The Agile Cliché: Using Flexible Stereotypes as Building Blocks in the Construction of an Affective Lexicon

Tony Veale

Abstract Our affective perspective on a word is heavily influenced by the context in which it is used and by the features it is typically perceived to exhibit in that context. A nuanced model of lexical affect thus requires a feature-rich representation of each word's potential to mean different things in different contexts. To this end, we present here a two-level model of lexical affect. At the first level, words are represented as bundles of the typical properties and behaviors they are commonly shown to exhibit in everyday language. To construct these bundles, we present a semi-automatic approach to harvesting stereotypical properties and behaviors from the Web. At the second level, these properties and behaviors are related to each other in a graph structure that captures how likely one is to reinforce the meaning of another. We present an effective means of constructing such a graph from a combination of text n-grams and queries to the open Web. We calculate positive and negative potentials for each property in the graph, and show how these potentials can be used in turn to calculate an overall affective value for the higher-level terms for which they are considered stereotypical.

1 Introduction

Hamlet tells us that "there is nothing either good or bad, but thinking makes it so." This reasoning applies just as much to the words we use to label things as it does to the things themselves. In some contexts, for instance, "pride" denotes an admirable quality, but in others it denotes a deadly sin. Likewise, we praise go-getters, entrepreneurs and aspiring champions for their aggressiveness, but in many other contexts "aggression" denotes an unpleasant trait. Some words in English seem inherently positive or negative, but in reality there are very few words that cannot be given a reverse spin in the right context. Thus, words like *crazy, bad, wicked,*

Tony Veale
Web Science and Technology Division, KAIST, Yuseong, Korea. e-mail: tony.veale@gmail.com

sick and *evil* have all been re-engineered as positive descriptors in the vernacular of youth culture.

The sense inventories that lexicographers compile for a polysemous word offer a good approximation of the word's potential to convey meaning, but affect can operate across sense boundaries and even within individual senses, at the sub-sense level. Consider the word "baby", used to denote a human infant. In some contexts the word carries a positive affect: babies can be cute and adorable, curious and trusting, and an obvious target of love and affection, especially when asleep. Crying babies, however, can be selfish, whining, drooling, hissing, tantrum-throwing little monsters. Both views are stereotypical of human babies, and either can be intended when a speaker uses the term "baby" figuratively, whether to describe a beloved partner or an annoying colleague. This is a matter of conceptual perspective, not of lexical sense, and many other words exhibit a similar affective duality; "teenager" for instance can mean "whining brat" just as easily as "growing adolescent". The concepts *Baby* and *Teenager* are complex and multifaceted, and different uses in context may highlight different stereotypical behaviors of each. Their affective meaning in context is therefore not so much a function of which lexical sense is intended but of which behaviors are highlighted, and of the perceived affect of those behaviors.

Context can change the way we perceive the affect of a word or concept, and the language we use in context can reinforce this shift in perception, for language provides various means of putting an appropriate contextual spin on the perceived affect of a word. We might use an adverb with a strong affect of its own, as when we say someone is "impressively aggressive" or "disgustingly rich." We might tack the caveat "in a good way" or "in a bad way" onto the end of a description, or say "for better or worse" if we want to highlight both the positive and negative aspects of a word's meaning. Conversely, in ambiguous cases the addressee may seek clarification using the construction "good X or bad X?", as in "good strange or bad strange?" when someone has been described as strange. A pleasantly strange person may be novel, mysterious, exciting and unpredictable, whereas an unpleasantly strange person may be incomprehensible, troubling, alien and freaky.

Affective ambiguity is also found at the level of complex objects that are described in terms of these basic properties. President Barack Obama, for instance, is often criticized for acting like a "professor", though it would be an unusual dictionary that assigned a negative sense to this word. In this case, one assumes that it is the negative qualities of the stereotypical professor that are highlighted by the criticism. In turn, these negative qualities are the stereotypical traits that can be given the most negative spin, such as when scholarly objectivity and logicality are taken to be signs of emotional detachment.

In cases like "professor", we cannot rely on the lexicon to provide appropriate positive or negative senses for our words, for in practice most words can be given an affective spin in the right linguistic context. Rather, we should instead attempt to model and represent the stereotypical properties and behaviors on which different uses of the same word will derive their affective value in context. With a sufficiently rich behavioral model, we can determine the affect of a word like "baby" or "teenager" on a case-by-case and context-by-context basis, rather than wiring a one-size-fits-all measure of average affect directly into the lexicon. In short, we propose a two-level structure for a context-sensitive affective lexicon: a mapping of word-concepts to their normative stereotypical behaviors (e.g. *mewling, shrieking, drooling, sleeping* and *smiling*); and an affective profile of those behaviors (e.g. indicating the degree to which *shrieking* is unpleasant and *smiling* is pleasant). The affect of a word/concept in context can then be calculated as a function of the affect of its stereotypical behaviors that are primed in that context. We describe the construction of this two-level model of lexical affect in this paper. At the first level we capture the stereotypical properties and behaviors of commonplace ideas and the words that denote them. At the second level, we then calculate the perceived affect of a complex object – like *baby* or *professor* – as a function of those properties that are primed in context.

With these goals in mind, the rest of the paper assumes the following structure. We begin in section two with a discussion of related work in the field of lexical affect. In section three we then present a computational means of acquiring the stereotypical knowledge on which the current model is predicated. This knowledge is used in section four to estimate an affective value for each property and behavior in our representation, and for each complex object for which these properties and behaviors are considered stereotypical in everyday language. We outline how the two-level model can be used in an affective search application in section five. An empirical evaluation of the model is presented in section six, showing that good results are achieved on both of its levels. The paper concludes with a discussion of key issues in section seven.

2 Related Work and Ideas

In its simplest form, an affect lexicon assigns an affective score – along one or more dimensions – to each word or sense. The underlying lexicon may be a pre-existing resource that covers the bulk of a language , such as WordNet [7], or it may be a collection of sentiment-bearing words that aims to cover a small but relevant subset of the language. For instance, Whissell's *Dictionary of Affect* (or *DoA*) [23] assigns a trio of numeric scores to each of its 8000+ words to describe three psycholinguistic dimensions: *pleasantness, activation* and *imagery*. In the DoA, the lowest pleasantness score of 1.0 is assigned to words like "abnormal" and "ugly", while the highest, 3.0, is assigned to words like "wedding" and "winning". Less extreme words are assigned pleasantness scores closer to the DoA mean of 1.84. Though Whissell's DoA is based on human ratings, Turney ([19]) shows how affective scores can be assigned automatically, using statistical measures of word association in Web texts.

