

Web 画像による語義・概念の視覚的な提示に関する検討

Visualization of Senses and Complex Concepts by Web Images

林 良彦^{1*}
Yoshihiko Hayashi¹

¹ 大阪大学大学院 言語文化研究科

¹ Graduate School of Language and Culture, Osaka University

Abstract: Images, deemed language-independent to some extent, can be utilized as an effective medium for representing/disambiguating user's search intent in interactive information access systems, such as a cross-language information retrieval system. If a target query concept (e.g. beaver) could be decomposed into a set of salient concept-feature pairs (e.g. beaver builds dams; beaver chews on woods, etc.), the depiction of each concept-feature pair may greatly help manifest the user's search intent. Given this motivation, we conducted an investigation on the Web-imageability of complex concepts (concept-feature pairs), which the psycholinguistic semantic feature database developed by McRae and his colleagues provides. More specifically, this paper reports on the human assessment results on how the images acquired from the Web by means of an image search engine, given a concept-feature pair, can adequately deliver the meaning. The results clearly show that physical motions and feeding behaviors performed by animate beings were more adequately depicted in the acquired Web images. This paper also discusses possible utility of visual images in interactive information access systems, and further argues mutual grounding of visual and linguistic information.

1 はじめに

単語が指示するアトミックな概念、さらには、語句により想起される複合的な概念が画像や記号により視覚的に提示できれば、インタラクティブな情報アクセスにおいて、ユーザの意図の明確化を支援する手段として利用できると考えられる。特に、画像や記号が(ある程度までは)言語独立であることを考えれば、言語横断情報検索のような「言語の壁」を越える必要があるタスクにおいて、有用な手段となると考えられる。

このような考えに基づき、本稿の著者らは、言語横断情報検索において、クエリ単語の翻訳曖昧性解消の手がかりとして Web 画像¹を用いることを提案し [7]、その有効性について評価を行った [9]。その結果、(1) 適切なクエリ単語の翻訳を選択するために画像手がかりは有用であるが、(2) 検索過程の効率性の向上には必ずしも寄与しない、ことが分かった。また、クエリ単語の表す概念が Web 画像によりどの程度表現されるかについての調査を行い、(1) 形状を有する具体的な事物が

Web 画像により適切に表されることが多いだけでなく、(2) 抽象的な概念も自然現象やある種の動作を表すものは、その状況を想起させる視覚的表現が Web 画像として取得できる場合が多いことを確認した [8]。

以上の一連の研究²では、品詞が名詞である単独の単語を調査の対象としていた。しかしながら、たとえば具体的な事物(のクラス)を指示する単一の名詞であっても、我々はその事物が有する顕在性の高い特徴をいくつか心的に思い浮かべている、あるいは、心的に保持していると考えられる。例えば、「蛙」といえば、色が「緑色」であり、「ピョンと跳ねる」という行動が特徴的といったことが「蛙」の心的な意味表現の構成に含まれていると考えることができる。

本研究では、単語が指示する概念の有する一つの特徴を意味属性と呼ぶ。また、ある概念が有するある意味属性、すなわち、概念と意味属性の組み合わせ(例: 蛙はピョンと跳ねる)を概念意味特徴と呼ぶ。

ここで、概念意味特徴は対象概念を主体として含む一種の複合概念と考えることもできることを注意しておく。また、概念は、それに関する概念意味特徴の集合により規定されると見なすことができるので、各概念意味特徴を表す、あるいは、想起させる非言語メディ

*連絡先: 大阪大学大学院・言語文化研究科・言語情報科学講座
〒560-0043 豊中市待兼山町 1-8
E-mail: hayashi@lang.osaka-u.ac.jp

¹Web から取得できる画像、特に画像検索エンジンにより取得できる画像を本研究では Web 画像と呼ぶ。

²NTT コミュニケーション科学基礎研究所との共同研究による。

アによる感覚的表現 (画像や音などのメディア情報) が適切に得られるならば、それらのメディア情報もやはりインタラクティブな情報アクセスなどにおける支援の手段として利用できることが期待できる。

ところで、従来より認知心理学、あるいは、心理言語学の分野では、単語を刺激として被験者に提示し、それに対する反応を収集することにより意味属性の体系を構築する試みが行われてきた [3]。そこで本稿では、そのような心理言語学的なアプローチにより収集された概念意味特徴から検索クエリを構成し、これを用いた Web 画像検索により収集された Web 画像が当該の概念意味特徴をどの程度、視覚的に適切に表しうるか (以下、概念意味特徴の Web 画像適合性) について調査した結果を報告する。

