Dances with Words

Carlo Strapparava and **Alessandro Valitutti** and **Oliviero Stock**
ITC-irst, Povo, I-38050, Trento, Italy
{strappa, alvalitu, stock}@itc.it

Abstract

Animated text is an appealing field of creative graphical design. Manually designed text animation is largely employed in advertising, 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 advertising, internet applications and intelligent interfaces.

1 Introduction

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 the actor's contribution to the play. With the body's 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 we will show that through *automatic* detection of the affective meaning of texts using state of the art NLP techniques, we can consequently animate the words that compose them. In automated text animation the text itself is capable of augmenting its expressivity and of *moving* in an autonomous way.

Animated text is widely employed in creative graphical design. Manually designed text animation has been employed for a long time in movie titles, television advertisements, and web pages. Nevertheless there are applied 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). Once the affective load of a sentence is detected, 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). This semantic similarity function is acquired automatically in an unsupervised way from a large corpus of texts, and it allows us to put into relation concepts and emotional categories.

We suppose that, if text animation is semantically consistent with text content, the communication of the affective meaning is more effective. Specifically, 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. Section 2 presents resources and functionalities for the recognition of affective terms. An affective hierarchy as an extension of the WordNet-Affect lexical database was 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. Section 3 introduces text animation (i.e. kinetic typography) and the development of a flexible scripting language to describe and dynamically generate text animation. Section 4 shows an evaluation and Section 5 discusses some conclusive remarks.

2 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 their semantic relation with emotional concepts. While some words have emotional meaning with respect to the individual story, for many others the affective power is part of the collective imagination (e.g. words "mum", "ghost", "war" etc.).

We are interested in this second group, because their affective meaning is part of common sense knowledge and can be detected in the linguistic usage. 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 *et al.*, 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 "killer"

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 `horryfyingly#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 *et al.*, 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 constraints (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.

This fine-grained affective lexicon selection can open up new possibilities in many applications that exploit verbal communication of emotions. For example, [Valitutti *et al.*, 2005] exploited the semantic connection between a generic word and an emotion for the generation of affective evaluative predicates and sentences.

2.1 WORDNET-AFFECT and the Emotional Categories

WORDNET-AFFECT is an extension of the WordNet database [Fellbaum, 1998], including a subset of synsets suitable to represent affective concepts. Similarly to what was done for domain labels [Magnini and Cavaglià, 2000], one or more affective labels (*a-labels*) are assigned to a number of WordNet synsets. In particular, the affective concepts representing an 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://wndomains.itc.it`. See [Strapparava and 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.

2.2 Computing Lexical Affective Semantic Similarity

There is an active research direction in the NLP field about sentiment analysis and recognition of semantic orientation from texts (e.g. [Turney and Littman, 2003; Liu *et al.*, 2003; Mihalcea and Liu, 2006]). In our opinion, a crucial issue is to have a mechanism for evaluating the semantic 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 LSA [Deerwester *et al.*, 1990], second order relations among terms and 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 full British National Corpus[1].

SVD is a well-known operation in linear algebra, which can be applied to any rectangular matrix in order to find correlations among its rows and columns. SVD decomposes the term-by-document matrix $\mathbf{T}$ into three matrices $\mathbf{T} = \mathbf{U}\mathbf{\Sigma_k}\mathbf{V}^T$ where $\mathbf{\Sigma_k}$ is the diagonal $k \times k$ matrix containing the k singular values of $\mathbf{T}$, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k$, and $\mathbf{U}$ and $\mathbf{V}$ are column-orthogonal matrices. When the three matrices are multiplied together the original term-by-document matrix is re-composed. Typically we can choose $k' \ll k$ obtaining the approximation $\mathbf{T} \simeq \mathbf{U}\mathbf{\Sigma_{k'}}\mathbf{V}^T$. More specifically, in the experiments for this paper we use the matrix $\mathbf{T}' = \mathbf{U}\mathbf{\Sigma_{k'}}$, whose rows represent the term vectors in the reduced space, taking into account the first 100 dimensions (i.e. $k' = 100$).

LSA can be viewed as a way to overcome some of the drawbacks of the standard vector space model (sparseness and high dimensionality). In fact, the LSA similarity is computed in a lower dimensional space, in which second-order relations among terms and texts are exploited. The similarity in the resulting vector space can be measured with the standard cosine similarity. Note also that LSA yields a vector space model that allows for a *homogeneous* representation (and hence comparison) of words, word sets, sentences and texts.

