我々は後者の方法を用いる。これは、コーパスという事例を用いることにより、新しい意味に対応する類義語の抽出と、既存のソーラスでは捉えにくい単語間の関係¹を捉えることを期待するためである。

以下、次節では、上記iの類義語抽出方法について述べ、2節で、上記iiの優先度の算出方法について述べる。

2.1 類義語の抽出

我々は、係り受け関係を用いて単語間の類似度を算出する方法を考案した。これは「ある名詞と同じ格関係を持つ動詞どうしは、意味的に類似している」という仮説に基づいている。

まず、EDR 日本語共起辞書から「NをV」の関係にあるもの（『を』格）を取り出し、以下の式(1)を用いて単語間の類似度を算出した。

$$\text{sim}(w_1, w_2) = \frac{N}{\sqrt{N_1} \sqrt{N_2}} \times \sum (I(w_1, w) + I(w_2, w)) \quad (1)$$

¹ たとえば、既存のソーラスでは「鳥」と「飛行機」の関係を「生物と無生物」という観点からしか捉えておらず、「飛ぶ」という観点から見たときの類似した関係を捉えることができない[9]。

ここで、 N, N_1, N_2 はそれぞれ、 w_1 と w_2 に共通して係り受け関係を持つ単語数、 w_1 と係り受け関係にある単語数、 w_2 と係り受け関係にある単語数である。 w は、 w_1 と w_2 の共通して係り受け関係にある単語である。

また、 $I(w_1, w_2)$ は、 w_1 と w_2 の相互情報量²[13]であり、次式で表す。

$$I(w_1, w_2) = \log \frac{P(w_1, w_2)}{P(w_1)P(w_2)} \quad (2)$$

表1は、式(1)を用いて、動詞「開く」と類似度の高い動詞上位5語を提示したものである。

w_1	w_2	$sim(w_1, w_2)$
開	開催	10.66
開	催	4.99
開	開け	3.71
開	閉じ	3.52
開	つく	2.67

結果は、人間の直感と合致した単語が上位に現れていると思われる。

この方法の有効性を検証するため、ベクトル間の余弦（ベクトルのなす角）を用い、単語間の類似度を算出した結果と比較した。ベクトル間の余弦を用いた方法は、単語 w_1 とその共起単語、単語 w_2 とその共起単語がとる相互情報量が類似しているほど、 w_1 と w_2 が類似しているという仮定に基づき、式(3)でその類似度を算出する。

$$sim(w_1, w_2) = \frac{\vec{w}_1 \cdot \vec{w}_2}{\|\vec{w}_1\| \|\vec{w}_2\|} \quad (3)$$

\vec{w}_i は単語 w_i をベクトル化したものであり、次式で表す。

$$\vec{w}_i = (I(w_i, w_1), I(w_i, w_2), \dots, I(w_i, w_{n-1}), I(w_i, w_n))$$

また、 $\|\vec{w}_i\|$ は、 \vec{w}_i の大きさ（ノルム）であり、 $\vec{w}_1 \cdot \vec{w}_2$ は、 \vec{w}_1 と \vec{w}_2 の内積を表す。

表2は、式(3)を用いて、動詞「開く」と類似度の高い動詞上位5語を提示したものである。ただし、計算量を削減させるため、 w はコーパス

中、出現回数18以上の修飾語となる自立語³、 w_2 はコーパス中、出現回数39以上の動詞⁴とした。

表2：「開」と類似度が高い動詞

w_1	w_2	$sim(w_1, w_2)$
開	い	0.69
開	し	0.67
開	行	0.64
開	出	0.62
開	つ	0.60

表1と表2を比較すると、表1の方がより人間の直感に近いと言え、係り受け関係が単語間の類似性を捉えるうえで、有効な手がかりになることがわかる。したがって、式(1)により算出される値の高い単語を順に類義語とみなし⁵、単語の意味を記述することを試みた。

