

言い換え表現抽出に基づく関連用語収集手法の
シラバス検索システムへの応用

Automatic Collection of Related Terms from Syllabus Data
and its Application to the Retrieval of Syllabi

1) 大学評価・学位授与機構 評価研究部

Faculty of University Evaluation and Research, National Institution for
Academic Degrees and University Evaluation

2) 大学評価・学位授与機構 学位審査研究部

Faculty for the Assessment and Research of Degrees, National Institution for
Academic Degrees and University Evaluation

3) 京都大学 学術情報メディアセンター

Academic Center for Computing and Media Studies, Kyoto University

芳鐘冬樹¹⁾ 井田正明¹⁾ 野澤孝之¹⁾

宮崎和光²⁾ 喜多 一³⁾

Fuyuki, YOSHIKANE Masaaki, IDA Takayuki, NOZAWA

Kazuteru, MIYAZAKI Hajime, KITA

Abstract

In this paper, we discuss a method for automatically collecting related terms of a seed term, such as narrower terms which represent more specific concepts, from syllabus data. The proposed method is based on the technique of extracting term variants from corpora with morpho-syntactic rules. By applying it to the syllabus retrieval system, we demonstrate the usefulness of the proposed method.

1. はじめに

今日、大学では、(特に学内で生産される)学術情報の整理・発信が、研究・教育の振興およびその国際通用性の観点から重要な課題であり、各大学は、図書館機能の改善などの取り組みを通して、学術情報発信の機能の整備を進めている(文部科学省研究振興局情報課、2003)。その中で、

教育情報として非常に重要な位置を占めるのがシラバスである。シラバスは、学生に対して授業の目的や計画などの情報を提供する仕様書としての機能や、担当教員から学生に履修方法・規定を伝達する機能を有している(大学セミナー・ハウス、1999)。近年、ウェブを通して、シラバスを広く社会に公開している大学も多く、シラバスは、授

業内容を示す資料として、受験生が大学を選択する際に用いられ、また、大学における教育の外部評価や、大学評価・学位授与機構における学位授与審査の際の修得単位の内容確認にも利用されている。

現在、ウェブ上で公開、あるいは CD-ROM で配布されるなどして、シラバスは電子化されていることが多い。そこで、上で述べたような様々な用途での利用において、情報技術による支援が期待される。シラバスのデータベース化（井田ら、2004）、シラバスからの専門用語抽出（芳鐘ら、2004a）とそれに基づくシラバスのクラスタリングによる教育課程の構造分析（野澤ら、2005）や科目分類支援（Miyazaki ら、2004）など、情報技術によりシラバスデータを活用する様々な研究が既になされている。

シラバスが含む情報の大部分は、科目名や授業概要・計画など、専門用語を含む言語記述による項目であり（井田ら、2004）、それらを扱うとき考慮しなければならないのが関連用語の存在である。1つ1つの専門用語は、基本的に、それぞれ独自の概念を表すものと考えられるが、全ての用語が各々全く独立した概念を表すわけではなく、多くの用語には、何らかの関係がある用語（同じような概念を表す同義語や、より広い／狭い概念を表す上位語／下位語、あるいはその他の関連語）が存在している。そのような異なる用語間に関する知識は、用語が集まってできているシラバス同士、つまり、シラバスによって表される科目同士の関係の理解に役立てることが可能である。例えば、関連用語を全く別の用語として扱うのではなく、グループ化するなどの関連付けを行えば、シラバスに基づく科目のクラスタリングや分類のパフォーマンス向上が期待できる。また、シラバスの検索において、検索語に關係する用語と關係のタイプについての知識が利用できれば、

検索の再現率の向上だけでなく、出現する用語の關係のタイプから科目間の關係を整理して、検索結果を提示することで、ユーザの利便性の向上も期待できる。そこで、本研究では、ある用語と關係する用語を、關係のタイプごとに、シラバスから収集する手法の検討を行う。さらに、提案手法を応用して作成したシラバス検索システムについて報告する。

以下、本稿では、第2節で収集対象とする関連用語のタイプについて、そして第3節で収集の手法について具体的に述べ、最後に、第4節で試作したシラバス検索システムの概要について説明する。

