

インタラクティブな言語横断検索における画像手がかりの有効性

Effectiveness of Visual Clues in Interactive Cross-Language Information Retrieval

林 良彦

Bora Savas

永田昌明

大阪大学大学院言語文化研究科 NTT コミュニケーション科学基礎研究所

hayashi@lang.osaka-u.ac.jp, bsavas@gs.lang.osaka-u.ac.jp, nagata.masaaki@lab.ntt.co.jp

1 はじめに

クエリ翻訳に基づくインタラクティブな言語横断検索において、クエリ語に対する適切な訳語をユーザに選択させるために、訳語候補に関する非言語的な画像手がかりを提示するインタフェースの検討を行なっている。このような画像手がかりの有効性に関する評価実験の結果によれば、画像手がかりの提示は適切な訳語の選択率の向上に有意に寄与したのに対し、訳語選択過程の効率化(時間短縮)に関する効果は限定的であった。本稿では、この評価実験について詳しく述べるとともに、クエリ語の持つ語義の特性と画像手がかりの有効性との関係についても検討する。

2 インタラクティブ言語横断検索

クエリ翻訳型の言語横断検索では、原言語のクエリ語が適切に目的言語に翻訳されることが必要である。このため、訳語の適切性をインタラクティブに確認する検索インタフェースが考えられるが、一般には、ユーザが適切なレベルの目的言語の知識を有することは期待できないので、適切な支援手段が必要となる。本研究の対象であるインタラクティブ言語横断検索インタフェース [2] は、クエリ語に対する適切な訳語をユーザに選択させるために、訳語候補に関する非言語的な画像手がかりを提示する。より具体的には、外部の画像検索エンジンを利用して各訳語候補に対するサムネイル画像を収集し、ユーザに提示する。このようにして収集した画像を本稿では「Web 画像」と呼ぶ。

図1に「ソウルのバス」という日本語クエリに対する検索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つの訳語候補があり、「バス」に対しては、「buss」、「bath」、「bass」の3つの訳語候補が得られている。マウスを各訳語候補に重ねることにより、対応する Web 画像が表示される。

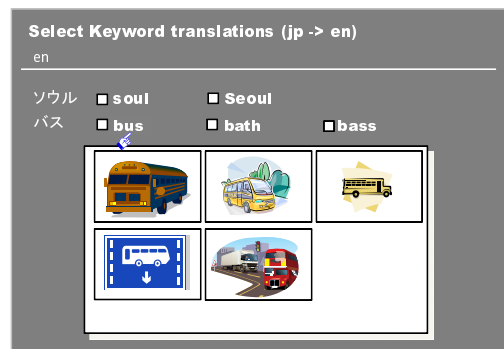


図 1: Displaying Web images with check-boxes.

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。検索インタフェースは、選択された単語群を基にターゲットの検索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。

3 画像手がかりの有効性評価

ユーザによるクエリ語の訳語選択のための手がかりとして Web 画像がどの程度有効であるかを、(a) 適切な訳語の選択に対する寄与、(b) 効率的な訳語の選択に対する寄与、の2つの観点から実験的に評価した¹。なお、今回の実験は日英方向の言語横断検索を想定したものである。

3.1 評価実験

実験のモード: 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の手がかりとしては、画像手がかりの他にクエリ語の訳語候補に対する逆翻訳が考えられる。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、訳語候補自体を提示することが有

