

オブジェクトの外観情報の Web マイニング

服部 峻[†] 手塚 太郎[†] 田中 克己[†]

[†] 京都大学大学院 情報学研究科 社会情報学専攻 〒 606-8501 京都府京都市左京区吉田本町

E-mail: †{hattori,tezuka,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし 我々はこれまで、地物（建築物）という特定領域のオブジェクトを対象に、その名称が与えられた場合に、その外観に関する言語的な記述を Web からマイニングし対応付ける手法を考案して来た。しかしながら、これまでの提案手法は、地物という領域だけに依存した手法ではないため、他の領域に属するオブジェクトの外観情報の抽出にも適用できる可能性があるとして我々は考えた。そこで、本論文ではまず、同手法をそのまま用いて、地物に限らず、動植物や製品など、一般のオブジェクトの外観情報を同様に高精度に抽出することができるか否か、検証実験を行って考察する。その上で、一般のオブジェクトの外観情報を Web から抽出する手法の改良を試みる。

キーワード 情報抽出, Web マイニング, 外観情報, 印象語, 感性語.

Discovering Appearance Information from the Web

Shun HATTORI[†], Taro TEZUKA[†], and Katsumi TANAKA[†]

[†] Department of Social Informatics, Graduate School of Informatics, Kyoto University

Yoshida-Honmachi, Sakyo, Kyoto 606-8501, Japan

E-mail: †{hattori,tezuka,tanaka}@dl.kuis.kyoto-u.ac.jp

Abstract We have proposed a method to extract linguistic description about a geographical feature's appearance by mining the Web based on its name for route guidance or universal access so far. This method does not, however, depend on the specific domain (geographical features) and so we found that this method might be also applicable to the other domains (general objects). In this paper, first, we perform some experiments and discuss their results, in order to examine whether or not this method can also accurately extract appearance information for general objects such as animals, plants or manufacturing products. Based on the experiments, we propose an improved method to extract linguistic description about a general object's appearance for its name by mining the Web.

Key words Information Extraction, Web Mining, Appearance Information, Impression Word, Look and Feel.

1. はじめに

近年、情報検索技術の発展により、情報を調べたい対象であるオブジェクトの名称が検索時に分かっている場合には、その名称をそのまま検索クエリとして入力さえすれば、ある程度十分に満足できる検索結果を得ることができるようになって来た。しかしながら、検索対象のオブジェクトの名称が検索時に分かっている場合、対象オブジェクトの様々な特徴（外観や印象など）を検索クエリとして入力しただけでは、その対象オブジェクトに関する情報を効率良く、高精度に検索することは容易ではない。また、検索対象のオブジェクトの名称を検索者が未知の場合には、検索結果ページから、その対象オブジェクトの名称を見付け出すことも非常に困難である。

情報検索技術の進歩と同時に、モバイル計算機環境の整備も進み、携帯電話などにより広く一般に普及して来ており、モバ

イル環境で情報検索を行う機会も年々増加して来ている。例えば、街中や自然の中など、様々な場所で出遭ったオブジェクトに関して、その名称、さらには、その詳細情報を調べたいという要求がある。今日の携帯電話の多くにはデジタルカメラ機能が搭載されており、情報を調べたい対象オブジェクトの写真を撮影することが可能な場合には、その写真を画像認識することで、対象オブジェクトの名称を識別するというアプローチ^(注1)が考えられるが、不特定領域の一般オブジェクトを画像認識する実用的な手法は著者らの知る限り未だ考案されていない。

対象オブジェクトの写真を撮影できない場合にも対応するためには、写真を撮影するのではなく、情報を調べたい対象オブ

(注1): OLYMPUS, “スティッキーズ!,” <http://4965.jp/>

NTT-IT, “MagicFinder/O,” <http://obje5.csquash.jp/v-portal/TechIndex>, “PicLin,” <http://www.piclin.jp/>

ジェットの外観（外見）を検索者が言語を用いて表現し、その記述を入力として、システムが対象オブジェクトの名称を識別するというアプローチが有用であり、オブジェクトの名称と、その外観に関する言語的表現との対応関係を構築しておく必要がある。我々は、これまで、地物名だけから成る従来の道案内文に対して、各地物の外観に関する言語的記述を付加し、より分かり易い道案内文を生成するために、地物の外観情報を Web から抽出する手法を考案して来た [1]。また、文書中の地物画像に対して、その周辺テキストから特定した地物名だけでなく、その地物の外観に関する言語的記述も Web 全体からマイニングして来て画像に対応付けることで、画像中に写った地物の外観を説明する代替テキストとして利用したり、外観語によって地物画像を検索したりすることも可能になる [2]。

これまで我々が考案して来た手法は、地物の外観情報の抽出が目的ではあったが、地物という特定領域に特化した手法ではなかった。そこで、本論文ではまず、地物以外の一般のオブジェクトに対して同手法を適用し、地物の場合と同様に十分に高い精度で抽出できるか否かを検証するために実験を行った。しかしながら、これまでの手法を一般の対象オブジェクトにそのまま適用したのでは良好な結果を得ることができなかったため、これらの検証実験の結果を考察した上で、一般のオブジェクトの外観情報を抽出するために改良した手法について述べる。

2. 関連研究

2.1 評判情報の抽出

Web や Blog 文書などを解析し、名称が与えられた対象オブジェクトの評判・意見情報を抽出する手法に関する研究が盛んに行われている。小林ら [3] は、評価対象表現、属性表現、評価表現の共起パターンを利用することで、これら領域依存の表現を効率的に収集する手法を提案している。藤村ら [4] は、文を構成する上で主要な単語のみを用い、文節の n-gram による素性を提案し、Web 文書から抽出した評判情報を肯定・否定・非評価に判定する分類器を構築している。鈴木ら [5] は、ある対象の評価表現として属性名と評価語のペアを Blog 文書中から抽出し、ラベル付きデータだけでなくラベル無しデータも学習に組み込んだ Semi-Supervised な学習手法により、肯定・否定・非評価に分類している。また、Web サービス「blogWatcher 3.0b」において、評判情報検索として実装されている。赤木ら [6] は、商品やサービスなどに対して、どのような観点から典型的に評価されているか、評価属性の名称の抽出を試みている。