Liu *et al.* ([10]) also present a multidimensional affective model that uses the six basic emotion categories of Ekman ([4]) as its dimensions: *happy, sad, angry, fearful, disgusted and surprised*. These authors base estimates of affect on the contents of Open Mind, a common-sense knowledge-base ([17]) that was harvested from the factual contributions of volunteers on the Web. These contents are treated as sentential objects, and a range of NLP models is used to derive affective labels for the subset of contents (approx. 10%) that appear to convey an emotional stance. These labels are then propagated to related concepts (e.g., excitement is propagated from rollercoasters to amusement parks) so that the implicit affect of many other concepts can be determined.

For reliable results on a large-scale, Mohammad & Turney [13] and and Mohammad & Yang [14] used the *Mechanical Turk* to elicit human ratings of the emotional content of different words. Ratings were sought along the eight primary emotional dimensions identified by Plutchik [16]: *anger, anticipation, disgust, fear, joy, sadness, surprise* and *trust*. Automated tests were used to exclude unsuitable raters, and in all, 24,000+ word-sense pairs were annotated by five different raters. Thus, words that suggest fearful contexts, like “threat”, “hunter” and “acrobat”, are all assigned a significant score on the *fear* dimension, while “disease” and “rat” score highly on the *disgust* dimension.

Strapparava & Valitutti [18] provide a set of affective annotations for a subset of WordNet’s synsets [7] in a resource called *Wordnet-affect*. The annotation labels, called *a-labels*, focus on the cognitive dynamics of emotion, allowing one to distinguish e.g. between words that denote an emotion-eliciting situation and those than denote an emotional response. Esuli & Sebastiani [5] also build directly on WordNet as their lexical platform, using a semi-supervised learning algorithm to assign a trio of numbers – *positivity, negativity* and *neutrality* – to word senses in their newly derived resource, SentiWordNet. (Wordnet-affect also supports these three dimensions as a-labels, and adds a fourth, *ambiguous*). Esuli & Sebastiani [6] improve on their affect scores by running a variant of the PageRank algorithm (see also [12]) on the implicit graph structure that tacitly connects word-senses in WordNet to each other via the words used in their textual glosses.

These lexica attempt to capture the affective profile of a word/sense when it is used in its most normative and stereotypical guise, but they do so without an explicit model of stereotypical meaning. Veale & Hao [20] describe a Web-based approach to acquiring such a model. They note that since the simile pattern “as ADJ as DET NOUN” presupposes that NOUN is an exemplar of ADJness, it follows that ADJ must be a highly salient property of NOUN. The authors of [20] harvested tens of thousands of instances of this pattern from the Web, to extract sets of adjectival properties for thousands of commonplace nouns. They show that if one estimates the pleasantness of a term like “snake” or “artist” as a weighted average of the pleasantness of its properties (like *sneaky* or *creative*) in a resource like the DoA [23], then the estimated scores show a reliable correlation with the DoA’s own scores. It thus makes computational sense to calculate the affect of a word-concept as a function of the affect of its most salient properties. Veale [22] later built on this work to show how a property-rich stereotypical representation could be used for non-literal matching and retrieval of creative texts, such as metaphors and analogies.

Both Liu *et al.* [10] and Veale & Hao [20] argue for the importance of commonsense knowledge in the determination of affect. We incorporate ideas from both while choosing to build mainly on the latter in this paper, to construct a two-level model of the affective lexicon. We focus chiefly on the determination of positive/negative affect, but we will also show how the two-level model can use the *halo effect* ([3]) to support an open-ended range of affective connotations. This will prove especially useful in tasks such as affective text retrieval (e.g. Veale and Hao [21] describe an affective news retrieval system), as it allows users to concoct their own ad-hoc mood filters to suit the needs of a particular query.

Veale & Hao [20] make the simplifying but unjustified assumption that all stereotypical properties are adjectival in nature, and work from adjectival properties (as inventoried by WordNet) to the nouns that exemplify them by successively binding ADJ in the Web query “as ADJ as a NOUN” to different adjectives. The resulting enfilade of queries is sent in rapid succession to the search engine Google. All bindings for NOUN are then automatically extracted from the results before being manually inspected. Here we instead use the *like*-simile patterns“VERB+*ing* like a NOUN” and “VERB+*ed* like a NOUN”, the preferred simile patterns to describe behavior. At the first level, these patterns are used to acquire a model of stereotypical properties and behaviors from the Web. At the second level, a graph organization for these properties and behaviors is also derived, again using the Web as a corpus, and this graph is used to estimate the pleasantness and unpleasantness of each vertex as a function of the pleasantness and unpleasantness of its adjacent vertices. The resulting structure can be used as a richly-featured affective lexicon that supports different kinds of stereotypical reasoning, or it can be used to augment existing ontologies – whether those built on formal foundations such as DOLCE [8], or those built on more lightweight semantic foundations such as WordNet [7] – with an additional layer of commonsense knowledge as to how everyday word-concepts are stereotypically and affectively understood.

3 Finding Stereotypes on the Web

Similes leverage the evocative power of stereotypes to exemplify a descriptive property. Conversely, stereotypes are learned, spread and perpetuated by their constant use in similes. Veale & Hao [20] exploit this symbiotic relationship to acquire a feature-rich representation of many everyday concepts from the Web.

Before performing another large-scale trawl of the Web, we first conduct a pilot study on the Google n-grams [2], a database of contiguous n-word strings ($1 \leq n \leq 5$) with a Web frequency of 40 or higher. For example, the pattern “VERB+*ing* like a NOUN” matches over 8,000 4-grams, while “VERB+*ed* like a NOUN” matches almost 4,000. However, we find here a good deal of empty behaviors, such as *acting* (as in “acting like a baby” rather than “acting like an actor”) and *looking* (as in “looking like a fool”). Indeed, just three empty behaviors – *looking/looked* and *seemed* – account for almost 2,000 n-gram matches. Others, like *walking* and *eating*, are too general and merely allude to a stereotypical behavior (as in “walking like a penguin”) rather than explicitly providing the specific behavior (e.g. *waddling*). Sifting through the n-gram matches yields a few hundred nuggets of stereotypical insight, such as “circling like a shark”, “salivating like a dog” and "clinging like a leech". Our pilot study reveals that most instances of the *like*-simile patterns are not so specific and informative, making a large-scale Web trawl with these patterns impracticable.

