

## 専門用語の分野基礎性に関する一考察

内 山 清 子<sup>†1</sup>

本研究では、論文や技術文書などに多く含まれる専門用語について、ある特定分野の基礎的で必要不可欠な用語の度合い（分野基礎性）を判定するための観点と指標について考察する。観点は、その用語を提示する側（教える側）と獲得する側（学ぶ側）の側面があり、両側面から議論する。指標は、優先度（用語を学ぶための順序）、経年推移（専門用語が用いられる年度と頻度）、親密度（分野における専門用語の親しみやすさ）、分野偏り度（下位分野における出現傾向）、語彙構成度（派生専門用語の生産性）、定義明確度（定義付けをする度合い）の6つについて整理した。この観点と指標に基づいて、各種データを用いて分析をし、考察を行った。

### A Study for Determining a Basic Terminology

KIYOKO UCHIYAMA

In this paper, we address an issue for determining a basic terminology in a specific field. A large amount of terminology are included in academic papers and technical documents, however, there are no clear criteria how important or useful for learner such as the student and researcher in other field. We analyze the basic terminology from the point of view of teaching and learning, and 6 criteria; priority, transitional variance, familiarity, degree of bias in subcategories, productivity of technical terms and definition.

#### 1. はじめに

特定の専門分野をこれから学ぼうとする学生や他の分野の研究者が、教科書や Wikipedia のような概要が書かれている書物を読む時に、必ず専門用語が含まれ、その内容を理解する必要がある。その分野を理解するためには、その分野に出現する用語の理解が不可欠で

ある。しかし、専門外の人がある専門用語を知ろうとする時、その用語が、どの程度の重要性を持っているのかが曖昧である。つまり、その専門用語の分野における重要性、基礎性、先端性の度合いがわからないため、どの専門用語に焦点を当てて学ばよいか明確ではない。また、概要を把握したい場合は、全ての用語を網羅する時間がないため、基礎的な要素だけを選んで学びたいものである。その時に、その分野で核となる基礎的な用語がわかっていたら、効率的に学んだり、検索することが可能になる。

そこで、本研究では、専門用語の分野基礎性を定義するために、様々な観点と指標から議論を行う。ここでの分野基礎性とは、その分野における最低限抑えておくべき基礎的で必須であることを示す。分野基礎性が高い用語は、その用語を知らなければ、その専門分野のことを理解することができない、他の専門用語も理解することができない用語とする。一方、分野基礎性が低い用語は、基礎性の高い用語の理解を深めた上で、その知識を利用しなければ理解することができない専門性の高い用語であるとする。当然、分野基礎性が高く専門性が高い語もある。

以下、論文の構成は、本研究に関連する研究との比較と、分野基礎性の観点と指標の整理、具体的なデータを用いて分野基礎性を解析した結果とその結果に基づいた考察を述べる。本研究では、専門用語の抽出は行わず、ある特定分野において専門用語であると認識されている用語を対象とする。

#### 2. 関連研究

##### 2.1 重要語抽出

先行研究で行われてきた重要語抽出<sup>1)</sup>は、与えられた文書や文書集合から、その文書や文書集合に特徴的な語を重要度が高い順に抽出することである。重要度は出現確率に加えて、接続する単語の種類や回数によって重要度を計算することで決定される。本研究における分野基礎性と重要度との関連としては、分野基礎性の高い語は「重要度」も高いはずだが、与えられる文書集合に依存してしまう。たとえば、特定分野の事典や教科書などの文書集合を与えた場合は、「重要度」が高い語の集合の中に、分野基礎性が高い語が含まれている率が高いと予想され、一方で、論文などの文書集合を与えた場合は、分野基礎性が高い語はすでに分野全体に浸透し平均的に使用されるために「重要度」がそれほど高くないと考えられる。

##### 2.2 専門用語の専門性、分野判定

専門用語の専門性を推定する研究としては、技術者がプレゼンテーションの時に専門外の人に対して、専門用語を使用せずに平易な用語に置き換えて、効果的に技術成果を伝えるた

