人工知能学会 インタラクティブ
情報アクセスと可視化マイニング研究会(第6回)
SIG-AM-06-05

利用者の状況に応じた用語解説抽出システムの提案とその実現に向けた検討

Proposal and Consideration for Extraction of Term Explanation Depending on Context

本間 康允[1*]　渋木 英潔[2]　森 辰則[2]
Kosuke Homma[1]　Hideyuki Shibuki[2]　Tatsunori Mori[2]

[1] 横浜国立大学大学院環境情報学府
[1] Graduate School of Environment and Information Sciences, Yokohama National University
[2] 横浜国立大学大学院環境情報研究院
[2] Faculty of Environment and Information Sciences, Yokohama National University

Abstract: In this paper, we propose and discuss extraction of term explanations depending on context. When we read some documents, we encounter some unknown terms. Although there are various explanations of a term, it is important to give users good explanations suitable for their context for understanding the unknown terms. We pay attention to brace expressions which appears frequently, and study those expressions in explanation of them. We also discuss a system that extracts explanations for terms from the Web.

1　はじめに

我々が各種文書を読み進める際、未知の用語に遭遇する機会は少なくない。我々は未知の用語に遭遇したとき、辞書で調べる、Web で調べるといった手段で解決を図る。しかし辞書に存在するのは、未知の用語に対する定義文が主であり、それが常に文書を読み進めるときの助けになるとは限らない。一方で、Web 上には各種用語に対して定義文は勿論、様々な言及が存在し、用語解説として利用出来るものも多い。近年の Web の発展により、我々が未知の用語に遭遇した際、検索エンジンは用語に対する言及を容易に調査することを可能にした。しかし、それらの言及の数は膨大で、様々な文脈で述べられている。更に言及の内容も多種多様で、用語の定義を述べている言及、用語の別名を述べる言及、用語が指し示す事物の具体例を述べる言及などが存在する。それらに対して、利用者は読み進めている文書に適した言及を見つけ出す必要がある。例えば「前駆体」とは三省堂大辞林によると「一連の生化学的反応過程の中で着目したある物質よりも前の段階にあって、一ないし数段階の反応によってその物質に変わりうる物質。」と説明される。しかし、文書中で「ダイオキシンの前駆体」が現れた際に、上記説明では「前駆体」の定義しか分からない。一方で、Web 上の文書には「ダイオキシンの前駆体」に対して「クロロフェノール、クロロベンゼンなど」「T3CB」といったダイオキシンに限定した「前駆体」への言及が存在する。利用者はその時々に応じて、適した言及を選択する。しかし利用者が保有する知識と読み進めている文書によって、利用者に適した言及は変化するため、用語と利用者の状況に応じた言及は一対一対応で決めることが難しいと考えられる。

そこで本稿では、未知の用語の理解支援を目的として、Web 上に存在する用語に対する言及を用語解説とみなし、それらを抽出し、整理して提示するシステムを検討する。また、その実現に向けての検討を行う。用語解説は、定義を述べる用語解説だけでなく、書き手によって注釈づけられた補足的な用語解説も存在し、それは括弧表現で現れるときもあれば、文書の最後に用語に対する注釈を記述しているときもある。我々は、書き手によって注釈付けられた記述が未知の用語への理解を深めるのに有効であると考えた。本稿では、用語解説が現れる表現の中でも、注釈付けを行う典型的な表現として、括弧表現による注釈記述に注目する。しかし、括弧表現による注釈記述もまた様々であり、その現れ方の調査を行う必要がある。

本稿では想定するシステムの実現のために、まず括弧表現による注釈記述を調査し、その現れ方と用語に対する注釈記述の役割について分析を行う。次に、想

*連絡先： 横浜国立大学大学院環境情報学府
〒 240-8501 神奈川県横浜市保土ケ谷区常盤台 79-1
E-mail: ammoh.k@gmail.com

定するシステムの概略を示し、上記の調査と分析に基づいて、想定するシステムに必要な処理と本稿で実験を行う処理を述べる。その後実験を行い、その結果について考察を行う。

2 関連研究

Web 上から用語解説を抽出する研究は、これまでにも多数行われている。

土橋ら [1] の WWW 検索エンジンを利用した用語解説文抽出システムでは、「X とは Y である」といった説明文抽出テンプレートを複数作成し、検索エンジン Google を用いて得られた文書群に対して適用することで、用語説明文の獲得を試みている。特徴的な表現を利用して用語に対する補足を得ようとする点では、本稿で述べるシステムと同じだが、利用者が読み進めている文書には注目していない点に違いがある。

