

言語横断情報検索における画像手がかりを用いた インタラクティブな翻訳曖昧性解消の評価

林 良彦^{1,a)} 永田 昌明² サワシュ ボラ¹

受付日 2011年12月20日, 採録日 2012年4月7日

概要: インタラクティブな言語横断情報検索において, Web から取得した画像がクエリの翻訳曖昧性解消のための手がかりとしてどのように有用であるかを実験に基づき評価した. 8名の被験者が参加した, 92語からなるクエリセットによる評価実験の結果によれば, 言語的な手がかり(逆翻訳)と比べて, 画像手がかりは適切な訳語の選択率向上に有意に寄与したのに対し, 訳語選択に要する時間の短縮に関する寄与は有意ではなかった. このクエリセットは, クエリ語の語義が持つ意味特徴や親密度といった特性が画像手がかりの得られやすさとの間にどのような関係を有するかについて, 人手によって評定したデータから構成したものである. この評定データからは, 抽象概念であっても特定の意味特徴を持つ場合には適切な画像手がかりが得られやすいこと, 原言語において高い親密度を持つ語義に対する訳語に対しては適切な画像手がかりが得られやすいこと, Web から画像を取得する場合, 特に固有名に関するバイアスが問題になる場合があることなどが分かった. これらの結果は, インタラクティブな言語横断情報検索インタフェースの設計に有用な指針を与える.

キーワード: 言語横断情報検索, インタラクティブシステム, 翻訳曖昧性, 語義, Web 画像

Interactive Resolution of Translation Ambiguities with Web Images for Interactive Cross-language Information Retrieval

YOSHIHIKO HAYASHI^{1,a)} MASAOKI NAGATA² BORA SAVAS¹

Received: December 20, 2011, Accepted: April 7, 2012

Abstract: Interactive search environments are gaining more attractions than ever, particularly in the setting of cross-language information retrieval, where the query language and the document language are different. This paper, through experiments, shows that images acquired from the Web toward a translation of a user-issued query term can serve as an effective clue in selecting proper translations; hence they may be effective in improving the cross-language search effectiveness. This paper further discusses that the semantic feature and the sense familiarity of a query term could affect the utility of the associated images obtained from the Web.

Keywords: cross-language information retrieval, interactive system, translation ambiguity, word sense, Web image

1. はじめに

世界各地から発信される多様な情報をユーザの母国語を

用いてアクセスする手段の実現が望まれている. このための核となる技術は, 言語横断情報検索 (Cross Language Information Retrieval: 以下, CLIR) と呼ばれる. CLIRにおいては, ユーザによるクエリの言語 (以下, クエリ言語) と検索対象となる文書の言語 (以下, 文書言語) が異なるため, 通常の単言語の情報検索手法を適用する前段階において, 言語の違いに対処することが必要になる. その方法論として主流となっているのは, ユーザのクエリを文書

¹ 大阪大学
Osaka University, Toyonaka, Osaka 560-0043, Japan

² NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories, Souraku, Kyoto
619-0237, Japan

a) hayashi@lang.osaka-u.ac.jp

言語に翻訳するクエリ翻訳のアプローチである。対訳辞書を用いたクエリ翻訳においては、未知語をどのように扱うか、また、辞書中に複数存在する訳語の中からユーザが意図する意味に最もよく適合するものをいかに選択するかが大きな課題となる。本研究では、クエリ語の多義性に起因する訳語の曖昧性解消という後者の課題を対象とする。

この課題に対しては、訳語が目的言語において有する文脈や共起情報を利用するなどの言語処理的な観点からの多くの研究がなされてきた [17] が、一方で、積極的にユーザの介入を取り入れたインタラクティブなシステムの研究も行われている [20]。インタラクティブなシステムにおいては、ユーザが曖昧性の解消のための主要な資源であると位置づけられるため、システムはユーザが適切な判断を行えるような有効な手がかりを提示することが望まれる。特に CLIR においては、ユーザが文書言語における十分な能力や知識を有していない可能性を前提として、用いる手がかり情報のタイプを定める必要がある。

以上のような観点から、本論文では、言語に対する依存性が低い手がかりとして、クエリ語の訳語に対して Web から取得した画像（以下、Web 画像）を利用することを考える。Web 画像が翻訳曖昧性解消の手がかりとして有効に働くことを確かめるためには、まず、(a) 意図する語義を反映する適切な Web 画像がどの程度得られるかを明らかにし、次に、(b) 適切な Web 画像がある程度得られるという状況において Web 画像がどの程度有効であるかを検証する必要がある。

以下では、クエリ語の語義を持つ特性と Web 画像との関係を調査した結果を 3 章において報告する。これは、上記の (a) の課題に対応する。また、上記の (b) の課題に対応して行った有用性の評価実験について 4 章で述べる。この評価実験においては、3 章に述べた結果に基づいて設定したテストクエリセットを用いた。なお、これらの議論の前提として、想定する CLIR インタフェース [14] の概要を 2 章で述べる。また、関連する研究状況と今後の方向性について、それぞれ 5 章、6 章にまとめる。

2. 画像を手がかりとするインタラクティブな CLIR

Web 画像をユーザによる訳語選択の手がかりとして用いるインタラクティブな CLIR インタフェース [14] のシステム構成を図 1 に示す。

まず、ユーザにより入力された原言語によるクエリを解析し、内容語をクエリ語として抽出する。次に、対訳辞書などを利用することにより、各クエリ語に対する訳語を得る。一般にクエリ語は多義を有するため複数の訳語が得られるが、ユーザが意図する語義に対応する訳語を選択できることが重要である。そこで、候補となる各訳語をクエリとして外部の画像検索エンジンに送ることにより、ユー

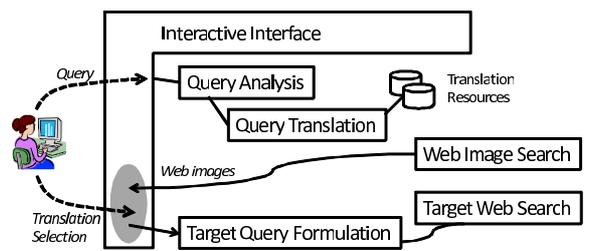


図 1 インタラクティブな CLIR システムの構成
Fig. 1 Organization of an interactive CLIR system.



