意味的情報フィルタリング機能を有する 目的指向型外国語学習システムの実現

佐々木 朋美† 佐々木 史織‡ 清木 康††

†慶應義塾大学総合政策学部 〒252-8520 神奈川県藤沢市遠藤 5322
‡慶應義塾大学政策・メディア研究科 〒252-8520 神奈川県藤沢市遠藤 5322
††慶應義塾大学環境情報学部 〒252-8520 神奈川県藤沢市遠藤 5322

E-mail: †tsasaki@mdbl.sfc.keio.ac.jp, ‡sashiori@mdbl.sfc.keio.ac.jp, ††kiyoki@sfc.keio.ac.jp

あらまし 本稿では，外国語の自律学習において重要と考えられる習得度，到達目標，興味・関心情報を「学習コンテキスト」とし，個人の学習コンテキストに応じたネットワーク上の外国語学習向きマルチメディア・コンテンツを学習資源として自動的に選択・提供する目的指向型学習システムの実現方式を示す．本方式は，(1)ユーザの入力したサンプルから自動的に学習コンテキスト（習得度レベル・到達目標レベルを表す数値，および，興味・関心分野を表すメタデータ語群）を抽出する学習コンテキスト抽出機能，(2)外国語学習向きマルチメディア・コンテンツから自動的にメタデータ（難易度を表す数値，および，内容の専門分野を表すメタデータ語群）を抽出するメタデータ自動生成機能，(3)ユーザの言語習得度と到達目標を判定し，コンテンツを選択する学習レベル・フィルタリング機能，(4)学習者の興味・関心，専門分野情報を反映したコンテンツを提供する意味的情報フィルタリング機能から構成される．本方式の特徴は，学習者が自らの学習レベルや到達目標レベル，および，興味・関心に合致すると判断したサンプル・コンテンツを入力することのみにより，自身の外国語習得度や到達目標が自動的に判定され，興味対象・専門分野に応じた学習資源の自動的な獲得が可能となる点である．

キーワード 外国語学習，自律学習，意味的連想検索，パーソナライゼーション，情報フィルタリング

A Goal-Oriented Foreign Language Learning System with Semantic Information Filtering

Tomomi SASAKI† Shiori SASAKI‡ and Yasushi KIYOKI††

†Faculty of Policy Management, Keio University 5322 Endo, Fujisawa, Kanagawa, 252- 8520 Japan
‡Graduate School of Media and Governance, Keio University 5322 Endo, Fujisawa, Kanagawa, 252- 8520 Japan
††Faculty of Environmental Information, Keio University 5322 Endo, Fujisawa, Kanagawa, 252- 8520 Japan

E-mail: †tsasaki@mdbl.sfc.keio.ac.jp, ‡sashiori@mdbl.sfc.keio.ac.jp, ††kiyoki@.sfc.keio.ac.jp

Abstract In this paper, we present a goal-oriented language learning system for autonomous learning in foreign language study that enables the automatic publication of studying materials. As "Study Context", the user's purpose for studying, proficiency and interests will be reflected in the automatic selection and publication of multimedia contents (for foreign language study) on the network. Our system consists of four functions: (1) An automatic Study Context extractor that uses samples from a user's input as target data (2) An automatic metadata generator for multimedia contents (3) A study level filtering function that selects contents by determining the user's present proficiency and target levels (4) A semantic information filtering function to deliver language materials according to the student's interest or specialized field. By our method, users will be able to acquire learning materials that match their purpose for studying, proficiency and interests automatically, with the input of sample contents.

Keyword foreign language study, semantic associative search, personalization, information filtering

1. はじめに

外国語学習に関する研究においては，外国語学習には「自律学習」が肝要であると指摘されている．[1] 自律学習とは学習者が教師に頼らずに，教材，学習を開始する時間，学習の方法などについては自分で選択しておこなう学習のことであり，その成否については自分の責任とする学習方法のことをいう．

従って，自律学習には(1)視覚・聴覚的情報や非言語的現象による情報を含む自然なコンテンツ，(2)各学習者の学習履歴・学習進度に応じたコンテンツ，(3)各学習者の目的，到達目標，興味・関心情報を反映させたコンテンツが必要と考えられ，これらを提供するための学習支援システムが求められている．

また，外国語学習に関する研究においては，学習者

が自然な言語データにアクセスすることで，学習が前進することが指摘されている．[2]自然な言語データとは，視覚・聴覚的情報や，発話の状況，話者の表情や聞き手の反応などの非言語的現象による情報を含む言語データであり，外国語学習支援においてはこうした数多くの自然な言語データにアクセスできる環境を学習者に提供することが求められている．

近年のインターネットの普及とマルチメディア・コンテンツの爆発的増加に伴い，マルチメディア教材開発やCALL(Computer Assisted Language Learning)システム[3]の教育現場への導入が進み，外国語学習の環境や方式が多様化している．

しかし，従来の教材作成においては，能力レベルや学習目的に沿った教材のカスタマイズが困難である点や，多様化する学習目的に対応しきれない点が課題である．

関連する研究として，学習者の操作ログを記録して教師支援を目的とした学習者履歴を利用したシステムの研究[4]や，学習用マルチメディアコーパスの構築の研究[2][5]が挙げられる．文献[5]は母集団を特定した用例コーパスなど，ターゲットを絞った形で行われている点から学習目的型であるといえる．また，文献[6]のように教材作成のオーバーヘッドを軽減することを目的とした教材の自動生成の研究も行われている．