また Web から用語に関する説明情報を収集して事典的コンテンツを生成する研究に、藤井ら [3] の Web マイニングによる事典的コンテンツの構築と多様なアクセス手法がある。藤井らは Web の事典的利用を目的とした検索システム CYCLONE を提案しており、獲得した説明情報を分野で分類することで多義語の説明の区別や、用語の関連語を視覚的に提示することを可能にしている。本稿で述べるシステムと非常に似ているため、CYCLONE については 4.3 節にて詳細に述べることとする。

括弧表現による注釈記述に関する研究としては、中山ら [6] の括弧表現の抽出・分類に関する研究がある。中山らは丸括弧や鉤括弧といった括弧表現の用法をそれぞれ 16 種類と 3 種類に分類し、用語に対する括弧表現の位置や用法ごとの特徴によって自動分類を試みている。

括弧表現内に存在する多様な用法の中でも、言い換え表現に着目した研究には、岡崎ら [7] の言い換え可能な括弧表現の抽出法がある。岡崎らは新聞記事を対象に括弧表現を言い換え可能性の観点から分類を行っており、その中で、括弧表現による注釈記述には言い換えが成立しない関係が存在し、文脈に応じて多様な注釈記述が見られると述べている。本稿では、中山ら [6] 同様に括弧表現の現れ方や用語との関係を言い換えに限らず調査する。また特に括弧表現内の記述に重点を置いて考察を行う。

3 括弧表現の調査・分析

本節では、用語への注釈付けとして典型的な括弧表現に着目し、用語とそれに付随する括弧表現による注釈記述の現れ方について調査を行う。

3.1 調査対象

括弧表現内に現れる内容は多種多様であり、括弧表現の抽出と分類を行った中山ら [6] の研究では頭文字、場所、所属といった 16 種類の用法で丸括弧を分類した。また鉤括弧については会話、強調、題目の 3 つに分類している。本稿で我々が検討するシステムの目的は用語に対する用語解説を抽出して、整理して提示することである。そのため、用法が限られる鉤括弧については触れないこととし、括弧表現の中でも用法の多い**丸括弧表現**に限定して調査を行うことにする。

中山ら [6] の研究からも、また我々が持つ括弧表現に対する経験的な知見に拠っても、括弧表現の使われ方は様々であることは明らかである。またその使われ方は話題によって異なることが推測され、使われ方それぞれの数も話題によって変わると考えられる。

上記の考察から、特定の話題について記述している Web 文書を話題毎に数個集めた文書群を調査対象とし、そこに現れる括弧表現とそれにより説明が加えられる語や句に対して調査を行う。調査を行った Web 文書を表 1 に示す。以降に現れる文章の例は、特に記載がない限り、これらが引用元である。

表 1: 調査対象文書

国立環境研究所 地球温暖化第一部「地球温暖化とは？」: http://www.nies.go.jp/escience/ondanka/ondanka01/index.html
国立環境研究所 オゾン層の破壊-過去・現在・未来- : http://www.nies.go.jp/escience/ozone/index.html
レアメタルリサイクル技術 - 環境技術解説｜環境展望台 : http://tenbou.nies.go.jp/science/description/detail.php?id=62
日本年金機構 年金について 年金制度全般 : http://www.nenkin.go.jp/n/www/service/index.jsp
厚生労働省年金局 年金財政ホームページ : http://www.mhlw.go.jp/topics/nenkin/zaisei/
あなたの企業年金、お忘れではありませんか?—企業年金連合会 : http://www.pfa.or.jp/nenkin/callcenter/
わかりやすい政治課題解説 法律-日本政治.com : http://nihonseiji.com/policy

3.2 括弧表現の現れ方と機能

ここでは具体例を交えて、括弧表現の現れ方と機能を分析する。調査対象の Web 文書群に存在する括弧表

現の例を表 2 に示す。　括弧表現は書き手による注釈

表 2: 括弧表現の例

塩素原子をまったく含まない「ハイドロフルオロカーボン（**HFC**）」で、第 2 世代の代替フロンとよばれています。
これらを燃やすとダイオキシンの生成を助ける塩素が発生します。これが、ダイオキシンの 前駆体（**クロロフェノール、クロロベンゼンなど**）と結びついてダイオキシンが合成されるのです。
横浜金属（**株**）
これに対し日本自由法曹団は「(**公益及び公の秩序は**)『公益の福祉』とは異なり、抽象的な価値を...

