浮世絵表題の多言語検索をめざしたシソーラス構築法に関する研究

代表研究者 山田 奨治 国際日本文化研究センター研究部助教授

共同研究者 早川 聞多 国際日本文化研究センター研究部助教授

1 研究調査の目的

最近の高度情報化社会の進展に伴って、美術工芸品等の文化財をデジタルコンテンツとして電子化し、ネットワークを通じて閲覧するといった、あたらしい利用法が可能となってきた。また、インターネットとWWWに代表される、広域性や利便性をもったメディアの登場により、国境・言語を越えた情報の交流の機会が格段に増加している。しかし、このような状況では、情報アクセス時における言語の壁が大きな問題である。問題は言語の壁のみならず、コンテンツ表題表現における文化的・専門的知識の欠落も、検索精度向上のためには障害となる。これらの問題に対処するために、ネットワーク上に存在する日本美術（とくに、浮世絵情報）を言語や文化的・専門的知識に依存せずに検索閲覧可能とする、「多言語情報ゲートウェイ」の構想を示す。当面想定するアクセス言語は英語と日本語で、このゲートウェイの主な機能的役割は、つぎの通りである。

1) コンテンツに付随するキャプション等の日本語テキストを英語に自動翻訳しデータベース化する。(自動翻訳には、用例に基づく翻訳手法が有力である。)

2) 利用者の検索要求を解釈して、上記のデータベース中から適切なものを検索して一覧を示し、利用者の選択により、画像付きのコンテンツに導く。

将来においてこのような機能を実現するためには、つぎのような観点から、対象となるコンテンツに関する語と語の類義関係を記述したシソーラスが必要である。

- ・用例にもとづく翻訳手法では、用例データベースのなかから翻訳対象表現ともっとも類似した対訳用例を見いだす必要があるが、そのさいに語の類義関係を示す情報が重要な手がかりとなる。

- ・利用者が指定した検索語のみで検索すると、所望するコンテンツのうちの一部しか得られない等の問題があり、検索目的を満足するためには、文化的・専門的知識にもとづいて指定された類義語または関連する語も含めて検索をおこなう処理が必要である。

このような要求を満たすシソーラスの構築方法について研究するのが、本研究の目的である。

本研究で目的とするような、検索および翻訳に役立つシソーラス構築は、「浮世絵」という比較的範囲の明確な適用対象で、実証的な研究をおこなうことにより、実用的および学術的な見地から、じゅうぶんな知見を獲得しうる研究テーマと位置づけることができる。

前述した情報流通の広域化・多様化を背景として、国際日本文化研究センターでは日本固有の優れた美術・文化財コンテンツを海外に発信するための方式を検討するため、「日本美術データベース」を構築し、内外の日本研究者に限って公開している。このデータベースに蓄積された浮世絵表題情報をベースに、今後は自動翻訳技術を活用できる方向性を探りたい。ほとんどが複合名詞である表題を自動翻訳するためには、用例にもとづく翻訳手法が適しているが、この手法では翻訳対象表現ともっとも類似した対訳用例が抽出できるかがおおきな課題である。この類似用例を見いだす処理では、既存の汎用的な類語辞典を用いるのが一般的であるが、翻訳精度の点で問題を抱えているのが現状である。また、検索においても、入力キーワードだけを用いて検索したのでは、検索結果にもれが生じるため、シソーラス等を用いて類義語に展開する必要がある。

日本美術データベースでの実践経験にもとづいて、検索や自動翻訳のために不可欠な知識源として、対象固有分野のシソーラスの構築を目指すに至ったものである。

情報検索の分野において、目的とする検索結果を導くために、シソーラスの参照が有効であることが指摘されているが、検索に特化したシソーラスはあまり例がない。たとえば国立民族学博物館で開発されたマルチメディア検索システムにおいては、収蔵品名表現のシソーラスが作成されているが、汎用性のあるものではない。

一般に入手可能な日本語シソーラスとして、(1)類義語新辞典(角川書店)、(2)分類語彙表(国立国語研究所)、(3)日本電子化辞書(EDR)の概念辞書、(4)NTTの日本語語彙体系、がある。(1)および(2)は、人間が文書を書く際の手助けとなることを目的として作られていると考えられ、(3)および(4)は言語を計算機で扱うために作成され