2.2 順序集合の構成

前節の係り受け関係を用いる方法により抽出された類義語に対する文脈への優先度を計算し、類義語の順序集合を構成する方法について述べる。

たとえば、単語 w からなる文 $w_1 \dots w_k \dots w_n$ があり、 w_k の類義語として抽出された語が $S_1, \dots, S_j, \dots, S_m$ であったとする。このとき、基本的には式(4)により、それぞれの類義語の優先度を計算するが、 w_k と係り受け関係にある単語との相互情報量は5倍する。これは、 w_k と係り受けにある単語が、文中の他の単語よりも強い関係にあると思われるためである。

$$Ps_j = \sum_{i=1}^{n(i=k)} I(w_i, S_j) \quad (4)$$

この式で得られた各優先度を比較し、類義語の順序を決定する。

3. 実験

実験は、EDR日本語コーパス中の「を開く」を含む672文に対して行った。類義語は、前章2節の方法で上位30語を抽出した。

以下は実験結果の一部であり、ここでは、獲得された順序集合の上位5語を提示している。

³ 自立語の平均出現回数17.4を基準とした。

⁴ 動詞の平均出現回数38.2を基準とした。

⁵ 一般に反義語と認識される語も含まれているが、ここではこれを類義語の特殊な場合であると考え、他の類義語と同等に扱う。

² 本研究では、EDR日本語コーパスから、1文を窓の大きさとし相互情報量を算出した。

- ア) 10日に東商、17日に日商が臨時会員総会を開いて正式決定する。
 {開催, 予定, 催, 発足, 兼ね}
- イ) G・ガリレイが手作りの望遠鏡で見いだした金星の満ち欠け現象などは、地動説への強力な検証を与え、天文学への突破口を開いた。
 {見つけ, あげ, 開け, 行, 見}
- ウ) この時の売り上げを元手に、2DKのマンションを借りて、事務所を開いた。
 {借り, 開設, 設け, 兼ね, 見つけ}
- エ) 雨がひどくなってきたので、ぼくは手に持っていた傘を開いた。
 {広げ, 持, 行, 見, くぐ}
- オ) 1949年の革命後、中国政策の定住政策で農地を開いて農耕生活をするようになった。
 {切り開, 借り, し, する, 広げ}

ここに示した例から、文脈に沿った類義語と認識できる語が、ほぼ上位に出現していることがわかる。特にイ)のように、「見つける」という単語が文脈によっては「開く」の類義語として認識できる例もあげられ、我々の提案する方法が、既存のソーラスでは捉えにくい単語間の関係を捉えていると言えるだろう。

4. 評価

4.1 クラスタリングによる評価

EDR 日本語コーパスから抽出した「を開く」を含む 672 文から得られた「開く」の意味 672 個をクラスタリングすることにより、同様な意味で使われているもの同士がまとまりをなすかどうかを評価する。

クラスタリングの方法は、CLINK 階層クラスタ分析法を用いた。これは、クラスタ間の距離に基づき、2つのクラスタ R と S を併合して新しいクラスタ T とするという手順によりクラスタをつくった場合、新しいクラスタ T と他のクラスタ U との距離 d_{TU} を

$$d_{TU} = \max(d_{ru}, d_{su}) \quad (5)$$

$$r \in R, s \in S, u \in U$$

により定義する手法である。ただし、r, s, u はクラスタ R, S, U に含まれる対象である。これは、対象 r, s, u における超計算距離に基づいてクラスタ間の距離を定義することである。

各文から得られた類義語の優先度を各値とした 30 次元のベクトルとみなし、クラスタリングを行った。また、クラスタ間の距離 (親近

度) は、ベクトル間の余弦により算出した。(式(1)参照)