記述であるため、語や句が現れた後に現れることが多い。しかし、前に現れて語や句に対して説明をつける場合も存在する。本稿では括弧表現内の記述と関係を持つ語や句を**用語**と定義して扱うこととする。また文末に現れて、注釈を加える場合もあるが、本稿ではこれは**対象外**とし、用語の前後に現れる括弧表現について調査を行う。

次に括弧表現内の記述について詳しく述べる。ハイドロフルオロカーボンの例では、括弧表現内の記述は直前の用語に対しての言い換え表現を与えている。これは中山ら [6] の研究の分類では、外来語の頭文字に当たる用法である。次に前駆体の例を見ると、括弧表現内の記述は、前駆体として考えられる化学物質のうちダイオキシンの前駆体となる化学物質だけに限定して具体的な例を与えている。そのためこの記述は、前駆体が現れる他の文章において、常に前駆体を正しく説明するとは限らない。つまり、この括弧表現内の記述は、それが含まれる文書の文脈に依存していることが分かる。それに対してハイドロフルオロカーボンの括弧表現内の記述である HFC は単なる言い換え表現であるため、どういった文脈でもハイドロフルオロカーボンの言い換えとして機能する。

以上のように、括弧表現内には**補足説明**となる記述が存在することが多いが、その一方で、括弧表現内の記述には定型表現が入る場合もある。横浜金属の例がそれに当たる。他に（有）や（独）、電話番号などが存在するが、このような定型表現は補足説明ではない。

以上の観察から、括弧表現内の記述について次のことが言える。

1. 括弧表現内の記述には、補足説明とそれ以外が存在する
2. 用語と補足説明の関係は複数の種類が存在する
3. 補足説明の中には文脈に依存したものが存在する

次の小節で、2 について調査対象の文書群に見られた用語と補足説明の関係を述べる。3 については、5 節において詳細に述べる。

3.3　用語と括弧表現内の記述の関係

調査対象の文書群を観察した結果、用語と括弧表現内の記述が持つ関係は少なくとも次のように分類された。そのうち補足説明になるものは以下の通りである。

1. **同一**

   補足説明が用語に対して同一の事物を指す関係を持つことを、本稿では**同一**と呼ぶことにする。

   塩素原子の例では括弧表現内の補足説明には

フロンは安定な物質で対流圏では分解されません。しかし、成層圏にまで達すると、強い紫外線によって分解され、塩素原子（**Cl**）を放出します。
一般永住者（**一定の要件を満たして永住許可申請をし、許可され、日本国に永住している外国人のこと**）には参政権が与えられていない。

   言い換え表現が与えられ、一般永住者の例では括弧表現内の補足説明には用語の定義が与えられている。

2. **属性値**

   本稿では、補足説明が用語のある属性に値を与えているような関係を**属性値**と呼ぶことにする。岡崎ら [7] の研究においても、これらの括弧表現に対して同様の分析を行っており、これらを属性と呼んでいる。　豊羽鉱山の例では、括弧表現内

レアメタルは、他の鉱物の副産物として産出されることが多く、例えば世界最大のインジウム鉱山であった日本の 豊羽鉱山（**北海道**）…
Ta コンデンサ（**携帯電話、デジタルカメラ、パソコン**）

   の補足説明には用語の存在する場所を示す北海道が与えられている。また、Ta コンデンサの例では括弧表現内の補足説明は Ta コンデンサの使用先を与えている。これらは場所や用途といった、用語が持つ属性に対して、値を与えていると考えることが出来る。

3. 例示

以下に示す二つの例では、どちらも括弧表現内の補足説明は用語に対して具体例を与えている。つまり用語によって定まる部分集合に対して、補足説明はその集合の要素または下位概念となる部分集合を列挙する関係を持っていることが分かる。本稿ではこの関係を**例示**と呼ぶことにする。

非化石燃料や新エネルギー技術（**たとえば、太陽エネルギーやバイオマスエネルギー**）を利用する社会...