<sup>†1</sup> 国立情報学研究所

めに用語の専門度推定を行ったものがある<sup>2),3)</sup>。この研究では「専門用語ではなく意味の近い平易な用語を用いる」ことを前提としているため、専門用語内での専門度については専門外の人から見て比較的専門的な用語であるか、かなり専門的な用語かの2段階に分けている。

また、専門用語が含まれる文書中における分野の割合に基づいて、用語の分野判定を行う手法がある<sup>4)</sup>。これは対象となる用語が出現する文書を収集し、用語が一般の文書よりも専門分野の文書に偏って用いられる度合いを測定して、専門用語であるかどうかを判定する。

これら二つの研究では、広い範囲における一般語と比較した専門性と、専門分野における用語の分野判定を行っており、専門性という軸における度合いを判定している研究だと考えられる。本研究では、専門性という軸ではなく、分野基礎性という軸における度合いを判定する。専門性と分野基礎性の違いは、専門性は一般語や他の分野との比較における度合いを軸とするが、分野基礎性では同じ分野内における基礎性・必須性を軸として考える点である。

### 2.3 関連用語収集問題

与えられた専門用語からそれに関連する専門用語を自動的に収集する方法が提案されている<sup>5)</sup>。この研究では入力キーワードとなる専門用語と強く関連する専門用語を収集・出力することであり、専門用語辞典を自動的に生成するタスクである。このタスクで行った実験では、対象分野から代表的な書籍を複数用意し、どの書籍にも出現する索引語を専門用語(キーワード)として採用している。このキーワードは本研究における分野基礎性が高い語のイメージに近く、分野において核となる専門用語と位置づけることができる。本研究では、このキーワードの選定が重要なポイントとなる。

## 3. 専門用語の分野基礎性の判定に必要な観点と指標

### 3.1 教える側と学ぶ側の観点

専門用語の分野基礎性を考える場合、その分野を熟知している研究者が教えるという分野の概念を提示する立場と、これからその分野を学びたい学習者として分野の概念を獲得する立場の二つがある。教える側からは、分野について説明しなければならぬ時に考慮すべき点として、まずその分野が他の分野とどう関連しているかという位置づけ、分野全体の概要を記述する時に必要となる最低限知っておいてもらいたい専門用語の選定することが重要となる。この時に選定される用語が分野基礎性の高い用語となる。この用語は、一般的な語で説明できるもので、その用語を使ってさらに詳しい専門用語を説明するような汎用性の高い用語だと考えられる。教える立場で考えると漏れないように、より多くの専門用語を提示

したいと考えるが、一方で学ぶ側・概念を獲得する側からすると、できるだけ効率的に順序良く学べることを望む場合が多い。これらの需要と供給のバランスによって、専門用語の分野基礎性が決定できると考える。

学ぶ立場には複数の段階と、段階に応じた目的と必要な知識や欲しい情報の種類がある程度決まっている。大雑把に以下の4段階に分類した。

- (1) 一般、大学学部生、他の研究分野の研究者(分野の概要を知りたい)  
Wikipedia レベルの情報 コアな情報のみ
- (2) 大学学部生(これからその分野を専門にする学生)  
分野の成り立ちも含めた詳細な概要
- (3) 大学院修士(修士論文テーマ探し)  
最新動向も踏まえた、比較的広い範囲での詳細な情報
- (4) 大学院博士(博士論文テーマ探し)  
過去の詳細な研究成果も含めた、狭い範囲での詳細な情報

(1)が最も分野基礎性が高く、他の分野と比較した場合には専門性が高い、つまり専門性をカバーする範囲が広く、一方で(5)が分野基礎性は低く、より狭い範囲における専門性が高い段階であるとする。(4)(5)の段階にすれば、だいたいその分野の傾向を把握できているので、(4)(5)レベルの分野基礎性の詳細な段階分けはあまり考えない。それよりも、まず初心者(上記の1から3の段階)に対して、どのような専門用語を提示するのかということに焦点を絞る。初期の学習段階では、検索のための適切なキーワード選びからして難しいことである。分野基礎性が明確に示された専門用語を適切なキーワードとして入力することによって、目的に合った情報を得ることが可能である。以下では、分野基礎性を判定するのに関連する指標を整理する。