¹すでに小規模な実験の結果を [2] で示しているが、今回はよりコントロールされた環境下で、規模(クエリ数、被験者数)を拡大して評価実験を行った。

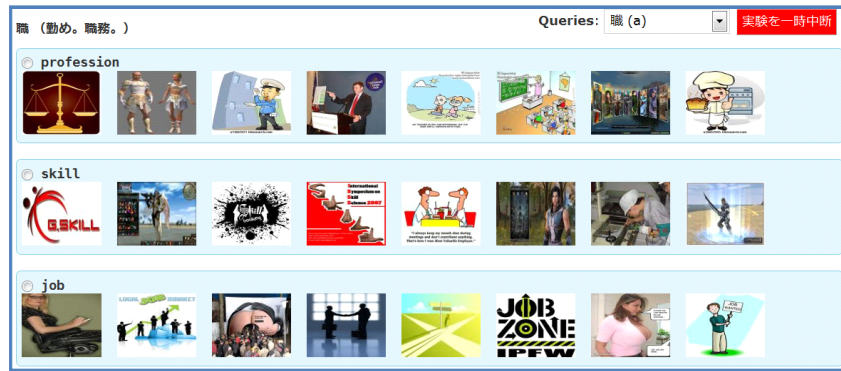


図 2: Screen shot of the evaluation interface (for mode-3).

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): 言語的手がかりである逆翻訳結果 (日本語) のみを提示する。画像手がかりの有効性を議論する際のベースラインとする。
- mode-2 (画像手がかり): 非言語的手がかりである Web 画像のみを提示する。このケースは、ユーザが目的言語に対する知識を有せず、言語的な手がかりが有効でないケースをシミュレートする。
- mode-3 (画像手がかり+訳語候補): Web 画像に加え、訳語候補 (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。

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、3.2節で述べるクエリセットを用い、図 2(画像手がかり+訳語候補の mode-3 の表示) に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候補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²した。

ここで、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。なお、語義の区分や語義の定義文は、基本語データベース Lexeed [5] を利用した。

²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。

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+訳語候補)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。

3.2 クエリセット

クエリセット中の各クエリ語 (日本語) に対し、その語が持つ語義に応じて、互いに異なる3つの訳語候補 (英語) をあらかじめ設定する。また、これらの英訳語候補をクエリとして画像検索エンジン (Google Images) を利用し、1つの訳語候補ごとに8件の Web 画像 (サムネール) をあらかじめ収集しておく。

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。例えば、画像で表現することが難しいことが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 (例えば、時代背景が大きく異なるような概念) ばかりを対象とすれば、評価が困難となる。このような要件をクエリセットを生成することは自明な過程ではないが、今回は、4節で述べる「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個の日本語単語をクエリ語として選択した。

- 被験者にとってなじみが薄い概念を避けるため、Lexeed における語義別親密度³を利用し、これが

³語義の「なじみの程度」を40名の被験者による評定実験によ

平均値程度以上である語義のみを対象とする。

- 上記のような語義を3つ持つ日本語単語であって、
- 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、各英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 [6] 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。

3.3 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、mode-1(逆翻訳)における平均正解数は68.875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言える。

一方、mode-3(画像+訳語候補)における平均正解数は77.5であり、mode-2との差は統計的に有意 ($p < .01$) であった。これより、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候補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、group-b:71.4 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言える。

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上で有効であると言える。この結果は、以前に報告した小規模な実験の結果 [2] と合致している。

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6793.955(ms)、mode-1(逆翻訳)における平均選択時

間を数値化したもの。その値域は1~7であり、Lexeedに収録されている45,951件の語義に対しては、平均値:4.65、標準偏差:1.09であると報告されている。

間は6951.114(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になる。この要因としては、被験者によっては選択時間のクエリによるバラツキが非常に大きいことが挙げられる。

一方、mode-3(画像+訳語候補)における平均選択時間は5346.774(ms)であり、mode-2(画像)との差は統計的に有意 ($p < .05$) 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候補を追加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、順序による馴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。

クエリの難易度と選択時間: 「多くの被験者が正解したクエリは、(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」と予想される。そこで、この傾向を mode-1 と mode-2 それぞれで調べた。より具体的には正解数によるランキングと選択時間によるランキングを Kendall の順位相関係数 τ により評価した。まず、mode-1(逆翻訳)では $\tau = -0.276$, ($p < .001$) となり、言語的な手がかりの場合においてこの傾向が裏付けられた。さらに、mode-2(画像)では $\tau = -0.502$, ($p < .0001$) であり、画像手がかりを用いた場合にその傾向がさらに顕著であることが分かった。このことは、画像手がかりがかなり有効に働くクエリが存在したことを示唆する。