2. 関連用語のタイプと収集対象

表す概念から見た用語間の關係のタイプとして、同義關係、階層關係、関連關係が挙げられる（日本工業規格、1991）。さらに、階層關係は類種關係（包含關係）、全体部分關係、例示關係に、関連關係は同一カテゴリに属する語の關係と、異なるカテゴリに属する語の關係に分けられる（表1を参照）。本研究では、これらのうち、同一カテゴリに属する関連語、類種關係の下位語、そして、それらの関連語・下位語および元の用語の同義表現を収集の対象とする。本稿で「関連用語」と言えば、同義表現、関連語、下位語のすべてを含むものとする。

このような対象の制限は、後で述べるように形態的／統語的な言い換えの認識のみに基づく本手法の限界によるものだが、逆に、これらを他と区別して収集できるという利点があると言える。同一カテゴリに属する関連語だけを収集し、共通するカテゴリを上位語として系統的に整理できれば、単に関連語を集めるよりも精緻な知識が得られるものと考えられる。下位語については、分野による差はあろうが、階層關係の中で類種關係が

表1 用語間の關係のタイプ

關係のタイプ	サブタイプ	例
同義關係		「カメラ」－「写真機」
階層關係	類種關係（包含關係）	「カメラ」－「動画カメラ」
	全体部分關係	「自然科学」－「生物学」
	例示關係	「高速鉄道」－「新幹線」
関連關係	同一カテゴリに属す語の關係	「動物図鑑」－「植物図鑑」
	異なるカテゴリに属す語の關係	「出版」－「図書」

占める割合は少なくないという報告もあり¹⁾、ある程度のカバレッジは得られると考えられる。

また、関連用語収集の出発点とする入力用語（検索システムにおける検索語）は複合語に限定している。これも言い換え認識上の制限である。本研究の手法は、一般語、専門用語を問わず適用可能だが、専門用語は複合語が大部分を占めるため（Ishii, 1987）、専門用語に対して特に効果がある。例えば、人工知能分野の『人工知能大辞典』（Shapiro & Eckroth, 1987）では、3,869語中3,245語（約84%）が2要素以上から成る複合語である。表2に、複合語3,245語の品詞パターンの内訳をまとめた。ただし、N は名詞、TPN は名詞性接尾辞、S は助詞を表す。2名詞から成る複合名詞（N N）が全体の約40%を占めること、そして（N N）は他の多くのパターンの基礎になっていることが、表から確認できる。本研究は、関連用語収

集の最初のステップとして、この（N N）を入力用語として想定した手法の検討を行う。

3. 収集手法

関連用語の自動収集手法に関しては、文書内の共起情報に基づくものなど、様々な研究が存在する（Chen ら、1995；永松&田中、1996；小原ら、2004）。しかしながら、それらの多くは、関連用語の収集にとどまり、関係のタイプの区別までは行っていない。区別を行っているものも、収集源が辞書の語義文などに限定されていたり、区別の範囲が上位語や下位語に限定されており（鶴丸ら、1992；佐藤&佐々木、2003）、必ずしも十分とは言えない。本研究の手法では、複合語の形態的／統語的な言い換えに着目し、元の用語を言い換えた異形（を加工したもの）を、元の用語に関係する用語として抽出する。複合語の形態的／統語的な異形に限定されるが、言い換える分類に基づいて、関係のタイプ（同義表現、下位語、関連語）を区別しつつ、関連用語を収集することができる。

本研究が提案する関連用語の収集手法は、テキストからの言い換え表現（用語異形）のルールベース抽出に基づく。用いる言い換えルールは、Yoshikane ら（2003）、Kageura ら（2004）のルールを拡充したものであり、それらを言い換え認識システム Fastr（Jacquemin, 1994）にインプリメントして、言い換え表現の抽出を行う。作成した212のルールは、(i) 機能語の付加／削除、元の構成要素の品詞変化、主要部の入替のみが生じる