### 3.2 基礎性判定の指標

#### 3.2.1 優先度

学問を学ぶ時には、学ぶ優先順位がある程度決まってくる。自然言語処理の場合は、形態素解析を学んだ後で、構文解析、構文解析を学ぶといった優先度のことである。こうした学習において共通して初期の段階で優先的に教えられる項目は特に重要で、分野基礎性が最も高い用語として絶対的な尺度であると考えられる。たとえば、複数の教科書に共通する用語や同じ研究分野の大学講義等で複数の先生が共通して初期に教える用語などは優先度が高い語であると考えられる。

### 3.2.2 経年推移度

昔は論文等で頻繁に使われていた用語が年数を経るごとに頻度が減ってきたり、その反対に増えてきたりする用語がある。分野基礎性が高い語は、年数が経っても平均してある一定以上の頻度を保って出現し、分野基礎性が低い語は突然爆発的に使われたとしてもある時期に落ち着いて、以後使われなくなったりする等、出現頻度に安定性がないと考えられる。

### 3.2.3 親密度（頻度）

重要な語は、文章で繰り返し使われる語であり、様々な指標に頻度情報が含まれているのはその理由からである。論文や書籍などで頻繁に使われる用語は重要であるから繰り返しされ、繰り返されることによってその語に親しみを感じ、より使われるようになる。分野基礎性においても同様のことが言え、分野において重要な概念であるため繰り返し出現する用語は、その分野の研究者が親しみ、馴染みのある語として頻繁に用い繰り返される。頻度が高い語は、分野において親密度が高い語であると言える。この傾向から、頻度情報は分野基礎性においても重要な指標であると考えられる。

### 3.2.4 下位分野偏り度

同じ分野でもいくつかの下位カテゴリに分類することができ、言語学の場合は形態論、統語論、意味論、語用論などが下位カテゴリとなる。自然言語処理の場合は、言語学、人工知能の他、人工知能の一分野としての機械学習、自然言語処理の応用システム（情報検索、機械翻訳、質問応答）、認知科学等複数の分野が関連している。特に言語学は、ある言語現象について自然言語処理を用いて検証することがあるため、関連が深い。また、応用分野では自然言語処理技術を用いていることから、これらの応用にも基礎的な部分が共通している。

複合領域の分野および下位カテゴリの用語をどの程度分野基礎性が高い用語に含めるかも難しい問題である。特定研究分野全体に広く一定した頻度で用いられている用語は分野基礎性が高く、一方で、一つの下位カテゴリの中でしか使われない用語は、分野基礎性はそれほど高くないと考えられる。

教える側から考えると、隣接領域の中でも基礎的な用語について万遍なくカバーして提示するのが必要である。一方、学ぶ側では、背景知識を持った人間であれば、たとえば、言語学を専攻している学生は、言語学の知識があるのでそれは知る必要がなく、その他の領域の知識を学ぶ必要がある。自分の知識に欠けている下位分野の用語の中から基礎性が高いものから学ぶと効率が良い。その場合、対象とする分野と下位分野の関連性の強さを基準にして考える必要がある。言語学であれば形態論、統語論、生成文法のうち、自然言語処理との関連の強い用語が自然言語処理の分野基礎性の高い用語に含まれることになる。つまり、下位

分野単独における基礎専門性ではなく、上位分野との関連性が強い語が重要な語となる。

### 3.2.5 語構成度

分野基礎性が高い用語はその用語単独でも多く出現するが、前や後ろに様々な語が接続して多くの新規複合語を構成していることが予測できる。たとえば、「機械翻訳」という用語の場合、後ろに「システム」が結合して「機械翻訳システム」、前に「統計的」が結合して「統計的機械翻訳システム」など、様々な派生の専門用語および新しい複合語（いずれは専門用語として認識されるものも含む）を生成することができる。この基準は重要度計算の時でも利用されているが、どれだけ語と接続する可能性があるのかで、その基となる用語の重要性が計算できる。ある用語が新規の専門用語を構成している数が多ければその用語の概念は重要であると考えられる。