図 2 クエリ訳語に対する画像手がかりの提示
Fig. 2 Displaying translation candidates with Web images (for “bus”).



図 3 クエリ訳語に対する画像手がかりの提示
Fig. 3 Displaying translation candidates with Web images (for “bass”).

ザによる訳語選択のための画像手がかりを得る。

図 2, 図 3 に「ソウルのバス」という日本語のクエリに対する CLIR インタフェースの情報提示の例を示す。「ソウル」に対して、“soul”, “Seoul” の 2 つの訳語があり、「バス」に対しては、“bus”, “bath”, “bass” の 3 つの訳語が得られている。マウスを各訳語に重ねることにより、対応する Web 画像が表示されるので、ユーザはこの画像を参考にして、適切な訳語をチェックボックスにより選択する。図 2 では、乗り物としての “bus” にふさわしい画像が得られているので、この語義を意図している場合は、無条件にチェックを付けることができる。一方、図 3 では、少なくとも “bass” には楽器と魚の 2 つの語義があることが示さ

れるが、ユーザの意図する語義が“bass”に対応しうるかどうかは判断することができる。

検索インタフェースは、選択された訳語群を基にターゲットのサーチエンジンに対する検索クエリを生成し、最終的な検索を実行する。本インタフェースは、特定の外部エンジンに固有の処理は行わないため、画像サーチエンジン、ターゲットのサーチエンジンともに Web 上で利用可能な適当なものを利用することが可能である。

3. 語義と Web 画像

クエリ翻訳の曖昧性解消のための手がかりとして画像を用いる主な理由は、直感的であること、言語に対する依存性が低いことである。画像に表出されている情報は、もちろん例外はあるものの、多くの場合は言語や文化に独立な普遍的な情報であると期待できる。しかしながら、つねに画像手がかりが有効であるわけではなく、適切な検索インタフェース実現の観点からは、クエリ語の特性、特に意図される意味（語義）の観点から、クエリ語に対する語義-訳語ペアごとに、それに対する画像手がかりの有用性が予測できることが望ましい。

また、このような画像手がかりをどこに求めるかという問題も存在する。現在のところ、抽象的な語義にも対応するような幅広い画像手がかりを収集・整理した画像データベースは存在しないことから、画像検索エンジンなどを用いて動的に収集する Web 画像を適切に用いる方法を開発することが望まれる。

そこで、Web 上で利用可能な画像検索エンジンを単純に用いることにより、意図する語義を反映する適切な Web 画像がどの程度得られるかを調査した。調査は以下に示す考え方・手順により行い、その結果を基に「語義・画像適合性評価データセット」と呼ぶデータセットを構築した [13]。次節で述べる Web 画像の有用性を評価する実験に用いたテストクエリセットは、本データセットを基に設定された。

3.1 利用する言語資源

画像認識の分野においては、画像に現れている事物が何であるかを認識しようとする一般物体認識 [23] の研究が活発化している。このような研究の促進のために、たとえば、英語の語彙データベースである WordNet の語義概念 (synset) に対して Web 画像を対応付ける資源 (ImageNet) の構築 [6] が進められており、このような資源に基づいてテストクエリセットを構成することが考えられる。しかしながら、ImageNet ではいわゆる具体物の属性を持つ名詞のみが対象とされているうえ、語義が持つ「なじみの程度」といった特性と関連付けることができない。また、日本語を対象としたものではないという問題もある。

そこで、日本語において基本的であると考えられる単語を語義分けし、各語義に対する語義別親密度を付与した基

本語データベース [3] を利用し、ここから評価の対象とする日本語単語を選択し、これらに対して Web 画像を収集することとした。ここで語義別親密度とは、語義の「なじみの程度」を数値化したもので、その値域は 1~7 である。基本語データベースに収録されている 45,951 件の語義に対しては、その平均値は 4.65、標準偏差は 1.09 であると報告されている*1。

また基本語データベースにおいては、各語義に対してその定義が日本語で付与されているため、これをそのまま次節で述べる評価実験において使用することができる。しかし、基本語データベースに付与されている語義は階層的な分類体系によっていないので、大まかな意味の特徴による傾向などを調べるのが難しい。そこで、日本語語彙大系 [15] の階層的な意味属性体系へ対応付けした結果 [8] を用いた。

3.2 語義・画像適合性評価データセット

データセットの構築は以下のような手順による。

まず、語義親密度が 2.0 以下の語義を除去した後の語義数が 2 から 5 である名詞を基本語データベースから抽出した。この条件は、(1) 数多くの語義を有する語義曖昧性の高い語を避ける、(2) 情報検索において多く用いられると思われる名詞に品詞を限定する、という理由による。この結果として、2,380 語（全部で 5,540 語義）が抽出された。次に、これらの各語義ごとにその意味に適合する英訳語を付与した。訳語の付与にあたっては、複数の日英対訳辞書から得た訳語の中から人手による選択を行った。さらに、これらの訳語をクエリとして画像検索 (Google Images) を行い*2、検索結果の上位 20 件のサムネイル画像の中に語義を視覚的に表現している画像が含まれている度合いに基づき、各語義に対する画像適合度 (0 から 3 の 4 段階) を人手により付与した。

