Proceedings of the ASME 2014 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC/CIE 2014
August 17-20, 2014, Buffalo, New York, USA

DETC2014-35288

TOWARDS EXTRACTING AFFORDANCES FROM ONLINE CONSUMER PRODUCT REVIEWS

Amanda Chou, L.H. Shu*
Dept. of Mechanical and Industrial Engineering
University of Toronto, 5 King's College Road
Toronto, ON, M5S 3G8, Canada
*Corresponding author: shu@mie.utoronto.ca

ABSTRACT

We examined online product reviews as a source of novel affordances. Certain affordances may only be discovered through extended use across various environments. User-generated reviews may thus contain unique insights. We analyzed online consumer product reviews from Canadian Tire, one of Canada's largest retailers. We determined properties of this collection of reviews and commonalities between valuable reviews. In addition to typical challenges associated with natural-language processing, e.g. word-sense disambiguation, we identify characteristics of online consumer reviews that create additional challenges. These challenges include the use of 'wild English' and sarcasm in online reviews.

We first present criteria to define and more objectively identify novel affordances from review content. Next, k-means clustering reveals that a combination of syntactical features and high frequency word percentages can separate descriptive from non-descriptive review content. Finally, we identified cue phrases that may indicate higher likelihood of affordance content in a review. Despite existing obstacles, the substantial volume of available online product reviews has potential to become a valuable source of affordances and feedback for designers and retailers alike.

1 AFFORDANCES

We begin with our understanding of the concept of affordances before summarizing the significant body of others' work on affordances.

1.1 Our understanding of affordance versus function

While there is little consensus on the definition of an affordance (Burlamaqui and Dong, 2014), we begin by quickly clarifying what we mean by the term *affordance*, while differentiating it from what we mean by *function*.

Any flat horizontal surface affords the placing of items on it. Such flat surfaces may exist to satisfy a closely related function. For example, bookshelves and tabletops are flat horizontal surfaces that exist for the function of supporting items. However, many flat surfaces arise from other functions but still afford the placing of items on top, which could have neutral, positive, or negative effects. For example, a flat horizontal surface may result from an enclosure but include a ventilation grill, as with some dehumidifiers, heaters, etc. Items placed on top of such flat surfaces would degrade intended functions such as ventilation, or if on top of heaters, may constitute fire hazards.

Yet casual removal of common affordances may lead to reduced product satisfaction. For example, product geometry that arbitrarily prevents the resting of objects on top thwarts a common use that may unnecessarily frustrate the user. Figure 1 shows a non-flat top surface of a hot-drink dispensing machine. The tape holding down a container of drink stirrers demonstrates not only an expected affordance, but that the geometry frustrates the user in this expected affordance.

Figure 1. Expected affordance thwarted by geometry

Affordances may thus arise from properties like geometry, material and color, or their interactions, and may exist as side effects of design decisions made. They may be either desired or undesired, a designation that is context-specific in itself: a desired affordance in one context may be harmful in another.

In this paper, we focus on product affordances that may serve some novel use, in the sense that they were unintended by the designer. We refer to intended uses as functions. Other researchers emphasize the interaction component of affordances, or what one could physically do in a more basic sense (e.g., push on a door), as opposed to satisfying some higher-level need (e.g., allow access to pantry). We make no distinction between the two.

1.2 Origins and history of affordances

Affordances originated in perceptual psychology as an idea that objective and measurable action possibilities exist for every object or environment (Gibson, 1979). This concept has since evolved in areas such as artificial intelligence, human-computer interaction as well as product design.

Don Norman popularized the notion of affordances for product design with his book *The Design of Everyday Things* (1988). Norman noted that the design of an artifact provided visual cues to its user regarding its action possibilities. It follows that functions, or designer-intended affordances can be made obvious through design, and mitigation of undesired affordances should be considered. In other words, the designer could (and should) anticipate the ways in which users may interact with their product in different environments.

1.3 Affordances and the design research community

Maier and Fadel (2006, 2009a, 2009b) have worked to systematize affordance-based methods in the design process. User needs identified through methods such as surveys and focus groups are converted into affordances, which are then used to inform design decisions. Affordances of prototypes are then analyzed to iteratively improve the design.

Brown and Blessing (2005) have shown the value of affordances in product redesign. Cormier et al. (2013) focus on desired affordances that translate from user needs. Kim et al. (2013) used activity analysis to support design for affordance. Srivastava and Shu (2013) explored the transfer of affordances to enable resource-efficient behavior. Burlamaqui and Dong (2014) point out that affordance is an inconsistently used term, and work towards developing a more useful framework.

2 AFFORDANCE IDENTIFICATION AND NOVELTY

Maier and Fadel (2007) note the potentially infinite number of affordances, since any given system can interact with infinitely many artifacts and environments. We therefore considered the type of affordances on which to focus. We are interested in the ability of affordances to describe obscure or novel usage opportunities. While non-obvious product capabilities may be highly context-specific, they may still point to exciting potential for product enhancement, or the need for redesign to avoid undesired use possibilities.