### 3.3 定義明確度

定義明確度として、分野基礎性が高い用語はまず定義を明確に述べるが行われる。たとえば「形態素解析とは、与えられた文を形態素の単位に分割し、その文法機能（一般には品詞および活用情報）を同定する処理を言う」というように手掛かり語「とは」を用いて定義付けを行う。このように「AとはB」のAに当たる用語は分野基礎性が高いと考えられる。上記の6つの指標のうち、定義明確度を除いた5つの指標について表1に示す。個々の指標は、それぞれが相互に関連しているものとする。また指標の度合いは前述した学術側の段階に合わせ4段階としている。

## 4. 分 析

分野基礎性の尺度について見てきたが、実際のデータではどのような傾向があるのかを調べるために分析を行った。研究分野として馴染みのある「自然言語処理」を対象とした。

### 4.1 優先度の分析

優先度を調べるために大学で行われている講義資料と一般に出版されている書籍を参照

表 1 分野基礎性の指標

Table 1 Sections and sub-sections in which list-like environments are used (example of table).

分野基礎性	優先度	経年推移度	親密度	分野内の出現分布	語構成度
(1) 非常に高い	最優先 1	安定（高頻度）	高頻度	偏りが無い	多数の用語生成
(2) やや高い	優先 2	安定（低頻度）	やや高頻度	少し偏る	やや多数の用語生成
(3) やや低い	順序性なし	差が激しい	やや低頻度	2,3 の分野だけに偏る	やや少数の用語生成
(4) 非常に低い	順序性なし	不安定（低頻度）	低頻度	1 分野にだけ偏る	用語生成しない

した。まずはじめに自然言語処理に関する講義資料へのリンク<sup>\*1</sup>から、任意の3つの講義資料に含まれるキーワードをピックアップして共通しているものを調べた。リンク切れを除いたが、古い講義資料へのリンクであったため、最新のデータではない。次に書籍についても1996年から2007年の間に出版された新旧合わせた5冊<sup>7)-11)</sup>から共通する索引語を調べて、頻度を数えた。表2に講義資料、表3に書籍の索引語の共通語リストを、頻度順に上位20語を示す。

共通語リストを見てみると、「形態素解析」「構文解析」「意味解析」は分野基礎性が高く、どの教科書や講義にも出てくると予想したが、「形態素解析」以外はどれかが欠けている状態であった。特に「構文解析」は書籍によっては「統語解析」と表現していることもあり、共通して出現していなかった。また、比較的共通する語の数が少なかった（全てに共通している語は講義資料では4語、教科書では8語）。これは、講義や教科書はその分野の基礎を教えるためのものであるが、講義者や著者の研究トピックの違いによっても講義をする視点が多少違っており、なかなか共通した用語を見つけることができなかった。対象とした書籍や資料の数が少ないことも原因であると考えられ、講義資料についてはもう少し検索対象を広げる必要がある。

#### 4.2 経年推移度の分析

次に経年推移度について、論文データを使って調べた。情報処理学会発行誌掲載論文データから自然言語処理研究会の研究報告で発表された1993年度から2006年度までの14年間1411件の論文抄録データを利用した。経年推移は各年度ごとに抄録を形態素解析器 MeCab で語に区切り、名詞が連続する2語の単位で頻度を数えた。頻度10以上の2語基単語をストップワードでフィルタリングした結果、468語の用語候補となった。表4は頻度順に上位20語を並べ、その20語が何年間出現し、その合計頻度を示している。年数14年間継続して出現する用語は当然頻度も高いが、年で平均しても出現する数が高かった。14年間継続して出現し、合計頻度も高い用語は分野基礎性が高いと考えられる。分野に基礎的な用語は初期段階から出現し、流行が一時期あったとしてもその後も持続的に出現している。用語を細かく見て行くと、自然言語処理の分野では、基礎技術としての「形態素解析」「構文解析」「言語モデル」、応用システムとして「機械翻訳」「情報検索」「質問応答」「音声認識」に利用されて、「曖昧性」や「未知語」といった問題を解決してきた分野であることが良く分かる。「固有表現」は固有表現抽出が1999年度から新たな研究課題として出てきてい