本稿の具体的な構成は以下のとおりである。まず 2 節では、心理言語学的な意味属性の体系、および、概念意味特徴のデータを提供するものとして利用するデータベース [12](以下、McRae のデータベース) について説明する。次に 3 節では、概念意味特徴の Web 画像適合性の調査の概要について述べる。主要なアイデアは、情報検索における性能評価指標を用いて概念意味特徴の Web 画像適合性を評価する点にある。続く 4 節では、Web 画像適合性の評価結果について検討する。より具体的には、検索クエリ生成の方略、概念意味特徴の特徴タイプ、概念意味特徴の構成要素の意味的な分類の観点からの検討を加える。また 5 節では、Web 画像をインタラクティブ情報アクセスにおける手がかり情報とする可能性や、言語と画像を相互にグランディングする可能性について議論する。

2 McRae のデータベース

2.1 概要

McRae のデータベース [12] は、基本レベル (basic level)³に属すると考えられる、生物・無生物に関する 541 種類の英語名詞により表される概念に対して、多数 (700 名以上) の被験者が心理実験的設定のもとで与えた英語による記述に基づいており、結果として、2,526 種類の意味属性と、これに基づいて上記 541 の概念に対して付与された総計 7,526 の概念意味特徴が提供されている。

本データベースにおいては、ターゲットとなる各概念の規定・記述に関する豊富な情報を有しているが、その中で最も基本的な情報は、各概念に対してある概念意味特徴が何人の被験者によって付与されたかという一種の頻度データである。

³日常生活においてオブジェクトを分類する際に参照されていると考えられる概念階層におけるレベル (抽象すぎず詳細すぎない) を指す。

表 1: 概念 beaver に対する記述の一部。

意味属性	BR ラベル	頻度
a_mammal	<i>taxonomic</i>	6
beh-_builds_dams	<i>encyclopaedic</i>	20
beh-_chews_on_wood	<i>visual-motion</i>	5
beh-_cuts_down_trees	<i>encyclopaedic</i>	7
found_on_the_nickel	<i>encyclopaedic</i>	6
has_a_tail	<i>visual-form_and_surface</i>	24
has_sharp_teeth	<i>visual-form_and_surface</i>	24
hunted_by_people	<i>function</i>	7
is_brown	<i>visual-colour</i>	19
is_furry	<i>tactile</i>	18

例として、概念 beaver に対するデータの一部を表 1 に示す。表の各行は一つ概念意味特徴に対応する。これによれば、ビーバーは例えば、「哺乳動物の一種 (a_mammal) であり」、「ダムを作る (beh-_builds_dams)」といった特徴があるが、これらの概念意味特徴はそれぞれ 6 人、20 人により与えられている。

2.2 意味属性キーワード

表 1 に例を示したように、多くの意味属性は、a_、has_、is_ などの一定の統制がなされた接頭辞部分と "mammal", "sharp_teeth", "brown" といった比較的自由的な単語 (列) による部分からなる。本稿では、接頭辞部分を意味属性キーワード⁴と呼ぶ。McRae らによれば、これらのキーワードは、意味属性をタイプ化するために導入されている。

各意味属性キーワードの意味はほぼ自明と思われるが、a_ (上位・下位関係)、has_ (全体・部分関係)、といった通常のオントロジ的な関係だけでなく、様態や機能を表す関係や、さらには、連想的な関係も含まれており、通常の語彙的オントロジの言語資源には必ずしも含まれない豊かな情報が含まれている。これらの中で、beh_-_beh_-_ というキーワードは、それぞれ生物・無生物オブジェクトの典型的な振る舞いを表すメタキーワードである。すなわち、「ダムを作る」(beh-_builds_dams) という意味属性は、ビーバーを規定する上で顕在性の高い⁵振る舞いに関する意味属性である。

2.3 Rrain Region (BR) ラベル

公開されているデータにおける付加的な情報として興味深いものに、概念意味特徴への 2 つの体系による分類ラベルの付与がある。そのうちの一つは、Cree と

⁴公開されているデータからは、96 種の意味属性キーワードを抽出したが、これらの中には誤りと思われるものもあるほか、整理統合が必要と思われるものも散見される。

⁵実際、データベース中でこの意味属性を持つ概念は beaver のみである。

表 2: BR ラベルとその頻度分布.