2.1 Limitations of existing methods

Surveys, interviews and focus groups have been used to gain insights about user needs during the information collection phase in design projects. However, survey respondents are unlikely to offer detailed answers about a wide range of products. Interviews allow researchers to capture details on the informants' extended interactions with a product, potentially revealing unexpected affordances. However, such interviews (and focus groups) are both time- and cost-intensive, may be impractical for collecting large amounts of data, and may depend on the interviewer's skills. In addition, discussions have been shown to polarize participant opinion, and consumers may not always be honest or open in person (Morgan, 1996).

2.2 Identifying novel affordances through product features

With respect to design innovation, uncommon or obscure affordances are of most interest. McCaffrey and Spector (2012) noted that innovations often stem from obscure features, which give rise to unexpected functions or affordances that can be exploited. Specifically, in a design experiment, they used the many features of a burning candle to conceptualize a self-snuffing candle by exploiting its mass-losing nature. Placed on one side of a balance, the candle will rise as it burns and wax drips off the balance, eventually reaching a snuffer above.

While examination of a product's features can be used to identify affordances, it is limited in contextual scope. It may be difficult to imagine the varying environments where products can be used, and thus, the tendency to use them in certain ways. For example, it may be obvious that a lockset with a numeric keypad affords entry without a key. However, it may be less obvious that a numeric keypad better affords use by feel rather than by sight, as it is more difficult to align and insert a key by feel alone. Such an affordance could become clear only after the lockset is installed in a location that is visually obstructed, e.g., by a drainpipe when used to secure below-porch storage, shown in Figure 2. Both numeric and keyed locksets could enhance this affordance by improving ability to operate by touch alone. We note that many numeric locksets already light up when activated, in correct anticipation of operation in the dark.

Figure 2. Lockset (circled) to secure under-porch storage

2.3 Motivation for our approach

We are interested in affordances with respect to existing products that may or may not arise from user needs. Specifically, we wish to discover latent affordances previously unknown, or not salient, to the designer, rather than address all affordances that should be considered. The affordances suggested in product reviews are likely to be extensions of those originally considered in the design of the product.

Such affordances are perhaps more useful for redesign or design improvement, consistent with Brown and Blessing (2005), who note that affordances should complement the functional view of design proposed, e.g., by Pahl and Beitz (2007). Brown and Blessing recognize that "other potential positive functions, as well as negative functions, might not be identified during the design process, but only during the use phase, due to unexpected modes of employment, user intentions, or constraints." We believe that user-generated product reviews may contain precisely this kind of information.

3 PROCESSING ONLINE CONSUMER REVIEWS

With increasing retailers and distributors turning to web-based consumer reviews, a vast amount of consumer insight is now publicly available. The informants share their thoughts in the comfort and privacy of their own environments. The task requires less effort and structure, and the volume of information is greater than that available using more formal information-gathering methods. Additionally, the online nature enables easier access for researchers, allowing for data aggregation and observation of trends across several product types and features. However, the unguided review environment also leads to significant amounts of irrelevant information when seeking useful affordance-related insights. It is thus worthwhile to develop methods to extract potentially useful information.

While online reviews may not have been used to systematically discover affordances, natural-language processing of online review text has been performed for a variety of applications. These studies both inform our approach and highlight the associated challenges.

3.1 Use of "wild English"

Natural language used in online reviews is inherently difficult to process computationally. Often referred to as 'wild English', review text is frequently plagued with improper grammar, spelling, punctuation usage, and slang. Heavy preprocessing of text is often necessary before part-of-speech (POS) tagging and parsing algorithms can be applied. One example of a non-trivial problem is sentence boundary detection, which involves identifying which periods actually act to end sentences. This is necessary for POS-tagging and dependency parsing, which apply a consistent set of features to find patterns in text. These tasks come with disambiguation problems of their own: e.g., *tire* can be both *a rubber covering around a wheel* (noun) and *to exhaust* (verb).

3.2 Text classification and sentiment analysis

Our aim falls under the broad category of text classification, as we wish to separate useful, affordance-related content from the non-useful. One common classification task is sentiment analysis, i.e., classifying text by opinion. Sentiment analysis has been applied in a variety of domains, including movie reviews, political debates, restaurant reviews, and product reviews (Dalal and Zaveri, 2013). Since online reviews have a vital impact on buyers' decisions, it is often to a retailer's advantage to present review summaries, which involves sentiment analysis. This requires machine learning techniques that extract and weight features corresponding to various opinion levels. At the most basic level, the aim is to identify whether a review is negative or positive, i.e., perform a binary classification. A possible insight is that use of negations increases the likelihood that the review is negative. Challenges in the task include sarcasm, implicitness, and subtlety, all of which are difficult to detect automatically, e.g. "this movie is an excellent choice if you want to waste your evening".

3.3 Deception detection

Ott et al.'s deceptive opinion spam detection algorithm aims to classify hotel reviews as truthful or deceptive, again via machine learning techniques (2011). Using POS-tagging, they found that extensive first person usage, for example, is often indicative of a deceptive review. This study, amongst others, suggests that syntax can often hint at the tone or intention of a reviewer. Our focus is not on truthfulness, as even reviews planted by suppliers may provide useful affordances. However, we did consider how various parts of speech or specific words could indicate the affordance-containing potential in a review.

3.4 Product review opinion summarization

Dalal and Zaveri (2013) studied opinion summarization of product reviews. They did this by extracting product features as N-(noun) and NP-(noun-phrase) tagged words of high frequency, and then examined what followed to determine the sentiment. Essentially, this is feature extraction followed by sentiment classification, or feature-based sentiment analysis (Liu, 2010), e.g., "the battery life is amazing."