オゾン層破壊と関連のある物質（**硝酸、二酸化窒素、亜酸化窒素、硝酸塩素、メタン、水蒸気、エアロゾルなど**）の濃度が測定されました。

4. 限定

以下の例では、括弧表現内の補足説明は、用語によって定まる集合の中でも特に文章中に合致する集合内の要素や下位概念となる部分集合を与える関係になっている。例示と似ているが、こちらは文脈によって一つに定まっており、列挙する関係を持っていない。本稿ではこのような関係を**限定**と呼ぶことにする。

噴射剤については特殊な用途を除いて炭化水素（**液化天然ガス：LPG**）

回転の状態が変わるとそのエネルギーの差を電波（**ミリ波**）として放出します。

5. その他

以下の例では、補足説明は用語に対して付加的な説明、非制限的な説明を与えている。どちらも前述したような同一、属性値、例示、限定のうち、どれにも当てはまらない。このような非制限的な説明を与える補足説明の関係を**その他**と本稿では呼ぶことにする。

表 1 に示す 31 鉱種（**ただし、レアアース（希土類）は 17 鉱種を総括して 1 鉱種とする**）をレアメタルと規定しています。

こうした緊張関係の中、権利の保護（**＝新たな創造への投資のインセンティブ**）と 利用の促進（**＝創造の果実の社会による享受や再創造へ**

補足説明以外では次のような記述が見られた。

1. 定型表現

以下の例は、日常生活において広く用いられる括弧表現である。括弧表現内の記述が用語に対して持つ関係は、会社の種類という属性に対して値を与えているため属性値が適していると考えられる。しかし広く用いられる表現であるため特に区別して、**定型表現**と本稿では呼び、補足説明には含めない。

（独）石油天然ガス・金属鉱物資源機構希少金属備蓄部

2. その他

括弧表現内の記述が補足説明ではなく、定型表現でもない場合**その他**と呼ぶことにした。以下の例では、用語に相当する部分が曖昧なため下線は付けない。

朝日新聞 DIGITAL「（**ネット選挙がわからん！**）政党や候補者の戦術、どう変わるんじゃ」2013.04.23.

上記の補足説明が用語に対して持つ関係の分析に基づいて、調査対象の文書群の中に現れる括弧表現内の記述について人手で分類を行った。その結果を表 3 に示す。なお「対象外」は 3.2 節で述べたように、用語の前後に現れない場合である。

表 3: 現れ方の分類

補足説明	同一	168
	属性値	60
	例示	15
	限定	32
	その他	13
補足説明以外	定型表現	19
	その他	3
対象外		50
総数		360

この結果から、括弧表現内の補足説明の中でも同一が非常に多く存在することが分かる。これは同一と分類する基準が、用語が指している事物と同じであるとしたため、用語の頭文字の省略や言い換え表現、定義文を含めていることが原因だと考えられる。従って、同一に関しては更に細分化した分類を行う余地があると思われる。また括弧表現内の補足説明は括弧表現全体の 80%となり、これは括弧表現内に補足説明が十分に存在することを示している。このことは括弧表現内から様々な補足説明が抽出できることの裏付けになっていると言える。

4　目標とする用語解説抽出システム

3 節では括弧表現について調査と分析を行った。その中で補足説明と分類される記述を、本稿では**用語解説**と呼び、これを整理して提示するシステムを考える。

本節では、本稿の目標とするシステムの概要とその構成について述べる。4.1 節では目標とするシステムのタスク設定を行い、4.2 節では目標とするシステムの実装上の課題について述べる。

