Web-imageability of the Behavioral Features of Basic-level Concepts

Yoshihiko Hayashi
Graduate School of Language and Culture, Osaka University
1-8 Machikaneyama, Toyonaka, 5600043 Osaka , Japan
hayashi@lang.osaka-u.ac.jp

Abstract

The recent research direction toward multimodal semantic representation would be further advanced, if we could have a machinery to collect adequate images from the Web, given a target concept. With this motivation, this paper particularly investigates into the *Web-imageabilities* of the behavioral features (e.g. "beaver builds dams") of a basic-level concept (`beaver`). The term Web-imageability denotes how adequately the images acquired from the Web deliver the intended meaning of a complex concept. The primary contributions made in this paper are twofold: (1) "beaver building dams"-type queries can better yield relevant Web images, suggesting that the present participle form ("-ing" form) of a verb ("building"), as a query component, is more effective than the base form; (2) the behaviors taken by animate beings are likely to be more depicted on the Web, particularly if the behaviors are, in a sense, inherent to animate beings (e.g., motion, consumption), while the creation-type behaviors of inanimate beings are not. The paper further analyzes linguistic annotations that were independently given to some of the images, and discusses an aspect of the semantic gap between image and language.

Keywords: Imageability of complex concept; Semantic feature norm; Semantic gap

1. Introduction

If the meaning carried by a linguistic expression is properly represented with non-linguistic media, the representation can be utilized in several types of applications, such as cross-language information retrieval (Hayashi et al., 2009) and language learning (Wang, 2010) systems. Recent attempts to integrate visual properties into semantic representation (Silberer et al, 2013) are highly promising, in the sense that such an approach is perceptually grounded (Barsalow, 2008). This direction toward multimodal semantic representation would be further advanced, if we could have a machinery to collect adequate images, given a target concept, from the Web.

Given this motivation, we conducted an investigation into the *Web-imageabilities* of complex concepts. In this investigation, a complex concept is denoted by an English expression (e.g., "beaver builds dams"), and comprises a basic-level nominal concept (`beaver`) and a semantic feature (`builds_dams`) for designating one of the salient behavioral properties of the target concept. Here, the term Web-imageability denotes how adequately the images acquired from the Web (henceforth, Web-images) deliver the intended meaning of a complex concept.

2. Semantic Feature Norms

2.1. Overview of McRae's Database

We utilized the well-known set of semantic feature norms provided by McRae et al. (McRae et al., 2005) (henceforth, McRae's database) as a source for extracting behavioral features of basic-level concepts. This database provides a total of 7,526 semantic feature norms assigned to 541 living and nonliving basic-level concepts, each organized on the basis of psychological experimental data collected from a large number of participants.

Table 1 exemplifies some of the semantic features given to describe `beaver`. Each row in the table shows a salient semantic feature of the target concept, as well as the number of participants who employ the feature. Also shown in Table 1 are Brain Region (BR) Labels, each of which roughly classifies semantic features from the perspective of brain function localization (Cree and McRae, 2003). As the frequency distribution (Table 2) demonstrates, perception-related categories, notably visual ones, are frequently observed in McRae's database.

Semantic feature	BR Label	Freq.
`a_mammal`	*taxonomic*	6
`beh_-_builds_dams`	*encyclopaedic*	20
`beh_-_chews_on_wood`	*visual-motion*	5
`beh_-_cuts_down_trees`	*encyclopaedic*	7
`found_on_the_nickel`	*encyclopaedic*	6
`has_a_tail`	*visual-form_and_surface*	24
`has_sharp_teeth`	*visual-form_and_surface*	24
`hunted_by_people`	*function*	7
`is_brown`	*visual-colour*	19
`is_furry`	*tactile*	18

Table 1: Semantic feature norms for describing `beaver`.

