

# A semantic-affective compositional approach for the affective labelling of adjective-noun and noun-noun pairs

Elisavet Palogiannidi<sup>2,3</sup>, Elias Iosif<sup>1,3</sup>, Polychronis Koutsakis<sup>4</sup>, Alexandros Potamianos<sup>1,3</sup>

<sup>1</sup> School of ECE, National Technical University of Athens, Zografou 15780, Athens, Greece

<sup>2</sup> School of ECE, Technical University of Crete, Chania 73100, Crete, Greece

<sup>3</sup> “Athena” Research and Innovation Center, Maroussi 15125, Athens, Greece

<sup>4</sup> School of Engineering and Information Technology, Murdoch University, Australia

epalogiannidi@isc.tuc.gr, p.koutsakis@murdoch.edu.au, {iosife,potam}@central.ntua.gr

## Abstract

Motivated by recent advances in the area of Compositional Distributional Semantic Models (CDSMs), we propose a compositional approach for estimating continuous affective ratings for adjective-noun (AN) and noun-noun (NN) pairs. The ratings are computed for the three basic dimensions of continuous affective spaces, namely, valence, arousal and dominance. We propose that similarly to the semantic modification that underlies CDSMs, affective modification may occur within the framework of affective spaces, especially when the constituent words of the linguistic structures under investigation form modifier-head pairs (e.g., AN and NN). The affective content of the entire structure is determined from the interaction between the respective constituents, i.e., the affect conveyed by the head is altered by the modifier. In addition, we investigate the fusion of the proposed model with the semantic-affective model proposed in (Malandrakis et al., 2013) applied both at word- and phrase-level. The automatically computed affective ratings were evaluated against human ratings in terms of correlation. The most accurate estimates are achieved via fusion improving the performance of the compositional model up to 5% and 4% for NN and AN, respectively.

## 1 Introduction

Affective analysis of text aims at eliciting emotion from linguistic information and it can be relevant for a wide range of applications such as sentiment analysis (Pang and Lee, 2008; Rosenthal et al., 2014; Rosenthal et al., 2015), news headlines analysis (Strapparava and Mihalcea, 2007) or affective analysis of social media (Quercia et al., 2011; Celli, 2012; Rosenthal et al., 2014; Rosenthal et al., 2015).

Word-level affective lexica can be created automatically with high accuracy (Turney and Littman, 2002; Strapparava and Valitutti, 2004; Esuli and Sebastiani,

2006; Malandrakis et al., 2013; Palogiannidi et al., 2015). Affective lexica are required in hierarchical models that combine words’ affective ratings for estimating affective ratings of larger lexical units, e.g., phrases (Turney and Littman, 2002; Wilson et al., 2005), sentences (Malandrakis et al., 2013; Rosenthal et al., 2014; Rosenthal et al., 2015) and whole documents (Pang et al., 2002; Pang and Lee, 2008). A sentiment classification approach on movies review documents was proposed in (Pang et al., 2002). Word-level semantic representations constitute the core aspect of DSMs typically constructed from co-occurrence statistics of word tuples. Such representations are the building block for models of larger lexical units, e.g., phrases and sentences, following the principle of semantic compositionality (Pelletier, 1994). They are meant to address a number of properties that are relevant to the compositional aspects of meaning, namely, “linguistic creativity”, “order sensitivity”, “adaptive capacity”, and “information scalability” (Turney, 2012). Compositional approaches have been reported for the estimation of compositional structures semantic similarity (Mitchell and Lapata., 2008; Mitchell and Lapata, 2010; Baroni and Zamparelli., 2010; Georgiladakis et al., 2015). A combination of a symbolic and a distributional compositionality approach was investigated by (Clark and Pulman, 2007), while an approach for compositional neural networks was presented in (Hammer, 2003). A recursive neural network model that learns compositional vector representations for phrases and sentences at any length and syntactic type was proposed by (Socher et al., 2012) and showed that it can be used for the prediction of sentiment as well.

In this work, we propose a compositional semantic-affective model, that is applicable to continuous affective spaces with one or more dimensions. This model is applied for the affective estimation of AN and NN pairs. The semantic-affective models are motivated by the assumption that “*semantic similarity implies affective sim-*

ilarity”, while the proposed compositional model is motivated by the CDSM proposed by (Baroni and Zamparelli., 2010), that focuses on adjective-noun composition and represents adjectives as functions and nouns as vectors. The affective ratings estimated by the compositional model are compared and combined with the ones obtained by the non-compositional semantic-affective models. Similar fusion schemes to (Georgiladakis et al., 2015) that aim to capture the compositionality degree of each word pair are investigated.