BR ラベル	頻度
<i>visual-form-and-surface</i>	2,336
<i>visual-color</i>	424
<i>visual-motion</i>	339
<i>tactile</i>	245
<i>sound</i>	142
<i>taste</i>	84
<i>smell</i>	24
<i>function</i>	1,517
<i>encyclopaedic</i>	1,417
<i>taxonomic</i>	730

McRae により提案された脳領域階層分類 (brain region taxonomy) [2] と呼ばれる体系によるものである. 表 1 の中央のカラムはこの体系におけるカテゴリのラベル (以下, **BR ラベル**) を示す. この体系には 10 種類 (概念階層を表す taxonomic も含んで) のカテゴリが含まれるが, そのうちの 7 つは知覚に関するものであり, 特に視覚に関するものは 3 種 (form-and-surface, color, motion) に細分類されている.

表 2 に BR ラベルのデータベースにおける頻度分布を示す. 延べ数においても概ね半数が知覚に関するカテゴリであり, 特に視覚的なカテゴリがその大部分を占めていることが分かる. 表 1 のビーバーの例における「木を齧る ("chews on wood")」振る舞いは視覚的な動作であるのに対し,「ダムを作る ("builds dams")」振る舞いは百科事典的知識として分類されている.

3 概念意味特徴の Web 画像適合性の調査

図 1 に意味属性の画像適合性の調査の概要を示す. 以下, 各ステップについて説明する.

3.1 概念意味特徴の選出

今回の検討においては, beh_, および, inbeh_ の意味属性キーワードを持つ意味属性のみを対象とした. すなわち, 生物・無生物の何らかの動作・振る舞いに関する概念意味特徴を調査した. この背景には, 動作・振る舞いがある程度は画像化されやすいであろうという予想があり, また, どのような動作・振る舞いが Web 画像として取得可能であるかという興味がある.

この条件により, 218 種の概念に対する 535 件の概念意味特徴を今回の検討の対象とした. 表 3 に, 535

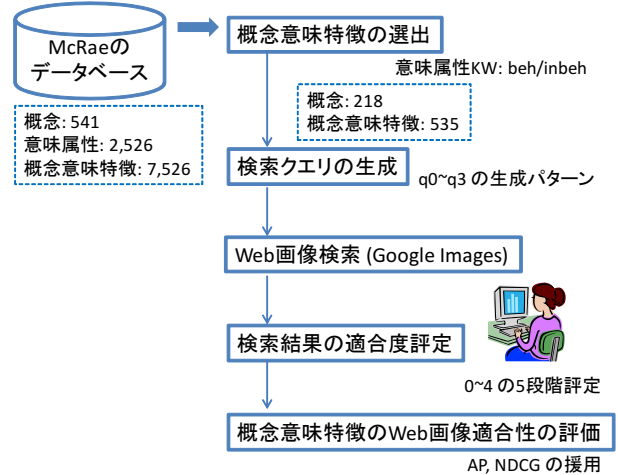


図 1: 概念意味特徴の Web 画像適合性の調査.

表 3: 概念意味特徴 (535 件) の分布.

	beh	inbeh
視覚的な BR ラベル	268 (f.a)	33 (f.c)
その他のラベル	151 (f.b)	83 (f.d)

件の概念意味特徴の beh_/inbeh_ の区別, BR ラベルが視覚に関するものか・それ以外かの区別の組み合わせによる分布を示す. 約半数のものが生物による動作・振る舞いに関するものであり, かつ, 視覚的であると分類されていることが分かる. 本稿では, 表における f.a から f.d を概念意味特徴の特徴タイプと呼ぶ.

3.2 Web 画像の収集:画像検索クエリの生成方略

与えられた概念意味特徴に対して, 以下の 4 つのパターンにより検索クエリを生成し, Web 画像検索エンジンを用いて Web 画像を収集する.

- q0: 概念を表す名詞のみを用いる ("beaver")
- q1: McRae のデータベースに与えられている記述に基づく ("beaver builds dams")
- q2: q1 において動詞を現在分詞形 (-ing 形) とする ("beaver building dams")
- q3: 動詞を現在分詞形 (-ing 形) とし, 概念名詞を後置する ("building dams beaver")

検索クエリにより得られる検索結果はもちろん利用する検索エンジンに依存するが, 上記の 4 つのパターンは概ね以下のような考えに基づいている.

- q0 および q1 はある意味でのベースラインである



図 2: Web 画像と適合度評定の例 ("beaver building dams").