2.2 話題語の抽出

ユーザから与えられた主題語に対して、Web 上でどのような話題が在るかを呈示し、検索クエリの曖昧性の除去を支援する等を目的とした、対象オブジェクトの話題語を抽出する手法に関する研究も数多く行われている。小山ら [7] は、主題語と詳細（内容）語の関係に基づく話題構造を、Web 文書の論理構造を活用して抽出する手法を提案している。タイトル中の語は一般に文書の全体的な話題を表し、本文中の語は主題について詳細に記述していると考え、主題語 A をタイトル中に含むページ群で頻繁に現れる語 B が見付かれれば、語 B は語 A を詳細化

するために用いられていると見なす。文書の構造を考慮しない主題語 A と語 B の共起確率と、主題語 A をタイトル中に含む場合の語 B の共起確率とを比較検定し、後者の共起確率の方が統計的に有意に高くなる語 B を、主題語 A を詳細化する語として抽出している。野田ら [8] は、ある主題語 A の話題となる語 B は、日本語による記述において、「(主題語 A) の (語 B)」という形式で現れることが多いことに着目し、各候補語に対して、主題語に対する話題語としての相応度を測る評価式を導入して、典型的な話題語の抽出を試みている。

3. 地物の外観情報の抽出を目的に考案した既存手法の一般オブジェクトへの適用

本章ではまず、地物の外観情報を Web からマイニングするために考案して来たこれまでの抽出手法について説明する。その上で、一般オブジェクトへの適用可能性を検証するために行った実験結果について述べ考察する。

3.1 地物の外観情報の抽出手法

地物の外観情報として、主に「色」「形状」「大きさ」「造り（建築様式）」「建て（階数）」などの外観属性が挙げられる。最も単純には、各外観属性の値となりうる候補を辞書として用意しておき、その候補集合の中で、対象の地物の名称との単純共起頻度（或いは係り受け共起頻度）に基づいて、候補の重み付けをする手法が考えられる。しかしながら、ある地物の外観を言語的に表現する際、「東京タワー」のように、その地物全体に対して記述されることが多い地物もあるが、地物は一般に様々な部分要素から構成されており、その部分要素に依って外観的な特徴が異なれば、その部分要素毎に、その色や形状、大きさなどの外観情報が記述されることになる。従って、ある地物名に対して、その構成要素毎に、その構成要素の外観に関する言語的記述を抽出できる必要がある。この要件は、一般オブジェクトにおいても同様に言える。そこで、我々がこれまでに考案して来た手法では、ある地物名に対して、まず、それを構成する要素の名称を Web から抽出し、その上で、その構成要素毎に外観に関する言語的記述を抽出し対応付ける。

3.1.1 地物の構成要素名の抽出

ある地物を構成する部分要素の名称のリストを Web マイニングする手法として、コーパス文書集合において「(地物名)の～」に高頻度で続く名詞に注目するという手法が考えられる。しかし、日本語の文法における「の」は様々な目的で使用されるため、Google 等の Web 検索エンジンに「“(地物名)の”」という検索クエリを投げた検索結果を解析しても、地物の構成要素とは何ら関係のないキーワードも多数抽出されてしまい上手くない。野田ら [8] は、この手法で抽出されたキーワードを対象の話題語と呼んでいる通り、外観情報として必要な対象の構成要素名以外も大量に含んでしまう。一方で、我々は、「(地物名)の～」という記述が画像（写真）の周辺に在る場合、「(地物名)の～」に高頻度で続く名詞が対象の構成要素名を表していることが多いことを発見した。そこで、「(地物名)の」という検索クエリを Web 検索エンジンではなく画像検索エンジンに投げ、その検索結果を基に「(地物名)の～」に続く名詞

を形態素解析により切り出し、その地物の構成要素名の候補語とする。次に、各候補語に対して、対象の地物に対する構成要素名としての相応度を以下の式によって評価する。

$$weight_o(c) = df^{image}([o \text{ の } c]) + df^{image}([oc]) \quad (1)$$

ここで、 $df^{image}([o \text{ の } c])$ は、地物名と候補語とを連体助詞「の」により接続した [(地物名 o) の (候補語 c)] という検索クエリを画像検索エンジンで実行した結果の検索件数を表す。また、 $df^{image}([oc])$ は、地物名の後に候補語を直接連結した [(地物名 o)(候補語 c)] という検索クエリを画像検索エンジンで実行した検索件数を表す。そして、評価値が十分に高い候補語ほど、対象の地物の構成要素名として採択される。

3.1.2 地物の構成要素毎の外観情報の抽出

対象地物の名称とそれを構成する部分要素の名称が既知である場合に、その地物全体ではなく、その構成要素独自の外観に関する言語的な記述を Web から抽出する手法について述べる。基本的には「(地物名)の(外観記述)(構成要素名)」という表現形式に着目した手法である。この形式の表現を収集するため、[(地物名)の*(構成要素名)] という検索クエリを Web 検索エンジンに投げた検索結果を解析し、「(地物名)の」と「(構成要素名)」の間の文字列を切り出し、外観記述の候補とする。但し「*」は Google 等がサポートする「ワイルドカード」検索を意味する。次に、外観記述の各候補 v に対して、Web 検索エンジンに [(地物名)の(外観候補)(構成要素名)] という検索クエリを入力した結果の検索件数により、対象の地物 o の構成要素 c 毎の外観情報としての相応度の評価値とする。

$$weight_{o,c}(v) = df^{web}([“o \text{ の } vc”]) \quad (2)$$

ここで、 $df^{web}([“o \text{ の } c”])$ は、地物名と候補語とを連体助詞「の」により接続し、その文字列の後に構成要素名を直接連結した [(地物名 o) の (候補語 v)(構成要素名 c)] という検索クエリを Web 検索エンジンで実行した検索件数を表す。そして、[(地物名)の*(構成要素名)] という検索クエリを Web 検索エンジンで実行した検索件数に比べて、十分に大きな占有割合(比率)を持つ外観記述の候補だけを、対象の地物の構成要素毎の外観情報として採択する。

3.2 一般オブジェクトの外観抽出への適用実験

これまでに考案して来た上述の抽出手法は、一般のオブジェクトではなく地物の外観情報を抽出することで目的ではあったが、地物という特定領域のオブジェクトに特化した手法ではないため、地物以外の一般のオブジェクトに対しても同手法を適用し、地物の場合と同様に十分に高い精度で抽出できるか否かを検証するために行った実験について述べる。