In contrast, frequencies of nouns do not help with finding novel affordances. While the subject of an affordance description is most likely a noun, most descriptions of nouns are not interesting affordances; they refer merely to specified features. Hence, we require a more sensitive method.

3.5 Cue phrases

Edmunson (1969) first proposed classifying sentences with certain pragmatic words, postulating their presence can affect the probability of relevance of a piece of text. This was later extended to the Cue Phrase method (Teufel and Moens, 1997). We explore this method prior to more automated approaches.

3.6 Eventual need for labeled training data

Common to almost all such analyses is the use of labeled training data from which the characteristics of each category are learned. Movie reviews, for example, are inherently labeled as reviewers often provide a star rating to accompany the review text. Such ratings provide a scaled quantity that represents overall product satisfaction. In the case of Ott et al.'s opinion spam detection (2011), freelance writers were hired to compose deceptive reviews to be used as training data.

Because identifying affordances in natural-language text is a newer concept, no such training data available to us exist. For now, we manually evaluate potential affordance text and perform exploratory analyses.

4. CHARACTERIZING NOVELTY IN AFFORDANCES

We first clarify what we believe characterizes novelty in affordances, and developed the following set of criteria.

4.1 Distance from intended function

Our interpretation of novelty is that the affordance is ‘unexpected’. Thus, the first criterion is that the affordance cannot be easily inferred from the list of product features or specifications, where specifications indicate intended uses or functions. While this criterion is still somewhat subjective, as the ability to ‘infer’ varies between people, specific scenarios that could further indicate novelty include the below.

4.1.1 Affordance arising as a side effect

Products include features with intended functions; suggested affordances may arise from the inclusion of those features, but as side effects, and may thus be novel to the designer. One challenge is that often, the obvious functions are not explicitly stated in a features list. For example, no kettle packaging will denote that the product has a spout or a handle.

Undesirable affordances are often side effects. Obvious negative affordances arising in expected use cases are likely recognized by the designer, but perhaps could not be mitigated. Since undesired affordances are generally unintended, they are often side effects and thus likely to represent novel information.

Even if corresponding corpora detailing obvious functions of products exist, enabling automated matching, there are multiple ways to express each function. In other words, even if a database of product functions existed, matching “you can use it to transport fluids” with “can hold water” is non-trivial. Indeed, Hoffmann et al.’s (2008) suggested semantic model for product feature description was intended for use only within each product’s own lifecycle, and not across products.

4.1.2 Difference in context of use

As affordances describe interactions, a change in usage context or environment may elucidate latent uses. However, both intended versus expected usage contexts, as well as the level of departure from these contexts are ill defined. For example, a bucket used in different places to hold different things is generally not surprising. Using an upside-down bucket as a stool may be a leap from its intended function, but is perhaps too common in real-life to be novel. Nonetheless, differences in usage context may indicate novelty.

4.2 Repeatability

Some reviews refer to subjectively determined experiences that may not hold true universally. Reviews suggesting a change in mood, feeling, preferences or decisions during or after using a product were scrutinized. While we suggested criteria above, affordance novelty is ultimately subjective. In case-by-case examination to distinguish the novel from the mundane, not even the co-authors agreed on whether certain affordances are useful or surprising. We will explore the role of inter-rater agreement in the future.

5. CORPUS STATISTICS

We chose to analyze reviews for Canadian Tire due to the sheer volume of available reviews, as well as its diversity in product offerings. Canadian Tire, one of Canada’s largest retailers with over 487 stores, offers sports, leisure and home products, in addition to automotive parts and services.

We created a database of all freely available Canadian Tire online reviews by scraping the source websites, using *Python* and *MySQL*. We obtained 60,922 reviews, dating back to 2007. Each review is characterized by the following set of features:

- *reviewKey*: a unique numerical identifier for each review
- *productName*: product name
- *description*: product description
- *reviewText*: body of the review
- *stars*: star rating corresponding to review
- *avgStars*: average of all prior ratings

Additional information includes the number of reviews for the product, product department, category, subcategory, regular price, review date and review title.

5.1 Nomenclature

First, we define the following terms used in this paper.

Corpus: body of text; plural *corpora.*

Content words: words with more semantic content, such as nouns and verbs; also called open-class words (since new words are often added).

Function words: words with little meaning, often serving grammatical purposes, such as *the* and *onto*; also known as closed-class words (as words are rarely added to this class).

Parsing: in computational linguistics, the analysis of sentences involving the identification of its constituents and their syntactical relationships, often represented as a parse tree.

POS: part-of-speech, a syntactical linguistic category to which a word belongs, such as *noun* or *verb.*

POS tagging: identifying and labeling the part-of-speech for tokens in a given text.

Tokenization: breaking down sentences into ‘tokens’, i.e., words, punctuation, etc.

(Web) scraping: extracting data from the web by accessing webpages and processing their HTML code.

Stop words: words often removed prior to language processing due to little semantic content; many are function words.

