

Bringing the Text to Life Automatically

Carlo Strapparava and Alessandro Valitutti

ITC-irst, Povo, I-38050, Trento, Italy
{strappa, alvalitu}@itc.it

Abstract

Animated text is an appealing field of creative graphical design. Manually designed text animation is largely employed in advertisement, movie titles and web pages. In this paper we propose to link, through state of the art NLP techniques, the affective content detection of a piece of text to the animation of the words in the text itself. This methodology allows us to automatically generate affective text animation and opens some new perspectives for advertisements, internet applications and intelligent interfaces.

Introduction

Rosencrantz: What are you playing at?

Guildestern: Words. Words. They're all we have to go on.

– from “*Rosencrantz and Guildestern Are Dead*” (1990)

An actor reads a script. He reads those words with the intention of transforming cold print into living speech. Vocal inflections, tone of voice, gestures and facial expressions are all part of actor's contribution to the play. With the body subtle vibrations and frequencies, he expresses the hidden emotional meaning. We can say that, through his *interpretation*, he brings the script to life.

In this paper exploiting state of the art NLP techniques, we will show that through *automatic* detection of the affecting meaning of texts (e.g. news titles) we can consequently animate the words that compose them. In automated text animation the text itself is capable to augment its expressivity and to *move* in an autonomous way.

Animated text is appealing and widely employed. It is a field of creative graphical design in which there is increasing interest. Manually designed text animation is employed for a long time in movie titles, television advertisements, and web pages. Nevertheless there are applicative contexts in which *automated* text animation would be very useful (e.g. smart presentation of newspaper headlines or advertising slogans). As far as we know, there are no tools for the automated animation of texts.

In this paper, we approach the automated creation of text animation linking it to the lexical semantic content (in particular, to affective meaning). In particular, once detected

the affective load of a sentence, we check for words that are most semantically similar to emotional concepts and then we emphasize the affective meaning through an appropriate animation (i.e. different emotions have different animations).

We suppose that, if text animation is semantically consistent with text content, the communication of the affective meaning is more effective. In particular, we want to pay attention to the memorizability of text and how it increases with a consistent animation. We believe that, beyond the pleasantness, affective animations can increase the memorizability of text and, in particular, the semantic consistency between words and animations has a significant role in the memorization of headlines.

The paper is structured as follows. The next section presents resources and functionalities for the recognition of affective terms. An affective hierarchy as an extension of the WORDNET-AFFECT lexical database is developed in the first place. The next phase is the development of a semantic similarity function, acquired automatically in an unsupervised way from a large corpus of texts, which allows us to put into relation words and emotional categories. The Section “Text Animation” introduces text animation (i.e. kinetic typography) and the development of a flexible scripting language to describe and dynamically generate text animation.

Affective Semantic Similarity

All words can potentially convey affective meaning. Each of them, even those more apparently neutral, can evoke pleasant or painful experiences, because of its semantic relation with emotional concepts. While some words have emotional meaning with respect to the story of a particular individual, for many others the affective power is part of the collective imagination (e.g. words “mum”, “ghost”, “war” etc.).

We are interested in the words of the second type, because their affective meaning is part of common sense knowledge and therefore it can be detected from the way they are used. For this reason, we studied the use of words in textual productions, and in particular their co-occurrences with the words in which the affective meaning is explicit. As claimed by Ortony et al. (Ortony, Clore, & Foss 1987), we have to distinguish between words directly referring to emotional states (e.g. “fear”, “cheerful”) and those having only an indirect reference that depends on the context (e.g. words that indicate possible emotional causes as “monster”

A-Labels	Valence	Examples of word senses
JOY	positive	noun joy#1, adjective elated#2, verb gladden#2, adverb gleefully#1
LOVE	positive	noun love#1, adjective loving#1, verb love#1, adverb fondly#1
APPREHENSION	negative	noun apprehension#1, adjective apprehensive#3, adverb anxiously#1
SADNESS	negative	noun sadness#1, adjective unhappy#1, verb sadden#1, adverb deplorably#1
SURPRISE	ambiguous	noun surprise#1, adjective surprised#1, verb surprise#1
APATHY	neutral	noun apathy#1, adjective apathetic#1, adverb apathetically#1
NEGATIVE-FEAR	negative	noun scare#2, adjective afraid#1, verb frighten#1, adverb horrifyingly#1
POSITIVE-FEAR	positive	noun frisson#1
POSITIVE-EXPECTATION	positive	noun anticipation#1, adjective cliff-hanging#1, verb anticipate#1

Table 1: Some of emotional categories in WORDNET-AFFECT and some corresponding word senses

or emotional responses as “cry”). We call the former *direct affective words* and the latter *indirect affective words* (Strapparava, Valitutti, & Stock 2006).