しかし,いずれも学習者の学習状況とコンテキストに対応する学習資源の選択は実現されていない．

外国語，特に英語に対して多様化する学習ニーズ[7]に応えるためには，各学習者の学習目的を反映した自律学習用学習資源と効率的な学習環境の実現が求められる．本稿では，習得度，到達目標，興味・関心情報を「学習コンテキスト」と呼ぶ．そして，各学習者が各自の学習コンテキストに合致する資源を「学習目的」を反映した学習資源とみなし，それら学習資源を自動的に提供する方式およびシステムを提案する．

本方式は，(1)ユーザの入力したサンプル・コンテンツから自動的に学習コンテキストを抽出する学習コンテキスト抽出機能，(2) 外国語学習向きメディア・コンテンツから自動的にメタデータを抽出するメタデータ自動生成機能，(3) ユーザの言語習得度と到達目標を判定し，コンテンツを選択する学習レベル・フィルタリング機能，(4) 学習者の興味・関心，専門分野情報を反映した言語データを提供する意味的情報フィルタリング機能から構成される．

文章の難易度を分析する方式として，既に提案されている語彙の利用頻度データからなるワードリストの作成法[8]やリーダビリティ（可読性）を計算する公式[9]を学習レベル・フィルタリング機能に適用し，ユーザの言語習得度と到達目標に合致するコンテンツを提供する．また,意味の数学モデルに基づく意味的連想検索方式[10]～[14]を意味的情報フィルタリング機能として応用し，ユーザの興味・関心と合致するコンテンツを提供する．

関連研究として，文献[10]～[17]が挙げられる．言葉と言葉の相関量を計量する意味の数学モデル[10]～[14]は，情報資源の意味的な関連性を計量する方式である．また，情報フィルタリングシステムは，ユーザのプロファイルに応じて，情報の選択を自動的に行うシステムである[15] [16]．文献[17] は，ユーザの入力したサンプル・ドキュメントをプロファイルとして使用し，意味的連想検索方式を用いてプロファイルと関係の高いドキュメントを自動的・選択的に配信するシステムの実現方式を示している．

2. 本方式の概要

本方式は，ユーザがサンプル・コンテンツを入力することにより，自身の外国語習得度や興味対象・専門分野に応じた学習資源の自動的な獲得を可能とするものである．

本方式の特徴は，次の三点にまとめられる．

1) 学習者が現在の言語能力に対応すると判断し，興味・関心に合致すると判断するサンプル・コンテンツを入力することにより，その学習者の学習レベル，および，そのコンテンツの分野（専門分野）に合致する学習資源を獲得可能である．

2) 学習者が到達目標と判断するサンプル・コンテンツを入力することにより，学習者の学習レベル向上を促す学習資源を獲得可能である．

3) 専門分野の知識を反映させた複数のベクトル空間を設定し，各空間に写像された対象メディア・コンテンツと学習者のサンプル・コンテンツとのマッチングをとることにより，学習者の興味・関心分野に最も合致する対象メディア･コンテンツが獲得可能である．

2.1 意味の数学モデルおよび意味的連想検索方式

本節では，本方式によって実現される目的指向型外国語学習システムの(1)学習コンテキスト抽出機能, (2)メタデータ自動生成機能, (4)意味的情報フィルタリング機能に適用される，意味の数学モデルおよび意味的連想検索方式[10]～[14]について概要を述べる．

2.1.1 メタデータ空間 *MDS* の設定

特定領域の知識を網羅的に表現した相関行列の固有値分解を通して，その領域に関するメディアデータ間の意味的な関連性の計量を行うための正規直交空間(以下，メタデータ空間 *MDS*)を設定する．

2.1.2 *MDS* へ写像する対象メタデータベクトルの設定

メタデータ空間 *MDS*S へ写像するための，検索対象メディアデータのベクトル化を行う．統一的な特徴(feature)で構成される相関行列でメタデータを表現しベクトル化することにより，同一空間上における検索対象メディアデータ間の意味的な相関が距離計算により求められる．

検索対象メディアデータ P には，メタデータとして t 個の基本データ $w_1, w_2, \ldots, w_t$ が以下のように付与されていることを前提とする．

$$P = \{w_1, w_2, \ldots, w_t\} \quad (1)$$

各基本語は，ベクトル表現された特徴で表される．

$$w_i = (f_{i1}, f_{i2}, \ldots, f_{in}) \quad (2)$$

各検索対象メディアデータは，メタデータとして付与されている t 個の基本語が合成されベクトル表現された後，メタデータ空間 *MDS* へ写像される．

2.1.3 *MDS* の部分空間(意味空間)の選択

検索者が与える複数の単語からなる文脈をコンテキストと呼ぶ．このコンテキストを用いてメタデータ

空間 *MDS* に各コンテキストに対応するコンテキストベクトルを写像する．これらのベクトルは，メタデータ空間 *MDS* において合成され，意味重心を表すベクトルが生成される．意味重心から各軸への射影値を相関とし，閾値を超えた相関値(以下，重み)を持つ重み付き軸からなる部分空間(以下，意味空間)が選択される．

2.1.4 *MDS* の部分空間(意味空間)における相関の定量化

選択されたメタデータ空間 *MDS* の意味空間において，検索対象メディアデータベクトルのノルム，もしくは，検索者の与えたキーワードベクトルとの間の距離を，コンテキストとの相関として計量する．これにより，与えられたコンテキストと各検索対象メディアデータとの相関の高さを定量化している．この意味空間における計量結果は，各検索対象メディアデータを相関の高さに応じてソートしたリストとして与えられる．また，検索対象メディアデータを特徴づける特徴の数が多い場合に，どのような意味空間が選ばれても意味空間におけるメディアデータベクトルのノルムが大きくなる傾向がある．そのために，本来，文脈との相関が強いと考えられるメディアデータベクトルのノルムよりも特徴の数が多いメディアデータベクトルのノルムが大きくなってしまい，適切な計量が行われないことがある．そのため，メタデータ空間における検索対象メディアデータベクトルの 2 ノルムによる正規化が選択的に行われる．