5.2 Star ratings

We conducted an exploratory analysis to characterize the set of review star ratings from Canadian Tire. We used a randomly selected set of 3,046 reviews. Of interest is the agreement reviewers seem to have in their assignment of star ratings to specific products. A proportional odds model showed that with every unit increase in the average of all previous star ratings for a product, the odds of achieving a higher star rating in the next review increases by 185% (95% CI [157%, 216%], $p < 2e\text{-}16$). In other words, a rating where the previous average is 3 stars has 2.85 times the odds of landing in the 4-5 stars range versus a rating where the previous average is only 2 stars.

If consumers tend to have similar levels of satisfaction with products they review, automated review summarization may be able to determine whether the satisfaction level stems from the same reasons, e.g., certain affordances. However, star ratings are highly subjective and involve several factors weighted differently by individuals, e.g., quality and price-performance ratio. In addition, cognitive biases, e.g., anchoring and availability, may also contribute to such agreement. Nonetheless, it may be possible to examine whether affordances correlate with star ratings assigned.

We also found that high star ratings tend to be posted earlier than lower ratings. This could indicate early review planting to increase sales. Alternatively, product issues may arise after some time has passed. The latter supports the value of online user-generated reviews to especially uncover negative affordances, as some may only arise after extended usage.

6 CLUSTERING

Clustering is the grouping of data by similarity. Commonly used in data mining, the technique can be used to identify whether and how a set of data clusters naturally. In the absence of labeled training data, we perform exploratory clustering to separate the type of sentences reviewers produce. We used k-means clustering to identify a group of non-descriptive reviews.

6.1 Feature selection

To cluster review sentences, we first select a set of quantifiable features. Due to its ease of extraction, we started with syntactical features, i.e., POS tags. Review text was tokenized and tagged with a modified Brill POS tagger. Each review sentence is then converted to a feature vector of values, e.g., percentage of first-person pronouns of the sentence.

6.1.1 Purchase-related frequent words

In addition, we manually identified in the corpus many purchase-related reviews with little descriptive content, and the words that they commonly shared. Thus, we include as a feature the percentage of purchase-related tokens, shown in Table 1. In the next section, we identify frequent words unique to the corpus, and also include the percentage of these tokens. The final set of features used is shown in Table 3.

Table 1. Purchase-related tokens, ranked by % frequency

bought	0.34%	cheap	0.04%	bucks	<0.01%
sale	0.16%	expensive	0.03%	dollars	<0.01%
buy	0.16%	value	0.03%	dollar	<0.01%
price	0.15%	warranty	0.03%	paying	<0.01%
$	0.15%	deal	0.03%	pricey	<0.01%
purchased	0.13%	pay	0.03%	penny	<0.01%
worth	0.08%	cost	0.02%	overpriced	<0.01%
money	0.06%	cheaper	0.02%	retail	<0.01%
store	0.05%	paid	0.02%	discount	<0.01%
purchase	0.05%	inexpensive	0.01%	raincheck	<0.01%

6.1.2 Non-purchase-related frequent words

Manual processing reveals that the majority of reviews did not contain affordances. In addition to purchase-related content, generic descriptions such as 'great fit' and 'very durable' were also uninformative. Since affordances are product-specific, words frequent throughout the corpus are likely to indicate little about affordances. Conversely, since reviews revealing affordances tend to be insightful, we may reasonably expect for them to use relatively infrequent words. Thus, highly frequent words may be candidate stop words for identifying affordance-containing reviews using machine learning algorithms.

Python's Natural Language Toolkit (NLTK) 3.0 provides a list of *stop words*. By comparing NLTK's stop words with the most frequent words in the Canadian Tire corpus, we identified high-frequency words unique to the corpus.

Table 2 lists the 60 most frequent words (in decreasing frequency) in the Canadian Tire corpus that did not coincide with stop words from NLTK, which are predominantly function words. The most frequent word in our list is ranked 25th in frequency out of all 4,153,562 words in the corpus vocabulary. Even in this small range of vocabulary, Zipf's Law applies: there are many more infrequent than there are frequent words.

For clustering purposes, we use the percentage of tokens that are words in Table 2 as a feature. We removed purchase-related words from the list. Also removed were words such as *car* and *bike*, products Canadian Tire carries in large quantities, and *battery* and *light,* common components in many products.

Table 2. Sixty most frequent words unique to Canadian Tire Reviews, ordered by percent frequency

one	0.45%	got	0.14%	go	0.10%
great	0.42%	first	0.14%	best	0.10%
use	0.35%	put	0.13%	around	0.09%
good	0.34%	work	0.13%	long	0.09%
would	0.34%	two	0.12%	bit	0.09%
easy	0.26%	love	0.12%	never	0.09%
get	0.25%	years	0.12%	new	0.09%
like	0.25%	recommend	0.12%	enough	0.09%
used	0.22%	need	0.12%	keep	0.09%
product	0.22%	still	0.12%	made	0.08%
well	0.22%	last	0.12%	every	0.08%
time	0.18%	using	0.12%	find	0.08%
even	0.17%	quality	0.11%	take	0.08%
also	0.16%	small	0.11%	lot	0.08%
really	0.16%	could	0.11%	found	0.08%
works	0.15%	better	0.11%	thing	0.08%
little	0.15%	nice	0.11%	another	0.08%
set	0.15%	year	0.11%	way	0.08%
much	0.14%	unit	0.11%	clean	0.08%
back	0.14%	make	0.11%	old	0.07%

An example review heavy in high-frequency words and low in affordance value is:

Great boot for the $$$ about $40 less expensive than Bass Pro and just as good! Great work Canadian Tire!