4.1　目標とするシステムとそのタスク設定

本稿で目標とするシステムは、利用者が各種文書を読み進めている際に、未知の用語に遭遇する状況を想定する。目標とするシステムの概観図を図 1 に示す。

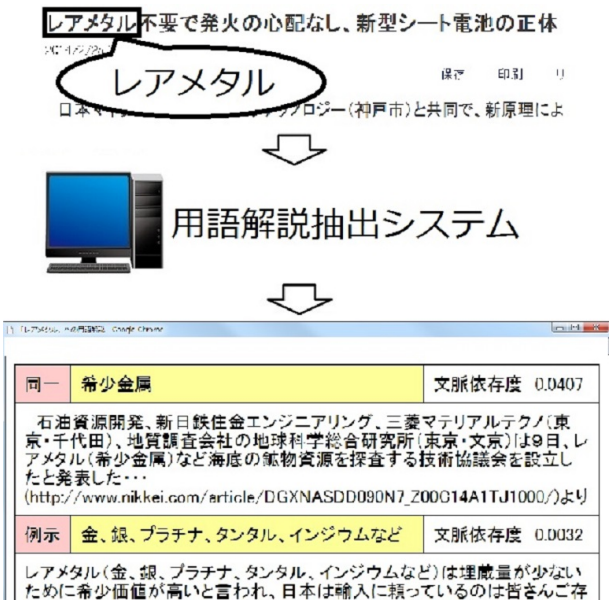

図 1: 想定されるシステムの入出力

目標とするシステムへの入力は次の二つである。

1. 利用者が読み進めている文書
2. その中に現れる利用者が調べたい文字列

ここで文字列は、利用者の指示によって与えられるが、本稿で呼ぶ用語に相当することが期待されるものとする。これらの入力を受けた後、同システムは Web 上から文書群を取得し、そこに存在する用語解説それぞれに処理を行い、図 1 に例示するような結果を出力として返す。各結果に表示される情報を表 4 に示す。

4.2　実装上の課題

目標とするシステムを実装する上での課題について述べる。同システムは利用者から用語（以下、**入力用語**と呼ぶ）と、利用者が読み進めている文書（以下、**入力文書**と呼ぶ）を入力として受け取る。その後、同システムは入力用語と入力文書の情報に基づいてクエリを作成する。それを元にテキスト集合を検索し、文書群を取得する。得られた文書群から用語解説を抽出し、用語解説が持つ役割や文脈依存度で整理を行う。そして整理された結果を利用者に提示する。同システムの実現に必要となる処理として以下のものが挙げられる。

表 4: 出力

他文書に現れる当該用語解説	用語と用語解説の関係
	用語解説が持つ文脈依存度

1. 用語解説が現れる文書の効率的な獲得
2. 用語解説が持つ役割の自動分類
3. 用語解説の文脈依存度による順位付け

1 については現在検討中だが、本稿で行う実験では検索エンジン Google を用いて Web 上に存在する文書群を獲得することとした。2 については、3 節にて用語解説が持つ役割の調査と分析を行った。しかし自動分類のためには、別途特徴表現などの調査が必要である。3 は、同システムが持つ特徴的な処理であり、新規性となりうる部分である。そのため本稿では、3 について 5 節にて順位付け手法の検討を行う。

4.3　CYCLONE との比較

用語への用語解説を集めて提示する点では、本稿で目標とする用語解説抽出システムは藤井ら [3] の CYCLONE によく似ていると言える。CYCLONE は Web から見出し語とその説明文章や関連語を収集し、事典コンテンツを自動構築し、更にコンテンツに対する多様なアクセスを可能にしている。本稿で目標とするシステムは、用語解説を抽出するに当たって、用語へ注釈を与える典型的な表現である括弧表現に注目することで、書き手による咀嚼がされた用語解説も抽出することを試みる。また用語解説が持つ役割をシステム側で分類しておくことで、利用者が求める用語解説を見つけ出しやすいようにしている。この点では同システムと CYCLONE が目指すことは同じであるため、今後参考にし検討する。また、CYCLONE では予め学習した分野モデルや大百科事典を基とした言語モデルなどを総合して、獲得した説明を分野に分類することで多義語の説明を区別する。同システムでは、用語解説が持つ文脈依存度を見ることで文脈依存性の問題に対応するとともに間接的に多義性の解消を試みる。これは用語が述べられる分野に特有の文脈が存在し、多数

の意味を持つ用語はその文脈によって意味が決まると考えたためである。

5　文脈依存度に応じた順位付け手法

目標とするシステムの一部の処理として、本節では文脈依存度に応じた順位付けを行う手法について述べる。

5.1　用語解説の文脈依存度

用語が現れるまでの文脈を明らかにし、用語解説の文脈依存度を判断するのであれば、係り受け解析や意味解析を行うべきだが、本稿では単純に文書内に現れる名詞群を文書の文脈を表現するものとして扱うこととする。

前述の通り、用語解説の中には文脈に依存するものも依存しないものも存在する。文脈依存度は候補の選別を行うための指標であり、解説候補間での相対的な大小が分かれば十分である。