まず、EDR 日本語コーパスから抽出した 672 文における「開く」の意味を以下の 9 つに分けた。括弧内の数字はその意味に該当するとと思われる文の数である。

1. (催し物を) 開く: 会議, 個展, etc. (461)
2. (関係を) 開く: 花, 道, 活路, etc. (76)
3. (店を) 開く: 学校, 事務所, 支店, etc. (51)
4. (物を) 開く: 本, 傘, etc. (42)
5. 口を開く (18)
6. 心を開く (12)
7. 目, 悟りを開く (6)
8. (土地を) 開く: 農地, 道, etc. (5)
9. まゆを開く (1)

672 文それぞれから得られた「開く」の意味が、以上に分類した意味 1~9 に、ほぼまとまりを成すことを期待し、672 個の意味を 9 つのクラスタに分割する。次に、例えば上記の意味 1 が最も多く含まれるクラスタに着目し、再現率 R と適合率 P を用いてクラスタリング結果の評価を行った。再現率とは、同じ意味の文がどの程度まとまりを成すかを表す指標で、式(6)で定義する。また、適合率は、クラスタ中にその意味をもつ文が占める割合であり、式(7)で定義する。

$$R = \frac{\text{着目したクラスタに含まれる、意味nの文数}}{\text{意味nの文数}} \quad (6)$$

$$P = \frac{\text{着目したクラスタに含まれる、意味nの文数}}{\text{着目したクラスタの文数}} \quad (7)$$

表 3: 意味 1~9 と再現率、適合率の関係

	R	P	R'	P'
意味 1	0.48	0.97	0.30	0.69
意味 2	0.55	0.49	0.34	0.11
意味 3	0.33	0.23	0.36	0.08
意味 4	0.33	0.27	0.36	0.06
意味 5	1.00	0.29	0.41	0.03
意味 6	0.75	0.18	0.45	0.03
意味 7	1.00	0.10	0.53	0.01
意味 8	0.60	0.03	0.57	0.01
意味 9	1.00	0.00	1.00	0.01
平均	0.67	0.28	0.48	0.11

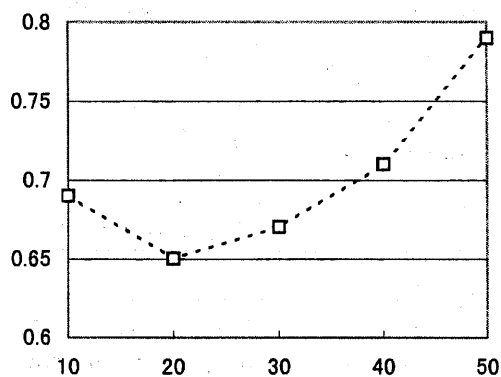
表3に結果を示す。ただし、 R' と P' は、672文を無作為に9つのクラスターに分割したときの平均値である。

平均値と実験から得られた値をそれぞれ比べると、ほとんどの意味1～9において、実験から得られた値の方が良い。このことから、同様な意味で使われているもの同士が、ある程度のまとまりをなすと言え、本手法の有効性がうかがえる。また、この結果から、本研究が提案する意味の記述方法は、クラスタリングによる語義定義に応用できると思われる。

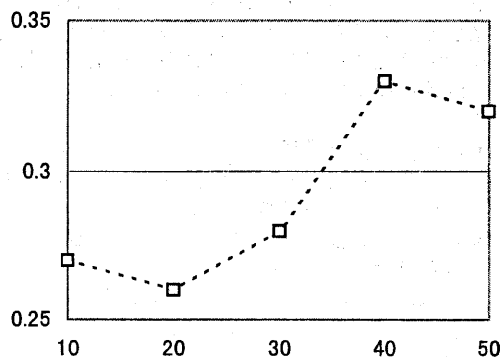
次に、このクラスタリングの手法を用いて、用いる類義語の数と、意味を表すベクトル作成方法の検証を行った実験と結果を示す。

4.1.1 用いる類義語の数の検証

クラスタリング実験から、用いる類義語の数に最適値があるかどうかを検証する。抽出された類義語の上位10語、20語、30語、40語、50語を用いて同様の実験を行い、それぞれの R 、 P の平均値を比較する。