表2 複合語の品詞パターン

品詞パターン	語数
N N	1297
N N N	419
N TPN N	213
N TPN	197
N N TPN	112
N S N	91
N TPN N N	71
N N N N	68
N N S N	43
N N TPN N	41
その他	693
計	3245

表3 言い換えルールの例

	規則例	言い換え例
(i-1)	$(X1 \rightarrow X2 X3) = (X1 \rightarrow X2 S4 X3)$ <X2 cat> = N NS NA, <X3 cat> = N NS NA, <S4 lem> = 'の'	生体[N] 計測[NS] → 生体[N] の[S] 計測[NS]
(i-2)	$(X1 \rightarrow NA2 X3) = (X1 \rightarrow NA2 MD4 X3)$ <X3 cat> = N NS NA	曖昧[NA] 情報[N] → 曖昧[NA] な[MD] 情報[N]
(i-3)	$(X1 \rightarrow NS2 N3) = (X1 \rightarrow N3 NS2)$	共有[NS] メモリ[N] → メモリ[N] 共有[NS]
(ii-1)	$(X1 \rightarrow X2 X3) = (X1 \rightarrow X2 S4 (A NA MD X TPN MD) + X3)$ <X2 cat> = N NS NA, <X3 cat> = N NS NA, <S4 lem> = 'の'	ネットワーク[N] 解析[NS] → ネットワーク[N] の[S] 動的[NA] な[MD] 解析[NS]
(ii-2)	$(X1 \rightarrow X2 X3) = (X1 \rightarrow (N X TPN) + X2 X3)$ <X2 cat> = N NS NA, <X3 cat> = N NS NA	情報[N] 検索[NS] → オンライン[N] 情報[N] 検索[NS]
(iii-1)	$(X1 \rightarrow X2 X3) = (X1 \rightarrow X2 ((SC L) (N NS NA)) + S4 X3)$ <X2 cat> = N NS NA, <X3 cat> = N NS NA, <S4 lem> = 'の'	生体[N] 計測[NS] → 生体[N] と[SC] 環境[N] の[S] 計測[NS]
(iii-2)	$(X1 \rightarrow X2 X3) = (X1 \rightarrow X2 S? X3 (L (N NS NA)) +)$ <X2 cat> = N NS NA, <X3 cat> = N NS NA	生体[N] 計測[NS] → 生体[N] 計測[NS] ・[L] 診断[NS]

もの、(ii) 修飾語として内容語が加わるもの、(iii) 元の構成要素と等位関係を持つ形で内容語が加わるもの、の何れかに分類できる。表3に、(i) (ii) (iii) それぞれの言い換えルールを例示した。Fastr では、主として統語的な制約しか記述できないため、すべての言い換え表現を誤りなく(再現率100%かつ精度100%で)抽出できるルールの作成は非常に困難である。本研究では、再現率と精度のトレードオフを考慮したルールを用いている(Yoshikaneら、2003)。

言い換えルールは、言い換え前後の文脈自由構文と、品詞(cat)・語彙(lem)に関する制約式から成る。この表では、NSはサ変名詞、NAは形容動詞語幹になる名詞、Nはその他の名詞、Aは形容詞、TPNは名詞性接尾辞、Sは助詞、SCは名詞接続助詞、MDは助動詞、Lは読点・区切り記号、Xは任意の品詞を表す²⁾。また、「+」などの正規表現もルールの記述に用いている。

- ・ (i) は、元の内容語の削除や新たな内容語の付加はなく、構成要素間の意味関係が保たれている((i-3)の例で言えば、言い換えの前後、どちらも「メモリ」が「共有」という動作の対象になっている)言い換えである。(i-1)(i-2)の異形は、必ずしも語の形をとらず、また、(i-3)の言い換えは、主要部入替を含むため、それらは厳密には同義語と呼べるものではないが、表層的な構造は変わっても、およその概念は保たれており、それらを含むシラバスは共通するトピックを扱っていると期待できる。そこで、これらの規則が適用される異形を、元の用語の同義表現(広義の同義関係にある語や句)として抽出する。(例:「生体計測」の同義表現として「生体の計測」を抽出)
- ・ (ii) は、修飾語の付加によって、より限定された概念を表す異形に言い換えるものである。これらの規則が適用される異形を、元の用語の類種関係にある下位語として抽出する³⁾。(例:「生体計測」の下位語として「生体物理計測」を抽出)
- ・ (iii) は、元の使用と共通の構成要素を含み並列関係を持つ用語を、元の使用と組み合わせるものである。これらの規則が適用される異形から、「並列関係を持つ用語」を分離して、元の使用の関連語として抽出する。(「生体計測」の