Instead we use a hypothesis-driven approach by first looking for attested mentions of a specific behavior with a given noun. Consider the noun *zombie*: searching the Google 3-grams for matches to the patterns "DET VERB+*ing zombie*" and "DET VERB+*ed zombie*" yields the following hypotheses for the stereotypical behavior of zombies (numbers in parentheses are frequencies of matching 3-grams):

{*decomposing(1454), devastating(134), shambling(115), rotting(103), ravaged(98), brainwashed(94), drooling(84), freaking(83), attacking(80), crazed(79), obsessed(73), infected(72), marauding(71), disturbed(65), wandering(64), reanimated(54), flying(52), flaming(52), revived(47), decaying(41), unexpected(40)*}

For each attested behavior in the Google n-grams we generate the corresponding *like*-simile, such as "decomposing like a zombie", and determine its frequency on the open Web. The corresponding non-zero frequencies for these behaviors in *like*-similes on the Web, obtained using Google, are as follows for *zombie*:

{*drooling(4480), wandering(3660), shambling(1240), revived(860), rotting(682), brainwashed(146), reanimated(141), infected(72), flaming(52), decaying(46), decomposing(8), attacking(7), flying(6), freaking(2), obsessed(3)*}

We also harvest all 3-word phrases that match the pattern "DET ADJECTIVE NOUN" in the Google 3-grams, where ADJECTIVE can match any adjective in WordNet. For each property ADJECTIVE that is attested for given noun NOUN in these 3-gram patterns, we generate the Web query "as ADJECTIVE as a NOUN" and dispatch that to Google also. Thus, for example, the 3-gram "a mindless zombie" yields the Web query "as mindless as a zombie", which occurs in hundreds of documents on the Web. The corresponding non-zero Google frequencies for *zombie* properties in *as*-similes on the Web are as follows:

{*slow(18200), scary(6550), hungry(3320), lifeless(2840), creepy(2710), mindless(890), emotionless(827), brainless(155), ravenous(81), strange(8), soulless(6), powerful(6), bizarre(2), bloody(2), brutal(2), unstoppable(2), cheesy(1), supernatural(1)*}

Unlike Veale & Hao [20] then, we do not use a relatively small (approx. 2000) set of queries that are made wide-ranging through the use of wild-cards, but generate a very large set of specific queries (with no wild-cards) that each derive from an attested combination of a specific property or behavior and a specific noun in the Google n-grams. We are careful not to dispatch queries that contain empty behaviors like "looking" or "acting", a list of which is determined during our initial pilot study with the Google n-grams. In all, we dispatch over 500,000 queries to Google, for the same number of attested combinations. No parsing of the Web results is needed, and we need record only the total number of returned hits per query / combination.

3.1 Web-derived Models of Typical Behavior

The 3-gram patterns "DET VERB+*ing* NOUN" and "DET VERB+*ed* NOUN" attest to the plausibility of a given noun-entity exhibiting a specific behavior, but they are only weakly suggestive about what is actually typical. As a basis for generating hypotheses about stereotypical behavior these patterns over-generate significantly, and less than 20% of our queries yield non-zero result sets when sent to the Web.

As shown by the *zombie* example above, some Web-attested behaviors are best judged as idiosyncratic rather than stereotypical. While *rotting, decaying* and *shambling* are just the kind of behaviors we expect of zombies, *freaking, flying* and *flaming* are ill-considered oddities that our behavior model can well do without. As one might expect, such oddities tend to have lower Web frequencies than more widely-accepted behaviors (like *drooling*), yet raw Web frequencies can be an unreliable guide to what is typical [9]. Note for instance how *decomposing* has a low frequency of just 8 uses on the Web (as indexed by Google).

Our Web data exhibits another interesting phenomenon. Consider the noun-entities for which the behavior *brainwashed* is attested, both in the 3-grams ("a brainwashed NOUN") and on the Web ("brainwashed like a NOUN"):

{*cult(1090), zombie(146), robot(9), child(7), fool(4), kid(4), idiot(3), soldier(2)*}

Since cults often use brainwashing, we can consider *cult* to be a stereotypical exemplar for this behavior. Zombies and robots, however, are not typically brainwashed, nor indeed are they even brainwash*able*. Rather, it is more accurate to suggest that the victims of brainwashing often resemble *robots* and *zombies*, and to the extent that brainwashing is made possible by being weak-minded, they can also resemble *fools, idiots, kids* and *children*. This appears to be an example of *ataxis* [1], insofar as *brainwashed* is a "migrant modifier" that more aptly describes the target of the simile than it does the vehicle (*robot* or *zombie*). In this case we can sensibly conclude that *brainwashed* is a figurative behavior of *robots* and *zombies* (since they typically act like a brainwashed person) and is the kind of association we want in our behavioral model. In contrast, it would not be sensible to include *brainwashing* as part of the behavioral description of *fools, idiots, kids, children* or even *soldiers* (though the latter is perhaps debatable).

Ultimately, the stereotypicality of a behavioral association is a pragmatic *gut* issue for the designer of a lexico-semantic resource, one that cannot be automatically resolved by considering Web frequency (or other statistical quantities) alone. As with the design of resources like WordNet, it is best resolved by asking and answering the question "is this an association that I would want in my lexicon?". For this reason, we filter the results of the Web harvesting process manually, to ensure that the final model contains only those qualities that a human would consider typical. In the end then, our approach is a semi-automatic one: automated processes scour the Google n-grams for hypotheses about typical behaviors and properties, and then seek supporting evidence for these hypotheses on the Web (in the form of *as*-similes and *like*-similes). Finally, a manual pass is conducted to ensure the model has the hand-crafted quality of a resource like WordNet.

It takes a matter of weeks to perform this manual filtering, but the stereotype lexicon that results from this effort has 9,479 different stereotypes, and ascribes to each a selection of 7,898 different properties and behaviors. In all, the new resource contains over 75,000 unique noun-to-property/behavior associations, which represents a significant extension to the 12,000+ associations first harvested for Veale & Hao's original resource [20]. The term *baby*, for instance, is associated with the following 163 properties and behaviors in this new, more comprehensive resource:

> {*delicate, squalling, weeping, baptized, adopted, startled, attentive, blessed, teeny, rocked, adorable, whining, bundled, toothless, placid, expected, rescued, treasured, new, sleepy, indulged, slumbering, weaned, pure, supple, helpless, small, sleeping, animated, vulnerable, wailing, cradled, kicking, soft, rested, bellowing, blameless, grinning, screaming, orphaned, cherished, reliant, thriving, loveable, guileless, mute, inexperienced, harmless, dribbling, unthreatening, nursed, angelic, bawling, beaming, naked, spoiled, scared, weak, squirming, blubbering, contented, smiling, wiggling, mewling, blubbing, sniffling, overtired, dimpled, loving, dear, tired, powerless, bewildered, peaceful, distressed, naive, wee, soiled, sucking, fussy, gurgling, vaccinated, heartwarming, pouting, constipated, drooling, quiet, wiggly, lovable, bare, weaning, suckling, cute, bald, whimpering, tender, pampered, incontinent, fleshy, charming, dependent, artless, fussing, flabby, babbling, warm, giddy, crawling, snoozing, hairless, cuddled, sweet, sobbing, squealing, wrapped, tiny, cooing, swaddled, laughing, toddling, fragile, innocent, moaning, gentle, terrified, precious, cranky, giggling, confused, pink, cuddly, fat, ignorant, snoring, young, howling, screeching, shrieking, trusting, shivering, napping, resting, frightened, fresh, loved, demanding, chubby, adored, appealing, happy, tame, relaxed, wriggly, rocking, wriggling, conceived, clean, content, smooth, crying, submissive, bumbling, sniveling*}