BR Label	Frequency
visual-form-and-surface	2,336
visual-color	424
visual-motion	339
tactile	245
sound	142
taste	84
smell	24
function	1,517
encyclopaedic	1,417
taxonomic	730

Table 2: Frequency distribution of BR Labels

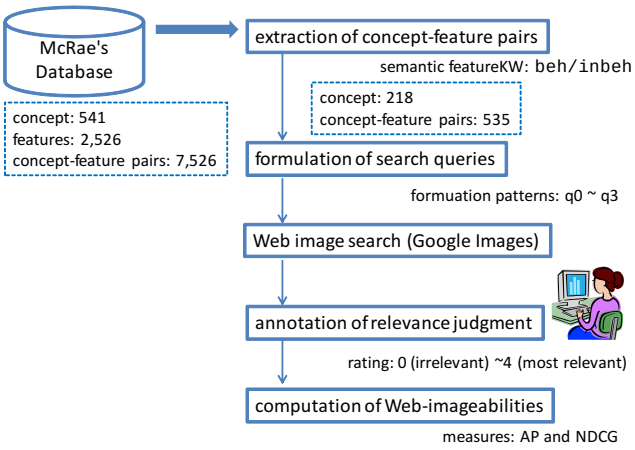


Figure 1: Overview of the Web-imageability assessment process.

2.2. Semantic Feature Keywords

As exemplified in Table 1, most of the semantic features are prefixed by predefined keywords or key phrases (e.g., "beh_- builds_dams"; "has_a tail"). These keywords and key phrases (henceforth, semantic-feature keywords) were introduced to classify semantic features into basic types. This paper focuses on two of the semantic-feature keywords introduced in McRae's database: beh_- and inbeh_-. The former signifies a behavior exhibited by animate beings (e.g., "beaver beh_- builds_dams"); while the latter denotes that an inanimate being does something seemingly on its own (e.g., "airplane inbeh_- crashes"). In McRae's database the number of semantic feature types with beh_- and inbeh_- amounts to 138 and 64, respectively.

3. Assessment of Web-imageability

Figure 1 gives an overview of the Web-imageability assessment process for assessing the Web-imageabilities of the behavioral features of target complex concepts. By looking at beh_- and inbeh_-, we extracted 535 concept-feature pairs (e.g., {beaver, builds_dams}) for 235 concepts from McRae's database.

Web image retrieval: Given a concept-feature pair, such as {beaver, builds_dams} and {accordion, produces_music}, we need to generate a query string to actually submit to a Web image search engine. This time, we employed four query formulation patterns (q0 through q3) by altering the verb form and the word order, as described below. Although we were well aware that the efficacy of query wording heavily depends on the nature of the search engine actually utilized, we wanted to explore an effective query pattern, if it exists, for the subsequent data analyses.

- q0: concept name alone ("beaver")
- q1: feature expression as given in the database ("beaver builds dams")
- q2: present participle verb form ("beaver building dams")

Figure 2: Examples of Web-images and their associated relevance ratings ("beaver building dams").

- q3: head (concept noun) final form of q2 ("building dams beaver")

Relevance judgment: We then used an annotator[1] to rate each of the retrieved Web images in terms of relevance to the meaning of a concept-feature pair. The judgment as to relevance was given on a 0-to-4 rating scale (from 0:irrelevant to 4:most relevant). We submitted each of the formulated 2,126 queries[2] to Google Images[3] and collected at most 15 images per query, yielding a total of 27,970 images, including duplicates. Figure 2 depicts examples of Web-images and their associated relevance ratings. Table 3 summarizes the overall results: it shows that virtually half (44.8%) of the Web images are considered relevant, when the relevance boundary is set between rating scores 1 and 2.

Measures of Web-imageability: We borrowed two IR-based performance measures (Manning et al., 2008), Average Precision (AP) and Normalized Discounted Cumulative Gain (NDCG), to measure the Web-imageability of a concept-feature pair (as represented by a query). Specifically, we regard the Web-imageability of a concept-feature pair to be higher than that of others if one or both of these performance measures are greater than those of its competitors.