図1 用語間の関係

関連語として、「生体・環境計測」から分離した「環境計測」を抽出)

Jacquemin (1996) は、フランス語を対象にして同様の手法で用語の収集を行っている。本研究は、そのアイデアを日本語に適用し、さらに Jacquemin (1996) では用いられていない外部からの修飾を含む言い換え((ii-2)など)も加えることで、より包括的な関連用語の収集を目指す。

テキスト(シラバス)から抽出した関連用語は、図1のような2系列のツリー形式に整理する。1つめは、入力用語(元の複合名詞)の主要部(後項)をルートとするツリー、2つめは修飾部(前項)をルートとするツリーである。入力用語を[a]の位置に、入力語の同義表現を[a']の位置に、下位語(およびその同義表現)を[c]の位置に出力する。以上は、2つのツリーに共通している。[b1][b2]の位置には、それぞれ、入力用語の主要部を共通のカテゴリとする関連語(およびその同義表現)、主要部の対象・目的などを表す修飾部を共通のカテゴリとする関連語(およびその同義表現)を出力する⁴⁾。2つめのツリーは、ルートと入力用語の主要部が一致しておらず、厳密には階層関係を表していないが、関連語との関係を示すために便宜的にツリー形式で表現することとした。

4. シラバス検索システム

前節で述べた関連用語収集手法に基づくシラバス検索システムを試作した。予め収集・情報抽出したシラバス集合を対象に、ユーザが入力した検索語、およびその関連用語を含むシラバスを検索し、検索語との関係のタイプごとに整理して表示するというものである。本研究の手法は、収集源

のテキストの種類を限定するものではないが、専門用語を多く含むシラバスへの適用は有効であると考えられる。本節では、作成したシラバス検索システムの概要と利用例、そしてシステムの評価実験について述べる。

4.1 システムの概要

大学のウェブサイトで公開されているシラバスページのファイル（多くは HTML 形式）をダウンロードし、それらを本システムの検索対象とした。シラバスには、「科目名」「達成される目標」「授業概要」「開講学期」「教員名」「授業計画」「教科書」「取得単位数」など多数の項目（井田ら、2004）が含まれる（大学によって情報量に差はある）。ただし、シラバスページの記述形式は大学によって異なっており、そのままでは大学の比較などに利用することが難しい。そこで、まず、シラバスページから情報を抽出するツール（渡辺ら、2004）を用いて、項目ごとに情報の抽出を行うことにより、シラバスデータを共通形式に変換した（共通の項目名のもとに情報を整理した）。抽出した情報のうち、「科目名」「達成される目標」「授業概要」「授業計画」「教科書」を授業の内容を表す項目と考え、本システムでは、これらを検索範囲（関連用語の収集源）としている。

シラバス検索の流れを以下に示す。

- ・ 検索画面（図 2）で、検索対象とする大学（シラバス集合）を選択し、検索語を入力する。
- ・ 検索を実行すると、上記のシラバスの項目から、Fastr によって検索語の言い換え表現が抽出される⁵⁾。
- ・ 言い換え表現は、言い換えルールの分類に応じて関連用語として整理され、用語間の関係を表すツリー形式で表示される（図 3）。
- ・ 見つかった関連用語の中から情報を見たいものを選ぶと、その用語と検索語の両方がシラバスに現れる科目の名前と、当該科目が属す教育課程（学科・コース）の一覧が表示される（図 4）。
- ・ 科目を一覧から 1 つ選ぶと、その科目のシラバスページが表示される。

図 4 に例を示したように、検索語が表すトピック一般か、よりスペシフィックなトピックか、あるいは関連トピックと合わせて扱われているかに従って分類して検索結果を表示することで、ユー