表2 講義資料で共通する用語

用語	頻度
形態素解析	3
構文解析	3
格文法	3
文脈自由文法	3
機械翻訳	2
LFG	2
GPSG	2
HPSG	2
機械翻訳	2
情報検索	2
最長一致法	2
深層格	2
ベクトル空間法	2
シソーラス	2
隠れマルコフモデル	2
2文節最長一致法	2
単一化文法	2
ベイズの定理	2
コーパス	2
句構造文法	2

表3 教科書に共通する用語

用語	頻度
意味ネットワーク	5
格フレーム	5
形態素解析	5
シソーラス	5
格文法	5
形態素	5
深層格	5
表層格	5
接辞	4
自然言語処理	4
要約	4
情報検索	4
索引語	4
機能語	4
再現率	4
内容語	4
意味解析	4
意味素性	4
活性伝播	4
文書集合	4

るにもかかわらず、頻度が高いのは、重要な問題として多くの研究者が取り組んでいるのが頻度から読み取れる。

#### 4.3 親密度

自然言語処理の低位分野に対して、専門用語の分布がどのようになっているのかを調べるために、デジタル言語処理事典<sup>6)</sup>に出現している用語の頻度を調べた。事典内のテキストを形態素解析器で単語に区切り、名詞が連続している塊を、一つの専門用語候補として抽出した。これまでの分析と同様にフィルタリングをした後、異なり語は一単語が5235語、複合語が9652語、合計で14887語が事典全体の名詞となった。事典の索引語は2692語あり、その索引語が事典テキスト中に出現する頻度を数えたが、索引語には一般語（「認識」、「目的」、「構造」等）も含まれているため、一単語はカタカナ文字だけカウントし、後は複合語を対象として頻度の高い順に並べた。表5はデジタル言語処理事典<sup>6)</sup>の索引語が事典のテキスト中に出現している頻度の上位20語を示している。事典の用語だけに偏りがあまりなく、分野全体を知る上で重要な語が上位に来ていることがよくわかる。

\*1 [http://www-lab25.kuee.kyoto-u.ac.jp/NLP\\_Portal/lecture.html/](http://www-lab25.kuee.kyoto-u.ac.jp/NLP_Portal/lecture.html/)

表 4 経年推移度 (情報処理学会 研究会報告抄録) 表 5 親密度 (デジタル言語処理学事典)

用語	頻度	年数	用語	頻度
コーパス	477	14	コーパス	341
機械翻訳	197	14	機械翻訳	110
形態素解析	188	14	目的語	100
類似度	149	14	意味役割	94
文字列	129	14	オントロジー	87
構文解析	126	14	意味論	85
情報抽出	118	14	情報検索	79
翻訳システム	108	14	タスク	79
情報検索	108	14	言語獲得	77
再現率	98	14	形態素解析	70
音声認識	97	13	音声認識	67
言語モデル	93	13	クエリ	66
固有表現	88	8	普遍文法	65
名詞句	86	13	主要部	63
質問応答	85	11	情報抽出	59
未知語	84	13	言い換え	56
適合率	80	14	アノテーション	52
正解率	77	14	言語知識	51
検索システム	73	13	文法化	50
専門用語	66	13	音韻論	50

#### 4.4 下位分野偏り度の分析

##### 4.4.1 出現章の重なり

自然言語処理の分野も含めて、他の分野でも下位分野がいくつかある。その下位分野に出現する分布をみるために、デジタル言語処理学事典を用いて調べた。言語処理学事典は5章から構成され、1章言語資源、2章基礎技術、3章統合技術システム・応用システム、4章言語科学の基礎、5章言語科学の展開となっている。これらの章を下位分野と見立て、それぞれの章における索引語の出現回数を調べた後、複数の章にまたがって出現するかどうかを比べてみた。複数の章に出現し、なおかつ総合的に頻度が高い語は分野基礎性が高い語であることがわかる。5つの章とも出現していた語は「統語構造」「統語論」「言語知識」「構文解析」「目的語」「言語情報」の6語であった。自然言語処理が、言語学の理論に基づいて発達してきたことや、事典の中で2章を割いて言語学について書かれているため、言語学の用語も含まれているのが特徴的である。4つの章にまたがって書かれていた語は「音声認識」「確率モデル」「言語能力」「機械学習」「機械翻訳」等を含めて30語あった。これらも言語学、基礎技術、応用に関連する基礎的な用語が平均的に出現していることがわかる。