4. Investigation into Web-imageability

This section investigates the Web-imageability results in terms of query formulation pattern and semantic composition of a concept-feature pair.

4.1. Query Formulation Pattern

Figure 3 compares the Web-imageability results, as measured by the IR-based performance measure, AP and NDCG, in terms of query formulation pattern. The figure clearly shows that, for the Google Images employed, q2-type queries, each of which uses present participle verb forms (e.g., "beaver building dams"), were significantly better than other query types ($p < 0.01$ both for AP and NDCG; Mann-Whitney U-test). These results prompted us to employ q2-type queries in the subsequent analyses.

[1]The annotation work was, in practice, divided among a group of annotators; however, the overall quality of the annotation was controlled by a supervisor. The authors of this paper did not participate in the work in either role.

[2]We could not formulate q3/q4-type queries for features with the "cannot + verb" pattern: One example is the feature "chicken cannot fly" for the target concept chicken.

[3]http://images.google.com/

Rating	0 (irrelevant)	1	2	3	4 (most relevant)
Number of images	6,705	8,734	4,025	1,910	6,596
Total: 27,970	15,439 (55.2%)		12,531 (44.8%)		

Table 3: Distribution of relevance ratings.

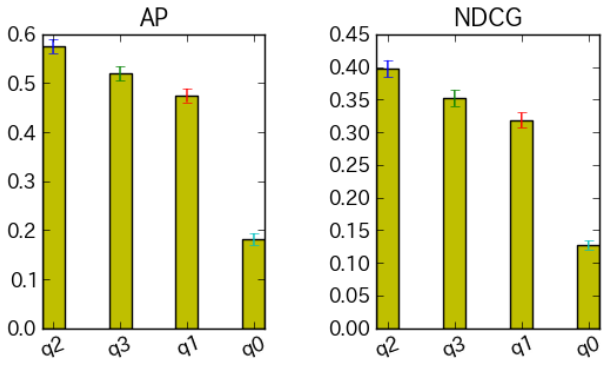


Figure 3: Comparison of Web-imageability by query formulation pattern.

Sense group	Number of noun concepts
animal	120
artifact	78
foot	11
plant	3
communication, substance	2
body, possession	1

Table 4: Distribution of the sense groups for the concept nouns.

4.2. Semantic Composition of a Concept-Feature Pair

In order to explore the potential relationships between the Web-imageability and the semantic composition of a concept-feature pair, this subsection presents the results of statistical analysis in which the WordNet lexicographer files[4] (LFs) were utilized as an inventory of semantic groups. Departing from its original purposes (Miller, 1998), the set of LFs has been utilized as a coarse-grained sense classification system in NLP (Kwong, 2012) and related fields. To achieve our objective, we manually assigned LF labels to the target concept nouns, as well as to the verbs appearing in concept-feature pairs.

Concept Nouns: Table 4 shows the distribution of sense groups for the targeted 218 concept nouns, while Figure 4 compares the IR-based performance measures. Although the results are somewhat different between the two measures, the LF-based sense groups can be divided into the higher performance group {*animal*, *food*, *substance*} and the lower performance group {*artifact*, *plant*}. In short, behaviors performed by animate beings tend to be more adequately depicted in Web images than the behaviors exhibited by inanimate beings.

[4] `http://wordnet.princeton.edu/man/lexnames.5WN.html`

Sense group	Number of feature verbs
motion	169
creation	117
consumption	108
communication	34
change	19
contact	16
competition, perception	15

Table 5: Distribution of the sense groups for frequent feature verbs.

Feature Verbs: Table 5 shows the distribution of sense groups for frequently occurred feature verbs, while Figure 5 compares the IR-based performance measures. Some sense groups were substantially frequent and exhibited significantly different tendencies from other sense groups: that is, {*consumption*, *motion*}, without a large range of variances, yielded higher performances; while {*creation*} constantly achieved the lowest performances in both IR-based measures. Major feature verbs belonging to the higher performance group are *consumption*: {{"eat," "chew," "drink," ...} and *motion*: {"fly," "swim," "crawl," "travel," ...}}. Conversely, the feature verbs belonging to the lower performance group are *creation*:{"produce," "build," "make," "give," "do," ...}, probably due to the fact that a created thing is not necessarily restricted to concrete things, as the example "music produces music" shows.