ている。ただし、(4)は電子化された媒体として入手できない。

このように、本研究のような人文科学的な領域でのシソーラス作成の先行例は、ほとんど見られない。

2 方法

本研究は、シソーラスの構築手法の研究である。シソーラスの作成では語の上位下位関係や類義関係などの決定をおこなう必要がある。この作業は、人間の考えによるところがおおきく、機械による自動化が困難な作業である。そのため、効率的にシソーラス作成をおこなうためには、まず基礎的検討をじゅうぶんにおこない、その後実際に応用した結果をフィードバックさせてシソーラスの改良をおこなうといった手順をとる必要がある。これらの作業を1年間で実施するのは困難であり、2年間の研究期間を設定して研究をおこなっている。

シソーラス作成においては、用例から得られる語の類義関係を示す情報が重要な手がかりとなる。このような語の類義関係に対しては視点の問題を考慮する必要がある。すなわち、ひとつの語も別の視点から見れば別の意味の側面に焦点が当たることになる。例えば、「女形(おやま)」「立役(たちやく)」「おたふく」「ひょっこ」という語は、「性別」という視点から見れば、「女形(おやま)」「おたふく」「立役(たちやく)」「ひょっこ」とそれぞれ分類され、「劇中における役割」という視点から見れば、「女形(おやま)」「立役(たちやく)」「おたふく」「ひょっこ」と分類されると考えられる。このように、シソーラスは実際には単純な木構造とはならず、あるノードに複数の親ノードが存在するのが一般的であるが、汎用的になればなるほど、視点の問題まで考慮した高精度なシソーラスの構築は現実的には不可能となる。そこで、本研究では、「浮世絵」という比較的範囲の明確な適用対象を設定し、視点も考慮したシソーラスの構築手法に関して、実証的な研究をおこなっている。

本研究では、まず、浮世絵表題に関する対訳用例データを作成し、シソーラス作成のための単語の抽出を一部実施した。そして、来年度以降の研究においてシソーラスの作成をおこない、その情報を用語辞書および用例データベースに反映させる。その後、浮世絵表題の検索、検索結果の分析、シソーラスの改良を繰り返しおこなうことで研究を進める。平成11年度は具体的には、浮世絵表題を対象にした用語辞書および対訳用例データベースの拡充をおこなった。具体的には、国際日本文化研究センター作成による日本美術データベース(非公開分を含めて約10000件)を分析して、浮世絵表題で使用される用語に関する対訳用例データベースおよび用語辞書の作成をおこなった。

平成12年度以降は、浮世絵の有識者の意見、参考文献等に基づきシソーラスを作成し、作成した用語辞書・対訳用例データベース中の各用語にシソーラス上の分類カテゴリーを付加していくことを予定している。

3 結果

平成11年度の研究成果として、(1)浮世絵表題の対訳用例データの作成と(2)浮世絵表題シソーラスの作成に向けた単語の抽出について述べる。

3.1 浮世絵表題の対訳用例データの作成

国際日本文化研究センターでは、従前より日本美術作品の画像データベースを整備し、その一部を公開している(<http://www.nichibun.ac.jp/dbase/japan.htm>)。公開データベースには文字情報の英訳も掲載してあるが、当該英訳は日本研究者向けの英訳となっているため、翻訳困難な用語はあえて翻訳せず、日本語音のローマ字表記にとどめているケースが多い。たとえば「見立て」は「Mitate」と表記するなどである。多言語検索のためのシソーラス作成を視野に入れた場合、このままでは対訳用例としての使用に耐えない。そこで、シソーラス作成をにらんで、英訳をやり直す作業をおこなった。たとえば「見立て」は、「Representation of ~」といった風に、翻訳困難な日本語概念を積極的に英語に置き換えるという方針を貫くことにした。英訳対象とした浮世絵表題は、画像公開分の約5000件に非公開分に加えて約10000件である。

浮世絵表題は、「扇屋内花扇」「鬼もん喜兵衛河原崎権十郎」のように、名詞が連続する複合名詞の形態をとるものが多い。それらをただ単に字句通りに翻訳したのでは、原意を表現できないことがほとんどである。「扇屋内花扇」は、「扇屋」という揚屋の遊女「花扇」の意味である。その場合は、「The Courtesan Hanaoogi of Oogiya」と訳さなければならない。同様に、「鬼もん喜兵衛河原崎権十郎」は、役者の河原崎権十郎が「鬼もん喜兵衛」を演じている浮世絵なので、「Kawaharazaki Gonjuuroo in the Role of Kimon Kihei」と訳さなければ、原意を伝えることができない。同じように、「大星由良之介」は劇中登場人物なので、「The Actor in the Role of Ooboshi Yuranosuke」と訳すのが正しい。