In order to manage affective lexical meaning, we (i) organized the direct affective words and synsets inside WORDNET-AFFECT, an affective lexical resource based on an extension of WORDNET, and (ii) implemented a selection function (named *affective weight*) based on a semantic similarity mechanism automatically acquired in an unsupervised way from a large corpus of texts (100 millions of words), in order to individuate the indirect affective lexicon.

Applied to a concept (e.g. a WORDNET synset) and an emotional category, this function returns a value representing the semantic affinity with that emotion. In this way it is possible to assign a value to the concept with respect to each emotional category, and eventually select the emotion with the highest value. Applied to a set of concepts that are semantically similar, this function selects subsets characterized by some given affective constraint (e.g. referring to a particular emotional category or valence).

As we will see, we are able to focus selectively on positive, negative, ambiguous or neutral types of emotions. For example, given “difficulty” as input term, the system suggests as related emotions: IDENTIFICATION, NEGATIVE-CONCERN, AMBIGUOUS-EXPECTATION, APATHY. Moreover, given an input word (e.g. “university”) and the indication of an emotional valence (e.g. positive), the system suggests a set of related words through some positive emotional category (e.g. “professor” “scholarship” “achievement”) found through the emotions ENTHUSIASM, SYMPATHY, DEVOTION, ENCOURAGEMENT.

These fine-grained kinds of affective lexicon selection can open up new possibilities in many applications that exploit verbal communication of emotions. For example, (Valitutti, Strapparava, & Stock 2005) exploited semantic connection between a generic word and an emotion for the generation of affective evaluative predicates and sentences.

WORDNET-AFFECT and the Emotional Categories

WORDNET-AFFECT is an extension of WordNet database (Fellbaum 1998), including a subset of synsets suitable to represent affective concepts. Similarly to our method for domain labels (Magnini & Cavaglià 2000), we assigned to a number of WordNet synsets one or more affective labels (*a-labels*). In particular, the af-

factive concepts representing emotional state are individuated by synsets marked with the a-label EMOTION. There are also other a-labels for those concepts representing moods, situations eliciting emotions, or emotional responses. WORDNET-AFFECT is freely available for research purpose at <http://wdomains.itc.it>. See (Strapparava & Valitutti 2004) for a complete description of the resource.

	# Synsets	# Words	# Senses
Nouns	280	539	564
Adjectives	342	601	951
Verbs	142	294	430
Adverbs	154	203	270
Total	918	1637	2215

Table 2: Number of elements in the emotional hierarchy.

Recently, we extended WORDNET-AFFECT with a set of additional a-labels (i.e. the *emotional categories*), hierarchically organized, in order to specialize synsets with a-label EMOTION. In a second stage, we introduced some modifications, in order to distinguish synsets according to emotional valence. We defined four additional a-labels: POSITIVE, NEGATIVE, AMBIGUOUS, NEUTRAL. The first one corresponds to “positive emotions”, defined as emotional states characterized by the presence of positive edonic signals (or pleasure). It includes synsets such as joy#1 or enthusiasm#1. Similarly the NEGATIVE a-label identifies “negative emotions” characterized by negative edonic signals (or pain), for example anger#1 or sadness#1. Synsets representing affective states whose valence depends on semantic context (e.g. surprise#1) were marked with the tag AMBIGUOUS. Finally, synsets referring to mental states that are generally considered affective but are not characterized by valence, were marked with the tag NEUTRAL.

Positive	Negative	Ambiguous	Neutral	Total
97	156	20	7	280

Table 3: Valence distribution of emotional categories.

Another important property for affective lexicon concerning mainly adjectival interpretation is the stative/causative

dimension (Goy 2000). An emotional adjective is said *causative* if it refers to some emotion that is caused by the entity represented by the modified noun (e.g. “amusing movie”). In a similar way, an emotional adjective is said *stative* if it refers to the emotion owned or felt by the subject denoted by the modified noun (e.g. “cheerful/happy boy”).