- 特に q_0 は、対象とする概念意味特徴が当該の概念において非常に顕在性の高いものであれば、概念名詞のみによって、当該の概念意味特徴に対してある程度の適合性を有する画像が検索できるのではないかと、という予測に基づく
- q_2 および q_3 においては、動作・振る舞いを直接的に表す動詞を現在分詞形 (-ing 形) としている。これは、画像 (特に写真) の持つ一般的な性質として、ある状況を描写するということがあり、そのために、Web コンテンツ中での画像に対する注釈・記述に動詞-ing 形が用いられているという可能性を考慮したものである

3.3 画像検索結果の適合度評定

上記によって生成した 2,126 件のクエリ⁶を Web 画像検索エンジン (Google Images) に送り、一つのクエリあたり最大 15 点、総計 27,970 点の画像を収集した。収集した Web 画像に対し、一名の評定者⁷により、0~4 の 5 段階評価 (0:不適合, 4:適合) により、検索された画像の概念意味特徴に対する適合度を評定した。

図 2 に検索結果と評定例を示す。また、評定結果の分布を表 4 に示す。適合度 2 以上のものを適合画像とするならば、半数近く (44.8%) の画像が適合画像という結果となった。以下の分析において、適合・不適合の 2 値判断が必要な場合は、この基準によっている。

3.4 概念意味特徴の Web 画像適合性の評価指標

ある概念意味特徴 (概念と意味属性の組み合わせ; 例: "beaver builds dams") に対して、前述のパターンにより生成したクエリを用いて収集した Web 画像集合がどの程度適切にその概念意味特徴を表しているかを、情報

⁶ q_1 パターンのクエリが "chicken cannot fly" のようになるものがあり、これらについては q_3/q_4 タイプのクエリを生成していない。

⁷実際には数名の評定者が分担して作業しているが、監督者が一定の基準となるよう調整を行っている。なお、評定者、監督者とも本論文の著者ではない。

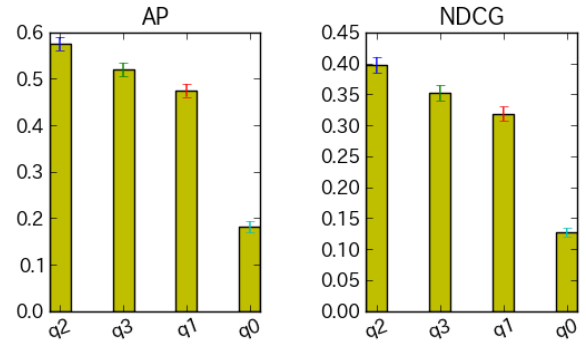


図 3: クエリタイプによる検索性能 (画像適合性) の比較。

検索においてよく用いられている平均精度 AP (Average Precision), および, NDCG (Normalized Discounted Cumulative Gain) の二つの指標 [11] により評価することとした。すなわち、これらの指標の値が高ければ、当該の概念意味特徴の Web 画像適合性は高いものとして扱うということである。

4 概念意味特徴の Web 画像適合性の評価

4.1 検索クエリ生成の方略

クエリ生成の方略ごとに見た画像適合性の比較を図 3 に示す。この図には、動詞の-ing 形を用いる q_2 タイプの画像適合性が他のタイプのクエリよりも有意に高い (AP, NDCG の双方について, $p < 0.01$; Mann-Whitney U-test) ことが示されている。このことから、本稿の以下の分析では、 q_2 タイプのクエリに対する結果について議論する。

ところで、図 3 には示されていないことであるが、ここまでの適合性判定をより甘い基準にし、適合度 0 の画像のみを不適合と考えると、 q_0 タイプのクエリの AP 値の結果が他のクエリタイプのものを上回る。このことは、特定の概念意味特徴が当該の概念において高い顕在性を有していることを示唆すると思われる。例えば、airplane という概念に対する q_0 タイプのクエリは、"airplane" であるが、これにより検索される画像集合には飛行機が飛んでいる画像が多く含まれる。もちろん、"airplane flies" という概念意味特徴はデータベースに存在するのであるが、わざわざこれを検索クエリにする必要はないことになる。

4.2 概念意味特徴の特徴タイプ

表 3 に示した f_a から f_d の特徴タイプから見た画像適合性の比較を図 4 に示す。

表 4: 評価された適合度の分布.