より具体的には、上位 8 件までに複数の適切な画像が得られている場合に画像適合度 = 3、これに限らず複数の適切な画像が得られている場合に画像適合度 = 2、1 つは適切な画像が得られている場合に画像適合度 = 1 とし、上位 20 件の画像の中に 1 つも適切な画像がない場合に画像適合度 = 0 とした。たとえば「学者」に対しては、1: 研究者、2: 学問を身につけた人の 2 つの語義があり、1 の語義 (訳語: scholar) に対する画像適合度は 3 となるが、2 の語義 (訳語: learned person) に対する値は 0 となった。

画像適合度の分布を表 1 に示す。全体的として、67.0% の語義-訳語ペアに対して、1 以上の画像適合度を得た。以下では、画像適合度が 0 のものを画像適合性: 無、1 以上の

*1 語義別親密度は、40 名もの被験者による評定結果を平均することにより算出された値である [3] ことから、個人に依存するバラツキが抑えられた平均的な傾向を示すデータとして、多くの心理言語学的実験で用いられている。

*2 2010 年 1 月に画像検索を実施。

表 1 画像適合度の分布

Table 1 Distribution of the relevance score.

画像適合度	語義数 (%)
3	1,803 (32.5)
2	1,114 (20.1)
1	794 (14.3)
0	1,829 (33.0)

表 2 分割表：14 の意味カテゴリと画像適合性の関係

Table 2 Contingency table: 14 semantic classes versus relevance groups.

意味カテゴリ	画像適合性：無	画像適合性：有
人	135 (<) 172.67**	388 (>) 350.33**
組織	14 (<) 19.81	46 (>) 40.19
施設	12 (<) 33.01**	88 (>) 66.99**
地域	25 (<) 43.58**	107 (>) 88.42**
自然	10 (<) 22.78**	59 (>) 46.22**
動物	10 (<) 43.91**	123 (>) 89.09**
植物	1 (<) 15.19**	45 (>) 30.81**
自然物	13 (<) 20.80*	50 (>) 42.20*
人工物	92 (<) 257.84**	689 (>) 523.16**
抽象物	270 (>) 212.28**	373 (<) 430.72**
人間活動	322 (>) 300.43	588 (<) 609.57
事象	94 (>) 78.24	143 (<) 158.76*
自然事象	55 (<) 78.57**	183 (>) 159.43**
抽象的關係	776 (>) 529.88**	829 (<) 1,075.12**

ものを画像適合性：有とする。

3.3 語義の意味分類と画像適合性

語義の持つ意味的な特徴と画像適合性との関係を調べるため、日本語語彙大系における 14 の上位レベルの意味カテゴリ*3との関係を χ^2 検定により調べたところ、全体としては意味特徴と画像適合性の有無には、有意な偏りがある ($p < .001$) ことが確認された。

χ^2 検定に引き続き残差分析の結果を表 2 に示す。『人』

*3 具体の大分類：人、組織、施設、地域、自然、動物、植物、自然物、人工物。抽象の大分類：抽象物、人間活動、事象、自然現象、抽象的關係。本文中では『』を用いて記す。

表 3 分割表：親密度レベルと画像適合度の関係

Table 3 Contingency table: familiarity levels versus relevance scores.

親密度レベル	画像適合度				語義数合計
	0	1	2	3	
c3 ($f \leq 3.0$)	55 37.31**	17 16.20	20 22.72	21 36.78**	113
c4 ($3.0 < f \leq 4.0$)	255 212.61**	97 92.30	123 129.50	169 209.59**	644
c5 ($4.0 < f \leq 5.0$)	781 685.05**	305 297.39	422 417.25	567 675.31**	2,075
c6 ($5.0 < f \leq 6.0$)	662 769.57**	329 334.08	486 468.72	854 758.63**	2,331
ch ($f > 6.0$)	76 124.46**	46 54.03	63 75.81	192 122.70**	377
合計	1,829	794	1,114	1,803	5,540

から『人工物』までの 9 つのカテゴリが『具体』の大分類、『抽象物』から『抽象的關係』までの 5 つのカテゴリが『抽象』の大分類に属する。表における各セルの上段は頻度の実測値、下段は期待値を示す。実測値の後に付された不等号は、期待値との差の正負を示し、期待値と実測値との隔たりに関する有意水準をマーク (* : $p < .01$, ** : $p < .005$) により示している。たとえば、『人』のカテゴリに関しては、全 523 件のうち画像適合性：無と評定されたものは 135 件であり、これは表 1 の分布による期待値 172.67 より有意に ($p < .005$) 少ないことを表している。

全体的な傾向としては、『具体』の大分類に分類される語義では期待値を上回る実測値が観測され、『抽象』の大分類に分類される語義では期待値を下回る実測値が得られている。このことは、予想されるとおり、具体的な意味概念に対する画像は得やすいの対し、抽象的な意味概念に対する画像は得にくいことを示唆している。

しかしながら、すべての意味カテゴリがこの傾向に従うわけではない。たとえば、『具体』に大分類される『組織』の意味カテゴリにおいて、画像が得られにくい語の例として、「財閥/financial combine」, 「二軍/farm team」などがあつた。その一方で、『抽象』の大分類に分類される語義においても、『自然現象』のように期待値を上回る実測値が観測される意味カテゴリや、『人間活動』のように差異が有意でない意味カテゴリが存在した。前者の例には、「雨」, 「風」や「波」といった気象現象に関連する語が含まれる。また、『人間活動』の意味カテゴリに関しては、それを行っている状況が画像となりやすい(例：「テニス」をしている画像)という性質によると考えられる。