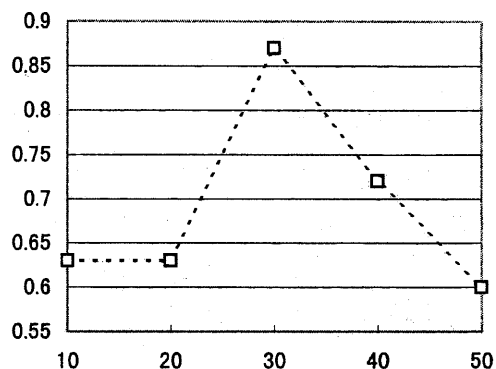
グラフ1：用いた類義語の数と再現率Rの関係



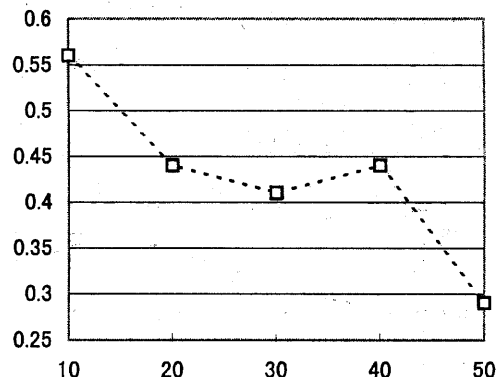
グラフ2：用いた類義語の数と適合率Pの関係

グラフ1、グラフ2に結果を示す。x軸は、用いた類義語の数を、y軸は、 R 、 P それぞれの平均値を表している。

グラフ1、グラフ2は、ともに右上がりの傾向を示しており、用いる類義語の数に比例して、再現率、適合率の値が高くなるように思われる。しかし、意味5～9は出現回数が少なく、スパースデータとしての問題を含んでいるため、意味1～4の文だけで同様の実験を行った。結果をグラフ3、グラフ4に示す。



グラフ3：用いた類義語の数と再現率Rの関係



グラフ4：用いた類義語の数と適合率Pの関係

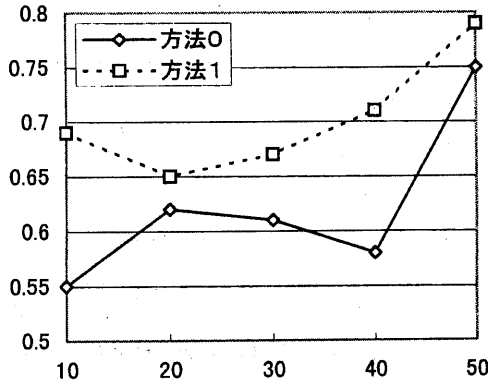
スパースデータを考慮すると、用いる類義語の数には、単語の意味をうまく分割するための最適値があると、この結果から判断できる。グラフ3からは、最も再現率の値が良くなる類義語の数が、20～40の間にあることが予想され、グラフ4からは、用いる類義語の数10の近辺に適合率が最も良くなる値が存在すると思われる。ただし、これらの値は動詞「開く」に依存する値であり、再現率、適合率の値が最も高くなる類義語の数は、単語によって異なるだろう。

4.1.2 意味を表すベクトル作成方法の検証

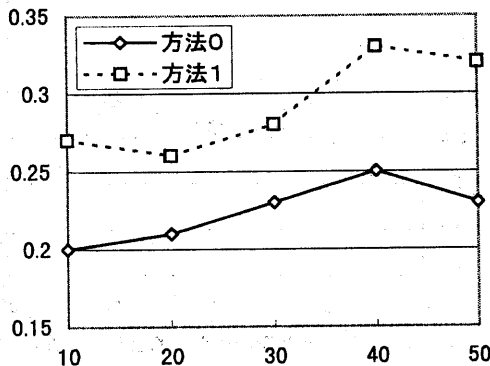
これまでのクラスタリング実験には、「を開く」と係り受け関係にある単語との相互情報量を5倍して作成したベクトルを用いた。本項では、この方法の有効性を検証する。