3.2.1 実験の方法と結果

既述の手法によって、一般のオブジェクトの外観情報を高精度に抽出できるか否かを検証するため、以下の 5 つのオブジェクト名に対して、まず構成要素名を抽出し、次に構成要素毎の外観情報を抽出する。公式ページや Wikipedia、画像検索の結果などを参考にして、外観が特徴的な各オブジェクトの構成要素名とその外観記述を列挙すると次のようになる。

- 巖島神社(地物):
 - 大鳥居: 朱色。楠の四脚造り。海の中に立つ。
 - 本殿, 廻廊: 柱や棟は朱塗りで、屋根は黒色。
 - 他の構成要素例: 平舞台, 高舞台, 能舞台など。
- カワセミ(動物, 鳥):
 - 嘴: 成鳥は長く、幼鳥は短い。オスは黒、メスは下だけ紅。
 - 腹: 成鳥の腹は橙色で、幼鳥はやや灰色味を帯びる。
 - 頭: 青緑色で大きく、鱗の様な模様がある。
 - 足: 短い。成鳥の足は赤く、幼鳥は黒っぽい。
 - 他の構成要素例: 目, 頬, 首, 背中, 羽翼, 尾など。
- ヘラクレスオオカブト(動物, 昆虫):
 - 角: 非常に長く太い。オスの成虫のみが有する。
 - 頭: 金髪のような毛が生えている。
 - 前翅: 高湿では黒色だが、低湿では黄褐色になる。
 - 他の構成要素例: 胸角, 頭角, 胸, 腹, 足など。
- 福寿草(植物, 花):
 - 花: 黄色。花径は 3cm ほど、花弁は多数。花期は初春。
 - 葉: 羽状に細かく分かれる。
 - 茎: 短い。当初は背丈が低いですが、暖かくなると伸びる。
 - 他の構成要素例: 根, 蕾, 芽など。
- Wii(工業製品):
 - 本体: 白色。据え置き型ハードの中では最も小さい。
 - リモコン: 白色。縦に長い。テレビリモコンのような形。
 - 他の構成要素例: ヌンチャク, センサーバーなど。

対象オブジェクトの構成要素名を抽出する実験においては、構成要素名の候補語として、[(地物名)の] という検索クエリを Google の画像検索エンジン^(注2)で実行した検索結果のスニペット中から「(地物名)の~」に続く名詞を形態素解析[9]により切り出す。画像検索のスニペットは非常に短いため、「(地物名)の~」という表現が全て入り切っていない場合があるが、その際は元文書を参照する。但し、「写真」「画像」などの語は、Google の画像検索エンジンが自動的に検索クエリから削除してしまい、また、構成要素名として明らかに相応しくないので、予め候補語から除外する。一方で、以上の手法で抽出した候補語リストには含まれなかったが、対象オブジェクトの構成要素名の適合解として列挙した語(表 1 で語の前に「*」を付している)も含めて評価値の計算を行っている。

次に、対象オブジェクトの構成要素毎に外観情報を抽出する実験においては、対象オブジェクトの構成要素名を期待通りに抽出できたと仮定し、各オブジェクトに対して相応しい構成要素名のいくつかに対して外観情報の抽出を試みる。その際、外観記述の候補として、[(地物名)の*(構成要素名)] というワイルドカード検索クエリを Google の Web 検索エンジン^(注3)で実行した検索結果のスニペット中から「(地物名)の」と「(構成要素名)」との間の文字列を切り出して採用している。「(構成要素名)」に係る文字列だけに限定すれば精度は向上するが、表 2 では参考のために載せている。

(注2): Google イメージ検索, <http://www.google.com/imghp?hl=ja>

(注3): Google ウェブ検索, <http://www.google.co.jp/>

表 1 一般オブジェクトの構成要素名の抽出への適用した結果の例

Table 1 Component Extraction for General Objects

オブジェクト名	抽出された構成要素名と評価値（降順ソート）
厳島神社	大鳥居 (346=207+139), 鳥居 (274=166+108), 回廊 (80=39+41), 風景 (73=71+2), 能舞台 (70=6+64), 境内 (61=31+30), 国宝 (56=1+55), 本殿 (47=9+38), 宝物館 (30=1+29), 歴史 (29=4+25), 中 (24=24+0), 鹿 (18=12+6), 高舞台 (8=4+4), *平舞台 (6=0+6), 由来 (6=5+1), 海中 (6=3+3), *廻廊 (4=0+4), 位置 (4=4+0), 信仰 (2=2+0), 現在 (2=1+1), 売店 (1=1+0), 床 (1=1+0), 光景 (1=1+0)
カワセミ	幼鳥 (114=50+64), 姿 (99=99+0), ホバリング (97=54+43), わかさま (91=91+0), 巢 (57=57+0), アルバム (53=50+3), メス (42=15+27), ダイブ (28=26+2), 巣立ち (28=21+7), 里 (22=21+1), 絵 (17=16+1), つがい (15=10+5), 羽 (14=13+1), 背中 (13=5+8), 大きさ (10=4+6), ムサシくん (9=9+0), 生息 (7=7+0), 繁殖 (6=5+1), 話 (6=6+0), 求愛行動 (6=6+0), 足 (5=5+0), 嘴 (3=2+1), 目 (5=2+3), 腹 (2=1+1), 頭 (1=1+0), 心配 (1=1+0), とき (1=1+0), 呼吸 (1=1+0), ステッカー (2=1+1)
ヘラクレスオオカブト	T シャツ (57=10+47), 飼育 (49=10+39), 角 (28=28+0), 幼虫 (18=10+8), 羽化 (10=10+0), バランス (9=9+0), サナギ (4=4+0), さなぎ (3=3+0), 成虫 (3=1+2), 標本 (2=2+0), 模型 (1=1+0), 魅力 (1=1+0), 交尾 (1=1+0), 仲間 (1=1+0), 亜種 (6=2+4), *頭 (0=0+0), *前翅 (0=0+0), *頭角 (0=0+0), *胸角 (0=0+0)
福寿草	群生 (190=96+94), 里 (176=175+1), 花 (174=171+3), 群落 (50=38+12), 蕾 (25=20+5), 芽 (17=17+0), 黄色 (16=11+5), 種 (11=11+0), 間 (11=11+0), お部屋 (5=2+3), 後 (4=4+0), 様子 (3=3+0), ページ (3=3+0), 代用品 (1=1+0), 生産地 (1=1+0), 子苗 (1=1+0), *茎 (1=0+1), *根 (1=0+1), *葉 (0=0+0)
Wii	ゲーム (698=119+579), 発売 (392=34+358), 名称 (241=4+237), 価格 (240=53+187), コントローラ (159=43+116), 発売日 (122=24+98), スマブラ (113=25+88), リモコン (98=4+94), ゼルダ (97=10+87), 本体 (84=1+83), 概要 (50=13+37), 値段 (42=17+25), 販売価格 (36=8+28), 注目ソフト (21=2+19), 隠し球 (20=10+10), 対応タイトル (10=3+7), センサーバー (6=2+4), 流行 (6=1+5), ボンバーマン (4=2+2), コンセプトビデオ (4=1+3), 野球ゲーム (3=1+2), 赤外線アレイ (2=1+1)