用語解説の文脈依存度の算出手法について、我々は二つの手法を考えた。以下では、その二つの手法について述べる。ただし、文脈依存度の値は用語解説候補の順位付けに使用するので、システムの出力においては文脈依存度の具体的な値よりも候補間の順位付けが重要である。後の評価でもその観点で行っている。

5.2　文脈依存度算出手法

5.2.1　手法 1

用語解説が抽出された文書の文脈が、入力文書の文脈にどれだけ近いかを測る。具体的には、入力文書と、用語解説が抽出された文書の双方を名詞の出現頻度を値とするベクトルとし、それらの cos 類似度をとり、文脈依存度として順位付けを行う。これを手法 1 と呼び、ベースラインとする。

5.2.2　手法 2

手法 1 では、用語解説が抽出された文書の文脈が、入力文書の文脈に近いほど文脈依存度が高くなる。しかし、入力文書に類似する文書において偶然、一般的で他の文脈でも通用する用語解説が現れる状況では、その用語解説に意図しない高い文脈依存度を与えてしまう。これを回避するには用語解説がどのような文脈に現れるかを調べる必要がある。

用語 X についた用語解説 Y、つまり「X(Y)」という表現が現れる文書群の文脈を考えると、典型的には次のような場合が考えられる。

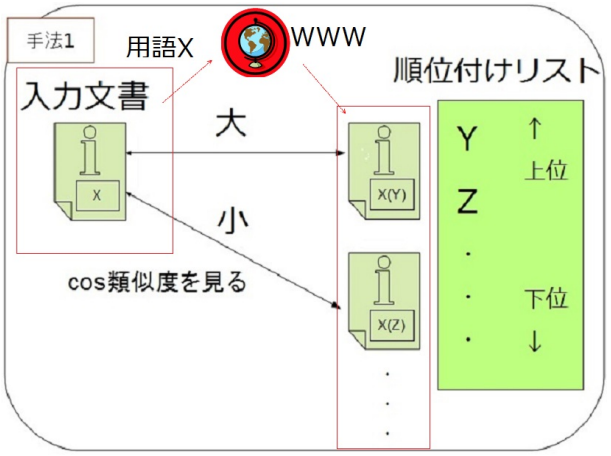

図 2: 手法 1

1. X(Y) が現れる文書群の文脈は多様
2. X(Y) が現れる文書群の文脈に偏りがあり、内容が入力文書に類似する

1 の時、用語解説 Y は文脈に依存しない用語解説と言える。様々な文書でそれぞれ違う文脈の下で現れるため、用語解説 Y は用語 X に対しての一般的な注釈付けと考えられる。2 の時、用語解説 Y は文脈に依存した用語解説と言える。特定の文脈に多く現れる用語解説 Y は、その文脈に依存した注釈付けと考えられるためである。

上記のことから、用語と用語解説の表現 X(Y) が現れる文書群の文脈と入力文書の文脈との類似度を見ることで当該用語解説が持つ文脈依存度を測ることができると我々は考えた。すなわち Y が一般的な用語解説であった場合それは多様な文脈の文書群に現れるので、上記類似度が低くなることを期待する。具体的には、検索エンジンを用いて用語と用語解説の表現 X(Y) について改めて検索し、その表現が現れる文書群を獲得した後、それらを一つの文書と考え、名詞の頻度を値とする文書ベクトルを作る。このベクトルと、入力文書に対応する文書ベクトルの間の cos 類似度を計算し、用語解説の文脈依存度とする。

6　評価実験

本節では、5 節にて述べた用語解説の文脈依存度の計算手法について評価実験を行う。6.2 節で用語解説を抽出するための文書群の獲得について、6.3 節で出力例と出力に対する評価結果を述べる。