A cursory glance at this list reveals a rich description of the stereotypical baby, one that incorporates pleasant and unpleasant behaviors in ample numbers. It makes little sense to reduce such a nuanced description to a single measure of lexical affect, or to parcel the description into separate senses, each with its own subset of behaviors. Instead, the partitioning of the description can be done on demand, and in context, to suit the speaker's meaning: if a term is used pejoratively, we focus on those qualities that are typically unpleasant (*sniveling, submissive, cranky, whimpering*, etc.); if the term is used affectionately, we focus instead on those that typically convey affection (*blessed, delicate, pure*, etc.); and so on. The affective rating of different qualities can be ascertained from any of the existing resources discussed earlier, with more or less success. Whissells DoA is perhaps the most limited, while Mohammad & Turneys eight-dimensional model of emotion [13] seems to possess the most nuance and power.

However, even basic properties and behaviors can be construed differently from one context to another. In some settings, for instance, *cunning* may be a positive description; in most others, it will likely be seen as negative. Many adjectival properties exhibit this duality of affect, such as *proud, tough, tame* and *fragile*, and the description of the stereotypical baby above contains many that could be used to compliment in one context and to insult in another.

For this reason, we concentrate next on the construction of a nuanced model of behavioral interaction, in which the affective profile of a behavior or adjectival property (and thus of the entity that exhibits that property or behavior in context) changes in response to how it is used by the speaker. This model, which forms the second stage of the two-level affective lexicon outlined in the introduction, will allow us to see the positive in properties like *trusting, cunning* and *demanding*, and the negative in properties like *proud, unthreatening* and *innocent*, as the context demands.

3.2 Mutual Reinforcement among Properties

In a representation as feature-rich as that for *baby* above, few features stand apart as truly unique. Some seem to mean much the same thing, while others form clusters of coherent, mutually-reinforcing properties and behaviors. Thus, *fat* reinforces *cuddly*, which reinforces *cute*, which reinforces *adorable*, which reinforces *lovable*, and so on. Intuitively, properties and behaviors that reinforce each other in this way are much more likely to share the same affective signature than those that clearly stand apart.

Yet to construct a support graph of mutually-reinforcing properties and behaviors, we need more than mere co-occurrence in the same stereotypical representation. We also need linguistic evidence to be certain of a link. Conveniently, this evidence can often be found in the Google n-grams, and if not there, then we may be forced to look for evidence on the open Web.

We begin by finding all Google 3-grams of the form "ADJECTIVE and ADJECTIVE" or "BEHAVIOR and BEHAVIOR", such as "cuddly and cute" or "swaggering and strutting". We then consider the number of stereotypes that contain both terms in their representation. If this number is non-zero, a bidirectional link is added between both in the support graph. If the number is zero, we try one more test on the open Web; this test, though time-consuming, is well-motivated, since the n-gram data attests to the possibility of a relationship. We generate the *as*-bracketed query "as ADJECTIVE and ADJECTIVE as" and use Google to determine how many times this pattern occurs in similes on the Web. This pattern works only for adjectival properties, and should be attested by Web evidence only if both adjectives work well together in the description of the same target concept.

Once constructed in this way, every vertex in the resulting graph structure, which we denote N, represents a different property or behavior. The neighboring vertices of a property or behavior p – which we denote $N(p)$ – constitute a set of similar, mutually-reinforcing properties or behaviors that occur in one or more of the same affective contexts as p. For example, the vertex corresponding to the property *cunning* has the following neighbors in N:

> {*insidious, cruel, shrewd, devious, daring, audacious, evil, powerful, artful, clever, strategic, dangerous, charming, calculating, farsighted, strong, wary, subtle, manipulative, wise, conniving, convincing, pragmatic, quick, fast, experienced, diabolical, mighty, greedy, swift, articulate, avaricious, determined, patient, canny, vicious, detailed, curious, deadly, resourceful, resilient, intelligent, cool, treacherous, beautiful, brutal, skilled, bloodthirsty, resolute, wicked, poisonous, dastardly, dishonest, deceitful, sexy, unfeeling, sneaky, mean, sly, smart, agile, bold, aggressive, graceful, deceptive, ingenious, insightful, selfish, unprincipled, inventive, shameless, good, secretive, careful, neurotic, heartless, despicable, brave,*

convoluted, slimy, sophisticated, exploitative, vindictive, disloyal, fluid, machiavellian, towering, brilliant, keen, violent, feared, suspicious, sinister, energetic, scheming, savage, merciless, cowardly, silent, tricky, astute, witty, nasty, free, pretty, lucid, unscrupulous, evocative, precise, seductive, cheating, nimble, versatile, malicious, courageous, virulent, playful, cautious, skillful, untrustworthy, uncaring, amoral, unmerciful, coarse, underhanded, spry, awesome, original, angry, devilish, vile, duplicitous, venomous, obnoxious, bland, fantastic, reclusive, cynical, shifty, stunning, relentless, crazy, funny, wry, loyal, reliable, twisted, effective, prepared, capable, dexterous, adroit, methodical, beguiling}

Guided by our intuition that the affective profile of p should be heavily influenced by the affect of its neighbors, we now consider whether the affect of p can be reliably estimated as a function of $N(p)$.

4 Estimating Lexical Affect

Since every edge in N represents an affective context, we can estimate the likelihood that a property p is ever used in a positive or negative context if we know the positive or negative affect of enough members of $N(p)$. Thus, if we label enough vertices of N with + or - labels, we can interpolate a positive/negative affect score for all vertices p in N.

To do this, we build a reference set $-R$ of typically negative words, and a set $+R$ of typically positive words. Given a few seed members of $-R$ (such as *sad, disgusting, evil*, etc.) and a few seed members of $+R$ (such as *happy, wonderful, pretty*, etc.), we easily find many other candidates to add to $+R$ and $-R$ by considering neighbors of these seeds in N. Veale [22] shows how large ad-hoc word-categories like these can quickly be constructed using flexible pattern-matching over the Google n-grams. After just three iterations in this fashion, we populate $+R$ and $-R$ with approx. 2000 words each.