適合度	0	1	2	3	4
画像件数	6,705	8,734	4,025	1,910	6,596
総計: 27,970	15,439 (55.2%)		12,531 (44.8%)		

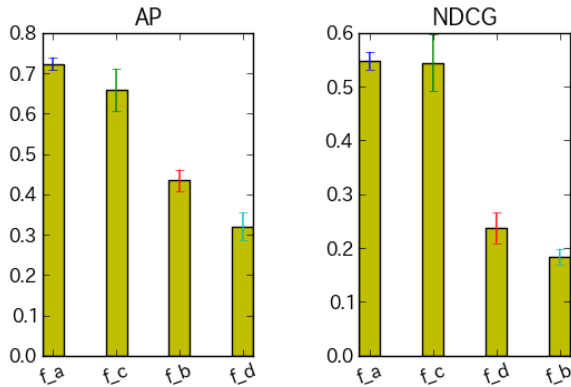


図 4: 概念意味特徴の特徴タイプによる検索性能 (画像適合性) の比較.

この図からは概ね以下のことが言える.

- 視覚に関する BR ラベルを有する概念意味特徴タイプ (f.a と f.c) の画像適合性がそれ以外の知覚ラベルを有する概念意味特徴タイプ (f.b と f.d) よりも有意に高い. このことは, McRae のデータベースで付与されている視覚に関する BR ラベルの妥当性を裏付けていると考えられる
- しかしながら, f.a と f.c の間の差は有意ではなく (AP については, $p = 0.192$; NDCG については, $p = 0.440$), 視覚に関する概念意味特徴を対象とする限り, 動作・振る舞いの主体が生物であるか無生物であるかという観点からは, 画像適合性に有意な差がない
- ただし, 無生物の主体の振る舞いを表す f.c タイプにおいては, 画像適合性のバラ付きが非常に大きい
- 視覚以外の知覚ラベルを有する概念意味特徴タイプ (f.b と f.d) に関しては, AP 値において両者に有意な差 ($p < 0.01$) が見られた

4.3 概念意味特徴の構成要素の意味的な分類

本節では, Princeton WordNet⁸ における lexicographer file⁹を粗い意味分類の体系として用い, 概念意

⁸<http://wordnet.princeton.edu/>

⁹<http://wordnet.princeton.edu/man/lexnames.5WN.html>

味特徴の構成要素 (概念を表す名詞, 動作・振る舞いを表す動詞, および, これらの組み合わせ) の意味的な特徴と Web 画像適合性との関わりを調べる.

lexicographer file は, そもそもは, WordNet を編纂する作業上の要請から設けられたもの [13] であるが, いくつかの研究 (例えば [10]) において, 粗い意味分類の体系として用いられている. 本研究でも lexicographer file を意味グループと考え, そのファイル名を意味グループのラベルとして用いる. このために, McRae のデータベース中の対象となる概念名詞, これらに対する概念意味特徴において使われる動詞に対して, 人手により意味グループラベルを付与した.

4.3.1 概念を表す名詞

表 5 に本稿で対象とする 218 種の概念名詞に対する意味グループの分布を示す. McRae ら [12] は, 生物 (living thing)・無生物 (nonliving thing) を研究の対象としたと述べているが, いくつかの概念名詞は, WordNet の分類においては, 抽象的なカテゴリ (communication, possession) に属する. 約 6 割強 (animal, plant に属するもの) が生物にあたり, それ以外の具体物に関するカテゴリが無生物にあたる.

表 5: 概念名詞の意味から見た分布.

意味グループ	概念名詞の数
animal	120
artifact	78
food	11
plant	3
communication, substance	各 2
body, possession	各 1

図 5 に概念名詞の意味グループ別により検索性能 (画像適合性) を比較する.

AP と NDCG の二つの指標によって傾向の違いが見られるが, 画像適合性が比較的高いグループ animal, food, substance と比較的低いグループ artifact, plant に分けることができ, 特に動物 (animal) は, 比較的安定して画像適合性が高いとみることができ一方で, 人工物 (artifact) は一貫して画像適合性が低い傾向にあることが分かる. この傾向は, 図 4 に示した傾向, すなわち, f.a, f.c タイプの概念意味特徴が f.b, f.d タイ

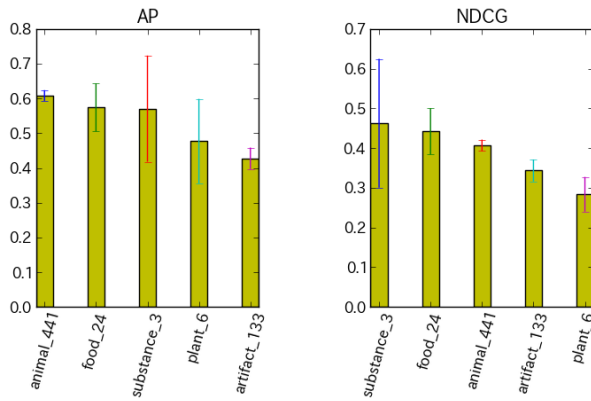


図 5: 概念名詞の意味グループによる検索性能 (画像適合性) の比較.