ザが、検索要求に適ったシラバス（科目）を探しやすくなるとともに、提示される関連用語は、次に検索語を修正する際の参考になり、シラバス検

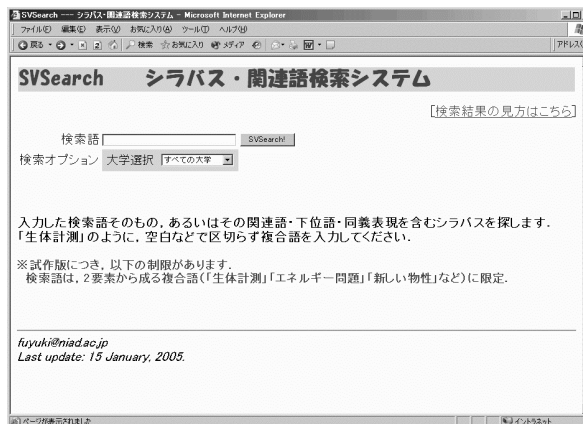


図 2 検索画面

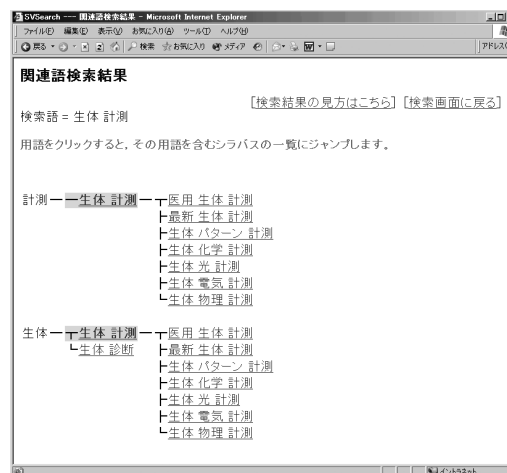


図 3 検索結果画面

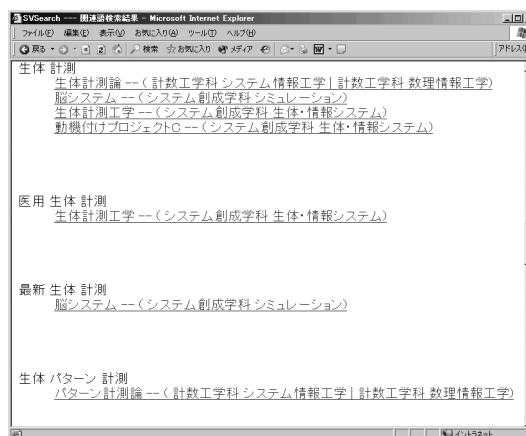


図 4 科目一覧

索の利便性の向上が期待できると考えられる。さらに、図示される関連トピックの系列に基づいて、それらを扱う科目群の体系をある程度把握することができ、そのトピックの教育に関する大学ごとの傾向の分析への活用も可能である。

4.2 システムの利用例

一例として、「エネルギー問題」を検索語とし、2つの大学（A・B）の工学部のシラバスを対象に検索を行った結果を下に示す。ただし、誤って抽出した用語（「エネルギーの合理」「力の問題」）は除いてある。抽出誤りについては4.4節で考察する。

A大学 検索語：エネルギー問題 <20>
 関連語：環境（の）問題 <10>、
 資源問題 <1>
 下位語：エネルギーの需給問題 <1>、
 エネルギー環境問題 <2>

B大学 検索語：エネルギー問題 <3>
 関連語：資源問題 <2>、
 人口問題 <1>

<>内の数字は、その用語と「エネルギー問題」の両方を含むシラバスの数を表す。検索結果画面で用語をクリックすると、科目名と開講学科・コース名の一覧が表示される（A大学は図5、B大学は図6）。例えば、この結果から、エネルギー問題は、A大学では環境問題と合わせて扱われることが多く（化学系・機械系など多様な学科で扱われている）、B大学では資源問題・人口問題と合わせて扱われる傾向がある（地球工学科で扱われている）ことが分かる。