This review contains only generic sentiment expressions, function words, pronouns, and price information. Because what constitutes an affordance remains unresolved, we cannot yet determine whether the words in the above lists are ‘stigma’ words (those that point to irrelevant content), or whether they hold different weights in their degree of stigmatism.

6.2 Clustering results

Since labeling reviews as affordance-containing versus not is labor-intensive, we did not have enough training data to perform supervised classification on the corpus. In other words, we do not have manually labeled data to allow machine learning algorithms to learn to differentiate between the two classes of review text. We therefore use k-means clustering, a relatively simple unsupervised learning algorithm. Since this is an exploratory analysis, we cannot definitively measure the success rate of the clustering in separating the useful from non-useful. Rather, we discuss findings that may help characterize review text based on the selected features. All reviews were used, which corresponds to 163,251 sentences.

We use the open-source WEKA machine learning tool developed at the University of Waikato. The tool accepts as input, files containing feature vectors for each instance of data (one sentence). We created two clusters using the k-means algorithm. In order not to overfit our data, we started with a small set of features and experimented with additions until we obtained meaningful clusters. Since sentence length and percentage of words from the word lists discussed above appeared to be obvious distinctions between the classes, we began with those. Experimenting with POS features, we noticed that proper nouns frequently co-occurred with uninformative text; perhaps planted reviews with generic descriptions mention the brand or product names more frequently. Since there is expected correlation between some features, it is not obvious whether all features contribute significantly to the usefulness of the clustering. While some review text is clearly not useful, evaluation is more difficult with some more descriptive reviews.

We generated two clusters that divided the text meaningfully. Notably, the cluster of sentences involving a higher percentage of purchase-related words tended to also have higher frequent-word usage, as well as shorter sentence length (Cluster 1, as shown in Table 3). Cluster 1 also uses fewer coordinating conjunctions, possibly indicative of simpler sentence structure. Cluster 1 accounted for 25% of the instances.

As no item in the feature set directly related to novelty, we have not extracted affordances. However, Cluster 2 may have higher probability of containing affordances as they are more descriptive. Table 3 shows the final set of features used, as well as the clustering output.

Table 3. Cluster attribute means

Feature	Full Data 163,251	Cluster 1 40,965 (25%)	Cluster 2 122,286 (75%)
coordinating conjunctions	3.1%	2.5%	3.3%
proper nouns	2.1%	3.7%	1.5%
adverbs	6.9%	8.1%	6.4%
purchase words	1.9%	2.6%	1.7%
frequent words	9.6%	22.4%	5.3%
sentence length, in tokens	16.6	12.3	18.1
number of characters	62.4	45.6	68.1

To assess the quality of our clusters, we considered internal validity measures which depend only on the dataset and resulting cluster assignments. External measures require comparisons to labeled data. In general, ideal clustering has high *cohesion*, or compactness, and *separation* (Halkidi et al., 2002). Cohesion measures how close the data points within a cluster are, and separation measures how distant points from different clusters are. The Silhouette measure (Rousseeuw, 1987) combines both cohesion and separation in a single value ranging from -1 to 1, where a value close to 1 is ideal.

We obtained average Silhouette values for each cluster, $s_{cluster1} = 0.533$ and $s_{cluster2} = -0.327$, which averages to $s = 0.103$. Cluster 1 generally has much more positive values than Cluster 2, indicating that we are conservative in the pruning of non-useful sentences. In other words, while many sentences in Cluster 2 may be similar to those in Cluster 1, those in Cluster 1 are different from those in Cluster 2 with relative confidence.

Below is a set of 2 consecutive reviews from the training data. Sentences in Cluster 2, which are more descriptive, are italicized, while Cluster 1 sentences represent the remainder.

> *I bought these tire for my 1991 Chevy S10 Blazer 4WD, I had them for 5000k they were great tires for the price, even for all the offroading I did with them they were actually pretty good for just an all terrain. But these are definitly not a Mud Terrain sic I traded for some more aggressive tires. The tires were spotless for the time I had them.* Great Product from BFG as usual.
>
> I installed a set of these on my Jeep Grand Cherokee. *They ride smooth, quiet and really corner well. I use them as a suumer tire and will buy them again in the future.* They are a great value for the price.

Cluster 1 tends to hold little descriptive content. We may be able to further refine the feature list to separate sentences or reviews at a finer level. Perhaps with Cluster 1 identified as ‘noise’, it may be easier to look for patterns within Cluster 2.

7. CUE PHRASES

In information retrieval, cue phrases can often hint at the relevance of a sentence or paragraph. We thus seek positive cues that point to interesting affordances as an alternative method. Therefore, we considered examining phrases that tend to suggest affordance content with their presence. To explore the viability of this approach, we manually identified such phrases. We next extracted all reviews containing each phrase to review their effectiveness as an affordance indicator.