3.4 語義の親密度と画像適合度

語義の持つ親密度と画像適合度の関係についても同様の検討を行った結果を表 3 に示す。表において、 f は親密度を表しており、 f の値によって 5 段階 (c3, c4, c5, c6,

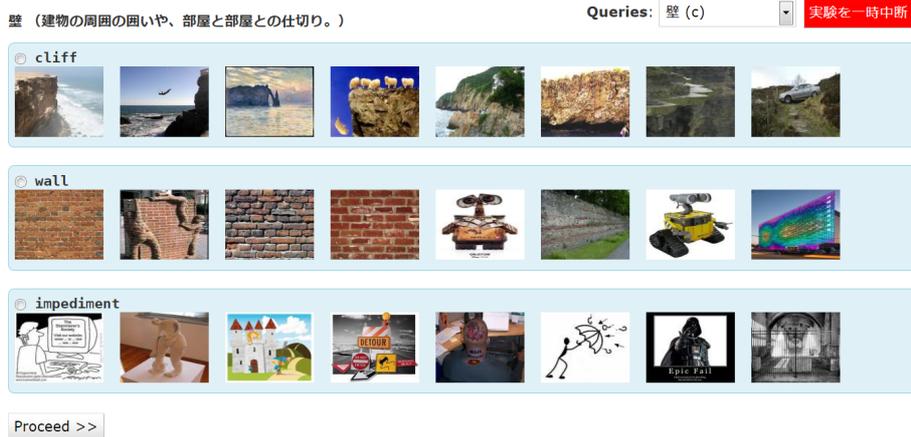


図 4 評価用インタフェースのスクリーンショット例 (mode-3)

Fig. 4 Screen shot of the evaluation interface (for mode-3).

ch) にレベル分けしている。

表 3 からは、親密度の高い (> 5.0) 語義は有意に画像適合度が高く、親密度の低い (≤ 5.0) 語義は有意に画像適合度が低いという明確な傾向を確認することができる。これは、日本語において高い親密度を有する語義に対する英訳語から得た Web 画像は日本語における語義との適合性が高い傾向にあることを示唆する。この原因については、さらに詳細な分析を行う必要があるが、親密度は心像性 (imageability) と高い相関を有するという報告 [21] と符合する結果となっている。ここで心像性とは、単語から喚起される種々の感覚イメージの思い浮かべやすさを表す主観的特性値である。

3.5 Web バイアス：固有名の問題

本研究では、既存の画像検索エンジンを利用して Web 画像を収集する。このため、その結果は Web 上の情報の分布を反映したものである。すでに単語語義解消問題において報告されている [11] ように、単語が本来的に持つ語義を反映した分布とは異なる可能性がある。

このような Web バイアスの問題は、画像検索のクエリとして利用される英単語が人名や商品名などの固有名として使われる場合に顕著である。たとえば、「機織り」という日本語クエリに対して得られる“weaver”という訳語をクエリとして画像検索を行うと、著名な女優の写真画像が多くヒットしてしまう。これは、ユーザが一般的な語義を想定し、一般語をクエリ語として用いる場合に問題となる可能性がある。

4. 画像手がかりの有用性の実験的評価

情報検索においては適切な検索結果が効率良く得られることが重要であり、有効性 (effectiveness)、効率性 (efficiency) が評価の観点となる。有効性の代表的な評価指標は、検索結果における精度、再現率である。一方、効率性

については様々な要因が関わるため標準的な評価指標は存在しない。

本論文では、特にユーザによるクエリ語の訳語選択における画像手がかりの有用性を評価するため、(a) 適切な訳語の選択率、(b) 訳語選択の過程に要する時間の 2 つを評価指標とする。適切な訳語の選択は、検索結果の有効性に直結する。また、訳語選択に要する時間は、検索過程の効率性に影響する 1 つの要因であると考えた。

4.1 評価実験の設定

実験用インタフェース・計測データ：上記 (a), (b) に示した 2 つの観点からの評価を行うため、後述するテストクエリセットを用い、図 4 にスクリーンショット例を示す評価実験用のインタフェースを用いて、(1) あらかじめ設定した正解の訳語を選択できたクエリ数 (以下、正解数) をカウントし、(2) 各クエリに対する訳語選択を行うまでの時間 (ミリ秒単位) (以下、選択時間) を計測^{*4}した。

図 4 に対応するクエリ例を表 4 に示す^{*5}。この例で、あらかじめ設定した正解は日本語のクエリ語が持つ語義の 1 つ (本例では語義番号 = 1) に対応しており、被験者には、正解として設定されている語義の定義文が提示される。すなわち被験者は、日本語による語義文を見ながら、提示される手がかりを参照することにより、その語義に合致したクエリ語の翻訳を選択することが求められる。これらの語義の区分や語義の定義文は基本語データベースによる。

実験のモード：画像手がかりのほかに、インタラクティブな言語横断検索における訳語選択のために有効である可能性のある手がかりとして、クエリ語に対して求めた訳語をクエリ言語に逆翻訳した結果が考えられる。すなわち、

*4 あらかじめ定められた時間 (30 秒) を経過すると自動的に次のクエリに移行する。このため、1 回あたりの選択時間の上限は 30 秒である。

*5 mode-3 における表示例。各語義に対する表示順はランダム化されるので、表における語義番号とは一致していない。

表 4 クエリの例:「壁」

Table 4 An example query: Kabe.