Computing Lexical Affective Semantic Similarity

There is an active research direction in NLP field about sentiment analysis and recognition of semantic orientation from texts (e.g. (Turney & Littman 2003; Mihalcea & Liu 2006)). A crucial issue is to have a mechanism for evaluating the similarity among generic terms and affective lexical concepts. To this aim we estimated term similarity from a large scale corpus. In particular we implemented a variation of Latent Semantic Analysis (LSA) in order to obtain a vector representation for words, texts and synsets. In the present work we will use this technique to build a vector representation in the LSA space for each emotional category in WORDNET-AFFECT. These emotional vectors are then used to calculate a semantic similarity between *general* terms and emotions.

In LSA (Deerwester *et al.* 1990), term co-occurrences in the documents of the corpus are captured by means of a dimensionality reduction operated by a Singular Value Decomposition (SVD) on the term-by-document matrix. For the experiments reported in this paper, we run the SVD operation on the British National Corpus¹.

The resulting LSA vectors can be exploited to estimate both term and document similarity. Regarding document similarity, Latent Semantic Indexing (LSI) is a technique that allows us to represent a document by means of a LSA vector. In particular, we used a variation of the *pseudo-document* methodology described in (Berry 1992). This variation takes into account also a *tf-idf* weighting schema (see (Gliozzo & Strapparava 2005) for more details). Each document can be represented in the LSA space by summing up the normalized LSA vectors of all the terms contained in it. Also a synset in WORDNET (and then an emotional category) can be represented in the LSA space, performing the pseudo-document technique on all the words contained in the synset. Thus it is possible to have a vectorial representation of each emotional category in the LSA space (i.e. the *emotional vectors*). With an appropriate metric (e.g. cosine), we can compute a similarity measure among terms and affective categories. We defined the *affective weight* as the similarity value between an emotional vector and an input term vector.

For example, the term “sex” shows high similarity with respect to the positive emotional category AMOROUSNESS, with the negative category MISOGYNY, and with the ambiguous valence tagged category AMBIGUOUS_EXPECTATION. The noun “gift” is highly related to the emotional categories: LOVE (with positive valence), COMPASSION (with negative valence), SURPRISE (with ambig-

ous valence), and INDIFFERENCE (with neutral valence).

In summary, the vectorial representation in the Latent Semantic Space allows us to represent in a *uniform* way emotional categories, terms, concepts and possibly sentences and full documents. See (Strapparava, Valitutti, & Stock 2006) for more details.

For example, Table 4 displays some news titles (taken from the CNN and Google News sites), the respective more similar affective category, the affective weight, and the word in the title most similar to that category. In the next section we will see that this functionality is the basis for indicating which words to animate and in which way.

Text Animation

Kinetic Typography

Kinetic typography is the technology of text animation, i.e. text that uses movement or other changes over time. It is part of graphical design. There are design schools currently investing effort for extending the range of applicative possibilities of text animation. Actually, kinetic text allows graphic designers to communicate more than just the basic meaning of the words. For example, one of the most interesting applications is instant messaging for animated chat: (Forlizzi, Lee, & Hudson 2003) describe a system where users are allowed to create personalized animations.

The advantage of kinetic typography consists in adding a further communicative dimension to simple text, combining verbal and visual communication, and providing opportunities to enrich the expressiveness of static texts. According to (Lee, Forlizzi, & Hudson 2002), kinetic typography can be used for three different communicative goals: capturing and directing attention of recipients, creating compelling characters (e.g. in a dialog), and expressing emotions. A possible way of animating a text is mimicking the typical movement of humans when they express the content of the text (e.g. “Hi” with a jumping motion mimics exaggerated body motion of humans when they are really glad. See Figure 1).

In our work, we explore the idea of establishing a link between lexical semantics of texts (automatically discerned through NLP techniques) and some kinetic properties exploited for animating the words. In particular in this paper, exploiting the affective semantic similarity introduced above, we consider affective connotation of texts. This holds particularly for “indirect affective words” (Strapparava, Valitutti, & Stock 2006). For example, these words may indicate possible emotional causes (e.g. “monster”) or emotional responses (e.g. “cry”). Thus kinetic typography allows us to make explicit the indirect affective meaning in order to automatically augment the affective expressivity of texts.

Development of a KT Tool

A first step of our work was the individuation of an appropriate tool for the authoring and visualization of text animations. In particular, we required an environment that allows us to realize animations in a very simple manner and to represent them in a easily exportable format. Further functionalities are necessary also for the automated composition of animations.

¹The British National Corpus is a very large (over 100 million words) corpus of modern English, both spoken and written (BNC-Consortium 2000).