これらのように、浮世絵表題の英訳は、すべてもとの浮世絵作品の画像を参照し、表題が伝えんとするところの意味を、できる限り正確に伝えるように配慮しながら翻訳作業を進めた。作品を参照しながらの作業であったので困難を極めたが、

平成11年度に約10000件の浮世絵作品の表題について、英訳を付与することができた。今後、翻訳の正確さの検証を進めなければならないが、シソーラス作成のための対訳用例データとしては、ある程度の規模を確保できたのではないかと考える。

付与した英訳の一部を参考1に添付した。なお今回作成した英訳は、浮世絵表題対訳データの暫定版として、インターネット公開を予定している。

3.2 浮世絵表題シソーラスの作成に向けた単語の抽出

浮世絵表題のシソーラス作成に向けた作業として、浮世絵表題に登場する単語の抽出を開始した。作業手順として、表題をなるべくおおきな概念で単語抽出をし、比較的数多くみられる種類の単語を集めて当座のラフな分類カテゴリーを作成して分類を進めた。現在の分類カテゴリーは、つぎの29種類になっている。

場所 固有の地名 風景 店名 遊女名 役者名 人名 その他の人名
動物名 植物名 品物 物語名 年中行事 年号 その他
見立て 寺社関係 歌詠み 職人 屏風 美人図 講 遊女 七福神
楽器 婦人 風俗 戦争 歌舞伎その他

これらの分類カテゴリーは、作業者の便宜的な理由によって定めたものなので、現段階では概念レベルが錯綜しており、このままシソーラスのカテゴリーとして使用することはできない。現在までに抽出した単語数は、4400単語である。単語リストの一部を参考2に添付した。

作業進行上の問題点として第一に、作品の題目と内容が一致しないケースがあることがあげられる。たとえば、題目に「A祭礼」とあっても、絵の内容は風景（祭礼ではない）としかみえないものがある。その場合Aという単語は祭礼名にするべきか、風景に分類するべきか簡単には判断できず、作品に関するより多くの情報を集めなければ分類を決定できない。第二に、浮世絵に女性が描かれている場合、それが市井の女性なのか遊女なのか決めたいという問題がある。これに関しても、作品に関するより高度な調査をしながら進める必要があるといえる。第三に、種々の雑多な内容が、とくに「風俗」「その他」といったカテゴリー蓄積されている。適当な時点でそこに蓄積されている単語を分析して、サブカテゴリーに分類する必要がある。第四に、現段階の分類が作業者の主観的判断によっている点に問題がある。作業者自身も作業を進めるうちに、概念分類の判断基準にゆらぎが生じており、ある程度の単語量を蓄積した段階で、これまでの経験を踏まえて概念分類を再度見直す必要があるように思われる。

今後、これらの単語をシソーラスとして構築していくうえにおいて、既存の日本語シソーラスを参考にしながら、できればある程度それと整合性を保ちながら浮世絵表題シソーラスを作成できることが理想的である。既存の日本語シソーラスとしては、国立国語研究所の分類語意表のようなものがモデルとして考えられる。今後は分類語意表を参考にしながら、浮世絵表題シソーラスの構成を構想していきたい。

4 今後の予定

幸いにも本研究に対して平成12年度も助成をお認めいただいた。それを受けて平成12年度は、浮世絵表題からの単語抽出とシソーラスへの作り込みの作業を続行する。作成したシソーラスは、当初から内容的な完全性を追及するのではなく、作業途中の暫定的なバージョンであってもそれを積極的に公開し、他研究者からのフィードバックを受けながら内容を改訂していく方針を取りたい。