表 2 一般オブジェクトの構成要素毎の外観に関する記述の抽出への適用した結果の例

Table 2 Visual Description Extraction for Each Component of General Objects

オブジェクト名	構成要素名	抽出された外観情報と評価値（降順ソート）
厳島神社	大鳥居	朱丹の (67), 巨大 (30), 朱の (24), 社殿と (21), 赤い (20), シンボルである (16), 社殿や (15), 朱色の (9), 遠浅に建つ (5), シンボル (5), シンボルの (5), 正面から見る (3), 超有名な (3), シンボルでもある (3), 舞台から (2), 海中 (1), 神々しい (1), 回廊から (1), 中でも (1), 境内から (1)
厳島神社	回廊	朱の (3), 西 (3), 鳥居と (2), 朱色の (1), 丹塗りの (1), 朱塗の (1), 赤い (1), きれいな (1), 西側 (1), ような (1), 社殿及び (1), 入口東 (1)
厳島神社	能舞台	海に浮かぶ (3), 西側回廊付近に建つ (1), 国宝の (1), 大鳥居や (1), 中ほどには (1)
カワセミ	羽	美しい緑の (20), 青い (7), 幼鳥が 2 (6), 美しい青い (5), ヒナ 2 (5), ヒナ 5 (5), コバルトブルーの (3), 腰の (3), 幼鳥が 4 (3), 緑の (2), 鮮やかな青い (2), 背中 (2), 拾った (2), あの (2), 透明感あふれる (1), 繊細な (1), ヒナ七 (1), 幼鳥も 2 (1), 雛 3 (1), 親子三 (1)
カワセミ	足	赤い (25), 短い (1), 雑に (1), 幼鳥たちは (1)
カワセミ	目	鳴き声で (8), 姿を (4), 姿が (3), 映像 (3), 鋭い (2), 黒い (2), 仕草に (2), 1 枚 (2), コバルト色に (2), 狩りも (2), 写真が (2), 一枚 (2), 愛くるしい黒い大きな (1), 真剣な (1), は (1), 凄 (1), ぐりん (1), 表情や (1), 普段の (1), ルリ色が (1), 青さに (1), 青い光が (1)
ヘラクレスオオカブト	角	長い (4), 下の (3), 様に (3), 上 (2), 中で最も (1)
ヘラクレスオオカブト	前翅	注目した表現形式に合致する Web 文書が無かったため候補語さえ無し。
ヘラクレスオオカブト	頭	同上。
福寿草	花	黄色い (137), 黄いろい (15), 小さな (5), 迎春の飾り (5), 蕾や (4), かわいい (3), 開花と (2), 可愛らしい (2), 初 (2), 黄金 (1), 群落が続く (1), 開いた (1), きれいな (1), 後に咲く (1), 群落も (1), 綺麗な (1), 春を告げる (1), お目出度い (1), 画像と (1)
福寿草	蕾	黄色い (2), 花芽や (2), 花も固く (2), 可愛い (1), 丸い (1), 小さな (1), 黄色い花が (1)
Wii	リモコン	Wii (9890), ヌンチャク (29), 追加 (28), スティック (22), ワイヤレス (11), 新型 (7), 持つ (6), フライング (5), 専用 (5), 付属 (4), 殺人 (4), 動かせる (3), ロング (3), 拡張 (1), 単体 (2), モーション (2), 赤外線 (2), 真似して (2), 片手 (1), ウィー (1), 欠陥 (1), マイ (1)
Wii	本体	コントローラ (301), ソフトも (60), ソフトやゲーム (40), ソフト (11), ロンチタイトル (11), ヌンチャク (4), 箱と (4), オプション (3), スタンドから (2), ハード (2), 中古 (1), 新品 (1)

3.2.2 実験結果の考察

一般オブジェクトを構成する要素名の抽出に適用した実験の結果(表1)を考察した結果、以下のような知見が得られた。

- オブジェクトによって、或いは、オブジェクトの種類によって、外観に関する話題が全ての話題集合に占める重要性が低いものもある。このため、画像の周辺テキスト中だけを対象にして「(オブジェクト名)の～」という表現形式に注目したとしても、文書の全体のテキスト中を対象にして「(オブジェクト名)の」に続くキーワードを採用するよりは改善されるが、そのオブジェクトの構成要素名以外の様々なキーワードを依然として含んでしまう。
- 地物、特に「巖島神社」のような寺社仏閣では、その外観や歴史に関する話題が多くを占めるため、その構成要素名を高精度に抽出することができる。
- 動植物は成長に伴って形態を劇的に変え、その呼び名が変わる。鳥である「カワセミ」に対する抽出結果では「幼鳥(若鳥)」「成鳥」など、昆虫である「ヘラクレスオオカブト」に対する抽出結果では「卵」「幼虫」「さなぎ」「成虫」などの語が出現する可能性が高い。また、雌雄を別する動植物に対する抽出結果は「雌(メス)」「雄(オス)」「つがい」などの語を含む可能性も高い。さらに、動植物は自律的に様々な行動をするため、それを撮影した写真の周辺テキストで行動名とも共起する。例えば、「カワセミのホバリング」「カワセミのダイブ」「ヘラクレスオオカブトの羽化」「ヘラクレスオオカブトの交尾」などが相当する。但し、これらの語は、対象オブジェクトの構成要素名としては不適であるが、対象の変態や雌雄、行動に依存した、より細分化した外観情報を抽出するために必要なパーツ語としては有用である。
- 工業製品では、ファッション性が重要視されるような製品(モバイル機器など)を除いて、「Wii」のように価格や機能性の方が重要視されるものでは、対象オブジェクトの構成要素名以外のノイズが非常に増え、その上、動植物とも異なり、より細分化した外観情報を抽出するためのパーツ語としても有用でないノイズが多くを占めてしまう。