2.2 基本方式およびシステム構成

次の四つの機能から構成される本方式の実現システムを図 1 に示す．

2.2.1 学習コンテキスト抽出機能

ユーザの入力したサンプル・コンテンツから自動的に学習コンテキストを抽出する．抽出される学習コンテキストは，1-a)学習目的，到達目標，言語習得度と 1-b)ユーザの興味・関心情報に分けられる．前者は，「リーダビリティ」と「語彙力」という指標を用いてユーザの現在の言語習得レベルと到達目標レベルを表す数値として抽出される．後者は，専門分野ごとに生成されたワードセットとユーザが入力するサンプル・コンテンツを構成する単語群とのパターンマッチングによって，専門分野別のメタデータ言語群として抽出される．

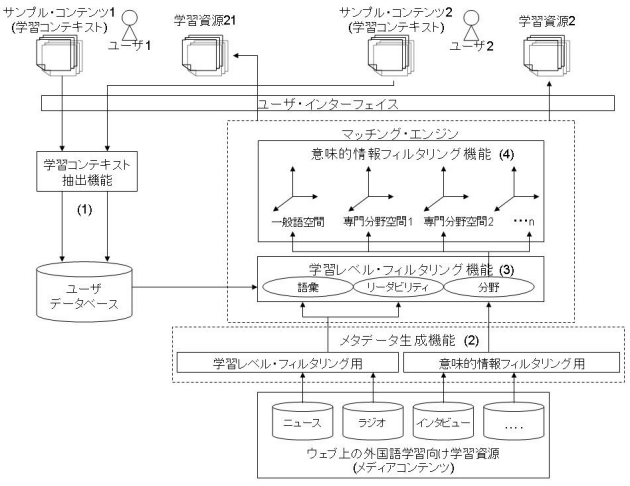


図 1 システム構成図

2.2.2 メタデータ自動生成機能

外国語学習向きメディア・コンテンツから自動的にメタデータを抽出する．抽出されるメタデータは，2-a)学習レベル・フィルタリング機能用メタデータ，および 2-b)意味的情報フィルタリング機能用メタデータに分けられる．前者は，(1)の学習コンテキスト抽出機能と同じ指標を用いて，コンテンツの難易度を表す数値として抽出される．後者は，専門分野ごとに生成されたワードセットと対象メディア・コンテンツを構成する単語群とのパターンマッチングによって，専門分野別のメタデータ言語群として抽出される．

2.2.3 学習レベル・フィルタリング機能

学習コンテキスト抽出機能によって抽出されたサンプル・コンテンツの 1-a)言語習得レベルと到達目標レベルを表す数値と，メタデータ自動生成機能によって抽出された 1-b)学習レベル・フィルタリング機能用メタデータ，すなわち対象メディア・コンテンツの難易度を表す数値をユーザ・データベースに格納し，リレーショナル・データベースを用いて対象ユーザの言語習得度と到達目標に合致するコンテンツを選択する．

2.2.4 意味的情報フィルタリング機能

学習コンテキスト抽出機能によって抽出された 2-a)サンプル・コンテンツの専門分野別のメタデータ言語群とメタデータ自動抽出機能によって抽出された 2-b)対象メディア・コンテンツの専門分野別のメタデータ言語群を対象とし，複数の専門分野別に生成されたメタデータ空間 *MDS* を用いて意味的連想検索方式[10]～[14]による相関量計量を行い，学習者の興味・関心，専門分野情報を反映した言語データを提供する．

3. 実現方式

本方式において，4 つの機能のうち，(1)学習レベル抽出機能，および，(3)学習レベル・フィルタリング機能は，ユーザが入力するサンプル・コンテンツと対象メディア・コンテンツからリーダビリティと語彙レベルの測定結果をメタデータとして抽出し，習得度レベルと到達目標レベルを閾値としたフィルタリングを行う．

また，(2)メタデータ自動生成機能，および，(4)意味的情報フィルタリング機能は，複数の専門分野別空間を並べ，ユーザが入力するサンプル・コンテンツと対象メディア・コンテンツから各空間別に生成された専門分野別メタデータを抽出し，各空間に写像されたサンプル・コンテンツと対象メディア・コンテンツとの相関量を計量し，相関量の高い順に表示する．

3.1 学習コンテキスト抽出機能

本機能は，ユーザのサンプル・コンテンツの入力に対して学習コンテキスト抽出を行い，メタデータを生成する機能である．(図 1 の(1))本方式においては，学習コンテキストを大別して(1)ユーザの習得度，到達目標 (2)ユーザの興味・関心の二種類に分類する．これらの学習コンテキストは，学習資源のメディアデータの属性を次のように設定し，自動的な抽出を行う．各属性につき具体例を図 2 に示す．

(a)マルチメディアデータより聴覚に関係する属性
(b)テキストデータより単語群に関する属性
(c)シチュエーションから抽出されるその他の属性

Aural	speech rate, pitch, volume, elevation
Text	**vocabulary, readability, field**
others	character structure, kansei information

図 2 外国語学習向けメディア・コンテンツの属性

ここでは，学習の目的を文章の読解力に限定し，テキストデータの(b)単語群に関する属性に焦点を当て，(1)のユーザの習得度，到達目標に対応する語彙力・リーダビリティと，(2)のユーザの興味・関心に対応する学習者の関心分野に特化してシステムを実現する．具体的には，以下のステップによって学習コンテキストの抽出を行う．