プの概念意味特徴よりも高い画像適合性を示すという結果と符合している. 以上より, Web からは動物による動作がより画像として得やすいという傾向があることが分かる.

4.3.2 動作・振る舞いを表す動詞

表 6 に本稿で対象とする 535 件の概念意味特徴において用いられる動詞に対する意味グループの分布を示す. この意味グループを用いて, Web 画像適合性の高いと思われる動詞の意味的な傾向を調べる.

表 6: 動作・振る舞いを表す動詞の意味から見た分布.

意味グループ	動詞の数
<i>motion</i>	169
<i>creation</i>	117
<i>consumption</i>	108
<i>communication</i>	34
<i>change</i>	19
<i>contact</i>	16
<i>competition, perception</i>	各 15

図 6 に動詞の意味グループ別により検索性能 (画像適合性) を比較する.

まず, *consumption*, *motion* の二つの意味グループは, これらに属する動詞の数も多く, また画像適合性も比較的安定して高い. *consumption* に属する代表的な動詞は, "eat", "chew", "drink" などであり, 生物の摂食行動に関連するものが多い. *motion* に属する代表的な動詞は, "fly", "swim", "crawl", "travel" などであり, やはり生物の基本的な行動である移動に関するものが多い. 表 5 に示したように, McRae のデータ

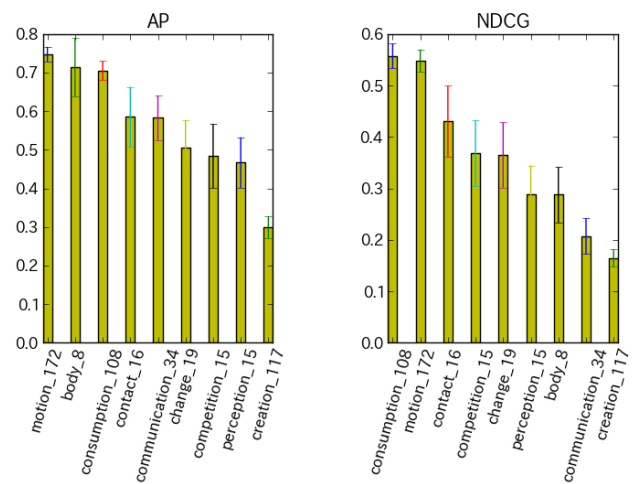


図 6: 概念名詞の意味グループによる検索性能 (画像適合性) の比較.

ベースには *animal* の意味グループに属する概念が多く含まれていが, 動物の摂食行動や基本的な移動動作は, 画像として Web 上から取得しやすい傾向にあることが分かる.

一方, *creation* の意味グループは, 頻度は高いものの画像適合性は低い傾向にある. この意味グループに属する代表的な動詞は, "produce", "build", "make", "give", "do" などである. これらの動詞による意味特徴においては, 焦点は create されるものに当てられると考えられるが, create されるものが具体的な形状を持つものとは限らない (例: "accordion produces music") ことから, 画像適合性は低い傾向にあるものと考えられる.

4.3.3 概念名詞と動作・振る舞い動詞の組み合わせ

最後に, 意味グループのレベルでみた場合の概念名詞と動詞の組み合わせで頻度の高いものを表 7 に示す.

表 7: 頻度の高い概念名詞+動詞の組み合わせ.

意味グループ組み合わせ	頻度
<i>animal+motion</i>	126
<i>animal+consumption</i>	114
<i>animal+creation</i>	63
<i>artifact+creation</i>	49
<i>animal+communication</i>	35

図 7 に概念名詞と動詞の組み合わせにより検索性能 (画像適合性) を比較する.