次に、オブジェクトの構成要素毎の外観抽出に適用した実験の結果(表2)を考察すると、以下のような知見が得られた。

- 「(オブジェクト名)の(外観記述)(構成要素名)」という唯一の表現形式にしか注目していないため、オブジェクト名とその構成要素名に対して満たす文書数がそもそも少なく、上位 k 件の適合率が高い場合でも、再現率(網羅性)は低く、表現形式に合致する文書数が0件であり、候補語さえ全く得られない場合も起こり得る。また、表現形式に合致する文書数が少ない場合、適合解と不適解との件数の差が小さいため、容易に逆転してしまう危険性を孕む。
- 地物の「巖島神社」の外観情報の抽出結果では、何れの構成要素名に対しても上位を適合解が占めている。しかし、評価値の大きい構成要素名に対して抽出された外観記述は十分な高精度、及び、網羅性となるが、評価値の小さい構成要素名、つまり、Web文書を作成する際の話題として取

り上げられ難い構成要素名では、注目した表現形式に合致する文書数がやはり少なく、網羅性が低く、かつ、不適解との評価値の差が小さいため適合率が揺らぐ危険性が依然として残っており、未だ完全とは言えない。

- 動植物では、外観に関する記述が確かに存在するにも関わらず、注目した表現形式で記述されるケースが非常に少ないため、網羅性が低く、かつ、不適解との評価値の差が小さいため適合率まで揺らいでしまう。表3を見ると、動植物は地物(神社)に比べて、注目した表現形式に合致する文書数と単純共起頻度との比率が約1/10程度しかないため、他の外観表現にも対応する必要性がある。例えば、「(構成要素名)は/が(外観記述)」といった表現形式でよく出現し、かつ、外観以外に関する記述が続く場合が少ないように見受けられた。
- 工業製品である「Wii」の例では、構成要素毎の外観記述の候補は多数得られるにも関わらず、その中に適合解を含んでさえいない。候補リストに少なくとも適合解が含まれない限り、評価値の計算手法をいくら工夫したとしても無意味である。表3を見ると、構成要素名「リモコン」に関しては、注目した表現形式に合致する文書数、及び、単純共起頻度に対する比率も、地物(神社)の構成要素名と比べても十分に大きい、それにも関わらず、候補中に適合解が含まれないということは、工業製品の外観情報は注目したものと全く異なる表現形式で出現するようである。「(構成要素名)は/が(外観記述)」といった表現形式で出現はするものの、動植物の場合とは異なり、その価格や機能など外観以外に関する記述が続く場合が多く、やはりノイズが多くなってしまふ。但し、候補リストに少なくとも適合解が含まれれば、評価値の計算手法を工夫することでフィルタリングできる可能性はある。

表3 注目した表現に合致する文書数 $df^*(o, c) = df^{web}(["o" * c])$ と単純共起頻度 $df(o, c) = df^{web}(["o" AND "c"])$ との比較
Table 3 Proportion of Frequency of Web Documents Matching the Focused Descriptive Form to Simple Co-occurrence

オブジェクト名, 構成要素名	$df^*(o, c)$	$df(o, c)$	比率
巖島神社, 大鳥居	195/	51600=	0.00378
巖島神社, 回廊	41/	19900=	0.00206
巖島神社, 能舞台	37/	11100=	0.00333
八坂神社, 境内	117/	130000=	0.00090
八坂神社, 鳥居	93/	62400=	0.00149
八坂神社, 西楼門	51/	1270=	0.04016
カワセミ, 羽	105/	271000=	0.00039
カワセミ, 足	29/	192000=	0.00015
カワセミ, 目	76/	555000=	0.00014
ヘラクレスオオカブト, 角	6/	13500=	0.00044
ヘラクレスオオカブト, 前翅	0/	53=	0.00000
ヘラクレスオオカブト, 頭	0/	25300=	0.00000
福寿草, 花	167/	263000=	0.00063
福寿草, 蕾	12/	31800=	0.00038
Wii, リモコン	27200/	1730000=	0.01572
Wii, 本体	862/	1810000=	0.00048

4. 一般オブジェクトの外観情報を抽出するための改良手法の提案

本章では、前章で考察した我々の従来手法の問題点を踏まえ、地物だけでなく、その他一般のオブジェクトの外観情報も高精度に抽出できるように改良した手法について述べる。

4.1 一般オブジェクトの外観情報の抽出手法

我々の従来手法の問題点の一つは、対象オブジェクトを構成する要素の名称を抽出する際、画像の周辺テキスト中における「(オブジェクト名)の～」という表現形式に注目したが、そのオブジェクトの全ての話題集合に対して外観が占める割合が小さいと、構成要素名以外のキーワードも同一の表現形式で多数出現してしまうことであった。そこで、本章で提案する改良手法では、対象オブジェクトの構成要素名を抽出する際に、オブジェクトに依存せず、外観情報を記述するのによく使われる修飾句をサンプルとして与え、対象オブジェクトについて記述された文書中で、その修飾句に係る名詞句、或いは、その修飾句に係る名詞句を候補語とする。具体的には、色名を表す形容詞(「赤い」など)と名詞(「赤色」など)、大きさ、長さ、太さを表す形容詞の対(「大きい」と「小さい」など)をサンプルとして用いる。この手法により、少なくとも、外観を表す修飾句と強く結びつく名詞句に限定することが可能である。また、単純に、内容テキストの何処かに対象オブジェクトの名称を含む文書中で、与えた修飾句と係り受けする名詞句を抽出したとしても、その修飾句と名詞句のペアが、対象オブジェクトの外観を説明しない場合も多いため、タイトルに対象オブジェクトの名称を含む文書中に限定して、与えた修飾句と係り受けする名詞句を、対象オブジェクトの構成要素名の候補語として抽出する。

「赤い」や「大きい」といった形容詞から成る外観表現に基づいて、対象オブジェクトを構成する要素の名称を求める場合、 $[intitle:“(オブジェクト名)”AND(形容詞の基本形)]$ という検索クエリを Web 検索エンジンに投げた検索結果のスニペットにおいて、「(形容詞の基本形)」に係る、又は、「(形容詞の基本形)」に係る名詞句を日本語係り受け解析器 [9] を用いて抽出し、注目している形容詞を外観記述として持つ対象オブジェクトの構成要素名の候補語とする。次に、各候補語 c に対し、対象オブジェクト o の構成要素名であり、かつ、形容詞 v を外観記述として持つか否かの相応度を次の式により評価する ($weight$ の添え字「 a 」は、「adjective」を表す)。

$$weight_{o,v}^a(c) = df_o^{web}(["c * v_b"]) + df_o^{web}(["c * v_c"]) + df_o^{web}(["v_b c"]) \quad (3)$$