Step 1: ユーザによって入力されたサンプル･コンテンツから，習得度，到達目標を抽出する．

ここではテキストデータに対して難易度測定を行う方法を示す．文章の難易度を決定する要素は語彙・文法・構文など多様かつ複雑であり，難易度を定量化することは困難であるため，要素ごとに難易度を測定する必要がある．

Step1-1: 文章の難易度を測る一つの指標として，単語の利用頻度データを用いて語彙の難易度を測定する．WordNet Stoplist[18]，A General Service List of English Words by West [19]，The Academic Word List by Coxhead [20]を用いて頻出単語が 1000 語，2000 語，3000 語に分かれた語彙リストを作成する．これら語彙リストの単語群が各サンプル・コンテンツに占める割合を Range Program[21]によって計算することで，各テキストの語彙レベルを測定する．

Step1-2: テキストのリーダビリティ（可読性）を測定する．リーダビリティのスコアは各対象サンプル・コンテンツ内の文の長さ，単語の長さ及び文構造の相対的な複雑さなど複数の変数を使用して，必要とされる読解レベルを客観的に測定する尺度である．ここでは，広範囲に利用されている Flesch Reading Ease Score(FRES）を利用し，0(難) から 100(易) の値によって計量結果を出す．この値を学習コンテキストのメタデータとして利用する．FRES は次の式によって算出される．

FRES = 206.835 - 84.6 * 音節数/ 単語数- 1.015 * 単語数/文章数

Step 2: ユーザによって入力されたサンプル･コンテンツから,興味･関心に関するメタデータを抽出する．

ここではテキストデータを対象として，興味・関心に関する情報をメタデータ語群として抽出する方法を示す．具体的には以下のステップを実行する．

Step2-1: 各テキストデータ中に含まれる語のうち，専門家が作成した対訳表を用いて各分野の用語（複合語を含む）を特定し，活用や語尾の変化などを統一する．

Step2-2: 専門分野ごとに生成されたワードセットと対象メディア・コンテンツを構成する単語群の集合和をとり，各専門分野空間に対応した分野別メタデータ言語群として生成する．

入力されたコンテンツ C からその中に含まれる n 個の単語 $w_1, w_2, \ldots, w_n$ を抽出する．複数の分野別空間 MDS^j（j=1~k; j は分野識別子である）の m 個の専門用語を $f^j_{r-1}, f^j_{r-2}, \ldots, f^j_{r-m}$ とするとき，空間 MDS^j におけるコンテンツ群 C に付与されるメタデータ語群 M^j は各空間の専門用語のみで構成される．したがって，ある専門的なコンテンツは，合致する専門分野空間に写像される場合，空間 MDS^j の m 個の専門用語との共通部分が多いため,メタデータ語群が多めに付与される．逆に分野が異なる空間に対して写像される場合は，コンテンツに付与されるメタデータ語群が少なく付与される．メタデータが付かない場合は相関量が 0 になる．専門性が高いコンテンツは，分野が合致する空間において相関量が高く計量され，選択されることが期待される．

3.2 対象メディア・コンテンツのメタデータ自動抽出機能

本機能は，ウェブ上からメディア・コンテンツを収集した後，コンテンツの難易度を判定し，一般語空間および各専門分野空間ごとに分野に依存した専門性の高いメタデータを自動的に生成する機能である．

Step 1: 対象メディア・コンテンツから学習レベル・フィルタリング機能用のメタデータを抽出する．

3.1 で記述されている，学習・コンテキスト抽出機能の Step1 と同様の指標および方法によって，対象メディア・コンテンツの難易度を数値として算出する．

Step 2: 対象メディア･コンテンツから意味的情報フィルタリング機能用のメタデータを抽出する．

3.1 で記述されている，学習・コンテキスト抽出機能 Step2 と同様の方法によって，対象メディア・コンテンツの専門分野別メタデータ語群を抽出する．

3.3 学習レベル・フィルタリング機能

ユーザプロファイルを抽出するためのユーザ・データベースと学習資源を有するマルチメディアコーパスを構築し，それを用いて学習レベルのフィルタリングを行う．ユーザの入力したサンプル・コンテンツに対して 3.1 のプロセスで生成したメタデータを参照し，各指標につき数値が最も低いサンプルを習得レベル，最も高いサンプルを到達目標レベルを示すものとし，外国語学習向けメディア・コンテンツに対する学習レベルのフィルタリングを行う．

また，専門分野の英語力を習得することが学習目標である学習者に対しては，各専門分野で流通する言語資料からなる学習資源が必要だと考えられる.同様に，基礎的な言語能力の習得が目標である場合は，一般的な言語資源を使った学習資源が必要だと考えられる．

従って，意味的情報フィルタリングにおいて複数の専門分野空間で検索を行うのか，一般語から構成される空間のみで検索を行うのかを判別することによって，上記のようにターゲットとなる学習資源のフィルタリングを可能とする．3.2 で提示したリーダビリティに関するメタデータ抽出法より, Reading Ease Score に相当する米国の学年レベル(Flesch-Kincaid Grade Level)を算出できる．このレベルに対してユーザが任意の閾値を設定することで自動判別機能を実現する．

3.4 意味的情報フィルタリング機能

本機能は，学習レベル・フィルタリング機能によって絞り込まれた対象メディア・コンテンツに対し，ユーザの興味・関心に合致した分野のメディア・コンテンツを自動的に選択し，提示する機能である．

Step1: 分野別に構築された複数の専門分野別空間

を設定する.

Step2: 3.2 のメタデータ自動抽出機能によって各空間別に生成された対象メディア・コンテンツの専門分野別メタデータをベクトル化し，各専門分野空間に写像する.

Step3: 3.1 の学習コンテンツ抽出機能によって抽出されたサンプル・コンテンツのメタデータをコンテキストとして与え，部分空間を選択する.