4.3 評価実験

評価実験として、前述の『人工知能大辞典』に記載された用語の関連用語をシラバスから抽出する実験を行った。検索対象のシラバスは、3大学の工学部、合計1,699科目のものをを用いた。また、入力用語としては、『人工知能大辞典』から2名

詞から成る複合名詞を20語選んで用いた（表4を参照）。入力用語選択の際には、シラバスに1度も出現していないものは除外している。



図5 検索結果の科目一覧（A大学）

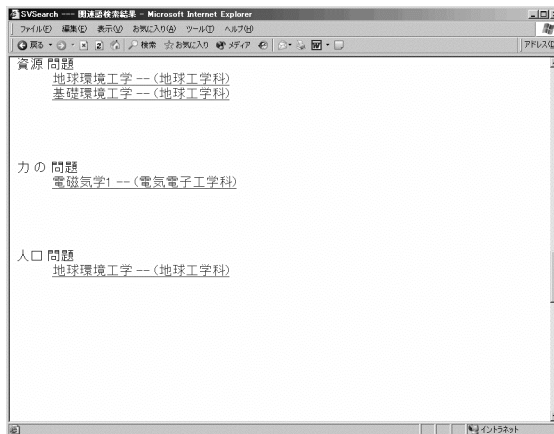


図6 検索結果の科目一覧（B大学）

表4 入力用語リスト

確率理論	定理証明
プログラム生成	画像解析
適応制御	運動制御
挙動予測	知識獲得
パターン認識	知識システム
階層構造	制御装置
音声合成	言語モデル
音声処理	移動ロボット
プロセス計画	言語処理
データ構造	ソフトウェア開発

表5 関連用語の抽出結果

同義表現			関連語[b1]			関連語[b2]			下位語			全体		
正	誤	精度	正	誤	精度	正	誤	精度	正	誤	精度	正	誤	精度
13	3	0.81	1	1	0.50	4	1	0.80	17	1	0.94	35	6	0.85

表5に、同義表現、関連語、下位語それぞれの抽出数と抽出精度を示した。約85%の精度で関連用語を抽出することができ、同様の手法でウェブページ一般を対象とした関連用語収集の結果の約77%（芳鐘ら、2004b）に比べ、高い精度が得られた。ウェブページ一般を対象にしてシステムが抽出を行った場合、大学のシラバスページに対象を限定した場合と比較して、誤字が多い、ページ検索の段階で誤って中国語のページが混ざることがあるなど、テキストの質が低く、それが抽出精度の低下の原因になっている。

関連語 [b1] については、低い精度しか得られなかった。ただし、ウェブページ一般を対象とした場合の [b1] の精度は90%を超えていたことから（芳鐘ら、2004b）、今回の結果はサンプル数の少なさに起因している可能性も高い。今後、サンプルを増やして検証したい。

また、学術論文などを対象とした場合に比べ、シラバスでは専門用語のバリエーションが少ないためか、量的な面では必ずしも満足のいく抽出結果は得られなかった。今後、言い換え規則の拡充などにより改良を試みたい。

4.4 誤り診断

本システムによる関連用語抽出の誤りのほとんどは、以下の3つに大別できる。

- (1) 形態素解析の誤りに由来するもの。
形態素への分割、あるいは品詞の判定に誤りがあると、言い換えルールが正しく適用されない。
- (2) 係り受け構造の抽出の誤りに由来するもの。
例：画像解析 → ×計測解析
「レーザー利用計測・画像解析」において、「計測」が「画像」とともに「解析」に係る（「{計測・画像} 解析」）と誤って認識した結果、「画像解析」の関連語として、誤って「計測解析」を抽出してしまったケース。
- (3) 構成要素間の関係が元の用語と異なるものを抽出したケース。
例：言語処理 → ×処理言語
前者は「言語を処理する（言語が処理の対象）」、後者は「何かを処理するための言語（言語が処理の手段）」であり、全く意味が異なっており、関連用語として正しくない。

これらは、ウェブページから関連用語収集を行った芳鐘ら（2004b）、および、論文抄録から異形抽出を行ったYoshikaneら（2003）の指摘と共通している。(1)は、用いる形態素解析システムに依存する。一方、(2)と(3)については、構文解析を行い並列構造を認識するなどの前処理を加えたり、言い換えルールを洗練したりすることで、ある程度は改善できると考えられる。