ここで、 $df_o^{web}(["q"])$ は o をタイトルに含み、かつ、 q を含む文書数を、 v_b は形容詞の基本形、 v_c は形容詞の連用形を表す。

一方、「赤色」といった名詞から成る外観表現に基づいて、ある対象オブジェクトを構成する要素の名称を求める場合、 $[intitle:“(オブジェクト名)”AND(名詞)]$ という検索クエリを Web 検索エンジンで実行した検索結果のスニペットにおいて、「名詞」に係る、又は、「(名詞)」に係る名詞句を係り受け解析器を用いて抽出し、注目している名詞を外観記述として持

つ対象オブジェクトの構成要素名の候補語とする。次に、各候補語 c に対して、対象オブジェクト o の構成要素名であり、かつ、名詞 v を外観記述として持つか否かの相応度を以下の式により評価する ($weight$ の添え字「 n 」は、「noun」を表す)。

$$weight_{o,v}^n(c) = df_o^{web}(["c * v"]) \cdot 2 + df_o^{web}(["v * c"]) \quad (4)$$

4.2 改善手法の評価実験と考察

3章で用いた5つのオブジェクト名「厳島神社」「カワセミ」「ヘラクレスオオカブト」「福寿草」「Wii」に対して、本章で提案した改良手法を適用すると表4のような結果となる。

まず、鳥の「カワセミ」に対する抽出結果について考察する。「カワセミ」に対して抽出された外観表現と構成要素名とのペアの全てを評価値の大きい順に列挙すると以下ようになる。ここで、対象の外観情報として相応しいペアを太字にしている。

青い/鳥 (217), 長い/時間 (168), 青い/画集 (147), 青い/空 (104), 短い/距離 (93), 赤い/下 (83), 短い/時間 (78), 青い/宝石 (77), 長い/くちばし (69), 白い/部分 (67), 赤い/嘴 (58), 赤い/実 (57), 青い/背中 (56), 長い/嘴 (55), 長い/間 (51), 青い/羽 (47), 白い/目の後 (43), 赤い/足 (40), 長い/レンズ (40), 短い/足 (36), 赤い/くちばし (35)...

我々の従来手法で構成要素名を抽出した結果(表1)では、上位を構成要素名以外のノイズが占めてはいたが、対象オブジェクトの変態、雌雄、行動名といった外観情報のパーツ語である場合も多かった。一方、本章の改良手法での抽出結果の上位には、構成要素名以外、しかも、対象オブジェクトの関連語でもないノイズ「時間」「空」「距離」など、また「カワセミ」の画集のタイトルである「青い/画集」「カワセミ」を漢字で書くと「翡翠」であることに起因する「青い/宝石」などが現れており、対象オブジェクトの構成要素名の抽出の精度が明らかに向上したとは言えない。しかしながら、従来手法で抽出できる適切な外観情報は最大でも、

美しい緑の/羽, 美しい青い/羽, 赤い/足, 短い/足, 鋭い/目, 黒い/目

程度であり、実際には構成要素名の抽出でノイズが多数あるため、これらの適合解の全てが上位に来るとは限らない。従って、改良手法での結果がノイズを含んだままであっても、より多くの外観情報をより正確に呈示できるように改善されたと言える。

ここで、実験結果を見渡すと、候補語のうち構成要素名以外のノイズの評価値を高めてしまっている項は、式(3)、式(4)各々の最後の項であることが分かる。そこで、最後の項を削除し、それぞれの評価値を以下のように再定義する。

$$weight_{o,v}^{a'}(c) = df_o^{web}(["c * v_b"]) + df_o^{web}(["c * v_c"]) \quad (5)$$

$$weight_{o,v}^{n'}(c) = df_o^{web}(["c * v"]) \cdot 2 \quad (6)$$

これらの評価式を用いて、「カワセミ」に対して抽出された外観表現と構成要素名のペア全ての評価値を計算し直し、さらに、

同じ外観表現に対して構成要素名の読みが同じものを併合した上で列挙すると、次のように上位の多くが適合解となる。

赤い/嘴+くちばし+クチバシ (97=51+35+11), 赤い/下 (82), 赤い/下嘴+下くちばし (52=28+24), 長い/くちばし+嘴+クチバシ (45=31+8+6), 長い/時間 (44), 白い/目の後 (43), 赤い/足 (29), 短い/時間 (24), 青い/背中 (20), 赤い/下のくちばし+下の嘴 (20=12+8), 短い/足 (15), 白い/糞 (10), 青い/空 (10), 赤い/実 (9), 短い/距離 (8), 青い/鳥 (6), 青い/羽 (6), 長い/レンズ (4), 白い/部分 (1), 青い/宝石 (1), 長い/間 (1), 青い/画集 (0)

対象オブジェクトの構成要素名の抽出としては十分に高い精度となっているが、対象オブジェクトの外観情報の抽出としては未だ不完全である。なぜならば、3.2.1 節で述べたように、「赤い/嘴」は、あらゆる「カワセミ」の外観を表してはならず、特に「成鳥」の「メス」の外観であるからである。従来手法での構成要素名の抽出結果に現れている変態や雌雄名を活用して、本章の改良手法に組み込むことを今後の研究課題とする。

同様に、昆虫の「ヘラクレスオオカブト」に対して抽出された外観表現と構成要素名のペアの全ての評価値を計算し直して大きい順に列挙すると、次のようになる。

太い/角 (12), 長い/角 (9), 長い/胸角 (7), 太い/胸角 (6), 黒色+黒い/前翅 (6=3+3), 細い/角 (5), 細い/胸角 (4), 大きい/昆虫ゼリー (3), 太い/頭角 (2), 黄色+黄土色/上翅 (1.5=1+0.5), ...

一方、従来手法で抽出されるペアを上位から列挙すると、

基本的/飼育 (10), 幼虫/飼育 (9), プリントが迫力の/Tシャツ (5), 長い/角 (4), 成虫と/幼虫 (4), 幼虫の/飼育 (2), 小ケース/飼育 (2), 幼虫が/羽化 (2), 成虫の通常/飼育 (1), 半袖/Tシャツ (1), ピックサイズ/幼虫 (1), オスが/羽化 (1), ...