Step4: 各専門分野空間の各部分空間においてコンテキストと対象メディア・コンテンツの相関量を計量し，相関量の高い順に表示する.

本方式の実現にあたり，異種の専門空間の一例として，情報通信分野の空間（以下IT 空間）と国際関係分野の空間（International Relations 空間,以下IR 空間），一般語の定義を用いた一般語の空間（以下，LG空間）を設定した.IT 空間生成にはオンラインの専門用語集[22]，IR 空間生成には当該分野で汎用的な専門辞書[23]を用い，一般語空間にはロングマン英英辞書[24]を用いた．各専門空間は，IT 空間が154次元，IR 空間が710次元，ロングマン辞書によるLG空間は1761 次元となっている.

4. 実験

ここでは広域ネットワーク上のテキストデータからなる学習資源を対象に，意味的情報フィルタリング機構を有する目的指向型外国語学習システムの実現可能性を検証する．3.1 の方式により実現した学習コンテキスト抽出機能およびユーザ・データベース，3.2 の方式において実現したメタデータ自動抽出機能，3.4 で提示した専門的な IT 空間，IR 空間，および一般語の LG 空間を用いて実験を行った．対象テキストデータとして，1)ロイター通信社の International News10 件[25]，2)Internet Headlines のオンラインニュース記事各 10 件[26]，3)文学作品 10 件[27]を設定した．専門家が実際に選んで IR 空間において関連が高いと予想される 1)を IR テキスト(ir_text1, ir_text2, .., ir_text10)と呼び,これを IR 空間において上位にランキングされるべき正解セットとする．同様に 2)を IT テキスト(it_text1, it_text2, .., it_text10)と呼び，IT 空間における正解セット，3)を LG テキスト(lg_text1, lg_text2, .., lg_text10)と呼び，LG 空間における正解セットする.

また，ユーザモデルとしては，図 3 のようにユーザの学習レベルと興味・関心の組みあわせは複数パターン考えられるが，本実験においては，具体例として結果の比較考察が容易な次の二つのユーザモデルを設定する.

・基礎的な言語習得度を持ち，かつ，文学に興味・関心を持っているユーザ 1(Ⅰ)

・言語習得度が高く，かつ，専門的な国際関係論との関連で外国語学習を行っているユーザ 2(Ⅳ)

	一般	専門的領域
学習レベル(基礎)	Ⅰ	Ⅱ
学習レベル(高度)	Ⅲ	Ⅳ

図 3 実験全体のデザイン

表 1 学習コンテキストとして抽出された ユーザ 1 とユーザ 2 の学習レベルに関するスコア

	リーダビリティ	語彙L1	語彙L2	語彙L3	その他
ユーザ1					
サンプル1	87.6	36.73	19.24	0.87	43.15
サンプル2	75.4	72.25	16.25	0.5	11
ユーザ2					
サンプル1	55.9	46.29	8.96	8.21	36.57
サンプル2	48.9	45.52	15.17	19.31	20

表 2 対象メディア・コンテンツの 学習レベル判定用メタデータ生成例

	リーダビリティ	語彙L1	語彙L2	語彙L3	その他
ir_text1	41.7	43.22	8.9	11.86	36.02
ir_text2	49.6	44.1	11.03	12.82	32.05
it_text1	43.2	39.05	5.92	14.2	40.83
it_text2	45.2	35.59	8.47	22.03	33.9
lg_text1	77	65.06	14.77	1.42	18.75
lg_text3	7.5	55.83	12.2	5.15	26.83

4.1 実験 1

4.1.1 実験 1-1：学習コンテキスト抽出機能の検証

本実験では，学習レベルの異なるユーザが入力した各サンプル・コンテンツから，学習コンテキストのうちの学習レベルのメタデータ抽出が機能することを示す．サンプル・コンテンツは，各ユーザが自分の学習レベルと合致していると判断し，また，興味・関心分野と合致していると判断するテキストである．語彙 L1-L3 は 1,000 語からなる頻出単語リスト[19]～[20]であり，L1 から L3 に上がるにつれて範囲度が上がる．表 1 は，学習レベルの異なるユーザ 1 とユーザ 2 が，習得レベルを示すサンプル 1 および到達目標レベルを示すサンプル 2 を入力した結果を示している．高い習得度を持つユーザ 2 からは，基礎的な語学力を持つユーザ 1 の入力したサンプルよりも低いリーダビリティの結果（難易度が高い）と，難易度の高い語彙群 L3 の値が高く算出されるという想定通りの結果が得られた．また，各ユーザの各サンプルのうち，到達目標レベルを示すサンプル 2 からは，習得レベルを示すサンプル 1 よりも可読性の値が低く（難易度が高く），語彙力に関しては難易度の高い語彙群 L3 の値が高いことが分かる.

4.1.2 実験 1-2：メタデータ自動生成機能の検証

本実験では，学習レベル判定用メタデータ生成機能が適切にメタデータを生成していることを検証する．表 2 は，学習資源となるテキストデータを対象に，リーダビリティと語彙レベル判定を行った結果を示している.難易度の高い専門用語を多く含む IR テキストおよび IT テキストのリーダビリティの値は一般語を多く含む LG テキストのリーダビリティよりも低く（難易度が高く），語彙力に関しては難易度の高い語彙群 L3 の値も高いことが分かる.