語義番号	語義	訳語	逆翻訳
1	建物の周囲の囲いや、部屋と部屋との仕切り.	wall	壁
2	登山で、険しく直立した山の岩肌.	cliff	崖
3	物事の進展を行き詰まらせるもの. 障害. 障害物.	impediment	障害

逆翻訳により得られる単語はユーザのクエリ言語によるものであるため、文書言語の知識がなくても、意図したクエリと大きく意味が異なる単語に逆翻訳された訳語を適切でない可能性が高いものとして排除することができる。また、ユーザが何らかの目的言語の知識を有している場合には、文書言語における訳語自体を提示することが有用である可能性もある。そこで、評価実験では以下の3つのモードを設定し、mode-1とmode-2、mode-2とmode-3の比較を行う。

- mode-1 (逆翻訳) : 複数の言語資源^{*6}を用い、1つ以上、最大10個の逆翻訳語(日本語)を検索し、これらを提示する。本論文の焦点である画像手がかりの有効性を議論する際の主な比較対象とする。
- mode-2 (画像手がかり) : 非言語的の手がかりであるWeb画像のみを提示する。このモードは、「システム側から訳語を提示されたとしてもユーザが目的言語に対する知識を有していないために提示された訳語が手がかりとして有効に働かない」というケースをシミュレートするためのものである。
- mode-3 (画像手がかり + 訳語) : Web画像に加え、訳語(英語)を提示する。ユーザが目的言語に対する知識をある程度有しているという状況を検討するための付加的な実験条件である。

被験者・実験の制御: 今回の評価実験における被験者は、日本語を母語とする8名(研究者3名、学生5名; 本論文の著者は含まない; いずれもある程度の英語力を有する)である。上記の3つの実験モードを前提とし、これらの被験者を以下の2つのグループ(group-a, group-b)に分けた。すなわち、mode-1(逆翻訳手がかり)を先に行うgroup-aとmode-2(画像手がかり)を先に行うgroup-bである。いずれのグループにおいても付加的な実験条件であるmode-3(画像 + 訳語)は最後に行った。なお、各被験者の各モードにおける実験において、クエリセット中のクエリが提示される順はランダムにコントロールした。

4.2 テストクエリセット

評価を適切に行うためには、適切なクエリセットにより実験を行う必要がある。たとえば、画像で表現することが難しいと予想されるような抽象概念ばかりを対象とした

り、被験者にとってなじみが薄い概念(たとえば、時代背景が大きく異なるような概念)ばかりを対象としたりすることは望ましくなく、実験が効率良く実施できる範囲内で適切なバランスを保つことが必要である。

このような要件を満たすクエリセットを生成することは決して自明なタスクではないが、前節で述べた「語義・画像適合性評価データセット」から以下に述べるような手順で92のクエリ語を選択した^{*7}。

- 基本語データベースにおける語義別親密度を利用し、これが平均値程度以上である語義のみを対象とする(基本語データベースは基本的な語を収録したものであるため、その中において平均以上の親密度を持つ語義を選択することにより、多くの被験者にとってなじみの薄い概念を排除できる)。
- 上記のような語義を3つ^{*8}持ち、
- 各語義ごとに互いに異なる英訳語を持つ単語を対象とする。
- さらに、「語義・画像適合性評価データセット」を用いて、各訳語に対して適切なWeb画像が得られる可能性があらかじめ分かっている単語を対象とする。

このようにして生成したクエリセットにおいて、各単語において正解として設定した語義の日本語語彙大系における意味的属性の分布は、『具体』: 53件、『抽象』: 39件であった。また、これらの語義親密度の平均値は、『具体』: 5.74、『抽象』: 5.59であった。

4.3 実験で提示したWeb画像の特性

図4に例を示したように、クエリ語の1つの語義に対応する英訳語ごとに8件のWeb画像(サムネイル)を被験者に提示した。これらのWeb画像は、画像検索エンジン(Google Images)を利用することにより、あらかじめ収集しておいたものである。正解として設定した語義に対して、提示したWeb画像がどの程度の適合性を持つかを、評価実験の被験者とは別の3名により評定した。評定のカ

^{*6} EDICT: <http://www.csse.monash.edu.au/~jwb/edict.html>, EDR 対訳辞書: http://www2.nict.go.jp/r/r312/EDR/J_index.html, Google 翻訳をこの順に適用した。

^{*7} 実際には、いわゆるライスケール用問題として、これらのほかに選択が容易と考えられる2つのクエリを用意した。評価実験においては、計94クエリに対して実験を行い、すべての被験者がこの2つのクエリに関して正解を選択できたことを確認した。実験結果の集計分析においては、これらの2つのクエリに対する結果を除いている。

^{*8} この条件は、評価実験において画面スクロールの必要性を避けるためのものである。後述するWeb画像の件数(8件)も同様の理由による(参照: 図4)。

表 5 実験で提示した Web 画像の適合性評定結果
Table 5 Distribution of number of relevant Web images.

適合画像件数	クエリ数 (過半数)	累積割合 (%)	クエリ数 (少なくとも 1 名)	累積割合 (%)
8	14	15.2	23	25.0
7	17	33.7	26	53.3
6	12	46.7	11	65.2
5	11	58.7	11	77.2
4	15	75.0	10	88.0
3	7	82.6	7	95.7
2	7	90.2	2	97.8
1	9	100.0	2	100.0
0	0	100.0	0	100.0

テゴリは適合/不適合の 2 値であるが、Fleiss の κ 指数による評定者間の一致度は、 $\kappa = 0.667$ と高く、安定した評定が得られた。

各クエリにおいて正解と設定した語義に対して提示した画像 (全 8 件) において、過半数の評定者が適合と判定した Web 画像の件数、少なくとも 1 名の評定者が適合と判定した Web 画像の件数のクエリセットにおける分布を表 5 に示す。この分布は本評価実験の目的に沿ったものになっており、表 1 に示した分布とは異なる。