参考1 浮世絵表題の対訳用例データ

八段目	Act Eight
(荒獅子男之助と仁木弾正)	[Actors in the Roles of Arajishi Otokonosuke and Nikki Danjoo]
蒲原	Kanbara
関	Seki
相州袖ヶ浦	Sodegaura, Sooshuu
上州妙義山	Myoogisan, Jooshuu
桜花の顔ばせ	A Face as Fair as a Cherry Blossoms
(見立源氏?)	[Representation of the Tale of Genji]
(縁先官女)	[Court Lady on a Veranda]
(見立源氏若菜下)	[Representation of the Tale of Genji: Wakana, Part II]
玉や内花紫	The Courtesan Hanamurasaki of Tamaya
梶原平三坂東彦左衛門	The Actor Bando Hikozaemon in the Role of Kajiwara Heizoo
平治景高中村芝翫	The Actor Namura Shikan in the Role of Heiji Kagetaka
丸海老屋内江川	The Courtesan Egawa of Maruebiya
源頼光	Minamoto Yorimitsu
上松江田源三	Edagenzoo at Uematsu
石山暮雪鈴木重幸	Suzuki Shigeyuki in Ishiyama on a Snowy Night
神功后皇	Empress Jingoo
(海中の為朝)	[Tametomo in the Sea]
安部の保名くずの葉	Abe no Yasuna and Kudzu Leaves (Kuzunoha)
子雪姫	Rat: The Courtesan Yukihome
午於美輪	Horse: The Courtesan Omiwa
戌白井権八	Dog: Shirai Gonpachi
猪早太忠澄	Inohayata Tadazumi
朝夷義秀雌雄鱷を捕ふ	Asahina Yoshihide Capturing a Crocodile
図	

篠塚伊賀守重広と栗生	Shinozuka Shigehiro, Lord of Iga and the Justice Kuryuuza
左衛門尉頼賢	emon Yorikata
尾上栄三郎	The Actor Onoe Eizaburoo
関羽駕細工	Scene from the Kabuki Play Kan'u (Guan Yu)
田舎女	Peasant Woman
うつぼざる	Scene from the Kabuki Play, Utsubozaaru (Monkey Quver)
玉屋内歌川	The Courtesan Utagawa of Tamaya
日本橋通一丁目略図	View of First Street, Nihonbashi
浅草川首尾の松御厩河	"Good Results Pine" at Ommaya Bank, Asakusa River
岸	
深川洲崎十万坪	Juumantsubo (10,000-Acre) Plain at Suzaki, Fukagawa
赤坂桐畑雨中夕けい	Evening View, Paulownia Plantation at Akasaka in Downpour
	r
鹿嶋踊	Kashima Dance
年の市勇初恋	First Love at the Year-end Fair
風流大和ならざらし	Bleaching Nara Sarashi Ramie Cloth in Yamato
扇屋花扇	The Courtesan Hanaoogi of Oogiya
丸海老屋内増花	The Courtesan Masuhana of Maruebiya
(緑短冊を読む女)	[Green]: [Woman Reading a Tanzaku]
(金魚を見る母子)	[Mother and Child Watching Goldfish]
(狨を抱く女)	[Woman Holding a Small Dog]
あらし山さくら	Cherry Blossoms at Arashiyama
(杯を持つ茶屋女)	[Teahouse Waitress Holding a Sake Cup]
扇屋内花扇	The Courtesan Hanaoogi of Oogiya
紫式部	Murasaki Shikibu
陸奥	Mutsu
(機織)	[Weaving]
(御殿山の花見)	[Flower-Viewing at Gotenyama]
(三保の松原道中)	[On the Road to Mihoo no Matsubara]
(暮の餅搗)	[Making Rice Cakes at the End of the Year]
玉屋内花紫	The Courtesan Hanamurasaki of Tamaya
弁天	Benten
春駒	Hobby Horse
玉屋内花紫夕顔	The Courtesans Hanamurasaki and Yuugao of Tamaya
行列廊の花鑓	Procession in the Pleasure Quarters
田舎育女ことふれ	Country Women Singing and Dancing