4.1.3 実験 1-3：学習レベル・フィルタリング機能の検証

実験 1-1 と実験 1-2 の結果を用いて，学習レベル・フィルタリング機能の検証を行う．表 3 は，対象テキストデータをリーダビリティの値の低い順にソーティングし，ユーザ 1 とユーザ 2 の各サンプルデータとを照合した結果を示している．基礎的な語学力を持つユーザ 1 に対しては習得レベル 87.6 と到達目標レベル 75.4 の間のリーダビリティの高い（難易度の低い）LG テキストが選択され，習得度の高いユーザ 2 に対しては習得レベル 55.9 と到達目標レベル 48.9 の間のリーダビリティの低い（難易度の高い）IR テキストと IT テキストが選択されていることが分かる．

4.2 実験 2：意味的情報フィルタリング機能の検証

4.2.1 実験 2-1

本実験では，ユーザの入力したサンプル・コンテンツから学習コンテキストのうち，興味・関心情報を抽出するための意味的情報フィルタリング用メタデータ生成機能が適切に機能していることを示す．表 4 に，メタデータ抽出例を示す．興味・関心が小説にあるユーザ 1 の入力したサンプルには LG 空間を構成する一般語のメタデータが多く付与され，興味・関心が国際関係にあるユーザ 2 の入力したサンプルには IR 空間を構成する専門用語が多く付与されていることが分かる．

4.2.2 実験 2-2

本実験では，対象テキストデータから，意味的情報フィルタリング機能用の専門分野別メタデータ語群が適切に抽出できることを検証する．表 5 は，メタデータ抽出の例を示している．IR テキストには IR 分野の専門用語が，IT テキストには IT 分野の専門用語が，LG テキストにはロングマン英英辞書の一般語が多く付与されていることが分かる．

表 3 学習レベル・フィルタリング機能の実験結果

対象メディア・コンテンツ		サンプルデータ	
:			
ir_text2	49.6	sample2	48.9
it_text5	50.1		
ir_text6	52.2		
lg_text8	53.2		
ir_text4	54.4		
ir_text5	55.9		
ir_text8	59		
it_text10	59.8	sample1	55.9
ir_text9	62.2		
lg_text2	64.7		
lg_text4	72.9		
lg_text3	74.3		
lg_text7	75.8	sample2	75.4
lg_text1	77		
lg_text6	78.2	sample1	87.6

ユーザ1
ユーザ2

また，対象メディア・コンテンツに対して，生成された専門分野また，全ての対象テキストデータ 30 件について空間別メタデータの占める割合を図 4 に示す．IR テキスト 10 件には IR 空間用メタデータの割合が，IT テキスト 10 件には IT 空間用メタデータの割合が，LG テキスト 10 件には LG 空間用メタデータの割合が比較的多くを占めていることが分かる．

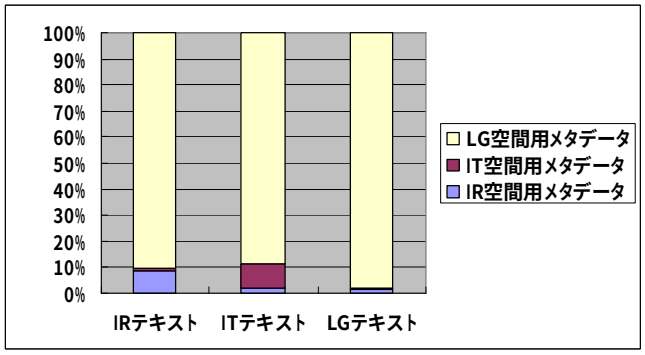
表 4 サンプル・コンテンツから抽出されたメタデータの一例

	分野別空間	サンプルに付与されたメタデータ
ユーザ1 - 興味関心:小説		
サンプル1 (習得レベル)	IR	
	IT	
	LG	name down bowl face mirror cry long begin call air whistle white stay pocket cross dark laugh point slow tower black cheek brush
サンプル2 (到達目標レベル)	IR	food aid international war treaty …
	IT	
	LG	bird son golden fox into sit morning find watch cage hear night apple away shoot good begin stand pass back …
ユーザ2 - 興味関心:国際問題		
サンプル1 (習得レベル)	IR	genocide war international justice authority militia immigration asylum
	IT	storage
	LG	charge crime member war time country …
サンプル2 (到達目標レベル)	IR	food aid international war treaty state power peace population boundary independence
	IT	
	LG	food government official year time world need international million foreign …

表 5 対象メディア・コンテンツの専門分野別メタデータ生成例

テキストID	IR分野の単語	IT分野の単語	一般語
ir_text1	negotiation region security aid peace conflict refugee force justice militia governance	monitor	talk round agreement violence group hold end official start peace attack main thousand year …
ir_text2	food aid international war treaty state power peace population boundary independence	link	report international call official into decision justice court son council end foreign add …
it_text1	technology	Web software Internet file link	sign deal use company agree show industry agreement copy music film into help month …
it_text2	security	hacker software network Internet	computer system threat protection good home pay risk need tell under business grow …
lg_text1		mouse	down fall begin time eat way well wonder either never right picture cat rabbit …
lg_text3	front actor realism		art beautiful long through able seem beauty picture form send face find …

図 4　対象テキストデータにおける専門分野空間別メタデータの占める割合 （IR テキスト 10 件，IT テキスト 10 件，LG テキスト 10 件）



4.2.3 実験 2-3

本実験では，マッチングエンジンとしての意味的情報フィルタリング機能が適切に機能することを検証する．

実験 2-3-1: サンプル・コンテンツのメタデータ一つ一つを個別にクエリとして与え，ランキングの上位 m 件のうち出現回数が n 以上の結果を求める．

閾値を m=10，n=10 と設定した場合の結果を表 6 に示す．結果として正解である IR テキストが 6 件，IT テキストが 1 件，LG テキストが 2 件，という妥当な結果が得られた．

実験 2-3-2: (a)サンプル・コンテンツのメタデータ全てクエリとして与える，(b)TF*IDF 値でランキングをした場合の上位 k 件のメタデータをクエリとして与える．