4.4 実験結果と考察

正解数：mode と被験者の 2 つを因子とする二元配置分散分析 (Tukey 法) を適用したところ、mode-2 (画像) における平均正解数は 73.125 (79.5%)、mode-1 (逆翻訳) における平均正解数は 68.875 (74.9%) であり、この差は統計的に有意 ($p < .001$) であった。これより、適切な訳語の選択に関しては、Web 画像手がかりの方が逆翻訳よりも有効であるといえる。

一方、mode-3 (画像 + 訳語) における平均正解数は 77.5 (84.2%) であり、mode-2 との差 (4.7%) は統計的に有意 ($p < .01$) であった。今回の被験者がいずれも一定の英語の能力を有していることから、ユーザが目的言語の知識を持つことが仮定できる場合、画像手がかりに加えて訳語を追加提示することにより、さらに多くの適切な訳語の選択が可能になることが分かった。

なお、group-a (mode-1 を先に実行) と group-b (mode-2 を先に実行) の被験者の間の結果は、group-a : 73.6 (80.0%)、group-b : 71.4 (77.6%) であり、 χ^2 検定の結果によれば統計的な有意差は見られなかった。つまり、正解数に関しては実験の順序による差はなかったといえる。

以上より、今回の実験の範囲内において、Web から収集した画像手がかりは、目的とする語義に対応した訳語をユーザが選択するうえで有効であるといえる。

選択時間：正解数の場合と同様の統計分析を行った。まず、mode-2 (画像) における平均選択時間は 6,793.96 (ms)、mode-1 (逆翻訳) における平均選択時間は 6,951.11 (ms)

であり、mode-2 の方がわずかに時間が短いもののこの差は統計的に有意ではなかった。つまり、効率的な訳語の選択に対する寄与という観点からは、画像手がかりの有効性は限定的であることが示唆される。この要因は、被験者によってはクエリによる選択時間のバラツキが非常に大きいことがある。

一方、mode-3 (画像 + 訳語) における平均選択時間は 5,346.77 (ms) であり、mode-2 (画像) との差 (1,447.19 ms) は統計的に有意 ($p < .05$) であった。すなわち、目的言語の知識が仮定できる場合、画像手がかりに加えて訳語を提示することにより、訳語の選択の効率性が向上する可能性が示唆された。しかしながら、今回の実験では、いずれの被験者も mode-3 を最後に実施しているため、順序による慣れの効果を否定しきれない。実際、統計的検定の結果からは、一番最初に実行したモードにより多くの時間を有するという傾向が示唆されているので、現時点では、画像手がかりに加えて訳語を提示することの効率性に関する有用性は断定できない。

クエリの難易度と選択時間：クエリの難易度を表す観測可能な指標として、評価実験において正解を選択できた被験者の数が考えられる。すると、たとえば、「多くの被験者が正解できたクエリはある意味で易しいクエリであり、このようなクエリに要する選択時間は短い」という予想が考えられる。そこで、この予想の妥当性を mode-1 (逆翻訳) と mode-2 (画像) の比較により検討した。具体的には、各クエリを正解被験者数 (降順) と平均選択時間 (昇順) によりランキングし、これらのランク系列の相関を Spearman の順位相関係数 ρ により評価した。その結果、mode-1 では $\rho = -0.3774$ ($p < .001$)、mode-2 では $\rho = -0.6674$ ($p < .0001$) となり、逆翻訳という言語的な手がかりよりも画像手がかりを用いた場合の方が正解被験者数と平均選択時間とのより強い相関があることが示唆された。このことは、画像手がかりがかなり有効に働くクエリが今回のクエリセットに存在したことを意味する。

表 4 に示した「壁」は、このようなクエリの典型例の 1 つである。実験においては、語義番号 = 1 を正解として設

表 6 クエリ難易度に関する指標の比較
Table 6 Comparison of query difficulty indices.

難易度指標	正解被験者数	平均選択時間
<i>nc</i>	1.0000	-0.6674
<i>idx</i>	0.6030	-0.5588
<i>maj</i>	0.4107	-0.5272
<i>one</i>	0.2641	-0.3443

定した。正解以外の語義に対しては適切な逆翻訳が得られているが、正解の語義に対する逆翻訳はクエリ語そのものとなっており、逆翻訳が効果的に働いていない。典型的には、クエリ語がカタカナ語である場合に逆翻訳もクエリ語と同等となる場合が多い。また、語義の1つ（『障害』の意味）は一種の意味拡張による抽象的概念であり、正解を含む2つの語義は画像的にかなり異なる点が画像手がかりが有効に働いた要因であると考えられる。

クエリの難易度を定める要因：正解と設定した語義以外の語義に対して、正解の語義に適合しうる Web 画像が検索される場合がある。実際、実験提示 Web 画像の評定の結果においては、正解として設定していない語義に対して検索された画像の中に一定の割合（今回は約 10%）で正解語義に適合する画像が得られていた。これらは、主に日本語・英語間の語義の粒度の差によるものと考えられるが、被験者の訳語選択の判断を攪乱するノイズとなりうる。

そこで、このような攪乱要因の影響を考慮するため、正解語義に対して得られた適合画像件数 x から正解でない2つの語義に対して得られた正解語義に対する適合画像件数 y を減算したクエリの難易度推定指標 $idx = x - y$ を設定し、他の2つのクエリ難易度の指標（評価実験における正解被験者数 nc 、提示 Web 画像の評定における過半数適合画像件数 maj ）との比較を行った。表 6 に各指標を用いた場合の正解選択者数、平均選択時間との Spearman の順位相関係数を示す。参考のため、提示 Web 画像の評定において、少なくとも1名の評定者が適合と判断した画像件数 one をあわせて掲載する。この結果から、判断の攪乱要因を考慮した idx は実測された指標である nc を最も良く近似する指標であることが分かる。