係り受け関係にある単語との相互情報量を5倍してベクトルを作成する方法（方法1）の他に、係り受け関係にある単語も文中の他の単語と同等に扱う方法（方法0）について、前項と同様の実験を行った。

その結果をグラフ5とグラフ6に示す。



グラフ5：用いた類義語の数と再現率Rの関係



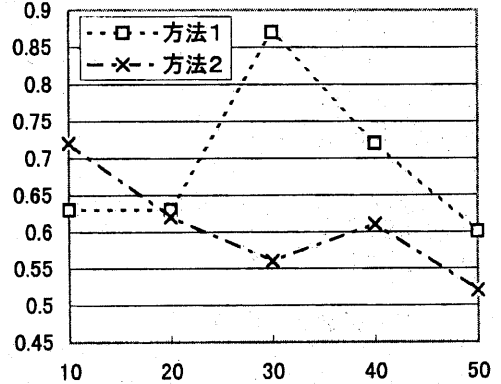
グラフ6：用いた類義語の数と適合率Pの関係

方法0と方法1は同じ傾向を示しているが、方法1の方が再現率、適合率ともに良い結果が得られており、係り受け関係が他の文中の単語よりも強い関係を持つことが、この2つのグラフからも言える。

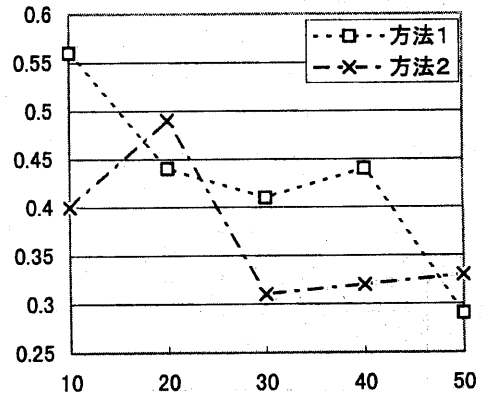
次に、係り受け関係がない単語の影響を考慮するために、係り受け関係の単語のみを用いてベクトルを作成し（方法2）、同様の実験を行っ

た。ただし、スパースデータを考慮し、出現回数のない意味5～9を持つ文については除外し、意味1～4を持つ文を用いて実験した。

その結果をグラフ7、グラフ8に示す。



グラフ7：用いた類義語の数と再現率Rの関係



グラフ8：用いた類義語の数と適合率Pの関係

この2つのグラフから、方法2で得られた結果は、方法1の結果と比べ、全般的に値が低くなっていることがわかる。このことから、係り受け関係がない単語も、単語の意味決定、獲得に影響を与えていると言えるだろう。

4.2 人手による評価

EDR日本語コーパスから抽出した101文の結果に対し、心理実験を行った。この101文は、前節に示した意味1～9を含むように選出し、被験者は大学生10人を対象とした。

問題は、文とそこから獲得された全30語の順序集合のうち上位5つの類義語を前章で示したように提示し、問題文中の「開く」と最も意味的に類似している、または最も対極の意味をな

していると思われる語を、その5語の中から選出させた。ただし、どの語も、類義語とも反義語ともみなすことができないと判断された場合は、選出しなくて良いとした。

その結果、個人差はあるが平均して全問題文の82%に対し、提示した5語のうちに類義語もしくは反義語として認識できる語が存在すると判断された。このことから、本手法により、文脈に沿った類義語の順序集合がほぼ形成されていると言える。

各意味に対する結果を表4に示す。

表4：人手による評価実験の結果(%)