経年推移度や親密度で上位に来ていた語(出現頻度が高い語)については、5つの章にまたがっているものもあるが、「形態素解析」などは3つの章にまたがっているのが特徴的であった。言語学の章では「形態素解析」が出てこないが、他の3章では出てくるといったものである。つまり言語学の章には出現しないで、他の3つの章に出現している語は自然言語処理の基礎技術に関連する語が多いと考えられる。一方で、言語学と基礎技術の部分に共通して出現するものは、言語理論を踏まえた技術であると言える。

##### 4.4.2 各章における頻度分布

表6に各章において頻度の高い用語を10語示した。下位分野の偏り度では、分野基礎性が高い用語をみることができ、各下位分野において頻度が高い語は、狭い範囲における分野基礎性が高い語であると言える。1章は言語資源に関連する項目が記述されていたが、「語彙的オントロジー」が一番高頻度であったが、他の章では全く出現していなかった。教える側からは各下位分野の頻度が高いものを一定量提示することによって、偏りを排除するが、学ぶ側の好み考えた場合は、学習する人のモチベーションや背景知識の違いによって偏りが生じる可能性がある。「自然言語処理の応用システムについて知りたい」人であれば、3章の統合技術システムにおいて頻度が高い語の概念を中心に学習すれば良いと考えられる。

表 6 章別の索引語頻度表

第一章		第二章		第三章		第四章		第五章	
語彙的オントロジー	43	訓練データ	41	情報検索	61	意味論	75	言語獲得	64
上位オントロジー	31	未知語	40	テキストマイニング	45	目的語	68	メタファ	47
語彙概念	27	形態素解析	39	音声認識	44	意味役割	57	普遍文法	47
言語資源	22	格フレーム	32	情報抽出	43	主要部	50	言語知識	31
国語辞典	21	論理式	31	機械翻訳	38	ラムダ項	49	言語能力	27
均衡コーパス	18	ツリーバンク	29	固有表現	37	優先順位	46	認知言語学	24
専門用語	13	並列構造	26	評価表現	24	真理値	43	言語機能	23
専門用語辞書	12	機械学習	23	類似度	24	音韻論	43	言語理論	23
フォントリソース	10	機械翻訳	23	マルチモーダル	23	命題論理	38	疑問文	23
コーパス言語学	10	意味解析	22	形態素解析	22	語形成	36	フレーム意味論	20

#### 4.5 語構成度の分析

デジタル言語処理学事典<sup>6)</sup>の索引語の語構成度を調べた。索引語の単語2グラムを用いて統計を出した。表7に基本となる単語2グラムおよびカタカナ語と結合して新しい用語に生成された数、結合用語例を示す。これらは、これまで見てきた分野基礎性が高い用語であり、語構成度も関連があることがわかった。語によって、前にだけ、あるいは後ろにだけ語が結合するものがあり、用語概念の特徴を表している。

表 7 語構成度の高い用語

用語	頻度	結合用語例
コーパス	32	均衡コーパス, 話し言葉コーパス
意味-論	25	語彙意味論, フレーム意味論
機械-翻訳	12	機械翻訳システム, 統計的機械翻訳
n-グラム	9	n グラムモデル, 単語 n グラムモデル
アルゴリズム	12	EM アルゴリズム, ブースティングアルゴリズム
句-構造	9	句構造文法, 主辞駆動句構造文法
言語-学	9	メタ言語学, 計算言語学
主要-部	9	右側主要部規則, 主要部先行型
曖昧-性	9	曖昧性解消, 構造的曖昧性
言語-情報	8	バラ言語情報, 多言語情報検索