For a property or behavior p can now define $N^+(p)$ and $N^-(p)$ as follows:

$$N^+(p) = N(p) \bigcap +R \tag{1}$$

For example, $N^+(cunning)$ denotes the following set of properties and behaviors:

{*shrewd, powerful, strong, subtle, wise, quick, mighty, articulate, intelligent, beautiful, daring, experienced, patient, fast, curious, cool, swift, detailed, skilled, resolute, witty, free, artful, careful, agile, brave, cute, canny, graceful, sophisticated, versatile, inventive, spry, fun, precise, bold, resourceful, keen, courageous, playful, determined, stunning, smart, seductive, astute, clever, strategic, towering, charming, ingenious, skillful, insightful, intricate, reliable, good, pretty, farsighted, nimble, pragmatic, lucid, brilliant, loyal, evocative, adroit, resilient, audacious, effective, awesome, capable, sexy, convincing, funny, fantastic, dexterous, methodical, beguiling, original, prepared, fluid, energetic, wry*}

$$N^-(p) = N(p) \bigcap -R \tag{2}$$

Thus, $N^-(cunning)$ denotes the following set of properties and behaviors:

{*insidious, cruel, devious, evil, dangerous, calculating, wary, manipulative, conniving, diabolical, greedy, avaricious, vicious, deadly, treacherous, brutal, bloodthirsty, wicked, poisonous, dastardly, dishonest, deceitful, unfeeling, sneaky, mean, sly, aggressive, deceptive, selfish, unprincipled, shameless, secretive, neurotic, heartless, despicable, convoluted, slimy, exploitative, vindictive, disloyal, machiavellian, violent, feared, suspicious, sinister, scheming, savage, merciless, cowardly, silent, tricky, nasty, unscrupulous, cheating, malicious, virulent, cautious, untrustworthy, uncaring, amoral, unmerciful, coarse, underhanded, angry, devilish, vile, duplicitous, venomous, obnoxious, bland, reclusive, cynical, shifty, relentless, crazy, twisted*}

That is, $N^{+}(p)$ is the set of neighbors of p that are known to be positive, and $N^{-}(p)$ is the set of neighbors of p that are known to be negative. We can now assign positive and negative scores to each vertex p in N by interpolating from the reference values in $+R$ and $-R$ to their neighbors in N:

$$pos(p) = \frac{|N^{+}(p)|}{|N^{+}(p) \bigcup N^{-}(p)|} \tag{3}$$

$$neg(p) = \frac{|N^{-}(p)|}{|N^{+}(p) \bigcup N^{-}(p)|} \tag{4}$$

For instance, the set $N^{-}(aggressive)$ contains 230 elements, while $N^{+}(aggressive)$ contains 201 elements. Thus, $pos(aggressive)$ is calculated to be 0.466 while $neg(aggressive)$ is calculated to be 0.534. In other words, *aggressive* is deemed to be more positive than negative, or to be more precise, (3) and (4) estimate that *aggressive* is more likely to occur in a negative descriptive context than in a positive descriptive context. In contrast, *cunning* is deemed to be slightly more positive than negative, given the number of positive and negative descriptive contexts that are captured in $N^{-}(cunning)$ and $N^{+}(cunning)$ as shown earlier. The properties *aggressive* and *cunning* are borderline cases, since each evokes a large number of descriptive contexts in which each could be viewed either positively or negatively. A property like *cynical*, however, is much more clear-cut: with 258 neighbors in $N^{-}(cynical)$ and just 38 in $N^{+}(cynical)$, $neg(cynical)$ is 0.87 while $pos(cynical)$ is just 0.13.

If a term S denotes a stereotypical idea and is described via a set of typical properties and behaviors $typical(S)$ in the lexicon, then:

$$pos(S) = \frac{\sum_{p \in typical(S)} pos(p)}{|typical(S)|} \tag{5}$$

$$neg(S) = \frac{\sum_{p \in typical(S)} neg(p)}{|typical(S)|} \tag{6}$$

Thus, (5) and (6) calculate the mean affect of the properties and behaviors of S, as represented via $typical(S)$. We can now use (3) and (4) to separate $typical(S)$ into those qualities that are more negative than positive (putting a negative spin on S) and into those that are more positive than negative (putting a positive spin on S):

$$posTypical(S) = \{p | p \in typical(S) \land pos(p) > neg(p)\} \tag{7}$$

$$negTypical(S) = \{p | p \in typical(S) \wedge neg(p) > pos(p)\} \tag{8}$$

Formulae (7) and (8) can be used to "spin" a concept positively or negatively in given context, to highlight only those qualities of S that support the chosen positive or negative viewpoint on S. For instance, the stereotype $terrorist$ has the following salient positive properties (numbers in parentheses indicate the value assigned by $pos(p)$ for each property p):

$posTypical(Terrorist)$ = {*committed(.826), daring(.733), networked(.8), sponsored(.833)*}

As we should expect, there are many more negative properties that are salient for the stereotype $terrorist$ (numbers in parentheses indicate the value assigned by $neg(p)$ for each property p):

$negTypical(Terrorist)$ =
{*hateful(.978), bad(.951), despicable(.98), harmful(.95), inhuman(.972), irrational(.916), odious(.97), horrid(.97), irresponsible(.916), depraved(.945), murdering(1.0), heinous(.951), hostile(.92), guilty(.954), misguided(.92), damaging(.89), bloodthirsty(.93), suspicious(.94), bigoted(.952), hated(.962), sickening(.969), callous(.928), raging(.917), appalling(.906), vicious(.86), deranged(.93), barbarous(.93), mindless(.866), unscrupulous(.89), threatening(.826), indiscriminate(.96), demonic(.98), wicked(.83), convicted(.96), destructive(.817), condemned(.97), pitiless(.9), crazed(.845), twisted(.815), alarming(.844), insidious(.84), merciless(.82), accused(.978), sinister(.79), dreadful(.89), diabolical(.85), devastating(.756), remorseless(.875), brainwashed(.893), shocking(.74), ruthless(.733), infamous(.828), menacing(.743), unforgiving(.768), evil(.904), hunted(.93), dreaded(.83), hardened(.764), disgruntled(.954), suspected(.925), fearsome(.712), armed(.685), imprisoned(.97), chilling(.638), prohibited(1.0), hating(1.0), criminal(.968), lethal(.628), wanted(.72), clandestine(.758), incendiary(.688), inflamed(.826), arrested(.961), captured(.888), masked(.78), feared(.627), shooting(.639), killing(.95), branded(.77), hooded(.916), banned(1.0), fanatic(.7), jailed(1.0), concealed(.68), targeted(.7), bombing(.961), fighting(.66), radical(.53), proscribed(.857)*}

Note how properties such as $armed$, $shooting$, $fighting$ and even $feared$ have lower negativity than unremittingly negative qualities like $murdering$ and $prohibited$. These lower scores reflect the greater possibilities for using these properties to impart a positive view of a topic, as when e.g. one desires to be $feared$ and $respected$.