以上より、CLIR インタフェースの設計においては、検索意図に関連する Web 画像を提示できるだけではなく、クエリ語の訳語選択を阻害する攪乱要因となる Web 画像を排除したり、適切にマークしたりするための機能を持たせることが有用であると考えられる。

5. 関連研究

クエリ翻訳型のインタラクティブ CLIR においては、文書言語の能力や知識が乏しいユーザを想定し、クエリ語の意味（語義）を明確化することにより、これを反映するクエリ語の翻訳を得るためのシステムの検討が多くなされて

きた。たとえば Keizai というシステムでは、クエリ語がそもそも有する語義の曖昧性をユーザに選択させるために、クエリ言語の辞書における語義区分を語義説明として提示する [19]。また、MULINEX というシステムでは、逆翻訳を行い、クエリ言語におけるクエリ語の擬似的な同義語集合を得ることにより、同様の効果を得ようとしている [5]。さらに文献 [18] は、クエリ語に対する訳語の使用文脈などを提示することにより、インクリメンタルに訳語を修正する枠組みを提示している。一方、非言語的な手がかりとして画像を利用する例は報告されていない。

本論文では、クエリ語の多義性に起因する翻訳曖昧性の解消のための手がかりとして Web 画像を利用した。この方向性を極めるためには、ある語義に対して与えられた画像が対抗する他の語義と区別するものであるかを判定することが必要になる。この問題は、語義そのものを表す画像を得るとい問題 [9] とは多少異なる可能性があるが、密接に関連している。さらに、画像曖昧性解消 (image sense disambiguation) [4], [7] や画像分野における画像注釈付け (image annotation) [12] の問題とも密接に関連している。画像注釈付けに関しては、画像と言語の同時モデルを利用する試み [16], [22] が活発化している。

6. おわりに

ユーザの介入によってクエリ翻訳を援助するインタラクティブな CLIR において、画像手がかりの有用性を実験的に評価した。

クエリ語の語義の特性と画像手がかりの得られやすさの関係に関する評定実験からは、ユーザにより意図されるクエリ語の語義が抽象概念であっても特定の意味特徴を持つ場合には適切な画像手がかりが得られうること、原言語において高い親密度を持つ語義に対する訳語に対しては適切な画像手がかりが得られやすいことなどが分かった。これらの知見は、インタラクティブな CLIR システムを実現する際に有用な指針を与える。たとえば、クエリ語の意味特徴や親密度、固有名としての使用状況と手がかりとしての適切性のデータを蓄積し、機械学習手法を適用することにより、画像検索により得られた Web 画像の手がかりとしての適切性を予測することが可能になる。4章で議論したように、Web 画像の適切性とクエリの難易度には明らかに相関が存在するので、適切性が低い場合には警告メッセージなどでユーザに別のクエリ語を試すよう注意を喚起すれば、インタラクティブなインタフェースの利用価値を向上させることができると考えられる。なお、クエリ語の意味特徴に関しては日本語語彙大系 [15] が利用できるほか、単語親密度については「日本語の語彙特性」[1], [2]、固有名の判別については DBpedia^{*9} (バージョン 3.7 で 360 万

*9 <http://dbpedia.org>

以上の実体を収録) や固有名識別技術・ツールが適用できる。日本語の語彙特性には語義別親密度は収録されていないが、約 10 万語に対して付与された単語親密度により、与えられたクエリ語の代表的な語義に対する親密度を近似することができる。

適切な Web 画像がある程度得られるという状況を想定した訳語選択の評価実験からは、画像手がかりは適切な訳語の選択率向上に有意に寄与することを確認した。その一方、訳語選択に要する時間の短縮に関する寄与はかならずしも明確ではなかった。これには様々な要因が関係していると考えられるが、特に問題となるのはユーザの判断を攪乱するような適合 Web 画像が意図しない語義に対しても得られるというケースである。この問題に対処するためには、各語義をより明確に反映する Web 画像集合を得る必要がある、このためには、Web 画像の検索において語義をより明確化するような関連語を付与する言語的な手段 [9]、さらには、ある語義に対して得られた Web 画像集合のコヒーレンスを評価する画像処理的な手法を適切に組み合わせていくことが有用と考えられる。

なお本論文の範囲では、実際の言語横断情報検索において画像を手がかりとして用いることの有用性までは実証できていない。たとえば、図 3 に示した例のように、クエリの訳語選択が適切に行われたとしても、その訳語は目的言語において多義性を有する可能性があり、これがターゲットとなる情報検索の有効性を損ねるという問題は残る。この問題に対しては、たとえば、クエリログを利用することによりクエリ作成自体を支援する方向性 [10] や、Web 画像とそれに付与されているスニペットなどの言語情報や画像の類似性に基づいて検索結果をクラスタリングすることにより整理する方向性などが考えられる。

謝辞 評価実験の設定について、心理学実験の立場からご助言いただいた NTT コミュニケーション科学基礎研究所の小林哲生氏、実験に参加いただいた被験者の方々に感謝します。

参考文献

[1] 天野成昭, 近藤公久 (編著): 日本語の語彙特性 第 1 巻 単語親密度, 三省堂 (1999).

[2] 天野成昭, 笠原 要, 近藤公久 (編著): 日本語の語彙特性 第 9 巻 単語親密度増補, 三省堂 (2008).