#### 4.6 定義明確度の分析

定義明確度では「とは」という手掛かり語を使った定義文を検索した。「～とは独立した」といった定義文ではない表現を削除し、194 語の定義付きの用語を抽出した。定義文では説明している文脈も考慮して記述しているため、非常にわかりやすく、辞書のように読むことができ、初心者に提示するために有用な情報であった。しかし、言語学の方々の定義が多く、技術的用语の説明が少なかった。今後は「とは」以外の手掛かり語を拡張して、分野基礎性の高い用語を含む文脈の抽出を行う予定である。

#### 5. おわりに

本研究では、専門用語の分野基礎性について 6 つの指標について論じ、その指標について実際のデータを用いて分析を行った。優先度については、扱ったデータ（書籍、講義資料）が不足していたこともあり、妥当なデータが得られなかった。経年推移度と、親密度については、ほぼ同じ結果となり、重要で基礎性が高い語は、分野が成立した当初から出現し、年を経ても一定した推移で総合的に頻度が高い用語であった。下位分野偏り度については、分野の下位分野での分布を調べ、下位分野に万遍なく出現している用語、語彙構成度では多くの複合語を生成する用語、定義文を明確に定義している用語がそれぞれ基礎性が高い語であることを確認した。今回はデータ全体における出現傾向から分野基礎性を議論したが、今後は語彙構成度を詳細に分析を行い、更に分野を拡張した上で、提案した指標の数値化を検討していく。

謝辞 情報処理学会刊行誌掲載論文データに関して、研究利用することを許諾していただいた、社団法人情報処理学会に感謝いたします。活発な議論をして下さった東京大学影浦峽

教授、名古屋大学佐藤理史教授、はこだて未来大学藤田篤准教授、広島市立大学難波英嗣准教授、東洋大学鈴木崇史専任講師、国立情報学研究所相澤彰子教授、宮尾祐介准教授、東京大学浅石卓真氏に深く感謝いたします。

#### 参考文献

- 1) 中川裕志, 森辰則, 湯本紘彰, 出現頻度と接続頻度に基づく専門用語抽出, 自然言語処理, Vol.10 No.1, pp.27-45, 2003.
- 2) 千田恭子, 篠原靖志, 奥村学, アンケートによる用語調査とWWW上の頻度分布を用いた用語の専門度推定, 言語処理学会第 10 回年次大会ワークショップ「固有表現と専門用語」発表論文集, pp.36-39, 2004.
- 3) 千田恭子, 篠原靖志, 奥村学, 技術成果を効果的に伝える表題作成支援手法: 開発と評価, 情報処理学会論文誌, Vol.46, No.11, pp.2728-2743, 2005.
- 4) 木田充洋, 外池昌嗣, 宇津呂武仁, 佐藤理史, ウェブを利用した専門用語の分野判定 電子情報通信学会論文誌, D, 情報・システム, J89-D(11), pp.2470-2482, 2006.
- 5) 佐々木靖弘, 佐藤理史, 宇津呂武仁, 関連用語収集問題とその解法 自然言語処理, Vol.13, No.3, pp.151-175, 2006.
- 6) 共立出版株式会社 デジタル言語処理学事典 自然言語処理学会編, 2010.
- 7) 長尾真, 佐藤理史, 黒橋禎夫, 角田達彦 岩波講座 ソフトウェア科学 15 自然言語処理 岩波書店, 1996.
- 8) 長尾真, 黒橋禎夫, 佐藤理史, 池原悟, 中野洋 岩波講座 言語の科学 9 言語情報処理 岩波書店, 1998.
- 9) 徳永健伸 情報検索と言語処理 東京大学出版会, 1999.
- 10) 荒木健治 自然言語処理ことはじめ 言葉を覚え会話のできるコンピュータ 森北出版, 2004.
- 11) 天野真家, 宇津呂武仁, 成田真澄, 福本淳一, 石崎俊 自然言語処理 (IT Text) オーム社, 2007.