となる。単純に上位 10 件の適合率を比較すると、従来手法では 0.1, 本章の改良手法では 0.7 であり、大きく改善していると言える。また、適合解の全集合を得ることが困難であり絶対的な再現率としての評価はできないが、従来手法での適合解は 1 個のみであり、本章の改良手法の方が 7 倍以上の再現率となる。

5. まとめと今後の課題

一般のオブジェクトの外観情報を抽出するという目標に対して、本論文ではまず、地物の外観情報を抽出するためにこれまでに考案して来た手法の適用可能性を検証するための実験を行い考察し、その上で、改良手法を提案した。我々の従来手法では、地物に対しては、その構成要素名、及び、構成要素毎の外観記述を高精度に抽出できるが、その他の一般オブジェクトに対しては、その構成要素名を精度良く抽出することができず、工業製品に至っては、正解である構成要素名を人手で与えたとしても、その構成要素の外観記述を抽出できなかった。しかし

ながら、本論文で提案した改良手法により、工業製品以外の一般オブジェクトに対しては、その構成要素名をより高精度に、かつ、網羅的に取得可能になった。工業製品に対しては、外観情報の抽出として使用に堪えるに十分な精度には未だ達してはいないが、外観記述の再現率がゼロから幾分か向上している。

今後は、抽出手法をさらに改良し、それを評価するために大規模な実験を行う。また、外観情報として、言語的な記述だけを抽出するのではなく、抽出した言語的記述に相応しい画像(写真)を精度良く検索する手法についても考案して行く。

謝 辞

本研究の一部は、文部科学省 21 世紀 COE 拠点形成プログラム「知識社会基盤構築のための情報学拠点形成」(リーダー: 田中克己, 平成 14~18 年度), 及び, 文部科学省研究委託事業「知的資産の電子的な保存・活用を支援するソフトウェア技術基盤の構築」, 異メディア・アーカイブの横断的検索・統合ソフトウェア開発 (研究代表者: 田中克己), 及び, 文部科学省科学研究費補助金特定領域研究「情報爆発時代に向けた新しい IT 基盤技術の研究」, 計画研究「情報爆発時代に対応するコンテンツ融合と操作環境融合に関する研究」(研究代表者: 田中克己, A01-00-02, 課題番号: 18049041) による。記して謝意を表す。

文 献

- [1] 服部峻, 手塚太郎, 田中克己, “サーチエンジンのメタデータを用いた道案内補助のための外観情報の抽出,” 情報処理学会 DBS 研究会 研究報告, Vol.2006, No.78, pp.169-176, 2006 年 7 月.
- [2] 服部峻, 手塚太郎, 田中克己, “Blog 文書中の地物画像を代替する言語的記述の抽出,” データベースと Web 情報システムに関するシンポジウム (DBWeb'06), pp.35-42, 2006 年 11/12 月.
- [3] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一, “テキストマイニングによる評価表現の収集,” 情報処理学会 自然言語処理研究会 研究報告, Vol.2003, No.23, pp.77-84, 2003 年 3 月.
- [4] 藤村滋, 豊田正史, 喜連川優, “文の構造を考慮した評判抽出手法,” 電子情報通信学会 第 16 回データ工学ワークショップ (DEWS2005) 論文集, 6C-i8, 2005 年 5 月.
- [5] Yasuhiro Suzuki, Hiroya Takamura, and Manabu Okumura, “Application of Semi-supervised Learning to Evaluative Expression Classification,” In Proceedings of the 7th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing'06), LNCS Vol.3878, pp.502-513, February 2006.
- [6] 赤木法生, 大島裕明, 小山聡, 田島敬史, 田中克己, “レビューページ例からの属性抽出に基づくレビューページ検索,” 電子情報通信学会 第 17 回データ工学ワークショップ (DEWS2006) 論文集, 2C-i10, 2006 年 6 月.
- [7] Satoshi Oyama, and Katsumi Tanaka, “Query Modification by Discovering Topics from Web Page Structures,” In Proceedings of the Sixth Asia Pacific Web Conference (AP-Web'04), LNCS Vol.3007, pp.553-564, April 2004.
- [8] 野田武史, 大島裕明, 小山聡, 田島敬史, 田中克己, “主題語からの話題語自動抽出とこれに基づく Web 情報検索,” 情報処理学会 DBS 研究報告, Vol.2006, No.78, pp.305-311, 2006 年 7 月.
- [9] Taku Kudo, and Yuji Matsumoto, “Fast Methods for Kernel-Based Text Analysis,” In Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL'03), pp.24-31, July 2003.
- [10] Michael J. Cafarella, and Oren Etzioni, “A Search Engine for Natural Language Applications” In Proceedings of the 14th International World Wide Web Conference (WWW'05), pp.442-452, May 2005.