We estimate a positive and negative affect score for each stereotype (using (5) and (6)) and for each of their properties and behaviors (using (3) and (4)), which produces an affect lexicon of over 16,000 words. For instance, $pos(Terrorist) = 0.178$ and $neg(Terrorist) = 0.822$. Overall, the mean positivity score is 0.517 (standard deviation = 0.313), while the mean negativity score is 0.483. In contrast, the mean positivity of the 1,977 words in $+R$ is 0.852 (standard dev. = 0.127), while the mean negativity of the 2,192 words in $-R$ is 0.813 (standard dev. = 0.154).

5 In the Mood for Affective Search

Thus far we have focused on a rather reductive view of affect as the potential of words to convey a positive or negative meaning. As shown in ([16],[13],[14]) other emotional dimensions can meaningfully be used to describe our affective perception of a word. Those authors show e.g., that some words convey sadness and fear to different degrees, while others suggest a degree of joy and even trust. While we do not explicitly distinguish different dimensions of mood or emotionality in the two-level model, the model does capture, through its network N of mood-bearing words, the emotional influence that the perception of one kind of property or behavior can have on the perception of other properties and behaviors. The effect is often called the *halo effect* in the psychological literature, wherein the perception of one positive quality, such as physical beauty or strength, can influence our perception of other qualities such as intelligence, honesty and leadership ([3]). Conversely, the network structure of N also supports reasoning under the so-called *devil effect*, wherein the perception of one negative quality (such as *angry*) can lead us to view the possession of related qualities (such as *aggressive*) more negatively ([15]). As such, the two-level model implicitly supports a whole lexicon of diverse but inter-connected mood types: every node in N evokes a halo of associated positive nodes, and a penumbra of associated negative nodes as well.

Consider that a word like *aggressive* implies a range of positive qualities that are captured in $N^{+}(aggressive)$ and a range of negative qualities that are captured in $N^{-}(aggressive)$. The halo of words in $N^{+}(aggressive)$ helps to convey the up-side of aggressive behavior (e.g. to be *aggressive* often implies that one is also *quick*, *energetic*, *vigorous* and *determined*) while the penumbra of words in $N^{-}(aggressive)$ evokes the down-side of aggressiveness (e.g. *aggressive* people are often *violent*, *angry*, *hostile* and *abusive*). We can, in effect, allude to a whole family of affective words with a single term like *aggressive*, or we can focus exclusively on wholly positive or negative word halos with polarizing labels such as $+aggressive$ and $-aggressive$. There is little need then to build aggressiveness or other moods into the lexicon as explicit dimensions of affect if we can use these labels to evoke the same word sets. Any one of 1000's of different words in the lexicon – or a combination thereof – can be used as mood filters to refer to ad-hoc families of affective words as the need arises.

Stereotypes themselves can also be used as powerful and expressive mood filters in the affective retrieval and ranking of documents. For instance, we can use the mood filter $+leader$ to rank documents by their relative density of words that convey positive leadership qualities, or $-terrorist$ to rank documents by their use of words that convey the many negative qualities of terrorists (which are enumerated above).

Pursuing this theme, we are now using the two-level lexicon to support affective text search over news content on the Web. In this application, users may use *+aggressive*, or *-sad*, or any property, behavior or stereotype (e.g., *+genius*, *-terrorist*) they consider apt, as mood filters to organize the retrieved document set. Currently, news articles are crawled from a dozen news sites (this number will grow) and their textual content is indexed using the *Lucene* system [11]. To allow for efficient document-level affect determination, any words that occur in the affect lexicon are stored in a separate document field. Queries to the system are separated into two kinds of query terms: regular query terms, which are unadorned words or phrases; and mood filters, which are terms prefixed either with $+$ or $-$ to indicate their affective polarity (such as *-proud* or *+exciting*). Regular query terms are used to retrieve matching documents from the indexing engine, before the mood filters are used to rank these documents by mood. The retrieved documents may be ranked by their relevance to the regular query term (e.g. as calculated by *Lucene*) or by their mood density (the proportion of words in a document that match the mood filter of the query), or by a weighted combination of both measures.

6 Empirical Evaluation

We shall now take a closer look at the affect scores for properties and behaviors in §6.1, before considering the scores estimated for stereotypes in §6.2. We then evaluate the performance of the affect lexicon on an affective separation task – in which the properties and behaviors of stereotypes in the reference set are separated into distinct positive and negative subsets – in §6.3.

6.1 Bottom Level: Properties and Behaviors of Stereotypes

If the intuition behind formulae (1) – (4) is valid, then we should expect that for every property or behavior p in $+R$, $pos(p) > neg(p)$, and conversely, $neg(p) > pos(p)$ for every p in $-R$. Recall that with (3) and (4) we model the problem of estimating affect as an interpolation task rather than as a learning task. Therefore, the *pos* and *neg* affect scores that are calculated for p are independent of whether or not p is in $+R$ or $-R$. So if we add p to a reference set, or remove p from a reference set, the same values for $pos(p)$ or $neg(p)$ will be estimated. Only the neighboring vertices of p can possibly be influenced by such a move, and the affect scores that are calculated for those neighbors cannot feed back into the scores calculated for p. It is thus reasonable to evaluate the intuition behind (1) – (4) using $+R$ and $-R$ as a gold standard.

When affect scores are calculated for the complete set of properties, behaviors and stereotypes in the lexicon, just five properties in $+R$ are given a positivity score of less than 0.5, leading those words to be wrongly classified as more negative than positive. The misclassified properties/behaviors are: *evanescent, giggling, licking, devotional*, and *fraternal*. These five words account for approx. 0.4% of the 1,314 adjectival properties in $+R$.

At the same time, twenty-six properties in $-R$ are given a negativity score of less than 0.5, leading those words to be wrongly classified as more positive than negative. The misclassified properties/behaviors are: *cocky, dense, demanding, urgent, acute, unavoidable, critical, startling, gaudy, decadent, biting, controversial, peculiar, disinterested, strict, visceral, feared, opinionated, humbling, subdued, impetuous, shooting, acerbic, heartrending, ineluctable*, and *groveling*. These twenty-six words account for approx. 1.9% of the 1,385 adjectival properties in $-R$.