ここまでの議論から予想されたとおり, 概念名詞:*animal*+動詞:*motion* の組み合わせが AP, NDCG 両方の

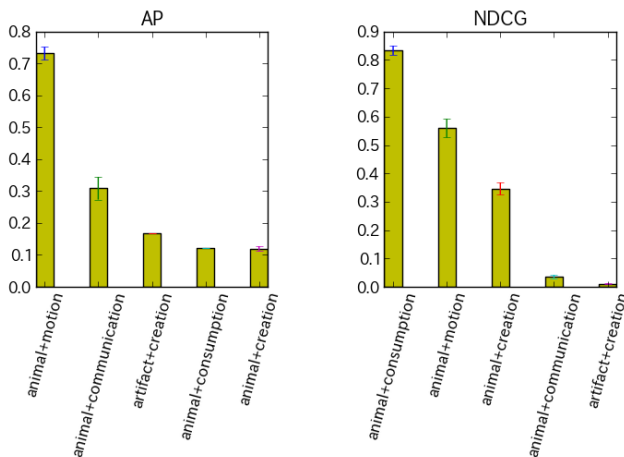


図 7: 概念名詞と動詞の組み合わせによる検索性能 (画像適合性) の比較。

尺度で安定して高い検索性能 (画像適合性) を示している。その一方で、概念名詞: *animal* + 動詞: *consumption* の組み合わせは、NDCG では高い画像適合性を示しているが AP では低いという傾向があり、検索された画像の中にはにおいては、必ずしも適合度が高くないものも混在しているという状況が想像される。

5 関連研究と議論

5.1 関連研究

近年、概念の具体性 (concreteness) [14] に代表されるような、心理言語学的な属性を自然言語処理や情報検索に導入しようという傾向が見られる [10, 17]。このような心理 (言語) 学的な属性の中で、最も興味深いものは、やはり McRae のデータベースで提供されているような意味属性¹⁰であろう。意味属性を用いて記述された概念意味特徴は、当該の概念に対して人々が持つ顕在性の高い情報を担っており、既存の言語資源からは得がたいタイプの情報を提供する。Silberer ら [15] は、McRae のデータベースによる意味属性の情報を従来からのテキスト情報と統合することにより単語の意味的類似度・連想度の精度が向上することを示している。

このような意味属性の持つ問題点の一つとして、意味属性が被験者実験により導出されるためにそのカバー率が低いことから、実用的な言語処理への適用が限定されることが挙げられる。この問題に対して、Baroni ら [1] は、大規模コーパスから有用な意味属性を抽出する手法を提案している。一方、Silberer ら [16] は、McRae

¹⁰semantic feature norm (意味属性記述) という用語が関連研究を含めて用いられているが、本稿では単に意味属性とした。

のデータベースにおける意味属性に相当する高次の視覚的属性 (visual attribute) を与えられた画像から抽出することを試みている。より具体的には、ImageNet [4] (WordNet 中の synset と対応する画像を収集したデータベース) における画像と McRae のデータベースにおける概念記述特徴を訓練データとして用い、視覚的属性抽出のモデルを学習している。

5.2 議論: 言語と画像の相互グラウンディング

言語表現の意味を画像に求めること、逆に画像が伝える意味を言語として表現することは、相互に関連した問題であり、また、情報アクセスという応用においても重要である。後者は、画像コンテンツに対する言語注釈を与えるという問題であり、情報アクセスの高度化に直結する課題である。前者に関連しては、例えば、辞書エントリに対して意味的に適合する画像を求めようとする藤田らの試み [6] や、Web 画像を取り込んでマルチメディア百科事典を構成しようとする藤井らの試み [5] があるが、これらの大きな意味でのグラウンディングを行うとする場合と、筆者らの以前の研究のように意味の区分のための手がかりとして画像を用いる場合とでは、意味を絶対的に表すものか、複数の意味を相対的に区分するためのものかという違いを考慮する必要がある。情報アクセスにおける手がかりとしては、より後者の条件に適合した画像を収集・判別する手法の開発が必要となる。

本研究で目指すところの一つは、概念意味特徴で表されるような概念を主体 (主語) とした複合的な概念を画像で表現する可能性の探求であり、主体となる概念名詞は生物・無生物を問わず具体物であった。しかしながら、特に他動詞を用いた概念意味特徴では、目的語として現れるものは必ずしも具体物とは限らない (例: *accordion produces music*)。このことから、抽象概念 (この例では *music*) に適合する、あるいは、それを想起させる画像を具体物を主体とした概念意味特徴に適合する画像からある程度は得ることができる可能性があると考えられる。