Noun-Verb Combinations: Table 6 shows the frequent noun-verb combinations, while Figure 6 compares the corresponding IR-based query performances. As expected, the combination *animal+motion* shows steadily high query performances, whereas the two IR-based measures surprisingly exhibit slightly different figures for the *animal+consumption* combination.

Sense group	Frequency
animal+motion	126
animal+consumption	114
animal+creation	63
artifact+creation	49
animal+communication	35

Table 6: Distribution of the frequent noun-verb sense combinations.

5. Analysis of Human-generated Annotations

An image which has been assessed as appropriately representing a certain linguistic meaning could be totally differently interpreted in different contexts, producing a kind of *semantic gap* between content and interpretation (Alm, 2006). To explore this issue in any way, we have collected

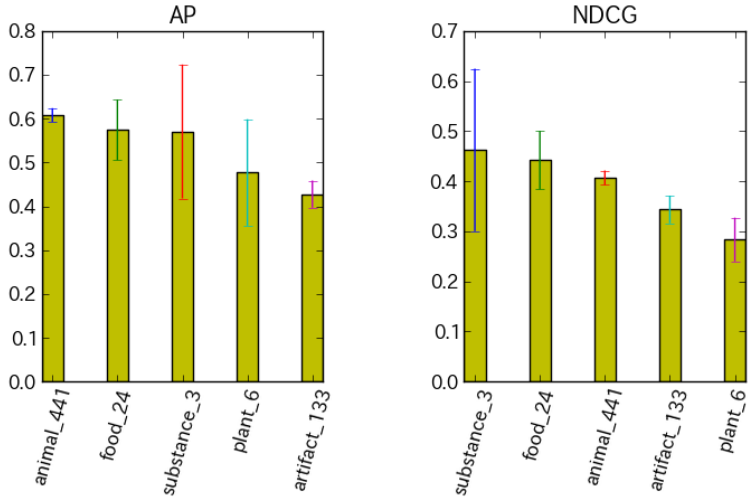


Figure 4: Comparison of query performances by sense group of concept noun.

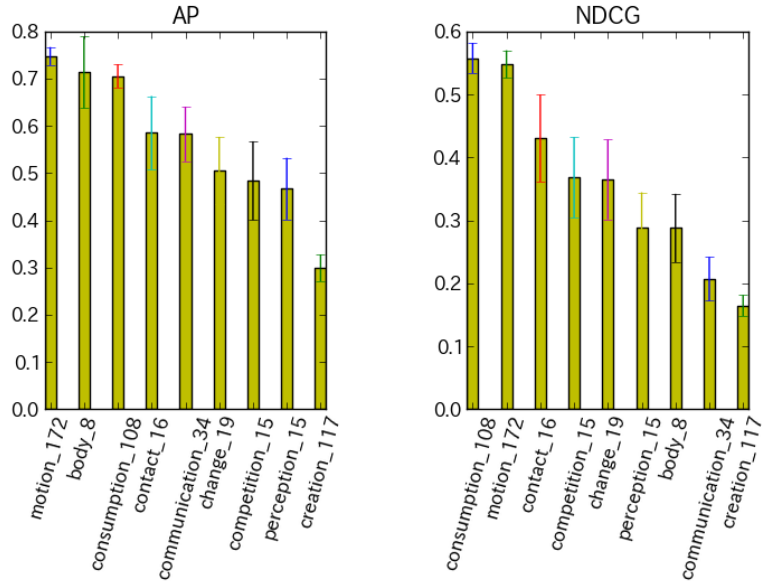


Figure 5: Comparison of query performances by sense group of feature verbs.

linguistic annotations for some of the acquired Web images, and analyzed the correlation between the imageability ratings and the semantic similarities calculated between the original semantic feature expressions and the acquired linguistic annotations.