意味1	96	意味6	46
意味2	87	意味7	70
意味3	89	意味8	76
意味4	74	意味9	20
意味5	92	平均	82

各意味と対応する類義語が上位に現れる率に差があることが、この表からわかる。特に、慣用的な表現である意味6、意味9における率が低い。慣用的な表現は、その意味を1語で表すことが困難であるため、その類義語の抽出も困難であることが、理由としてあげられる。しかし、意味5は慣用的な意味「口を開く」であるにも関わらず、高い値が得られている。これは、提示された5語のうちに「閉ざす」が上位に存在し、反対の意味の「口を開ざす」を形成するために、値が高くなったと思われる。

また、具体的な意味で用いられているにも関わらず、意味4の値は平均値よりも低くなっている。意味4では、「を」格を介して係り受けの関係にある単語が、類義語の提示に特に影響を与えると考えられる。したがって、係り受け関係にある単語と類義語の共起関係がスパースであれば、文脈に沿った類義語が抽出される可能性が特に低くなると言える。

たとえば、以下の文のように、「開く」と係り受けの関係にある単語「船底」が「開ける」などの類義語と共起関係にない場合には、文脈に沿った類義語の抽出は困難となる。更に、文中の他の単語と「開ける」などの類義語との共起関係も乏しい場合、文脈に沿った類義語の抽出は難しい。

この工法は、埋立材料をダンプカーやベルトコンベヤで水ぎわ線まで運び、海上では土運船を使用するもので、土運船は埋立予定地で船底を開いて土砂を海中に投下する。

{通, 作, する, 行, 設け}

次に、人手による評価の一部を以下に示す。太字は、括弧内の人数の被験者が最も文脈に沿った類義語であると判断した語である。

INF原則合意の勢に乗って、戦略核と宇宙の問題解決への突破口を開こうとしている。

{見つけ, 開け, あげ, し, 見} (5人)

2カ月に1回ほどのペースで勉強会を開くとともに、年1回親ばく会を開いて団結を高める。

{前, 通じ, つく, 通, 見} (3人)

こうした動きについて日本側関係者は、タイ企業により広い事業参画の道を開くよう求めるためのキャンペーン、と見ている。

{切り開, 閉ざ, 通じ, 開け, 広げ} (2人)

一般にシソーラスなどでは類義語として認識されにくい語が、本手法により提示され、人間により類義語として認識されていることが、この結果からわかる。このことから、我々の提案する方法を用いれば、既存のシソーラスでは捉えにくい単語間の関係を捉えることができると言える。

4.3 考察

人手により、文脈に沿った類義語もしくは反義語が上位5語までに出現する率は、82%であることが検証された。その率は意味によって異なり、類義語の抽出が難しい慣用的な意味における値が低くなっているが、クラスタリングによる再現率の値は比較的高いことから、本手法の有効性が示されている。また、既存のシソーラスでは捉えにくい単語間の関係を、本手法により捉えられることも示した。

クラスタリングの結果も、無作為に9つに分けたときの平均値よりも高い値が出ており、本手法の有効性がうかがえる。また、単語の意味獲得に用いる類義語の数には適切な値があること、係り受け関係にある単語は文中の他の単語よりも強い関係にあること、係り受け関係にない文中の他の語も意味の獲得に影響を与えていることが、クラスタリング実験により検証できた。

しかし、クラスタリング実験によって得られた再現率Rと適合率Pの値は、それほど高いとは言えない。これは、以下の3つの理由によると考えられる。

- ①「開く」と係り受けの関係にない単語との相互情報量がノイズとなる。
- ②他の格で相互情報量が高く、これがノイズとなる。
- ③「開く」と係り受けの関係にある単語と類義語が共起関係になく、相互情報量をとることができない。

本手法では、1文を単位とし、係り受け関係にない他の単語全てとの相互情報量を計算しているが、「開く」の意味を決定する際に影響する単語は、文中の単語全てではないというのが①の理由である。