Table 4 lists cue phrases that were identified as candidates for extracting useful reviews. For each phrase, a list of reviews containing the phrase was compiled. We then manually assessed a randomly selected portion of these reviews to compute the precision of candidate phrases as a metric for affordance-related review content. For each phrase, we evaluated 30 reviews or all retrievals, whichever was lower. While the affordance content may not occur in the same sentence as the cue phrase, we postulate that reviewers carrying a certain tone through semantic choices may be more likely to have included insightful content. Note that affordance comments apply to all mentioned products, and are not limited to the products being reviewed.

Table 4. Cue phrases that co-occur with affordances

Phrase	*Hits (total reviews containing phrase)*	*Precision**
As opposed to	51	32%
Can actually	74	29%
Doubles as	32	25%
Than usual	18	28%
Be sure to	131	< 10%
Surprisingly	157	< 10%
Warning	130	< 10%
I noticed	374	< 10%
Possible to	53	< 10%

* Precision is computed as the proportion of accurate hits: (hits that are reviews containing affordances)/(total hits)

Note that recall is difficult to compute since we would need to compute or precisely estimate the percentage of quality reviews within the entire database. If, say, 3% of reviews were relevant, then recall for each phrase would be computed as: (Hits x Precision)/(3% x 60,922).

Examples of reviews that contain cue phrases include:

1. **Product**: Cling sunshade, 2-Pc
   Description: Opaque sunshade attaches to any car window; Set of two 13½ x 21" (34 x 53 cm) vinyl shades; Durable static-cling material; Removable and reusable

   Review: This is the safest kind of shade, <u>as opposed to</u> *the kind that clips/suctions to the window which can become airborne in a collision*. Excellent product.

2. **Product**: Meguiar's Hot Rims® wheel cleaner
   Description: New formula is safe for all wheels; Easy use, just spray on and hose off; Keep your wheels looking like new; 710 mL
   Review: Excellent rim cleaner that removes brake dust without the need to scrub it. You <u>can actually</u> *hear the cleaner fizzing while its cleaning the Rims*. […]

3. **Product**: 3-step stepstool (shown in Figure 3)
   Description: 3 big steps; 200-lb (91 kg) load rating; Very durable; White powder-coated finish, steel material; Wide black tread on each step for traction; 4' high (1.2 m); Folds for easy storage

   Review: I *use this at public events to help children step up*. The *top bar is perfectly placed so kids will hold it to keep their balance*. The big steps provide good-sized targets and a very stable base. <u>Doubles as</u> a chair in a pinch. Folds fairly flat, about 6" deep.

Figure 3. Canadian Tire Three-step stepladder

4. **Product**: Rain-X® Anti-Fog
   Description: Anti-Fog eliminates and prevents fogging and steaming on windshields and visors; Improves visibility; Anti-static formula repels dust; For use on interior mirrors and glass 207 mL

   Review: The product also *attracts dust to windshield* so I have to clean the inside of the windshield much more often <u>than usual</u>.

We first confirm the adherence of the above examples to the criteria for affordance novelty outlined in Section 4. Examples 1 and 4 refer to negative affordances of which the designer may not have been aware. While clips or suction cups may be easily attached and detached, that attribute may be problematic in a collision, i.e., an intended affordance could become detrimental in a different context. The fizzing sound in Example 2 and attraction of dust in Example 4 are likely product side effects with affordance implications, e.g., not having to scrub and more cleaning than usual.

In Example 3, the top bar of the stepstool being an ideal height for children is another side effect example, also in a separate sentence from the cue phrase. Typical collapsible stepstools require the bar for handling and portability. However, the top bar is likely unintended for holding onto while on the stepstool, since it would only reach the leg area of an adult. Thus, we can consider the use of a stepstool as a podium for children as an affordance with some degree of novelty. A possibility for product improvement would be to embrace the potential secondary use of the top bar, e.g., by making it more applicable to a greater number of people. In general, children are good sources of affordances, as they have accumulated less functional-fixedness about how physical objects are to be used. Rather, they use objects in the ways allowed, or in some cases, ways not allowed.

The affordances identified in these examples are also repeatable, in accordance with the second major criterion for useful affordances. Below, we further examine each of the key phrases with greater than 10% precision.

7.1 *As opposed to*

In particular, we note that with *as opposed to*, useful information tend to neighbor verb phrases, rather than nouns. In the sunshade example above, *as opposed to* is followed by a noun phrase. However, the verb phrase, which *can become airborne in a collision*, representing an affordance, is embedded within it. The following is an example where *as opposed to* does not include an affordance:

> *I liked the horizontal pan for traditional looking loaf as opposed to the tower looking loaf from machines with only one kneading paddle.*

Here, *liked* is the only verb that appears, and its subject is the first person, rather than a product or its feature. We will explore the potential to further validate and specify cases where the phrase is useful.

7.2 *Can actually*

The phrase *can actually* is an indication of surprise in itself: it suggests that what follows is unexpected. Reviewers may adopt this tone for emphasis, often using it to describe the degree to which something unsurprising can be done (e.g., "can actually clean very well"). However, *can actually* does often lead to insightful comments, on both positive and negative affordances.