5.1. Linguistic Annotations

We have recruited two annotators who are fluent in English, and had them independently annotate 3,653 of the images already described in the previous section. In the annotation work, we have directed them to employ the original sentence patterns (e.g. Subj+Verb or Subj+Verb+Obj) as far as possible, but we have not forced them to observe any other restrictions.

Figure 7 displays two example images, both assessed as highly relevant (relevance rating:4) for the given semantic features. The annotations given by the two annotators were:

- (a) “cheetah hunts”:
  - Annotator 1: “cheetah chases prey” (remark: the direction for sentence pattern was not observed in this case)
  - Annotator 2: “cheetah runs”
- (b) “faucet leaks”:
  - Annotator 1: “water drips”
  - Annotator 2: “The faucet frips”

5.2. Correlation Analysis

We assumed that the semantic similarity between an original semantic feature (e.g. “cheetah hunts”) and an annotation (e.g. “cheetah runs”) was given by a weighted sum

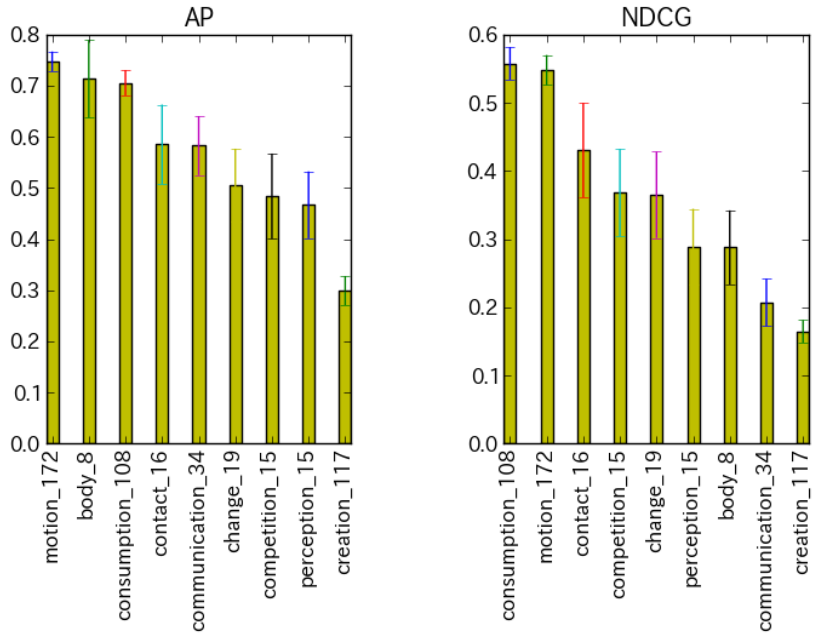
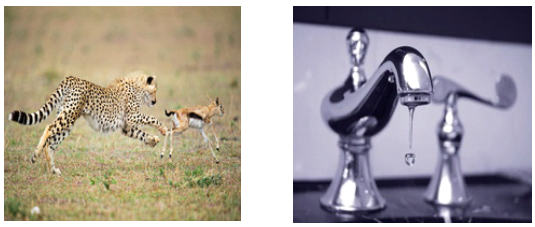


Figure 6: Comparison of query performances by combinations of concept noun and feature verbs.

	A1:sim-subj	A1:sim-verb	A1:sim-obj	A2:sim-subj	A2:sim-verb	A2:sim-obj
Lasso	-0.273	**1.011**	0.060	0.193	**0.685**	0.0
SVR	-3.446	**0.258**	-0.346	3.360	**1.638**	-3.387

Table 7: Weights for componential semantic similarities (A1: Annotator-1, A2: Annotator-2).