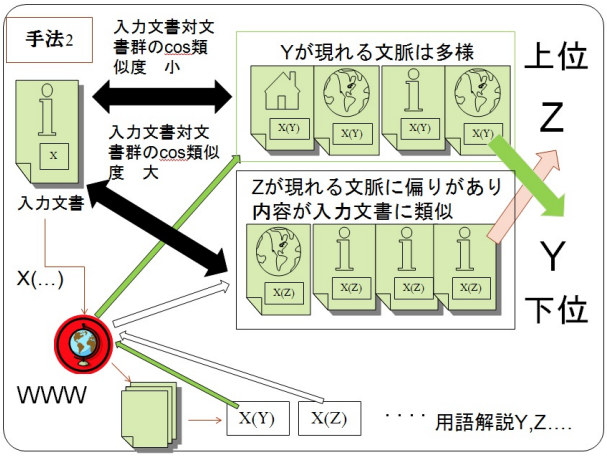

図 3: 手法 2

6.1 入力文書と対象とする用語

本実験で扱う任意に選択した 5 個の用語と、それが含まれる入力文書を表 5 に示す。オペやプレートは多

表 5: 入力文書と対象用語

対象用語	入力文書
オペ	http://www.bloomberg.co.jp/news/123-MHW5NK1A74E901.html
プレート	http://www.jishin.go.jp/main/yosokuchizu/kanto/p13_1_tokyo.htm
前駆体	http://www.kcn.ne.jp/~azuma/QA/IAQ/I026.htm
酵素	http://zeroone.searchnavi.jp/privacy2.html
微生物	http://www.kao.co.jp/pro/noro/how/p1.html

様な文脈で現れる用語である。前駆体、酵素、微生物は現れる文脈が限定されている。オペやプレートの場合に手法 2 で改善されることを期待する。

6.2 文書群の獲得と括弧表現内記述の抽出

ここでは、括弧表現が存在する文書群の獲得について述べる。本実験では、文書群を検索エンジン Google を用いて、Web 上から獲得する。想定されるシステムは、入力用語と入力文書を受け取る。しかし単純に入力用語が存在する文書を検索して得られる文書は非常に多く、その全てから抽出を行うのは現実的ではない。そこで抽出を行う文書の文脈をある程度制限することとした。まず、入力文書の単語頻度で降順のランキングを作成し、その中で上位 2 単語を、検索条件として用語に加えた。つまり、入力用語が X で、入力文書内の単語頻度のランキング上位 2 単語が A,B であった場合、検索エンジンに与えるクエリは「X A B」である。

上記の手順によって獲得した文書群から、括弧表現内記述を抽出する。ここでは、「X（Y）」と記述されている括弧表現のみを対象に抽出を行った。また補足説明以外もそのままで手法 1,2 を適用する。

6.3 実験結果

出力の具体例を表 6,7 に示す。

表 6: 手法 1

入力文章（一部）	用語解説
期待される効果は、マネタリーベースの供給拡大／加速によるデフレ期待の後退＝リフレ期待の台頭。心配される副作用は、国債バブルの助長、財政ファイナンス懸念による「悪い金利上昇」など。当座預金付利が撤廃されると、オペに応じる金融機関が減り、基金残高の積み上げが困難になる。	２月１日－８日
	こていきんりおぺ
	６月 14 日-22 日
	貸付利率 0.1 %
	６月 15 日-22 日
	８日－ 2010 年３月４日
	固定金利方式
	新型オペ
	短期国債の購入
	2010 年３月 31 日をもって完了
	オペレーション
	公開市場操作

表 7: 手法 2

入力文章（一部）	用語解説
期待される効果は、マネタリーベースの供給拡大／加速によるデフレ期待の後退＝リフレ期待の台頭。心配される副作用は、国債バブルの助長、財政ファイナンス懸念による「悪い金利上昇」など。当座預金付利が撤廃されると、オペに応じる金融機関が減り、基金残高の積み上げが困難になる。	公開市場操作
	オペレーション
	短期国債の購入
	２月１日－８日
	固定金利方式
	８日－ 2010 年３月４日
	2010 年３月 31 日をもって完了
	新型オペ
	こていきんりおぺ
	６月 15 日-22 日
	６月 14 日-22 日
	貸付利率 0.1 %

出力について、人手で評価を行った。入力文書の文脈に適合した用語解説が上位に存在することが重要である。著者以外の評価者 A,B の 2 名が別々に、手法 1,2 それぞれの出力について用語解説候補が、入力文書の文脈に適合しており用語への補足説明となると判断したなら○を、入力文書の文脈に適合してないが用語への補足説明となると判断したならば△を、用語への補足説明にもならないと判断したならば×を、抽出された用語解説候補それぞれにつけてもらうこととした。ただし各手法が候補に付与した文脈依存度の高い順に左から右にむけて候補の評価を並べている。左側に○が偏り、右側に×が偏るときの手法の出力が望ましい。