表 7 は，ユーザ 2 のサンプル 1 と 2 をクエリとし，(a)サンプルのメタデータ全てをクエリとして検索した結果と，(b)TF*IDF 値の高い上位 10 件のメタデータをクエリとして検索した結果を示している．(a)の場合，上位のサンプルテキスト 11 件のうち，正解である IR テキストが 6 件，また(b)の場合，12 件のうち正解である IR テキストが 8 件選択される，という妥当なフィルタリング結果を得た．

実験 2-3-1 および実験 2-3-2 の実験結果は，国際関係論との関連で外国語学習を行っているユーザ 2 の興味関心に対応しており，意味的情報フィルタリングによってユーザプロファイルのうち，興味・関心情報が反映されたテキストが正しく選択されていることが分かる．

4.3 実験 3：フィルタリング機能を組み合わせによる検証

本実験では，学習レベル・フィルタリング機能と意味的情報フィルタリング機能を組み合わせることにより，外国語学習向けメディア・コンテンツのうち，ユーザの学習目的・言語習得度に応じ，また興味・関心を反映した学習資源が獲得できることを検証する．

実験 1 において学習レベルによるフィルタリングがなされた学習資源用テキストに対して，実験 2-3 で提示した 3 通りの方法でクエリを与えて意味的情報フィルタリングを行う．学習レベル・フィルタリングでは，言語習得度を表すサンプル 1 の難易度と到達目標を表

表 6 サンプル・コンテンツのメタデータを個別にクエリとして与えた場合の結果上位 10 件

順位	メタデータ					
	genocide	war	international	justice	authority	・・・
1	ir_text4	ir_text3	lg_text3	ir_text2	ir_text2	
2	lg_text4	it_text4	ir_text2	lg_text3	lg_text3	
3	it_text4	ir_text2	ir_text6	ir_text6	ir_text6	
4	lg_text9	lg_text9	ir_text3	ir_text3	ir_text3	
5	lg_text6	ir_text8	ir_text1	ir_text1	ir_text8	
6	it_text6	ir_text6	ir_text8	ir_text8	it_text4	
7	it_text10	ir_text9	lg_text4	it_text4	lg_text9	
8	lg_text8	lg_text3	lg_text2	lg_text9	ir_text1	
9	ir_text3	ir_text5	it_text4	ir_text4	lg_text4	
10	it_text3	ir_text1	lg_text9	it_text8	it_text10	

表 7 (a)サンプル・コンテンツのメタデータ全て，(b)TF*IDF 値の高いメタデータをクエリとして与えた場合の実験結果

(a)		(b)	
テキストID	相関量	テキストID	相関量
lg_text3	0.272985	ir_text2	0.31787
ir_text2	0.271086	lg_text3	0.3058
ir_text6	0.225053	ir_text6	0.265024
ir_text3	0.21654	ir_text3	0.250804
ir_text4	0.185037	it_text4	0.225667
lg_text9	0.183101	ir_text8	0.203514
lg_text4	0.176111	ir_text4	0.203186
ir_text8	0.175202	ir_text9	0.181016
ir_text1	0.170362	ir_text1	0.17863
it_text4	0.168776	lg_text9	0.176376
lg_text2	0.150005	ir_text5	0.165541
:		lg_text4	0.157684
		:	

すサンプル 2 の難易度の間のテキストが選択され，さらに意味的情報フィルタリング機能によって，各ユーザが興味関心に合致したテキストが選択されることが期待される．

ユーザ 1 と 2 の実験結果を表 8 に示す．基礎的な語学力を持ち，興味・関心が小説にあるユーザ 1 には難易度の低い LG テキスト(lg_text7, lg_text6, lg_text1)が，言語習得度が高く，興味・関心が国際関係にあるユーザ 2 には難易度の高い IR テキスト(ir_text6, ir_text2, ir_text4, ir_text8)が上位に選択されていることが分かる．学習レベル・フィルタリングと意味的情報フィルタリング機能の組み合わせにより，各ユーザの学習コンテキストに応じて学習資源の自動選択・提供の実現可能性が確認された．

以上，実験 1，2，3 を通してサンプル・コンテンツと対象メディア・コンテンツを対象としたメタデータ自動抽出機能，学習レベル・フィルタリング機能と意味的情報フィルタリング機能を実装し，各機能ごとの実装実験(実験 1，2)および各機能を組み合わせた検証実験(実験 3)を行うことで，本方式による目的指向型外国語学習システムの実現可能性を示した．

表 8 学習レベル・フィルタリング機能と意味的情報フィルタリング機能を組み合わせたユーザ別の結果

意味的情報フィルタリング方法：	実験2-3-1		実験2-3-2(a)		実験2-3-2(b)	
ユーザ	テキストID	上位出現回数	テキストID	相関量	テキストID	相関量
ユーザ1	lg_text7	15	lg_text7	0.22661	lg_text7	0.159458
	lg_text6	9	lg_text1	0.21228	lg_text6	0.61385
	lg_text1	7	lg_text1	0.20757	lg_text1	0.160336
ユーザ2	ir_text6	15	ir_text2	0.27109	ir_text2	0.31787
	ir_text2	14	ir_text6	0.22503	ir_text6	0.265024
	ir_text4	11	ir_text4	0.18504	ir_text8	0.203514
	ir_text8	11	ir_text8	0.1752	ir_text4	0.203186
	lg_text10	7	lg_text8	0.12006	lg_text8	0.116188
	it_text8	6	it_text10	0.11493	it_text10	0.109036
	it_text5	0	it_text5	0	it_text5	0