For representing word sets and texts by means of a LSA vector, 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 and Strapparava, 2005] for more details). In practice, 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*), and consequently 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 (e.g. we can check how a generic term is similar to a given emotion).

For example, the noun "gift" is highly related to the emotional categories: LOVE (with positive valence), COMPASSION (with negative valence), SURPRISE (with ambiguous 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, generic terms and concepts (synsets), and eventually full sentences. See [Strapparava *et al.*, 2006] for more details.

For example, Table 3 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.

3 Text Animation

Kinetic typography is the technology of text animation, i.e. text that uses movement or other changes over time.

The advantage of kinetic typography consists in a further communicative dimension, combining verbal and visual communication, and providing opportunities to enrich the expressiveness of static texts. According to [Lee *et al.*, 2002], kinetic typography can be used for three different communicative goals: capturing and directing attention of recipients, creating characters, 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).

We explore the idea to have a link between lexical semantics of texts (automatically discerned through NLP techniques) and some kinetic properties exploited for animating the words. In this paper, we consider affective connotation of texts by exploiting the affective semantic similarity introduced above. This holds particularly for "indirect affective words" [Strapparava *et al.*, 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 the indirect affective meaning explicit in order to automatically augment the affective expressivity of texts.

A first step was the individuation of an appropriate tool for the authoring and visualization of text animations. In particular, we wanted to act in an environment that allows us to realize animations in a very simple manner and to represent them in a easily exportable format. Functionalities for the automated composition of animations were our specific concern. To this aim we considered the Kinetic Typography Engine (KTE), a Java package developed at the Design School of Carnegie Mellon University [Lee *et al.*, 2002]. It allows us to create a potentially wide range of animations. Taking this engine as a starting point, we first realized a development environment for the creation and the visualization of text animations. Our model for the animation representation is a bit simpler than the KTE model. The central assumption consists of the representation of the animation as a 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

[1]The British National Corpus is a very large (over 100 million words) corpus of modern English, both spoken and written (see `http://www.hcu.ox.ac.uk/bnc/`).

News Title	*Emotional Category*	*Affective Weight*	*Word with highest affective weight*
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 3: Some news titles and the respective emotional categories

(named *elementary kinetic behavior*) as a 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*), and exponential variation (*exponential*).

`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 4: Some elementary kinetic behaviors

The kinetic addition (K-ADD) of two animations with the same start time is obtained by 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, so that the ending time of the first is the starting time of the second. The kinetic loop (K-LOOP) concatenates an animation with itself a fixed number of times. In the development environment it is possible to freely apply these operators 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.

Figure 2: Jittering *anger*

3.1 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 rensponses (*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, SURPRISE with a sudden swelling of text, and finally SADNESS with text deflaction and getting squashed. Thus we annotated the corresponding emotional categories in WORDNET-AFFECT with these kinematic properties.

Figure 1 displays in detail the behavior of the anger emotion, showing the time-dependent composition graph 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 jitter behavior. 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 last 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.

3.2 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. 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 fully integrated one. Part of automatization depends on text form (in particular, length and punctuation), while another part (the main one) depends on