表 8,9 で、各用語の用語解説候補に対する評価結果を示す。

表 8: 評価者 A

用語	←上位　下位→	
前駆体	手法 1	×○××××○
	手法 2	×××○××○
オペ	手法 1	×××△×××△○×△○
	手法 2	○△○××××△×××△
プレート	手法 1	×○○×○×△×××○×△××××
	手法 2	○△○×△××××××○××××○
酵素	手法 1	○△×△×△
	手法 2	△×○×△△
微生物	手法 1	×××○×
	手法 2	×○×××

表 9: 評価者 B

用語	←上位　下位→	
前駆体	手法 1	×○×△△××
	手法 2	△×△○×××
オペ	手法 1	×△×××××○××△
	手法 2	△×○×××××△×××
プレート	手法 1	×○△△○×××××△××××△×
	手法 2	○×△×××××××△○×△××△
酵素	手法 1	○△△△××
	手法 2	△△○△××
微生物	手法 1	××○××
	手法 2	×××○×

7　考察

本節では 6 節で行った実験の結果について考察を行う。手法 1 に対する評価では、オペは下位に○がつくが、それ以外の用語について上位 5 件に一つは○が存在することが分かる。手法 2 ではオペも含めて、どの用語についても○が上位 5 件に一つは存在するようになっている。そのため手法 2 は、少なくとも今回選択した用語に対して、文脈依存度に応じた順位付け手法として安定した働きを見せている。

また△の推移を見ると、評価者 A でのオペ、評価者 B でのプレートといった多様な文脈で現れうる用語の場合は、手法 2 に期待するような文脈依存度の低い用語解説を下位に置く効果が現れている。これは 6.1 節で述べた期待通りだが、多様な文脈で現れうる用語に対してさらなる検証が必要である。

逆に手法 1 では、現れうる文脈が限定されるような用語「微生物」「前駆体」「酵素」について手法 2 より上位に○が来ていた。このことから、入力用語が現れる文脈についてトピック分析を行い、各手法と組み合わせるなどの手法の拡張が考えられる。

8　おわりに

本稿では、用語解説が現れる中でも、書き手による注釈付けが行われる典型的な表現として括弧表現に注目し、括弧表現内の記述の現れ方を調査した。それを受けて、利用者の状況に応じた用語解説抽出システムの提案を行った。また、その実現に向けて提案システムに必要な処理を整理し、用語解説に存在する文脈依存度について実験を行った。

その結果、手法 2 が今回対象とした用語のどれに対しても、上位 5 件のうちに一つは入力文書の文脈に依存した用語解説を含んだ出力を得ることが出来た。対象とする用語を更に増やして評価実験を行い、順位付け手法を検討すること、想定されるシステムの他の処理について検討を行うことが今後の課題である。

参考文献

[1] 土橋 惇一, 荒木 健治: WWW 検索エンジンを利用した用語解説文抽出システム, 情報処理北海道シンポジウム, pp.26-27, 2004

[2] 土橋 惇一, 荒木 健治: WWW 上の定義文における表現特徴を利用した用語説明文抽出のためのテンプレートの自動生成について, 言語処理学会第 11 回年次大会発表論文集, pp.791-794, 2005

[3] 藤井 敦, 伊藤 克亘, 石川 徹也: Web マイニングによる事典的コンテンツの構築と多様なアクセス手法, 電子情報通信学会技術研究報告, Vol.4, pp.31-36, 2004

[4] 藤井 敦, 伊藤 克亘, 石川 徹也: WWW は百科事典として使えるか？, 情報処理学会研究報告 自然言語処理研究会報告, 2002-NL-149, pp.7-14, 2002

[5] 藤井 敦, 石川 徹也: World Wide Web を用いた事典知識情報の抽出と組織化, 電子情報通信学会論文誌, VolJ85-D-, No.2, pp.300-307, 2002

[6] 中山 悟, 森田 和宏, 泓田 正雄, 青江 順一: 括弧表現の抽出・分類に関する研究, 言語処理学会第 16 回年次大会発表論文集, pp.379-382, 2010

[7] 岡崎 直観, 石塚 満: 言い換え可能な括弧表現の抽出法, 言語処理学会第 13 回年次大会発表論文集, pp.911-14, 2007

[8] 岡崎 直観, 石塚 満: 日本語新聞記事からの略語抽出: 人口知能学会全国大会論文集, Vol.21, pp.2G4-4, 2007