## 2 Semantic-Affective model

Semantic-affective models are employed in order to estimate affective ratings of AN and NN pairs. We focus on three continuous affective dimensions: *valence* (positive vs. negative), *arousal* (calm vs. activated) and *dominance* (controlled vs. controller). Affective ratings of each dimension are estimated by the semantic-affective model which was proposed in (Malandrakis et al., 2013). This model, that is an expansion of (Turney and Littman, 2002), is defined in (1) and it is applicable both to words as well as n-gram tokens. The assumption of (1) is that the affective rating of a lexical token can be estimated by the linear combination of its semantic similarities to a set of seeds weighted with the seeds’ affective ratings and trainable weights:

$$\hat{v}(t_j) = a_0 + \sum_{i=1}^N \alpha_i v(w_i) S(t_j, w_i), \quad (1)$$

where  $t_j$  is the unknown lexical token,  $w_{1..N}$  are the seed words,  $v(w_i)$ ,  $\alpha_i$  are the affective rating and the weight corresponding to the word  $w_i$  and  $S(\cdot)$  is the semantic similarity between two tokens.  $S(\cdot)$  is implemented within the distributional semantic models framework, following the assumption that tokens that occur in similar context tend to be semantically related (Harris, 1954). In this framework, each token is represented by a contextual feature vector that is formulated by the words that occur in a given distance from the current token in a corpus. The elements of the vectors are set according to a mutual information scheme as shown in (Palogiannidi et al., 2015). Then, the semantic similarity between two tokens is computed as the cosine of their feature vectors. The Affective Norms for English Words (Bradley and Lang, 1999) manually annotated affective lexicon was used for the selection of the seed words and Least Squares Estimation (LSE) was used for learning the weights  $\alpha_i$ .

### 2.1 Unigram and Bigram Affective Models

The unigram affective model ( $U$ ) is based on the assumption that the two words that constitute the word pair contribute equally to its affective content and it is defined as

follows:

$$\hat{v}_U(p) = \frac{\hat{v}(w_1) + \hat{v}(w_2)}{2}, \quad (2)$$

where  $p$  is the word pair, and  $\hat{v}(w_1)$  and  $\hat{v}(w_2)$  are the affective ratings that are estimated using (1). The bigram semantic-affective model ( $B$ ) handles each word pair as a single token, i.e.,  $\hat{v}_B(p) = \hat{v}(w_1 w_2)$ , where  $p$  is the word pair and  $\hat{v}_B(p)$  is the affective rating of the word pair that it has been estimated using (1) for the bigram lexical token  $w_1 w_2$ .

## 3 Compositional Affective Models

Word semantics have been represented efficiently with vector space models, as shown in (Androutsopoulos and Malakasiotis, 2010; Malandrakis et al., 2013), however representing the semantics of lexical structures larger than words is not trivial (Baroni and Zamparelli., 2010). The reason is that the meaning of complex structures derives from various compositional phenomena (Pelletier, 1994).

Semantic compositionality allows the construction of complex meanings from simpler elements based on the principle that the meaning of a whole is a function of the meaning of the parts (Partee, 1995). The key characteristic of compositionality is that the meanings of the constituent parts are combined into a single token (Mitchell and Lapata., 2008; Mitchell and Lapata, 2010). Compositional approaches in vector-based semantics can be modelled by applying a function  $f$  that acts on two constituents  $\mathbf{a}$ ,  $\mathbf{b}$  in order to produce the compositional meaning  $\mathbf{p}$ . Functions that were investigated by (Mitchell and Lapata., 2008; Mitchell and Lapata, 2010) are addition and multiplication. The additive compositional model takes the sum of the two vectors weighted with the appropriate weight matrices  $\mathbf{A}$  and  $\mathbf{B}$  respectively ( $\mathbf{p} = \mathbf{Aa} + \mathbf{Bb}$ ) and the multiplicative model is the projection of the  $\mathbf{ab}$  tensor product using a weight tensor  $\mathbf{C}$  ( $\mathbf{p} = \mathbf{Cab}$ ). These composition forms can be also simplified using the additive model with scalars instead of matrices. Similarly the multiplicative approach can be reduced to component-wise multiplication. Motivated by compositionality modeling (Baroni and Zamparelli., 2010) proposed an approach that focuses on adjective-noun composition according to which nouns are represented as vectors and adjective as functions.