表4 一般オブジェクトの構成要素名と典型的な外観表現とのセットの抽出結果の例

Table 4 Component Extraction for General Objects and Typical Visual Description

オブジェクト名	表現	抽出された構成要素名と評価値(降順ソート)
厳島神社	朱色	神社(23=10x2+3), 厳島神社(13=5x2+3), 大鳥居(12=1x2+10), 鳥居(10=3x2+4), 回廊(8=3x2+2), 柱(8=3x2+2), 全体(4=2x2+0), 神社内(4=2x2+0), 退色(1=0x2+1)
カワセミ	黒い	個体(31=0+0+31), くちばし(18=12+6+0), 嘴(13=8+4+1), クチバシ(10=8+2+0), オス(9=8+1+0), 全体(8=4+4+0), 顔(7=2+2+3), 足(5=4+0+1), 上下(5=3+2+0), 瞳(4=0+0+4), 線(4=0+0+4), 下クチバシ(3=2+1+0), 粉(3=0+0+3), 嘴の上下(2=2+0+0), 物(2=0+0+2), 物体(2=0+0+2), 下嘴(1=0+1+0), オメメ(1=0+0+1), ストライプ(1=0+0+1), 爪(1=1+0+0)
カワセミ	黒色	嘴(21=10x2+1), くちばし(6=3x2+0), オス(6=3x2+0), 上下(4=2x2+0)
カワセミ	白い	部分(67=0+1+66), 目の後(43=43+0+0), 花(36=1+0+35), 糞(29=1+9+19), 鳥(18=1+1+16), 雲(14=0+0+14), 色(12=1+3+8), 香り(12=0+0+12), 頭(10=7+1+2), 腹(10=2+3+5), ヒタキ(5=0+0+5), 羽毛(4=0+0+4), 物(4=0+0+4), 尾(2=0+0+2), 一部(1=1+0+0)
カワセミ	白色	先端(4=2x2+0), 先(4=2x2+0), 目の周り(2=1x2+0), 耳羽後方(2=1x2+0), 毛(2=1x2+0), 羽毛(1=0x2+1), 綿(1=0x2+1), ウツギ(1=0x2+1), 個体(1=0x2+1)
カワセミ	青い	鳥(217=4+2+211), 画集(147=0+0+147), 空(104=5+5+94), 宝石(77=1+0+76), 背中(56=9+11+36), 羽(47=5+1+41), 影(32=1+0+31), 色素(12=0+1+11), 実(8=2+0+6), 翼(6=1+0+5), 線(5=0+0+5), 羽自体(4=4+0+0), 体毛(3=0+0+3), 閃光(3=0+0+3), 体(1=0+0+1), 斑(1=0+0+1), 横斑(1=0+0+1)
カワセミ	赤い	下(83=66+16+1), 嘴(58=42+9+7), 実(57=0+9+48), 足(40=16+13+11), くちばし(35=25+10+0), 鳥(35=5+0+30), 下嘴(28=24+4+0), 下くちばし(24=20+4+0), 下のくちばし(12=10+2+0), クチバシ(12=11+0+1), 嘴の下(11=8+3+0), 下の嘴(8=8+0+0), メス(6=2+1+3), 雌(6=4+0+2), オス(6=3+1+2), あんよ(3=1+1+1), 下側(3=2+1+0), メスの下嘴(3=2+0+1), 嘴の下側(2=1+1+0), 基部(2=2+0+0), 嘴の下部(1=1+0+0)
カワセミ	黄色	鳥(11=2x2+7), 花(9=1x2+7), スイセン(6=1x2+4), くちばし(5=1x2+3), 下面(4=2x2+0), 羽(3=1x2+1), くちもと(2=1x2+0), 木々(2=1x2+0)
カワセミ	橙色	胸(14=7x2+0), 腹(12=6x2+0), 前後(12=6x2+0), 下面(9=2x2+5), 肢(6=3x2+0), 上腹(4=2x2+0), 下嘴(4=2x2+0), 目の後ろ(4=2x2+0), 下の嘴(2=1x2+0), おなか(1=0x2+1)
カワセミ	長い	時間(168=32+12+124), くちばし(69=10+21+38), 嘴(55=4+4+47), 間(51=0+1+50), レンズ(40=2+2+36), クチバシ(16=4+2+10), 話(10=1+2+7), 尾羽(7=4+0+3), 歴史(7=1+4+2), 縁側(5=0+0+5), 翼(3=2+0+1), 夜(2=0+0+2)
カワセミ	短い	距離(93=4+4+85), 時間(78=10+14+54), 足(36=11+4+21), 嘴(27=18+7+2), 尾(22=18+0+4), レンズ(16=0+1+15), 命(12=3+1+8), 尻尾(11=6+2+3), クチバシ(5=3+0+2), 道(4=0+0+4), 首(3=1+1+1), 文書(3=0+0+3), 話(3=0+0+3)
ヘラクレス...	黒い	斑点(4=0+0+4), 部分(4=0+0+4), 前翅(3=3+0+0), 点(1=0+0+1), 色紙(0=0+0+1)
ヘラクレス...	黒色	前翅(3=3+0), 全体(1=1+0), 成型色(1=0+1)
ヘラクレス...	白い	羽(1=1+0+0), 蛹(1=0+0+1), 点線(1=0+0+1), 布(1=0+0+1)
ヘラクレス...	白色	マット(4=0+4)
ヘラクレス...	青い	上翅(1=0+0+1), 風(1=0+0+1)
ヘラクレス...	黄色	上翅(1=1+0), 個体(1=0+1), 体(1=1+0)
ヘラクレス...	黄土色	前翅(1=1+0), 上翅(1=1+0), 縞々模様(1=0+1)
ヘラクレスオオカブト	大きい	サイズ(12=0+0+12), カブトムシ(5=0+0+5), 昆虫ゼリー(3=3+0+0), 黒斑(1=1+0+0), 大きさ(1=0+1+0), 顔(1=1+0+0), ヘラクレス(4=0+1+3), 奴(0=0+0+1)
ヘラクレスオオカブト	小さい	頃(82=0+0+82), ヘラクレス(2=0+0+2), 個体(2=0+0+2) 子供(2=0+1+1), もの(2=0+0+2), 容器(1=0+0+1), ケース(1=0+0+1), 虫(1=0+0+1), 角(1=0+0+1)
ヘラクレスオオカブト	長い	角(18=2+7+9), 胸角(7=1+6+0), 期間(6=0+5+1), 間(3=0+0+3), 傾向(2=0+0+2), 体長(1=0+1+0), カブトムシ(1=0+0+1), 寿命(1=1+0+0), 個体(1=0+0+1), 交尾(1=0+0+1)
ヘラクレス...	短い	足(1=0+1+0), 後脚(1=1+0+0), 角(1=0+1+0), 寿命(1=1+0+0), 方(1=0+0+1)
ヘラクレス...	太い	角(16=3+9+4), 胸角(6=1+5+0), 個体(4=0+0+4), 頭角(2=0+2+2), ミミズ(1=0+0+1)
ヘラクレス...	細い	道(8=0+0+8), 角(5=4+1+0), 胸角(4=3+1+0), 足(1=1+0+0), 特徴(1=0+0+1)
福寿草	黄色	福寿草(80=34x2+12), 花(71=12x2+47), 花びら(14=2x2+5), 花色(6=3x2+0), 花弁(5=0x2+5), おしべ(3=1x2+1), がく(1=0x2+1), 絨毯(1=0x2+1), 筒形(1=0x2+1)
Wii	白い	PS3(69=2+0+67), 戯言(47=0+0+47), 影(20=0+1+19), 部分(17=2+1+14), 箱(14=0+0+14), リモコン(14=2+0+12), 人(13=10+0+3), 目(11=0+0+11), 巨塔(8=0+0+8), ヤツ(7=0+0+7), 衣装(5=1+0+4), 壁(4=0+0+4), 筐体(3=0+0+3), 狼(3=0+0+3), 化粧箱(2=0+0+2), 丸(2=0+0+2), ゲーム機(1=0+0+1)