一方、シラバスを抽出源とした場合特有の問題として、「講義」「学生」など、シラバス中の頻出語がノイズになり、抽出の誤りを生じさせるという点が挙げられる。それらの語はストップワードにして、それらを含む用語は関連用語として抽出しない、といった方策が必要と考えられる。

5. おわりに

本稿では、テキストからの言い換え表現の抽出に基づく関連用語の収集手法の検討と、提案手法を実装したシラバス検索システムに関する報告を行った。本システムは、(1)シラバス検索そのもの、そして、(2)科目間の関連性の図示に基づく大学ごとの傾向の分析を目的としたものだが、本研究の関連用語収集手法は、(3)シソーラスの半自動構築などへの応用も可能である。

謝辞

本研究の一部は「科学研究費補助金若手研究（B）16700245」によるものであり、ここに謝意を表します。また、ご協力いただいた大学評価・学位授与機構「評価情報研究会（大学情報の構造解析による評価支援システム構築に関する研究）」参加者の皆様に謝意を表します。

注

- 1) 原田ら（1988）によれば、電気工学分野では、JICSTシソーラス中の階層関係のうち約45%が類種関係である。
- 2) 英語やフランス語などの欧米の言語とは異なり、日本語では、通常、語と語の間に区切り記号を置かないことから、語の単位や品詞体系の定義が難しく、必ずしもコンセンサスに至っていない。そのため、複数の定義が並存しているのが現状である（金田一、1982；村木、1985；松下、1989）。本研究の目的は実際的な応用システムの構築にあるため、語の定義をめぐる原理的な議論に立ち入ることを控え、処理に用いる形態素解析システム茶筌（松本ら、2000）に依拠して品詞体系を設定することにした。
- 3) 内容語が語中に挿入される場合、挿入後の結合パターンによっては、元の用語の下位語と見なせないケース

もある。例えば、「自動制御理論」学習は「自動学習」の下位語ではない。しかし、これも分野による違いはあろうが、情報処理分野の用語を調査した原田ら(1988)の報告によれば、そのようなケースは1割にも満たない。

- 4) 2名詞から成る複合名詞の構成要素間の関係には、並列、同格、修飾、格修飾があるが(石崎, 1987), ツリーを2系列出力するにあたって、並列関係の用語(「研究開発」など)は想定していない。そのような用語の場合、(iii)の言い換え規則は適用されず、関連語が抽出されないため、2つのツリーは同じものになる。
- 5) シラバスデータの形態素解析(形態素への分割と品詞の判定)には茶釜 version 2.3.2を用いている。

参考文献

Chen, H., Yim, T. and Fye, D. (1995) "Automatic thesaurus generation for an electronic community system," *Journal of the American Society for Information Science*, Vol.46, No.3, p.175-193.

大学セミナー・ハウス (1999)『大学力を創る：FD ハンドブック』, 東信堂：東京, 240p.

原田隆史, 細野公男, 田村俊作, 高柳敏子, 後藤智範, 岸田和明, 坂田亮子 (1988) “複合語の解析による語の上位 - 下位関係の自動抽出についての基礎的研究,” 三田図書館・情報学会研究大会, p.49-52.

井田正明, 野澤孝之, 芳鐘冬樹, 宮崎和光, 喜多一 (2004) “シラバスデータベースとシステム工学教育,” SSI2004 (計測自動制御学会システム・情報部門学術講演会2004), p.107-112.

Ishii, M. (1987) "Economy in Japanese scientific terminology," *Terminology and Knowledge Engineering '87*, p.123-136.

石崎雅人 (1987) “日本語複合名詞の解析,” 情報処理学会第35回全国大会, p.1315-1316.

Jacquemin, C. (1994) "Fastr: a unification-based front-end to automatic indexing," *Proceedings, Intelligent Multimedia Information Retrieval Systems and Management (RIA0'94)*, p.34-47.