4 語義と Web 画像

適切な検索インタフェース実現の観点からは、クエリ (語義-訳語) ごとに、それに対する画像手がかりの有効性を予測する手段を確立する必要がある。この問題を検討するため、「語義・画像適合性評価データセット」と呼ぶデータ [3] を構築した。

4.1 語義・画像適合性評価データセット

以下のような手順でデータセットを構築した。

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 Lexeed から抽出し (2,380 語, 5,540 語義)、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。英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候補の中から人手による選択を行った。次に、これらの英訳語をクエリとして画像検索 (Google Images) を行い、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⁴。

全体的な結果としては、33.0% の語義-英訳語に対しては、上記の手段で適切な Web 画像が得られなかった (適合度=0)。以下では、画像適合度が 0 のものを画像適合性:無、1 以上のものを画像適合性:有とする。

4.2 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⁵との関係を χ^2 検定により調べた。その結果、画像適合性の有無には有意に偏りがある ($p < .001$) ことを確認した。予想されるように、具体の大分類の配下にあるカテゴリは画像適合性:有 となるものが多いが、残差分析の結果からは、(1) 抽象の大分類であっても「人間活動」や「自然現象」のカテゴリは画像適合性:有 となりやすい、(2) 具体の大分類であっても「組織」などは有意に画像適合性:有 とはならない、などの興味ある結果が得られた。

4.3 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ところ、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が確認された。これより、親密度の高い概念はある程度、日英両言語で共通しており、また、これらの概念に対応する画像は Web から獲得しやすい傾向にあると言える。

4.4 「Web バイアス」

上記のような語義の特性とは別に、Web から検索によって収集する画像に特有な Web バイアスの問題も確認された。この傾向は、クエリとして用いる英訳

⁴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 に対する画像適合度は 3 (英訳語:scholar) となるが、2 (英訳語:learned person) に対する値は 0 となった。

⁵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。

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。この問題に対処するためには、固有名認識によるチェックを行うとともに、語義を適切に表す画像が得られるようなクエリ拡張を行う必要がある。

5 おわりに

ユーザの介入によってクエリ翻訳を支援するインタラクティブな言語横断検索において、適切な翻訳を効率的に選択できる支援となり得るかという観点から画像手がかりの有効性を実験的に評価し、特に前者に対する有効性を統計的に示した。今後は、クエリ (語義-訳語) ごとに画像手がかりの有効性を予測する手段を確立する必要がある。本研究において必要なことは、対抗する他の語義との差異を明確化する (相対的な) 画像手がかりを得ることであるが、語義を (絶対的に) 説明・表現するための画像をどのように得るかという問題 [1] と関連があり、さらには画像曖昧性解消 (image sense disambiguation) [4] の問題とも関連している。これらの研究の知見を参考にしながら、さらなる検討を進めたい。

参考文献

- [1] Fujita, S., and Nagata, M.: Enriching Dictionaries with Images from the Internet. - Targeting Wikipedia and a Japanese Semantic Lexicon: Lexeed -. *Proc. of COLING 2010*, pp.331-339 (2010).
- [2] Hayashi, Y., Savas, B., and Nagata, M.: Utilizing Images for Assisting Cross-Language Information Retrieval on the Web. *Proc. of WIRSS 2009*, pp.100-103 (2009).
- [3] Hayashi, Y., Nagata, M., and Savas, B.: Exploring the Visual Annotatability of Query Concepts for Interactive Cross-Language Information Retrieval. *Proc. of AIRS 2010*, pp.379-388 (2010).
- [4] Saenko, K.: Image Sense Disambiguation: A Multimodal Approach. PhD thesis, MIT (2009).
- [5] 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).
- [6] 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).