Example of a negative affordance:

> *It comes with two pads (Polisher / Wax) which are so cheaply made you can actually see through one of them, which means after one use the foam pad underneath will probably make contact with your paint.*

Sarcastic tones are also common in the usage of this phrase:

> *Absolute garbage for what they're supposed to do. The auger destroys the solidity of the ground as you twist it in (that's IF you can actually twist it in) negating the holding power of auger in the first place.*

In the above, even when used sarcastically, the phrase still identifies a lack of expected affordance or function. It may be necessary in these cases to implement detection of irony or sarcasm as a step towards understanding the information provided. While progress has been made in the area of sarcasm detection, González-Ibáñez et al. (2011) found that neither human judges nor machine learning techniques perform well in detecting sarcasm in tweets.

Surprise in tone also manifests itself in other forms:

> *But, I use it for shopping ..? Whats that you say...lol.. SHOPPING-YES I do shopping for the month, and now days slower than usuall, But with this cooler by myside I can take all the time I want putting in groceries has I buy.*

Here, a reviewer describes using a cooler as a grocery-shopping cart. Syntactically, it is difficult to identify a unique pattern, but it is clear that the reviewer expected to surprise the reader. Also note the poor grammar and spelling.

7.3 *Doubles as*

Doubles as was also a promising phrase, often indicating an alternative function. Unfortunately, in many cases it will refer to one that was intended or obvious, such as:

> *The only gripe I have is that the 1-cup measure doubles as a 3/4-cup measure.*

Although the phrase usually refers to an additional positive feature, and expectedly so, the above example demonstrates that it can refer to a negative feature as well. The reviewer laments that rather than including a separate ¾-cup measure, the product merely includes a ¾-cup indentation inside the 1-cup measure. We can infer that the problem here is the inability to level off when measuring ¾-cup of dry material. This example demonstrates that specific knowledge and experience are often required to interpret implied meanings. Nonetheless, should distinguishing between positive and negative affordances be of interest, we could consider established sentiment analysis techniques.

7.4 *Than usual*

Appearing in the context of an affordance, *than usual* often indicates that the product causes some deviation from the norm. This can refer to a result of using the product, e.g., "this style of kitty litter basket…using more litter than usual", or uniqueness in a product feature, e.g., "smaller than usual size." The latter may demonstrate how slight changes can alter affordances in a product. It may also highlight what consumers perceive as the norm, and which variations from the norm they welcome.

8 SUMMARY AND FUTURE WORK

Examining the Canadian Tire corpus as a source for insightful affordances, we identified several characteristics of the corpus, as well as methods to identify affordance-containing reviews and extract their features. We also suggest ways to objectively determine novelty of an affordance to allow for the automation of its classification.

We found that much of the content pertains to purchasing and pricing details, durability of the product and satisfaction levels, which was no surprise. We created word lists that were used as part of our feature set for k-means clustering of review sentences. This was performed to determine whether certain linguistic traits tend to co-occur in sentences, and whether they can indicate value in the text. We identified a portion of non-descriptive sentences through this method.

Upon manual inspection and preliminary parsing, we identified particular potential in the phrases *as opposed to*, *can actually*, *doubles as,* and *than usual*. We postulate that their usefulness will hold across most product review websites, as they stem from semantics and not specific characteristics of the Canadian Tire review environment or its product offerings. While these cue phrases give high precision, they fall short in recall rate. They do, however, provide confidence that mechanized approaches, e.g. a combination of POS tags, syntax and words, may identify more complex patterns.

In the future, a more general corpus can be obtained, with reviews from multiple retailer websites. We propose manually labeling reviews to indicate whether they contain affordance content, to serve as training data. Since we expect a mix of both insightful and non-insightful text within a review, a latent Dirichlet allocation (LDA) model will be considered. With this, we can move toward obtaining a generalized feature set to characterize user-generated affordance content in online product reviews. Ultimately, we hope to find patterns within the insightful subset of reviews that may elucidate the discovery of affordances to support product design.

ACKNOWLEDGMENTS

The authors are grateful for the financial support of the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Brown, D., Blessing, L., 2005, The Relationship Between Function and Affordance, ASME IDETC/CIE, Sept. 24-28, Long Beach, CA, USA, DETC2005-85017.

Burlamaqui, L., Dong, A., 2014, The Use and Misuse of the Concept of Affordance, 6th Int. Conf. Design Computing and Cognition, June 23-25, London, UK.

Cormier, P., Olewnik A., Lewis K., 2013, Towards a Formalization of Affordance Modeling in the Early Stages of Design, ASME IDETC/CIE, Portland, OR, Aug. 4-7, DETC2013-13170.

Dalal, M., Zaveri, M., 2013, Semisupervised Learning Based Opinion Summarization and Classification for Online Product Reviews, Applied Computational Intelligence and Soft Computing, 2013, 910706.

Edmunson, H.P., 1969, New methods in automatic extracting, Journal of the ACM, 16(2): 264-285.

Gibson, J., 1979, The Theory of Affordances: In the Ecological Approach to Visual Perception, Boston: Houghton Mifflin.

González-Ibáñez, R., Muresan, S., Wacholder, N., 2011, Identifying Sarcasm in Twitter: A Closer Look. Proc. 49th Ann. Meeting of the Assoc. for Computational Linguistics: Human Language Technologies, 2: pp. 581-586.

Guilford, J.P., Christensen, P.R., Merrifield, P.R., Wilson, R.C., 1960, Alternative Uses, Manual of administration, scoring & interpretation, Beverley Hills, CA: Sheridan Supply Co.