Jacquemin, C. (1996) "A symbolic and surgical acquisition of terms through variation." In: Wermter, S. Riloff, E. and Scheler, G. (eds.) *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing*. Heidelberg: Springer. p.425-438.

Kageura, K., Yoshikane, F. and Nozawa, T. (2004) "Parallel bilingual paraphrase rule for noun compounds: concepts and rules for exploring web language resources," *The 4th Workshop on Asian Language Resources (ALR-04)*, p.54-61.

金田一春彦 (1982)『日本語のしくみ(金田一春彦・日本語セミナー, 2)』, 筑摩書房：東京, 349p.

小原恭介, 山田剛一, 絹川博之, 中川裕志 (2004) “ウェブを利用した関連用語収集,” FIT2004 (第3回情報科学技術フォーラム), p.183-184.

松本裕治, 北内 啓, 山下達雄, 平野善隆, 松田 寛, 高岡一馬, 浅原正幸 (2000)『日本語形態素解析システム『茶釜』version 2.2.1使用説明書』, 奈良先端科学技術大学院大学松本研究室：生駒, 21p.

松下大三郎 (1989)『標準日本口語法』, 増補校訂修訂第3刷, 勉誠社：東京, 534p.

Miyazaki, K., Ida, M., Yoshikane, F., Nozawa, T. and Kita, H. (2004) "On development of a course classification support system using syllabus data." In: *Computational Engineering, I: Advances in Continuum Mechanics and Electromagnetics*. Nagano: Japan Society for Computational Methods in Engineering (JASCOME). p.311-318.

文部科学省研究振興局情報課 (2003)『学術情報発信に向けた大学図書館機能の改善について：報告書』, 文部科学省研究振興局情報課：東京, 110p. (online), available from <<http://www.soc.nii.ac.jp/anul/j/documents/mext/kaizen.pdf>>, (accessed 2005-1-17).

村木新次郎 (1985) “形態論の輪郭,” 国文学解釈と鑑賞, Vol. 50, No. 3, p.39-44.

永松健司, 田中英彦 (1996) “コーパスから抽出した係り受け共起情報に基づく類似度と文書検索における評価,” 情報処理学会研究報告, NL-116, p.73-78.

日本工業規格 (1991)『シソーラスの構成及びその作成方法：JIS X 0901-1991』, 日本規格協会：東京, 24p.

野澤孝之, 井田正明, 芳鐘冬樹, 宮崎和光, 喜多一 (2005) “シラバスの文書クラスタリングに基づくカリキュラム分析システムの構築,” 情報処理学会論文誌, Vol. 46, No. 1. (to appear)

佐藤理史, 佐々木靖弘 (2003) “ウェブを利用した関連用語の自動収集,” 情報処理学会研究報告, NL-153-8, p.57-64.

Shapiro, S. C. and Eckroth, D.; 大須賀節雄監訳 (1991)『人工知能大辞典』, 丸善：東京, 1316p.

鶴丸弘昭, 井上 淳, 日高 達, 吉田 将 (1992) “語義文からの階層関係の自動抽出,” 電子情報通信学会技術研究報告, NLC92-2, p.9-16.

渡辺将尚, 絹川博之, 井田正明, 芳鐘冬樹, 野澤孝之, 喜多一 (2004) “Web 上のシラバス情報の収集と XML 変換,” FIT2004 (第3回情報科学技術フォーラム), p.121-122.

芳鐘冬樹, 井田正明, 宮崎和光, 野澤孝之, 喜多一 (2004a) “シラバスからの専門用語抽出手法の検討,” 情報処理学会第66回全国大会, p.4/375-4/376.

芳鐘冬樹, 野澤孝之, 辻 慶太, 影浦 峯 (2004b) “ウェブからの関連語・下位語の収集手法の検討と検索システムへの応用,” 第52回日本図書館情報学会研究大会発表要綱, p.113-116.

Yoshikane, F., Tsuji, K., Kageura, K. and Jacquemin, C. (2003) "Morpho-syntactic rules for detecting Japanese term variation: establishment and evaluation," *Journal of Natural Language Processing*, Vol. 10, No. 4, p.3-32.