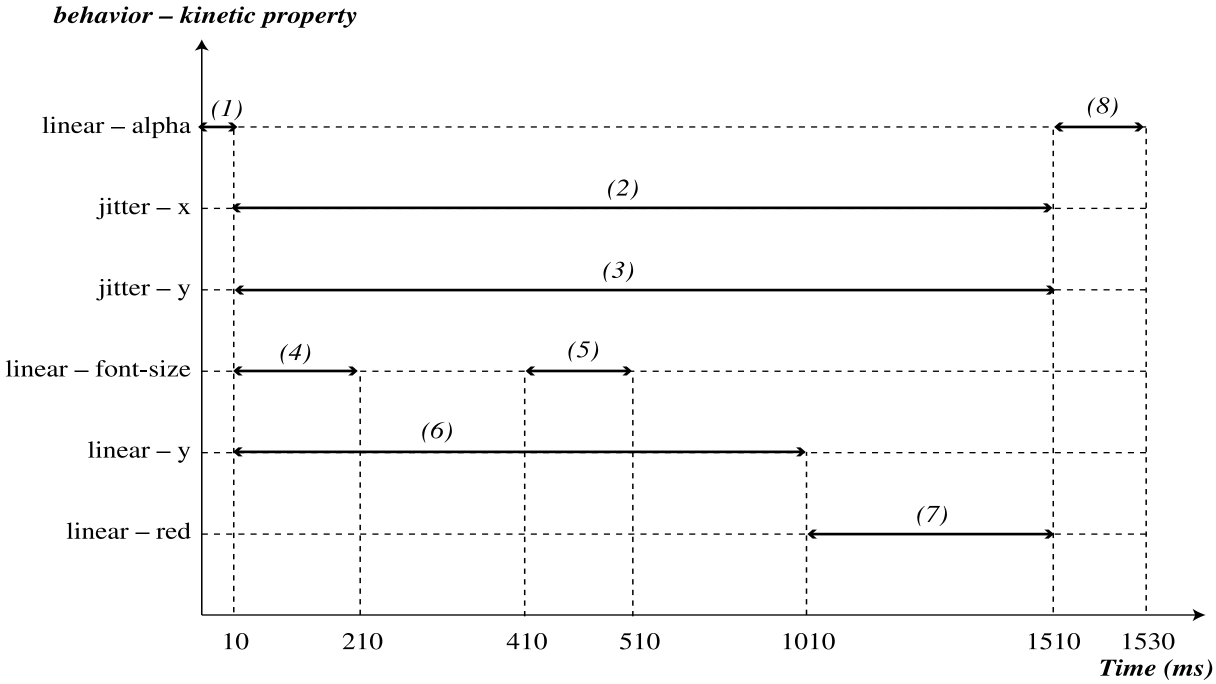

Figure 1: Kinetic behavior description for "anger" emotion

the lexical semantics of the text (e.g. the individuation of the most affectively relevant words).

In sum, the algorithm we follow in the case of headlines is[2]:

1. given a headline, using the lexical affective semantic similarity technique, check for the most similar emotional category (see Table 3);
2. mark the words in the headline that are closer to that emotional category;
3. assign to each word in the headline a neutral or an affective animation, corresponding to the affective weight of the word;
4. assemble a comprehensive animation script, and display the animated title.

Figure 3 shows a frozen image during a news title animation. As it is difficult to enjoy the animations on *static* paper, please visit the web page `http://tcc.itc.it/people/strapparava/DW-IJCAI` where some downloadable short movies are available.

Figure 3: A frozen moment during a title animation

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

4 Evaluation

We conducted an exploratory evaluation of memorization, affective coherence and pleasantness of the animated headlines. We involved ten people. The experiments were organized as follows.

- we showed to each person five static headlines in a serial manner (with an interval of 3 seconds between each. After a pause of a couple of minutes we asked the subject to recognize the five headlines among a list of 50 news titles. Afterwards we repeated this experiment with five animated headlines (of course with a different set of news titles). All the users were able to recognize the animated headlines faster, with a mean of 47.5% less time in finding the animated headlines.
- we asked the users to annotate a set of generated animations choosing from the following emotional labels: joy, fear, anger, sadness, surprise, and possibly no-label. The agreement with the automatic annotation was about 72%.
- 80% of the users declared they really appreciated the animated titles.

In addition, we created a set of "inconsistent" animations (e.g. some titles animated with a kinetic script not related to the respective emotional category). We repeated the first experiment. It is interesting to note that in this case the users performance in finding the headlines among a list of 50 news titles was even worse than the case of static headlines[3].

[3]We also asked the users to annotate these inconsistent animations. The general feeling reported was that of an annoying disorientation, in addition to the fact that the agreement was quite low