Though these results are not very surprising– after all, the elements of $+R$ and $-R$ were chosen to have an obviously positive or negative affect – they do validate the intuition in (1) – (4) that the affect of a property or behavior can be consistently estimated as a function of the other properties and behaviors with which it is used to form a coherent description.

6.2 Top Level: Stereotypical Concepts

The reference sets $+R$ and $-R$ also contain a significant number of nouns. The positive reference set $+R$ contains 478 nouns for which the stereotype lexicon provides a feature-level description, while $-R$ contains 677 nouns that are associated with specific stereotypical properties and behaviors. We can thus use these reference cases to evaluate the mean affect scores estimated for stereotypes in (7) and (8). If it is indeed sensible to average the positive and negative scores of the elements in $typical(S)$ to estimate a positive and negative score for a stereotype S, then we should observe $pos(S) > neg(S)$ for almost all stereotypes S in $+R$, and $neg(S) > pos(S)$ for almost all stereotypes S in $-R$.

When affect scores are calculated for the complete set of properties, behaviors and stereotypes in the lexicon, just sixteen positive stereotypes in $+R$ are assigned a positivity of less than 0.5, leading these stereotypes to be classified as more negative than positive. The misclassified stereotypes are: *patient, innocent, stable, rustic, giant, desire, expectation, heart, responsibility, sentiment, infant, toddler, fruitcake, giggle, sitcom*, and *granny*. These sixteen stereotypes account for approx. 3.3% of the 478 stereotypes in $+R$.

At the same time, just twenty-six negative stereotypes in $-R$ are assigned a negativity of less than 0.5, leading these to be classified as more positive than negative. The misclassified stereotypes are: *penitent, fire, regret, trial, opposition, accomplice, revenge, rebellion, enmity, debt, illusion, protest, drill, hide, wetland, dogma, disregard, revolt, jihad, handgun, grenade, sorceress, grudge, inquisition, duel* and *colonoscopy*. These twenty-six stereotypes account for approx. 3.8% of the 677 stereotypes in $-R$.

These results validate the guiding intuition in (5) and (6), namely, that the overall affect of a stereotype S (in a null context) can be reliably estimated as a function of the affect of its most typical properties and behaviors as represented by $typical(S)$.

6.3 Separating Words by Affect: Two Views

The reference sets $+R$ or $-R$ contain many of the properties and behaviors that are ascribed to each stereotype S in the stereotype lexicon, via $typical(S)$. We can thus use $+R$ and $-R$ as a gold standard for evaluating the separation of the properties and behaviors of a stereotype S into distinctly positive and negative subsets of $typical(S)$, denoted $posTypical(S)$ and $negTypical(S)$ in formulae (7) and (8).

The stereotype lexicon contains 6,230 stereotypes with at least one property or behavior in $-R \bigcup +R$, and on average, $-R \bigcup +R$ contains 6.51 of the properties and behaviors of each of these stereotypes (on average, 2.95 are in $+R$, 3.56 are in $-R$).

In a perfect separation we should obtain a positive subset that contains only those properties and behaviors in $typical(S) \bigcap +R$ and a negative subset that contains only those in $typical(S) \bigcap -R$. Viewing the problem as a retrieval task, whose goal is the accurate retrieval of distinct positive and negative subsets of $typical(S)$ for a given stereotype S using (7) and (8), we report the P/R/F1 results of Table 1. Note that the reported results are calculated as the macro-average of P/R/F1 scores for the separation process applied to the properties and behaviors of all 6,230 stereotypes in the experiment.

Table 1 Average P/R/F1 scores for the retrieval of pos. and neg. features from 6,230 stereotypes

Macro Average	Positive properties	Negative properties
Precision	.962	.98
Recall	.975	.958
F-score	.968	.968

In a complementary formulation of this problem, we must separate the list of stereotypes that exhibit a given property or behavior p into two distinct sets, the set of positive stereotypes that exhibit p and the set of negative stereotypes that exhibit p. The stereotype lexicon contains 4,536 properties and behaviors for which one or more of its associated stereotypes is in $-R \bigcap +R$. On average, each of these properties or behaviors is associated with 5.29 stereotypes in $-R \bigcap +R$ (on average, 2.06 are in $+R$ while 3.23 are in $-R$). Again viewing the problem as a retrieval task, of stereotypes rather than properties and behaviors, we report the results of Table 2. Note that the reported results are calculated as the macro-average of P/R/F1 scores for the separation process applied to the stereotypes associated with all 4,536 properties and behaviors in the experiment.

Table 2 Average P/R/F1 scores for the retrieval of pos. and neg. stereotypes for 4,536 features

Macro Average	Positive stereotypes	Negative stereotypes
Precision	.986	.965
Recall	.949	.982
F-score	.967	.973

As can be seen in Tables 1 and 2, the current model achieves very encouraging results, both on the property/behavior separation task and on the stereotype separation task. These tasks serve more than a purely evaluative function. The former (property/behavior separation) is performed whenever we wish to place a particular affective spin on a topic, such as when we use the words “baby” and “youngster” affectionately, or use the words “elite” and “professor” disapprovingly. The latter (stereotype separation) is often performed as part of the interpretation of a feature ascription : if someone describes you as “strange”, what might they be implicitly calling you if “strange” is meant negatively (a *weirdo*, or a *freak*, perhaps?), and what might “strange” describe if generously given the most positive interpretation (an *eccentric* or a *rarity*, perhaps)?

7 Conclusions

The chief innovation in this work is the imposition of a two-level structure onto the affect lexicon: the first level represents stereotypes as bundles of their most salient properties and behaviors; the second represents the relationship of these properties and behaviors to each other. The result is a lexicon that incorporates a great deal of common-sense knowledge of the world, for it is only through common-sense that a language user can fully appreciate the variability of a word’s affect from one context to another (see [10] for an expression of the same view).