Hoffmann, P., Feng S.C., Ameta G., Ghodous P., Qiao L., 2008, Towards a Multi-View Semantic Model for Product Feature Description, Collaborative Product and Service Life Cycle Management for a Sustainable World, 205-213.

Kim, Y. S., Hong, Y. K., Kim, S.R., Noh, J-H., 2013, User Activity Analysis for Design for Affordance, DS 75-5: Proc. 19th Int. Conf. on Engineering Design (ICED13) Design For Harmonies, Seoul, Korea, 5: pp. 31-38.

Liu, B., 2010, Sentiment Analysis and Subjectivity, Invited Chapter for the Handbook of Natural Language Processing, Second Edition.

Maier, J. R. A. and G. M. Fadel, 2003a, Affordance-Based Methods for Design, ASME IDETC/CIE, Chicago, Illinois, DETC2003-48673.

Maier, J. R. A., Fadel, G. M., 2006, Affordance Based Design: Status and Promise, Proc. IDRS'06, Seoul, South Korea.

Maier, J. R. A., Fadel, G. M., 2007, Identifying Affordances, Proceedings of ICED'07, Paris, France.

Maier, J. R. A., Fadel, G. M., 2009a, Affordance Based Design: A Relational Theory for Design, Research in Engineering Design, 20(1):13-27.

Maier, J. R. A., Fadel, G. M., 2009b, Affordance-Based Design Methods for Innovative Design, Redesign and Reverse Engineering, Research in Eng. Design, 20(1):225-239.

Halkidi, M., Batistakis, Y. and Vazirgiannis, M., 2002, On Clustering Validation Techniques, Journal of Intelligent Information Systems, 17(2/3), pp. 107-145.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009, The WEKA Data Mining Software: An Update; SIGKDD Explorations, Vol. e, Issue 1.

McCaffrey, T., and L. Spector., 2012, Behind Every Innovative Solution Lies an Obscure Feature, In Knowledge Management & E-Learning: An Int. Journal, 4(2):146-156.

Morgan, D. L., 1996, Focus Groups, Annual Review of Sociology, 22: 129-152.

Norman, D., 1988, The Design of Everyday Things, New York, NY: Basic Books.

Ott, M., Choi, Y., Cardi, C., Hancock, J. T., 2011, Finding Deceptive Opinion Spam by Any Stretch of the Imagination, Proc. 49th Ann. Meeting Assoc. for Computational Linguistics: Human Language Technologies, 1:309-319.

Pahl, G., Beitz, W., Feldhusen J., Grote K., 2007, Engineering design: a systematic approach, Vol. 3. New York, NY: Springer.

Rousseeuw, P.J., 1987, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20: pp. 53-65.

Srivastava, J., Shu, L.H., 2013, Affordances and Product Design to Support Environmentally Conscious Behavior, ASME J Mechanical Design, 135(10), 101006, pp. 355-363.

Taylor, A., Marcus, M., Santorini, B., 2003, The Penn Treebank: an overview, Text, Speech and Language Technology, Vol. 20.

Teufel, S., and Moens, M., 1997, Sentence extraction as a classification task, Mani and Maybury.

ANNEX A

THE PENN TREEBANK POS TAGSET

CC	Coordinating conj.	TO	infinitival *to*
CD	Cardinal number	UH	Interjection
DT	Determiner	VB	Verb, base form
EX	Existential there	VBD	Verb, past tense
FW	Foreign word	VBG	Verb, gerund/present pple
IN	Preposition	VBN	Verb, past participle
JJ	Adjective	VBP	Verb, non-3rd ps. sg. present
JJR	Adjective, comparative	VBZ	Verb, 3rd ps. sg. present
JJS	Adjective, superlative	WDT	Wh-determiner
LS	List item marker	WP	Wh-pronoun
MD	Modal	WP$	Possessive *wh*-pronoun
NN	Noun, singular or mass	WRB	Wh-adverb
NNS	Noun, plural	#	Pound sign
NNP	Proper noun, singular	$	Dollar sign
NNPS	Proper noun, plural	.	Sentence-final punctuation
PDT	Predeterminer	,	Comma
POS	Possessive ending	:	Colon, semi-colon
PRP	Personal pronoun	(	Left bracket character
PP$	Possessive pronoun	)	Right bracket character
RB	Adverb	"	Straight double quote
RBR	Adverb, comparative	‘	Left open single quote
RBS	Adverb, superlative	“	Left open double quote
RP	Particle	’	Right close single quote
Sym	Symbol	”	Right close double quote

ANNEX B

THE PENN TREEBANK SYNTACTIC TAGSET

ADJP	Adjective phrase
ADVP	Adverb phrase
NP	Noun phrase
PP	Prepositional phrase
S	Simple declarative clause
SBAR	Subordinate clause
SBARQ	Direct question introduced by *wh*-element
SINV	Declarative sentence with subject-aux inversion
SQ	Yes/no questions and subconstituent of SBARQ excluding *wh*-element
VP	Verb phrase
WHADVP	Wh-adverb phrase
WHNP	Wh-noun phrase
WHPP	Wh-prepositional phrase
X	Constituent of unknown or uncertain category
*	“Understood” subject of infinitive or imperative
0	Zero variant of *that* in subordinate clauses
T	Trace of wh-Constituent

Both tagsets from Taylor et al, 2003.