6 おわりに

言語表現が担う意味を言語独立性の高い非言語メディア情報を用いて適切に表現することができれば、インタラクティブな情報アクセスにおける有用な手がかりとして利用できるほか、コミュニケーション支援や言語学習などにも適用が可能であると考えられる。

本研究では、非言語メディア情報として画像を取り上げ、ある概念 (例: *beaver*) に対して心理言語学的手法により導かれた概念意味特徴 (例: *beaver builds dams*;

beaver chews on words) [12] がどの程度 Web 画像検索により収集できる Web 画像により表現されうるかを調査した。この結果、動物の摂食行動や、主に物理的な動き・移動に関する基本行動が Web 画像により表現されやすいことを確認した。また、英語による概念意味特徴において、より良い画像検索結果を得るためのクエリ生成手法について評価を行い、動詞の-ing 形を用いることが有用であることを確認した。

しかしながら、情報アクセスシステムに組み込める形の支援機能を実現するためには、ある概念意味特徴に基づいた検索クエリによって検索された画像が、その概念意味特徴をどの程度適切に表しているかを予測する手法の開発が必要であり、Web コンテンツの言語的な解析と画像解析を統合した手法の開発が必要である。

謝辞

本研究は、JSPS 科研費#24650123 (挑戦的萌芽研究:「画像による言語的意味のグランディングの可能性の探求」) の助成を受けた。

参考文献

- [1] Baroni, M., Murphy, B., Barbu, E., and Poesio, M. 2010. Strudel: A corpus-based semantic model based on properties and types. *Cognitive Science*, (34):222–254.
- [2] Cree, G.S., and McRae, K. 2003. Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology*, 132:163–201.
- [3] Cree, G.S., and Armstrong, B.C. 2012. Computational Models of Semantic Memory. In: Spivey, M.J, et al. (Eds), *The Cambridge Handbook of Psycholinguistics*, pp.259–282, Cambridge University Press.
- [4] Deng, J., Dong, W., Socher, T., Li, L.J., Li, L., and Fei-Fei, L. 2009. ImageNet: A large-scale hierarchical image database. *Proc. of CVPR 2009*, pp.248–255.
- [5] Fujii, A., and Ishikawa, T. 2005. Toward the automatic compilation of multimedia encyclopedias: Associating images with term descriptions on the Web. *Proc. of Web Intelligence 2005*, pp.536–542.
- [6] Fujita, S. and Nagata, M. 2010. Enriching dictionaries with images from the Internet. *Proc. of COLING 2010*, pp.331–339.
- [7] Hayashi, Y. and Savas, B. and Nagata, M. 2009. Utilizing images for assisting cross-Language information retrieval on the Web. *Proc. of WIRSS 2009*, pp.100–103.
- [8] Hayashi, Y. and Nagata, M. and Savas, B. 2010. Exploring the visual annotatability of query concepts for interactive cross-language information retrieval. *Proc. of AIRS 2010*, pp.379–388.
- [9] 林 良彦, 永田昌明, Bora Savas. 2012. 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. 情報処理学会論文誌:データベース, Vol.5, No.2, pp.26–35.
- [10] Kwong, O.Y. 2012. *New Perspectives on Computational and Cognitive Strategies for Word Sense Disambiguation*, Springer.
- [11] Manning, C.D, Raghavan, P., and Schütze, H. 2008. *Introduction to Information Retrieval*, Cambridge University Press.
- [12] McRae, K., Cree, G.S., and Seidenberg, M.S. 2005. Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, Instruments, and Computers*, 37(4):547–559.
- [13] Miller, G. 1998. Nouns in WordNet. In: Fellbaum, C. (Ed), *WordNet, An Electronic Lexical Database*, pp.23–46, The MIT Press.
- [14] Paivio, A., Yuille, J.C., and Madigan S.A. 1968. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, 76 (1, Part 2):1–25.
- [15] Silberer, C., and Lapata, M. 2012. Grounded models of semantic representation. *Proc. of the 2012 Joint Conference on EMNLP*, pp.1423–1433.
- [16] Silberer, C., Ferrari, V., and Lapata, M. 2013. Models of semantic representation with visual attributes *Proc. of ACL 2013*, pp.572–582.
- [17] Tanaka, S. Jatowt, A., Kato, M.P., and Tanaka, K. 2013. Estimating content concreteness for finding comprehensible documents. *Proc. of The Sixth ACM WSDM Conference*, pp.475–484.