The approach is a modular one, and one could in principle use any of the existing affect lexica to assign affect scores to the properties and behaviors of the second level. Nonetheless, we have shown that good results are achievable with the simple formulae in (3) and (4), and that these results are a good basis for estimating the affect of higher-level stereotypes in (5) and (6). In this first phase of the work, we have concentrated on building a stereotype lexicon in which each of the typical properties and behaviors ascribed to a word-concept are both justifiable and salient. That is, we have aimed for precision rather than recall in this foundation-building phase of development. However, coverage is also an important dimension of a stereotype lexicon. While we have shown that the two-level lexicon contains enough property ascriptions to ensure that the average overall affect of a stereotype S can be reliably estimated from the set $typical(S)$, we also need to quantify the likelihood that $typical(S)$ will contain any property or behavior that is commonly and consensually held to be salient of S. This remains a challenging goal of the work, especially given the lack of a gold standard against which the coverage of a stereotype lexicon can be quantified. For now, the *halo effect* (or conversely, the *devil effect*) can be used to infer the salience of an arbitrary property p to a stereotype S given the contents of $typical(S)$, since the relevance of p will be a function of $typical(S) \bigcap N(p)$. Interested readers who wish to exploit this early form of the lexicon can do so by contacting the author directly. As demonstrable improvements are made in the coverage of the lexicon, versions will be made publicly available for research purposes.

We have also demonstrated how a two-level affect lexicon might be used to understand how one topic can be viewed through the affective lens of another, as e.g. when we view *science as a religion*, *art as a science*, or *a leader* as a *pioneer* or a *tyrant*. In addition, the lexicon allows the users of an affective retrieval system to personalize their affective relationship to the words in a text or a query. For instance, a user can use *+cunning* or *-powerful* to specify that *cunning* should be viewed as a positive quality and *powerful* as a negative quality in the current retrieval context. The two-level lexicon represents a lightweight form of commonsense knowledge, albeit commonsense that is not explicitly axiomatized. Nonetheless, the affective and stereotypical dimensions of the two-level lexicon can always be used to enrich a more formal, axiomatized ontology like DOLCE (which, for instance, contains a rich hierarchy of social roles) with the kind of non-axiomatic pragmatic knowledge that one needs to reason effectively in social contexts.

Beyond the obvious applications of a fine-grained affective lexicon for classifying the polarity of texts – such as the texts retrieved using a Web search engine – the stereotypical perspective can also be used to find documents that exhibit a particular conceptual slant on a given topic (e.g. to retrieve documents that positively view *Apple as a cult*). In this sense our search engines really would support a form of a creative information retrieval (as defined in [22] and exploited in [21]), and allow us to see the best and worst in everything on the Web.

Acknowledgements This research was supported by the WCU (World Class University) program under the National Research Foundation of Korea, and funded by the Ministry of Education, Science and Technology of Korea (Project No: R31-30007).

References

1. Bolinger, D.: Ataxis. In Rokko Linguistic Society (ed.), Gendai no Gengo Kenkyu (Linguistics Today), Tokyo:1–17. (1988)
2. Brants, T. and Franz, A.: Web 1T 5-gram Version 1. Linguistic Data Consortium. (2006)
3. Dion, K, Berscheid, E. and Walster, E.: What is Beautiful is Good. Journal of Personality and Social Psychology, 3(24): 285-290. (1972)
4. Ekman, P.: Facial expression of emotion. American Psychologist, 48:384–392. (1993)
5. Esuli, A. and Sebastiani, F.: SentiWordNet: A publicly available lexical resource for opinion mining. Proceedings of LREC-2006, the 5th Conference on Language Resources and Evaluation, 417–422. (2006)
6. Esuli, A. and Sebastiani, F.: PageRanking WordNet Synsets: An application to opinion mining. Proceedings of ACL-2007, the 45th Annual Meeting of the Association for Computational Linguistics. (2007)
7. Fellbaum, C.: (ed.). WordNet: An electronic lexical database. Cambridge, MA: MIT Press. (1998)
8. Gangemi, A., Guarino, N., Masolo, C., Oltamari, A. and Schneider, L.: Sweetening Ontologies with DOLCE. Proceeddings of EKAW 2002, the 13th International Conference on Knowledge Engineering and Knowledge Management. Springer-Verlag, London, UK. (2002)
9. Kilgarriff, A.: Googleology is Bad Science. Computational Linguistics, 33(1):147–151, (2007)
10. Liu, H., Lieberman, H. and Selker, T.: A Model of Textual Affect Sensing Using Real-World Knowledge. Proceedings of the 8th international conference on Intelligent user interfaces, pp. 125–132. (2003)
11. McCandless, M., Hatcher, E. and Gospodnetić, O.: Lucene In Action (2nd Edition). Manning Publications. (2010)

12. Mihalcea, R. and Tarau, P.: TextRank: Bringing Order to Texts. Proceedings of EMNLP-04, the 2004 Conference on Empirical Methods in Natural Language Processing. (2004)
13. Mohammad, S. F. and Turney, P.D.: Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotional lexicon. Proceedings of the NAACL-HLT 2010 workshop on Computational Approaches to Analysis and Generation of Emotion in Text. Los Angeles, CA. (2010)
14. Mohammad, S. F. and Yang, T. W.: Tracking sentiment in mail: how genders differ on emotional axes. Proceedings of the ACL 2011 WASSA workshop on Computational Approaches to Subjectivity and Sentiment Analysis, Portland, Oregon. (2011)
15. Nisbett, R. E. and Wilson, T. D.: The halo effect: Evidence for unconscious alteration of judgments". Journal of Personality and Social Psychology (American Psychological Association) 35(4): 250256. (1977)
16. Plutchik, R.: A general psycho-evolutionary theory of emotion. Emotion: Theory, research and experience, 2(1-2):1–135. (1980)
17. Singh, P.: The public acquisition of commonsense knowledge. Proceedings of AAAI Spring Symposium on Acquiring (and Using) Linguistic (and World) Knowledge for Information Access. Palo Alto, CA. (2002)
18. Strapparava, C. and Valitutti, A.: Wordnet-affect: an affective extension of Wordnet. Proceedings of LREC-2004, the 4th International Conference on Language Resources and Evaluation, Lisbon, Portugal. (2004)
19. Turney, P. D.: Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In Proceedings of ACL-2002, the 40th Annual Meeting of the Association for Computational Linguistics, pp. 417–424. (2002)
20. Veale, T. and Hao, Y.: Making Lexical Ontologies Functional and Context-Sensitive. Proceedings of ACL-2007, the 45th Annual Meeting of the Association of Computational Linguistics, pp. 57–64. (2007)
21. Veale, T and Hao, Y.: In the Mood for Affective Search with Web Stereotypes. Proceedings of WWW-2012, the World Wide Web conference (demonstration track), Lyon. (2012)
22. Veale, T.: Creative Language Retrieval: A Robust Hybrid of Information Retrieval and Linguistic Creativity. Proceedings of ACL-2011, the 49th Annual Meeting of the Association of Computational Linguistics. (2011)
23. Whissell, C.: The dictionary of affect in language. In R. Plutchik and H. Kellerman (eds.) Emotion: Theory and research. Harcourt Brace, 113–131. (1989)