5. まとめと今後の展望

本稿では，外国語学習向きマルチメディア・コンテンツを対象として，意味的情報フィルタリング機能を有する目的指向型外国語学習システムの実現方法を示し，その実現可能性を検証した．本方式によって実現されたシステムを用いることにより，外国語学習者は個人の学習レベル，興味・関心分野に合致し，かつ，個人の学習レベル向上を促すような学習資源の自動的な獲得が可能となる．

今後の研究の課題として，本方式における学習レベル・フィルタリング機能の計量方法に関する考察および意味的情報フィルタリング機能の実験および，マルチメディアの特性を生かした聴覚的特性のメタデータ抽出とそれらに対応した学習レベル・フィルタリングの実現を行う．

6. 謝辞

本稿の執筆にあたり，多くのご助言を頂いた吉田尚史先生と鷹野孝典氏（慶應義塾大学政策・メディア研究科）に感謝いたします．

文　献

[1] 大木充，田地野彰，浅田健太郎，“自律学習と学習者の動機づけに対する CALL の有効性 - 自律学習支援環境の構築に向けて”，フランス語教育，32，pp87-100，2003.

[2] 佐藤滋，李相穆，“マルチメディアコーパスからのコロケーション情報の抽出と，その日本語学習支援への応用”，平成 11-14 年度科学研究費補助金特定領域研究 A「メディア教育利用」研究成果報告書平成１４年度研究計画，pp.239-244.

[3] 日本教育工学復興会外国語学習システム調査研究部，活用用例一覧，http://www.japet.or.jp/call_biz/2_jirei/index.htm.

[4] 三田泰正，藤岡健史，荻野哲男，高田秀行，上林弥彦，“学習履歴を利用した動的な問題提示を行う学習システムの提案”，電子情報通信学会第 15 回データ工学ワークショップ，Mar. 2004.

[5] 朝尾幸次郎，“英語学習者音声コーパスの作成と利用に関する研究『高等教育改革に資するマルチメディアの高度利用に関する研究』”，平成 13 年度科学研究費補助金（特定領域研究(A) A02: 外国語教育の高度化の研究）研究成果報告書，pp. 63-66.

[6] 佐野洋，“個人適合の学習教材自動生成を実現する語学教育システム”，電子情報通信学会第 13 回データ工学ワークショップ，Mar. 2002.

[7] 伊藤健二，“e-Learning の最前線”，情報処理，Vol.43, No.4, pp392-400, 2002.

[8] 染谷泰正，“AWK による語彙レベル分布計測プログラ WordLevel Checker”，http://www.someya-net.com/.

[9] Thomas Oakland, Holly B.Lane, “Language, Reading, and Readability Formulas: Implications for Developing and Adapting Tests”, International Journal of Testing, Vol. 4, Issue 3, pp.239-252, 2004.

[10] Kiyoki, Y., Kitagawa, T. and Hayama, T., “A metadatabase system for semantic image search by a mathematical model of meaning”, ACM SIGMOD Record, Vol. 23, No. 4, pp.34-41, 1994.

[11] Kiyoki, Y., Kitagawa, T. and Hitomi, Y., “A fundamental framework for realizing semantic interoperability in a multidatabase environment”, Journal of Integrated Computer-Aided Engineering, Vol.2, No.1, pp.3-20, John Wiley & Sons, Jan. 1995

[12] 清木康，金子昌史，北川高嗣，“意味の数学モデルによる画像データベース探索方式とその学習機構”，電子情報通信学会論文誌，D-Ⅱ，Vol.J79-D-Ⅱ，No.4, pp.509-519, 1996.

[13] 宮川祥子，清木康，“特定分野ドキュメントを対象とした意味的連想検索のためのメタデータ空間生成方式”，情報処理学会論文誌：データベース，Vol.40, No.SIG5(TOD2), pp.15-28, 1999.

[14] Sasaki, S., Kiyoki, Y. and Yakushiji T., “Semantic Space Creation and Associative Search Methods for Document Databases of International Relations”, Proceedings of the 7th IASTED International Conference on Internet and Multi- media Systems and Applications, pp.399-405, August 2003.

[15] 土方嘉徳，“情報推薦・情報フィルタリングのためのユーザプロファイリング技術”，人工知能学会誌，19, pp.365-372, 2004.

[16] 帆足啓一郎，松本一則，井ノ上直己，橋本和夫，“非適合プロファイルを利用した文書フィルタリング手法”，情報処理学会論文誌，42, 3, pp.507-517, 2001.

[17] 高松耕太，倉林修一，佐々木史織，清木康，“異種領域ドキュメント群を対象にしたコンテキストの動的計量を伴う選択的情報配信機構の実現”，研究報告書データベースシステム，No.2005-DBS-137.

[18] WordNet, http://wordnet.princeton.edu/.

[19] West, M., 1953 “A General Service List of English Words. London: Longman, Green and Co.”

[20] Coxhead, Averil, “A New Academic Word List”, TESOL Quarterly, 34(2), pp.213-238.

[21] Heatley, A., Nation, I.S.P. and Coxhead, A., “Range and Frequency programs”, 2002, http://www.vuw.ac.nz/lals/staff/Paul_Nation.

[22] F. Monster: “Computer glossary,” 2005.

[23] Evans, Graham and Newnham, Jeffrey: “Dictionary of International Relations, Penguin Books”, 1998.

[24] Longman Dictionary of Contemporary English, Longman, 1987.

[25] Reuters Technology and Science/ Internet News Headlines, http://today.reuters.com/news/newsChannel.aspx?type=internetNews.

[26] Reuters International News Headlines, http://today.reuters.com/news/newsChannel.aspx?type=worldNews.

[27] Project Gutenberg, http://www.gutenberg.org/.