参考2 浮世絵表題からの単語の抽出（作業途中）

場所	固有の地名	風景	店名	役者名	役名
浅草観世音	赤阪	薄雲	丸海老屋	荒木の直蔵	赤松四郎
浅草金龍山	あかさか	駅	赤蔦屋	嵐橋蔵	秋津嶋
浅草本願寺	浅草川	江戸名所築地浪除勝景	有松絞店	嵐団之助	悪の源太
芦辺東屋	浅草反甫	楓の下	家田屋	嵐りかく	遊君阿古屋
	麻楚山鳥居崎	篝火	海老屋	市川右団治	安達が原ノ ばば
吾妻橋	朝日山	楽屋三階	大かなや	市川緞十郎	安守勇蔵
市ヶ谷八幡	足利行道山	楽屋二階	扇屋	市川海老造	飯沼勝五郎
蔵島内侍	足柄山	くものかけはし	角町松葉や	市川こま蔵	生駒之助
田舎家	亜墨利加国	五ヶ国	ごふく店	市川鮭十郎	一寸徳兵衛
稲荷乃社	あたけ	小橋	さのや	市川眼十郎	井筒屋伝兵衛
稲荷社	愛宕山	菖蒲の前	芝屋	市川森之助	梅王丸
岩亀楼	吾妻の森	諸国名橋連理の梓	大丸屋	吉次	梅若
上野中堂ニツ堂	吾妻橋	聖堂	大文字屋	さの川万きく	遠藤武者
海浜	あら井	月下	田川屋	沢むら田の介	おかん
梅屋舗	荒井	月乃岬	玉屋	三代目瀬川路孝	おくみ
雲水福禅寺	あらし山	出口之柳	玉弥	三柘大五郎	お七
永代ばし	嵐山	西塔	中万字屋	三柘源之助	鬼童丸
大阪府鉄道寮	阿波	野分	丁子屋	実川延太郎	お祭佐七
ステン所					
岡崎駅	伊賀上野	橋間	つた屋	実川延之介	おまつり佐七
音無川堰	勢州	浜辺	鶴屋	実川鯨蔵	加藤清正
神奈川駅	石部	風景	丁子や	瀬川万蔵	神原佐五郎
亀戸天神	いしべ	ふじのゆき	てうじ屋	瀬川路孝	義太夫語竹

川口のわたし	石薬師	風呂場	橋本屋	中むら翫雀	房 吉三
川崎駅	石山	幻	兵庫屋	中村千之介	九紋龍
神田明神社	伊勢	山吹の里	豆茶屋	中むら福助	芸者おしゆ ん
祇園社	市ヶ谷	遊廓	三浦屋	中むら駒之助	芸者小糸
岸上	いちの谷	床下の場	紫若	中村弥之助	恋ぎぬ
金閣寺	一石ばし	雪中	木綿店	二世市川門 之介	斉藤龍興
金州城	猪王山	雪の三囲土手	柳原写真 所	萩野扇女	桜丸
草津駅	茨住吉	わたし場	やなぎや	坂東喜三郎	貞任
雲上	今戸ばし	井の頭の池弁財天 の社雪の景	吉田屋	坂東大吉	忍小僧弥太 郎
溪山	岩川	雨中	四ツ目屋	坂東養助	小鮎源五郎
玄関先	上州妙義 山	横浜商館繁栄之囃	若那屋	まつしまや 片岡我堂	白藤源太
孤亭	上野	吉原雪	岡本屋	むめや繁松	崇徳院
五百らかん寺さ ざみどう	上野山	吉野山之景	恵比寿屋 店先	河原崎三升	捨若丸
賽の河原	上野山した	近江八景	若松屋	坂東竹松	宗任
逆井のわたし	宇治	御座敷踊舞台	難波屋	市川姉蔵	曾我五郎
佐太村天満宮	宇治川	江戸姿八契	尾張屋	市川団六	袖萩
三十三間堂	碓井	江戸名所八景	料理茶屋	勝川又吉	平ノ清盛
三拾三軒堂	内川関屋 の里	四季八景	三浦屋	沢村宗之助	瀧口今助
三紅亭	宇津の山	新刻日光名勝十二 景	休み処こ じまや	沢村百之助	滝口今助
四天王寺	宇津之山	瀬田夕照	玉や	中村亀松	瀧夜しや姫
住吉神社	馬爾加国	石山秋月	ふじ屋	中村小伝治	天狗小僧霧 太郎
住吉本社	永代橋	窓下の戯れ	尾州楼	市川延升	常盤御前
関屋	江尻	其おもかげうつす夕 げしき	浅草かさ 屋	下り中むら 歌右衛門	新田梅次郎
善光寺	蝦夷	唐崎夜雨	つるや	三枅森蔵	二の瀬ノ源