<i>News Title</i>	<i>Emotional Category</i>	<i>Affective Weight</i>	<i>Word with highest affective weight</i>
Review: 'King Kong' a giant pleasure	JOY	0.78	pleasure#n
Romania: helicopter crash kills four people	FEAR	0.67	crash#v
Record sales suffer steep decline	SADNESS	0.61	suffer#v
Dead whale in Greenpeace protest	ANGER	0.69	protest#v

Table 4: Some news titles and the respective emotional categories



Figure 1: A jumping “Hi”

To this aim we considered the Kinetic Typography Engine (KTE), a Java package developed at the Design School of Carnegie Mellon University (Lee, Forlizzi, & Hudson 2002). It allows us to create a potentially wide range of animations. Nevertheless, this tool has some limitations with respect to our intended use. A problem is the lack of a graphical interface by which to edit a string of text and to animate it without acting at the code level. This is essential for the creative design of new animations. Another problem is the difficulty in representing animation structure in a simple format, to make them exportable to other animation tools. Finally, it is necessary to individuate the set of actions for automating the composition of animations.

In order to address these problems, we first built a development environment for the creation and the visualization of text animations. We implemented an interface that displays in real time a generic animation. Our model for the animation representation is a bit simpler than the KTE model. Our central assumption consists of the representation of the animation as composition of elementary animations (e.g. linear, sinusoidal or exponential variation). In particular, we consider only one operator for the identification of elementary animations (K-BASE) and three composition operators: kinetic addition (K-ADD), kinetic concatenation (K-JOIN), and kinetic loop (K-LOOP).

The K-BASE operator selects an elementary animation (named elementary kinetic behavior) as temporal variation of some kinetic property. Elementary kinetic behaviors correspond to a subset of dynamic variations implemented in KTE, for example linear variation (linear), sinusoidal variation (oscillate), exponential variation (exponential).

The kinetic addition (K-ADD) of two animations with the same start time is obtained adding, for each kinetic property of text, the corresponding dynamical variation of each single animation. The kinetic concatenation (K-JOIN) consists in the temporal shifting of the second animation, with the starting time of the first coinciding with the ending time of the second. The kinetic loop (K-LOOP) concatenates an

linear	linear variation
oscillate	sinusoidal variation
pulse	impulse
jitter	sort of “chaotic” vibration
curve	parabolic variation
hop	parabolic variation with small impulses at the endpoints
hop-secondary	derivative of hop, used as secondary effect to simulate elastic movements

Table 5: Some elementary kinetic behaviors

animation with itself a fixed number of times. These composition operators are implemented in the development environment so as to apply them for the real time building of new animations. Compositional structure of animations can be represented in XML format and then easily exported. Finally, an interpreter allows us to generate in real time the animation starting from its structural representation.

With these additional functionalities, we developed a vocabulary of text animations. In the following example, an animation representing a “jumping” word is shown.

```
(k-add
  (k-base :k-property 'y
          :behavior 'hop
          :duration 350
          :amount -80
          :starting-speed 1
          :ending-speed 1)
  (k-base :k-property 'y-scale
          :behavior 'hop-secondary
          :duration 350
          :amount 0.2
          :starting-speed 1
          :ending-speed 1)
  (k-base :k-property 'x-scale
          :behavior 'hop-secondary
```