(a) cheetah hunts (b) faucet leaks

Figure 7: Examples of linguistic annotations.

of componential semantic similarities. More precisely, the sentential semantic similarity was calculated by balancing similarities between subjects, verbs, and objects (if any): each of the componential similarities was calculated by applying Wu-Palmer's and Lin's methods (Bundanitsky and Hirst, 2006), and the optimized weight for each component was adjusted by applying a linear regression method (Lasso) (Tibshirani, 1996) and support vector regression (SVR) (Drucker et al., 1997).

By applying the regression processes, we obtained the following correlation ratios (in Pearson) between the series of image ratings and the componential similarities: 0.472 for Lasso and 0.556 for SVR respectively (for both methods, $p < 0.001$). These results show that there existed modest-level correlations between them, insisting that the linguistic annotations independently given to an "easier" image could be more similar to the original semantic feature expression than those given to an "harder" image.

Table 7 summarizes the obtained weights for the componential semantic similarities. As shown in the table, the similarities between verbs played a more prominent role in correlating the two modalities: image and language.

6. Concluding Remarks

This paper investigated into the *Web-imageabilities* of the behavioral features (e.g. "beaver builds dams") of a basic-level concept (`beaver`).

The primary contributions made in this paper are twofold: (1) "beaver building dams"-type queries can better yield relevant Web images, suggesting that the present participle form of a verb ("building"), as a query component, is more effective than the base form; (2) the behaviors taken by animate beings are likely to be more depicted on the Web, particularly if the behaviors are, in a sense, inherent to animate beings (e.g., motion, consumption), while the creation-type behaviors of inanimate beings are not.

Although these findings are limited to the concepts and the concept-feature pairs investigated in the presented work, the resulting resource can be utilized as part of training data for learning the imageability of Web-images relative to a given concept. Moreover, the presented work could initiate a new research direction that deals with the imageability of complex concepts, rather than atomic concepts, as a concept-feature pair in this paper can be seen as a kind of complex concept.

Furthermore, the correlation analysis discussed in the final

section revealed that the semantic gap could be relatively narrower for some of the Web images. Our future issues thus include the understanding the nature of such "easier" images and the "harder" images such as shown in Fig. 7. To advance this direction, we would incoporate established image features (such as SIFT), and consider linguistic theories of actions.

7. Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 24650123.

8. References

Alm, C.O. et al. (2006). Challenges for annotating images for sense disambiguation. *Proc. of the Workshop on Frontiers in Linguistically Annotated Corpora*, pp.1–4.

Barsalou, L.W. (2008). Grounded cognition. *Annual Review of Psychology*, 59:617–645.

Bundanitsky, A., and Hirst, G. (2006). Evaluating WordNet-based measures of semantic distance. *Computational Linguistics*, Vol.32, No.1, pp.13–47.

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.

Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A., and Vapnik, V. (2004). Support vector regression machines. In M. C. Mozer, et al. (eds.), *Advances in Neural Information Processing Systems 9*, pp.155–161.

Hayashi, Y., Savas, B., and Nagata, M. (2009). Utilizing images for assisting cross-Language information retrieval on the Web. *Proc. of Web Intelligence/IAT Workshops (WIRSS 2009)*, pp. 100–103.

Kwong, O.Y. (2012). *New Perspectives on Computational and Cognitive Strategies for Word Sense Disambiguation*, Springer.

Manning, C.D, Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*, Cambridge University Press.

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.

Miller, G. (1998). Nouns in WordNet. In: Fellbaum, C. (ed), *WordNet, An Electronic Lexical Database*, pp. 23–46, The MIT Press.

Silberer, C., Ferrari, V., and Lapata, M. (2013). Models of semantic representation with visual attributes. *Proc. of ACL 2013*, pp. 572–582.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Royal. Statist. Soc B.*, Vol. 58, No. 1, pp.267–288.

Wang, H-C. (2010). Using language-retrieved pictures to support intercultural group brainstorming. *Proc. of ACM GROUP '10*, pp. 351–352.