[3] 天野成昭, 小林哲生 (編著): 基本語データベース: 語義別単語親密度, 学習研究社 (2008).

[4] Barnard, K. and Johnson, M.: Sense disambiguation with pictures, *Artificial Intelligence*, Vol.167, pp.13-30 (2005).

[5] Capstick, J., Diagne, A.K., Erbach, G., Uszkoreit, H., Keisenberg, A. and Leisenberg, M.: A system for supporting cross-lingual information retrieval, *Information Processing and Management*, Vol.33, No.2, pp.275-289 (2000).

[6] Deng, J., Dong, W., Socher, T., Li, L.J., Li, K. and

Fei-Fei, L.: ImageNet: A large-scale hierarchical image database, *Proc. CVPR 2009* (2009).

[7] Fujii, A. and Ishikawa, T.: Toward the automatic compilation of multimedia encyclopedias: Associating images with term descriptions on the Web, *Proc. Web Intelligence*, pp.536-542 (2005).

[8] Fujita, S., Bond, F., Oepen, S. and Tanaka, T.: Exploiting Semantic Information for HPSG Parse Selection, *Research on Language & Computation*, Vol.8, No.1, pp.1-22 (2009).

[9] Fujita, S. and Nagata, M.: Enriching dictionaries with images from the Internet - Targeting Wikipedia and a Japanese semantic lexicon: Lexeed, *Proc. COLING 2010*, pp.331-339 (2010).

[10] Gao, W., Niu, C., Nie, J.Y., Zhou, M., Wong, K.F. and Hon, H.W.: Exploiting Query Logs for Cross-Lingual Query Suggestion, *ACM Trans. Information Systems*, Vol.28, No.2, pp.1-33 (2010).

[11] Gonzalo, J. and Verdejo, F.: Automatic acquisition of lexical information and examples, *Word Sense Disambiguation*, Agirre, E. and Edmonds, P. (Eds.), pp.253-273, Springer (2006).

[12] 原田達也: 大規模画像データセットを用いた自動アンテーション, *人工知能学会誌*, Vol.25, No.6, pp.785-793 (2010).

[13] Hayashi, Y., Nagata, M. and Savas, B.: Exploring the visual annotatability of query concepts for interactive cross-language information retrieval, *Proc. AIRS 2010*, pp.379-388 (2010).

[14] Hayashi, Y., Savas, B. and Nagata, M.: Utilizing images for assisting cross-language information retrieval on the Web, *Proc. WIRSS 2009*, pp.100-103 (2009).

[15] 池原 悟, 宮崎正弘, 白井 諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林 良彦: 日本語語彙大系, 岩波書店 (1997).

[16] Lavrenko, V., Manmatha, R. and Jeon, J.: A Model for Learning the Semantics of Pictures, *Proc. NIPS 2003* (2003).

[17] Nie, J.: *Cross-language Information Retrieval*, Morgan & Claypool (2010).

[18] Oard, D.W., He, D. and Wang, J.: User-assisted query translation for interactive cross-language information retrieval, *Information Processing and Management*, Vol.44, No.1, pp.181-211 (2008).

[19] Ogden, W., Cowie, J., Davis, M. and Ludvik, S.N.: Keizai: An interactive cross-language text retrieval system, *Proc. MT Summit VII Workshop on Machine Translation for Cross-Language Information Retrieval*, pp.41-46 (1999).

[20] Petrelli, D., Levin, S., Beaulieu, M. and Sanderson, M.: Which user interaction for cross-language information retrieval? Design issues and reflections, *Journal of American Society for Information Science and Technology*, Vol.57, No.5, pp.709-722 (2006).

[21] 佐久間尚子, 伊集院睦雄, 伏見貴夫, 辰巳 格, 田中正之, 天野成昭, 近藤公久: 心像性データベースの構築と概要, 第 8 回認知神経心理学研究会 (2005).

[22] Weston, J., Bengio, S. and Usunier, N.: WSABIE: Scaling up to large vocabulary image annotation, *Proc. IJCAI 2011*, pp.2764-2770 (2011).

[23] 柳井啓司: 一般物体認識の現状と今後, *情報処理学会論文誌: コンピュータビジョンとイメージメディア*, Vol.48, No.SIG 16 (CVIM 19), pp.1-24 (2007).



林 良彦 (正会員)

大阪大学教授 (大学院言語文化研究科・言語情報科学講座)。博士 (工学)。1983年早稲田大学大学院理工学研究科博士前期課程修了後、日本電信電話公社 (現 NTT) に入社。1994~1995年スタンフォード大学 CSLI 滞在研究員。2004年4月より現職。自然言語処理、情報アクセス、計算論的語彙意味論、言語資源・サービス基盤の研究に従事。電子情報通信学会、人工知能学会、言語処理学会、ACL、ACM SIGIR 各会員。



永田 昌明 (正会員)

1987年京都大学大学院工学研究科修士課程修了。工学博士。同年 NTT 入社。1989年3月から4年間 ATR 自動翻訳電話研究所へ出向。1999年4月から1年間 AT&T 研究所客員研究員。統計的自然言語処理の研究に従事。現在、NTT コミュニケーション科学基礎研究所主幹研究員。情報処理学会奨励賞 (1991年)、情報処理学会論文賞 (1995年)、人工知能学会研究奨励賞 (1995年) 等受賞。電子情報通信学会、人工知能学会、言語処理学会、ACL 各会員。



サワシュ ボラ

大阪大学大学院言語文化研究科博士後期課程在学中。2003年トルコ・チャナッカレ大学日本語学科卒業。自然言語処理、情報アクセスの研究に従事。Cludoq Inc. CEO を兼ねる。

(担当編集委員 吉田 尚史)