5 Conclusions and Future Work

This work has been about "giving life" to texts, automatically. The idea is that we can combine two main elements: (i) automatic recognition of the emotion evoked by a text and by specific lexical entries; (ii) an automatic way to produce animation with kinetic typography, given the text and the entries specifically marked in the first phase. 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. The mechanism involves a LSA-based similarity processing so that an emotion can be attributed to the text and to individual words. The realization of the animation is done through an underlying package for kinetic typography, so that the process develops automatically. Of course the first phase can be coupled with a different expressive mechanism, such as an embodied conversational agent with good prosody in its text-to-speech component and appropriate facial expression, but we think that the role of the written text is very important and we want to exploit its potential. Many things can be improved. For instance, a more thorough analysis of the text can lead to more elaborated structuring of the animation. Or personalisation can be brought in: for instance named entities can have a personalised valence that influences the analysis and animation. Or we can aim at recognizing irony in the text (for some preliminary work on recognizing humorous text, see [Mihalcea and Strapparava, 2005]) and express it appropriately. Basically, the presented work is meant to automatically produce what human graphic designers sometimes do for TV presentation of certain types of headlines, as in the examples of application of our work in this paper; or what is done on a much larger scale in the world of advertising. To emphasize the potential relevance of a system of this kind, it should be noted that Internet advertising was evaluated at $9.4 billion (8,000 million euro) in 2004 according to Kagan Research LLC. And growth is very fast: Google advertisement revenues went from 0 to 3,400 million euro in five years according to Business Week. This scenario calls for a strong role for computer-based intelligent technology for automatically producing novel appropriate advertisements. Future advertisements need to be flexible, and possibly depend on variable input, such as new, not previously processed text.

Acknowledgments

This work was developed in the context of HUMAINE Network of Excellence.

References

[Berry, 1992] M. Berry. Large-scale sparse singular value computations. *International Journal of Supercomputer Applications*, 6(1):13–49, 1992.

[Deerwester *et al.*, 1990] S. Deerwester, S. T. Dumais, G. W. Furnas, T.K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407, 1990.

[Ekman, 1977] P. Ekman. Biological and cultural contributions to body and facial movement. In J. Blacking, editor, *Anthropology of the Body*, pages 34–84. Academic Press, London, 1977.

[Fellbaum, 1998] C. Fellbaum. *WordNet. An Electronic Lexical Database*. The MIT Press, 1998.

[Gliozzo and Strapparava, 2005] A. Gliozzo and C. Strapparava. Domains kernels for text categorization. In *Proc. of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*, Ann Arbor, June 2005.

[Lee *et al.*, 2002] J.C. Lee, J. Forlizzi, and S.E. Hudson. The kinetic typography engine: An extensible system for animating expressive text. In *Proc. of ACM UIST 2002 Conference*, 2002.

[Liu *et al.*, 2003] H. Liu, H. Lieberman, and T. Selker. A model of textual affect sensing using real-world knowledge. In *Proc. of IUI 2003*, Miami, 2003.

[Magnini and Cavaglià, 2000] B. Magnini and G. Cavaglià. Integrating subject field codes into wordnet. In *Proc. of the 2^{nd} International Conference on Language Resources and Evaluation (LREC2000)*, Athens, Greece, 2000.

[Mihalcea and Liu, 2006] R. Mihalcea and H. Liu. A corpus-based approach to finding happiness. In *Proc. of Computational approaches for analysis of weblogs, AAAI Spring Symposium 2006*, Stanford, March 2006.

[Mihalcea and Strapparava, 2005] R. Mihalcea and C. Strapparava. Making computers laugh: Investigations in automatic humor recognition. In *Proc. of the Joint Conference on Human Language Technology / Empirical Methods in Natural Language Processing (HLT/EMNLP)*, Vancouver, October 2005.

[Ortony *et al.*, 1987] A. Ortony, G. L. Clore, and M. A. Foss. The psychological foundations of the affective lexicon. *Journal of Personality and Social Psychology*, 53:751–766., 1987.

[Strapparava and Valitutti, 2004] C. Strapparava and A. Valitutti. WordNet-Affect: an affective extension of WordNet. In *Proc. of 4^{th} International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, May 2004.

[Strapparava *et al.*, 2006] C. Strapparava, A. Valitutti, and O. Stock. The affective weight of lexicon. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy, May 2006.

[Turney and Littman, 2003] P. Turney and M. Littman. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems (TOIS)*, 21(4):315–346, October 2003.

[Valitutti *et al.*, 2005] A. Valitutti, C. Strapparava, and O. Stock. Lexical resources and semantic similarity for affective evaluative expressions generation. In *Proc. of the First International Conference on Affective Computing (ACII 2005)*, Beijing, China, October 2005.

(14% with the emotion of the text and 46% with the "wrong" emotion suggested by the animation, thus showing a small bias towards the latter).