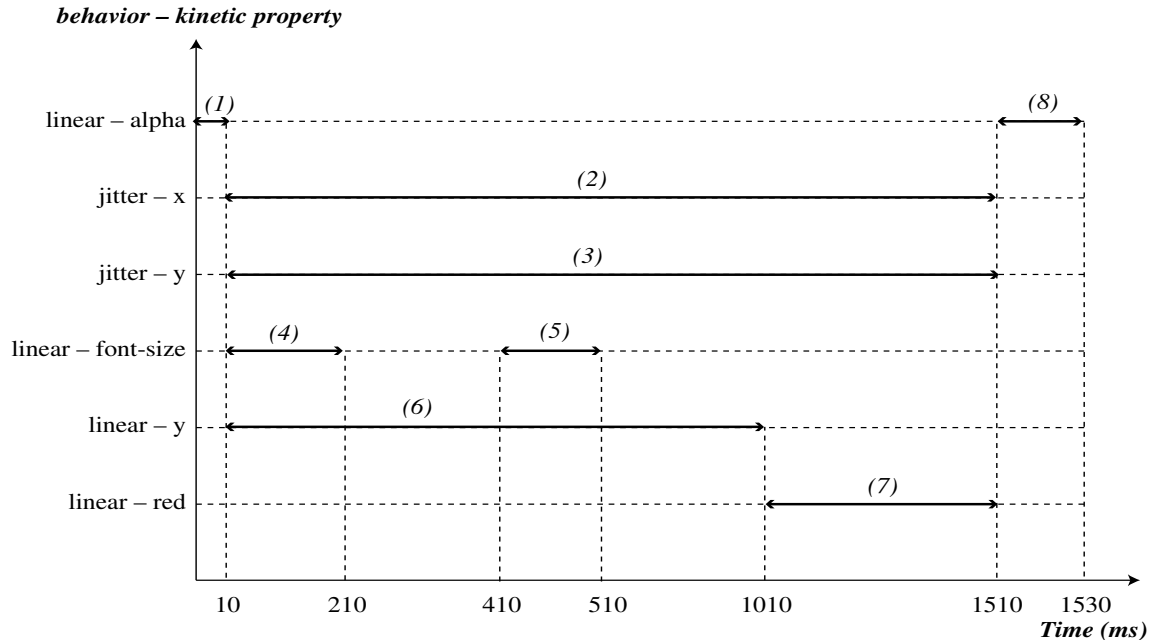


Figure 2: Kinetic behavior description for “anger” emotion

```
:duration 350
:amount -0.1
:starting-speed 1
:ending-speed 1)
```

In this case, the animation is composed through the addition (through K-ADD operator) of three elementary animations (individuated through K-BASE operator). In each of them, attribute *k-property* denotes the kinetic property to animate and attribute *behavior* identifies the corresponding elementary kinetic behavior.



Figure 3: Jittering *anger*

Affective Animation

After building the development tool, we selected a set of emotional categories and, for each of them, we created the

corresponding text animations.

In particular, we focused on five emotional categories: joy, fear, surprise, anger, sadness (i.e. a subset of Ekman emotions (Ekman 1977)).

The kinetic animation to associate to a fixed emotion can be realized imitating either emotional and physiological responses (analogous motion technique), or tone of voice. We consider only animations of the first type, i.e. we represent each emotion with an animation that simulates a particular emotional behavior. In particular, JOY is represented with a sequence of hops, FEAR with palpitations, ANGER with a strong tremble and blush, vsurprise with a sudden swelling of text, and finally SADNESS makes text deflating and getting squashed. Thus we annotated the corresponding emotional categories in WORDNET-AFFECT with these kinematic properties.

Figure 2 displays in details the behavior of the anger emotion, showing the graph time dependent composition of the basic animations. The string appears (1) and disappears (8) with a linear variation of the alpha property (that defines the transparency of a color and can be represented by a float value). The animation is contained between these two intervals and its duration is 1500 ms. The first component is a tiny random variation of the position (2) (3), represented by x and y kinetic properties, with behavior jitter. The second component consists of an expansion of the string (4) and a subsequent compression (5). The third component is given by a slow rise up (6). The latest component, before disappearing, is a color change to red (7). The whole behavior is then described and implemented using the scripting language introduced above.

In addition to affective animations, we also realized a set

of neutral ones, in order to visualize the part of text that is not related to emotions, for example to realize transition effects.

Automated Generation of Animations

Emotional and neutral animations are the results of creative design and constitute the basic ingredients for the automatic building of more complex animations. As in automatic cinematography generation (Callaway *et al.* 2005), this process can be regarded as an operation of script assembling. Here, the key idea is to automatize the composition of text animation through the automated recognition of the affective connotation and its representation via kinetic typography.

The animation algorithm is based on two steps: the automated recognition of the emotion and the representation of emotion by text animation. This is realized with the selection of the text fragments to animate, the association of the corresponding animations, and eventually the concatenation of component animations in a full integrated one. Part of automatization depends on text form (in particular, length and punctuation), while another part (the main one) depends on the lexical semantics of the text (e.g. the individuation of the most affectively relevant words).

In sum, the algorithm we follow to animate for example news titles is²:

1. given a headline, using the lexical affective semantic similarity technique, we check for the most similar emotional category (see Table 4). This is performed considering the vector of the whole headline and the vectors of the emotional categories;
2. using again the affective weight, we mark the words in the title that are closer to that emotional category;
3. we assign to each word in the title a neutral or an affective animation, corresponding to the affective weight of the word;
4. a comprehensive animation script is assembled, and the animated title is displayed.