次に②の理由であるが、たとえば「教室を開く」という文中の「開く」の類義語として「通」が抽出される。我々の提案する類義語抽出方法の性質から、例の「通」は「を」格をとる「通る」であり、「を」格をとらない「通う」ではない。しかし、本研究では、相互情報量計算用の共起辞書は、EDR日本語コーパスの表記のみから作成したため、「通」が「通う」と「通る」のどちらの語幹であるかを判別できない。つまり「教室」は「に」格を介した「通う」と共起回数が多いために、「通」との相互情報量を引き上げてしまう。

③は、データのスパース性の問題である。これは、コーパスなどの事例を基にした知識獲得のうえでは必ず起こりうる問題であり、いかに共起回数の少ない事例を他の事例から収集するかが、今後の課題となるだろう。

5. おわりに

本稿は、辞書やシソーラスを用いずに、より多くの観点から単語間の関係を捉え、意味を文脈に沿った複数の類義語で記述する方法を提案した。また、その有効性をクラスタリングと人手による実験で示した。

本手法は、他の品詞についての有効性も検証することにより、川村ら[9]が提唱する「分類観点によって変化する単語間の関係を捉えることができるシソーラス」の構築に応用できると考えている。

また、クラスタリング精度を向上させることによって、単語の語義定義に応用できると思われる。

参考文献

- [1] Wilks, Y. Fass, D. Guo, C. MacDonald, J. Platre, T. and Slator, B. : Providing Machine Tractable Dictionary Tools, In Pustejovsky, J.(ed.), Semantics and the Lexicon, pp.341-401, Kluwer Academic pub(1993)
- [2] 丹羽 芳樹, 新田 義彦: 単語ベクトルを用いた多義語の意味推定-共起ベクトルと定義距離ベクトルの比較-, 情報処理学会研究報告, 94-NL-102, pp.49-56(1994)
- [3] Hiro, K. Wu, H and Frugori, T. : Word-Sense Disambiguation with a Corpus-Based Semantic Network, Journal of Quantitative Linguistics, Vol.3, No.3, pp.244-251(1996)
- [4] Niwa, Y and Nitta, Y : Co-occurrence Vectors from Corpora vs. Distance Vectors from Dictionaries, In Proceedings of the 15th International Conference on Computational Linguistics (CLING-94), pp.304-309(1994).
- [5] Schutze, H : Word sense disambiguation with sublexical representations, In Workshop Notes, Statistically-Based NLP Techniques, AAAI(1992)
- [6] 福本 文代, 辻井 潤一: コーパスに基づく動詞の多義性解消, 電子通信学会技術研究報告, NLC94-24, pp.15-22(1994)
- [7] 平岡 冠二, 松本 裕治: 共起情報を用いた多義動詞の類別と名詞のクラスタリング, 言語処理学会第1回年次大会論文集, pp.149-152(1995)
- [8] 内元 清貴, 宇津呂 武仁, 長尾 眞: 動詞の語彙的知識獲得における類義語の用例を用いた多義性の類別, 情報処理学会研究報告, NL-94-101, pp.105-112(1994)
- [9] 川村 和美, 片桐 康裕, 宮崎 正弘: 語を種々の観点から分類した多次元シソーラス, 電子通信学会技術研究報告, NLC94-48, pp.33-40(1995)
- [10] Calzolari, N. and Bindi R. Acquisition of Lexical Information from a Large Textual Italian Corpus., In Proc. of the 13th International Conference on Computational Linguistics, pp.54-59, 3 (1990)
- [11] Hindle, D : Noun classification from predicate-argument structures., In 28th Annual Meeting of the Association for Computational Linguistics, pp.268-275, Pittsburgh, Pennsylvania. Association for Computational Linguistics, Morristown, New Jersey(1990)
- [12] Zernik, U. : Lexical acquisition: Exploiting online resource to build a lexicon, Uri Zernik(Ed.), pp.91-112, London, Lawrence Erlbaum Associates.(1991)
- [13] Church, K.W. and Hanks, P : Word association Norms, Mutual Information, and Lexicography, Computational Linguistics, Vol.6, No.1, pp.22-29(1990)