Figure 4 shows a moment during a news title animation. The algorithm classifies the headline as similar to SADNESS. In particular the verb “cloud” has a high affective weight with respect to this emotion, and thus it is consequently animated. As it is difficult to enjoy the animations on *static* paper, please visit the web page <http://tcc.itc.it/people/strappa/download/ATA> where some downloadable short movies are available.



Figure 4: A freezed moment during a title animation

²The system is implemented in Common Lisp, using Java for the final graphical rendering.

Conclusions and Future Work

In this work we developed a strategy for the automated text animation. It is based on linking the automatic recognition of the affective lexical meaning to an appropriate animation performed with kinetic typography. Each emotion in WORDNET-AFFECT (an extension of WORDNET with affective labels on synsets) is annotated with some *kinematic* properties that simulate a particular emotional behavior.

In the development of this idea, we had in mind the specific applicative context of persuasive communication. In this context, the communicative goal is not to express a given emotion nor to induce it, but simply to “evoke” it. In other words, the animated text should make recipients think of a specific emotion, and then associate it to the content of text. This can prove very useful in advertisement communication, in particular web advertisement. Actually on the web it is possible to have access to a large amount of text and to fine-tune the corpus on which the affective weight function is calculated.

As future work we aim at improving affective recognition and to enrich the library of textual animations to be linked to emotional concepts. Another possible development area concerns intelligent interfaces. Affective expressivity may be comparable to that of facial expression (used in embodied conversational agents) but at the same time it may be more effective to capture attention on textual content. Finally, the possibility to extend this research to the prosody and characters may be very stimulating.

References

- Berry, M. 1992. Large-scale sparse singular value computations. *International Journal of Supercomputer Applications* 6(1):13–49.
- BNC-Consortium. 2000. British national corpus <http://www.hcu.ox.ac.uk/bnc/>.
- Callaway, C.; Not, E.; Novello, A.; Rocchi, C.; Stock, O.; and Zancanaro, M. 2005. Automatic cinematography and multilingual NLG for generating video documentaries. *Artificial Intelligence* 165(1):57–59.
- Deerwester, S.; Dumais, S. T.; Furnas, G. W.; Landauer, T.; and Harshman, R. 1990. Indexing by latent semantic analysis. *Journal of the American Society for Information Science* 41(6):391–407.
- Ekman, P. 1977. Biological and cultural contributions to body and facial movement. In Blacking, J., ed., *Anthropology of the Body*. London: Academic Press. 34–84.
- Fellbaum, C. 1998. *WordNet. An Electronic Lexical Database*. The MIT Press.
- Forlizzi, J.; Lee, J.; and Hudson, S. 2003. The kinedit system. In *Proceedings of the Conference on Human Factors in Computing Systems*, 377–384. ACM Press.
- Gliozzo, A., and Strapparava, C. 2005. Domains kernels for text categorization. In *Proc. of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*.

- Goy, A. 2000. Lexical semantics of emotional adjectives. In Feist, S.; Fix, S.; Hay, J.; and Moore, J., eds., : *Proceedings of Student Conference in Linguistics. MIT Working Papers in Linguistics 37*. MIT Press.
- Lee, J.; Forlizzi, J.; and Hudson, S. 2002. The kinetic typography engine: An extensible system for animating expressive text. In *Proceedings of ACM UIST 2002 Conference*, 81–90.
- Magnini, B., and Cavaglià, G. 2000. Integrating subject field codes into wordnet. In *Proceedings of LREC-2000, the Second International Conference on Language Resources and Evaluation*, 1413–1418.
- Mihalcea, R., and Liu, H. 2006. A corpus-based approach to finding happiness. In *Proceedings of Computational approaches for analysis of weblogs, AAAI Spring Symposium 2006*.
- Ortony, A.; Clore, G. L.; and Foss, M. A. 1987. The psychological foundations of the affective lexicon. *Journal of Personality and Social Psychology* 53:751–766.
- Strapparava, C., and Valitutti, A. 2004. WordNet-Affect: an affective extension of WordNet. In *Proc. of 4th International Conference on Language Resources and Evaluation (LREC 2004)*, 1083 – 1086.
- Strapparava, C.; Valitutti, A.; and Stock, O. 2006. The affective weight of lexicon. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*.
- Turney, P., and Littman, M. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems (TOIS)* 21(4):315–346.
- Valitutti, A.; Strapparava, C.; and Stock, O. 2005. Lexical resources and semantic similarity for affective evaluative expressions generation. In *Proceedings of the First International Conference on Affective Computing & Intelligent Interaction (ACII 2005)*.