In this work we assume that composition occurs in the affective rather than the semantic space and thus, we combine the affective ratings instead of the semantic representation of a phrase’s constituent words. A compositional model that is applicable on continuous affective spaces with one or three dimensions is proposed. Results are shown for valence, however the proposed models are

$$\begin{bmatrix} 1 & \hat{v}(h_1) & \hat{a}(h_1) & \hat{d}(h_1) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \hat{v}(h_K) & \hat{a}(h_K) & \hat{d}(h_K) \end{bmatrix} \begin{bmatrix} w_{10} & w_{20} & w_{30} \\ w_{11} & w_{21} & w_{31} \\ w_{12} & w_{22} & w_{32} \\ w_{13} & w_{23} & w_{33} \end{bmatrix} = \begin{bmatrix} \hat{v}(m.h_1) & \hat{a}(m.h_1) & \hat{d}(m.h_1) \\ \vdots & \vdots & \vdots \\ \hat{v}(m.h_K) & \hat{a}(m.h_K) & \hat{d}(m.h_K) \end{bmatrix} \quad (3)$$

applicable to the rest affective dimensions as well<sup>1</sup>.

### 3.1 Proposed compositional approach

In this paper, we focus on word pairs, i.e., the *modifier* (first word) and the *head* (second word). A word pair (also referred as test pair)  $p$  is defined as:  $p = m.h$ , where  $m$  is the modifier and  $h$  is the head. The assumption of our affective compositional model is that the *modifier* modifies the *head's* affective rating in order to estimate the word pair's affective rating. Modifier's impact is defined by a weight coefficients matrix or weight scalars depending on the dimensions of the compositional model. The weight coefficients are learned in a distributional approach, according to which  $K$  pairs are extracted from a large corpus and serve as the training set of each modifier. The training pairs contain the same modifier with the current test pair and different head, i.e.,  $p' = m.*$ , where  $p'$  is the training word pair,  $m$  is the modifier and  $*$  indicates any head except from the test pair's  $p$  head. For each test pair  $p$ ,  $K$  training pairs  $p'$  are extracted.

The semantic-affective model of (1) is then employed to estimate the affective ratings of the training heads and the training pairs. The estimated affective ratings are incorporated in an LSE formulation where the ratings of the training pairs formulate the dependent variable and the ratings of the training heads formulate the independent variable. The proposed compositional approach we described is depicted in Figure 1. Training is performed

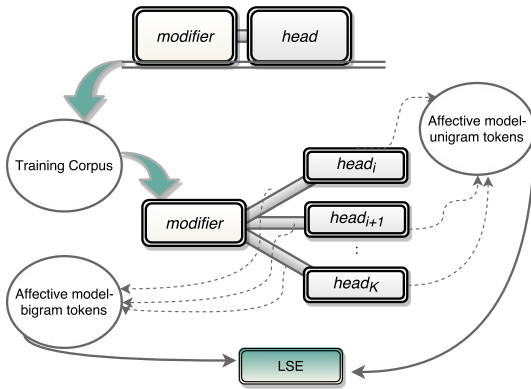


Figure 1: Compositional system overview.

separately for each modifier and the impact of each modifier is estimated through supervised learning, e.g., LSE. Finally, the continuous affective ratings of the word pair

<sup>1</sup>Evaluation results are not reported for the rest affective dimensions due to lack of groundtruth labels.

are estimated via an additive model. This model takes directly into consideration only the head, while the modifier's impact is captured by the weight matrix. The general compositional model is shown below:

$$\hat{v}_c(p) = \vec{\beta} + W\hat{v}(h), \quad (4)$$

where  $\hat{v}_c(p)$  is the compositional affective rating of the word pair  $p$ ,  $\vec{\beta}$  is the bias vector,  $W$  is the coefficients matrix and  $\hat{v}(h)$  is the affective rating of the head, estimated via (1). The compositional model may not always be the most appropriate for the estimation of a word pair's affective ratings. Thus, in order to measure the compositionality degree of each word pair the Mean Squared Error (MSE) of the model is measured. Specifically, the distance between the compositional affective ratings and the affective ratings is estimated via (1) and averaged over all training pairs:

$$MSE(p) = \frac{1}{K} \sum_{j=1}^K (\hat{v}(p'_j) - \hat{v}_c(p'_j))^2, \quad (5)$$

where  $MSE(p)$  is the MSE estimated for each pair  $p$ ,  $K$  is the number of the training pairs  $p'$ ,  $\hat{v}(p'_j)$  is the affective rating of the training word pair  $p'_j$  estimated using (1) for bigram tokens and  $\hat{v}_c(p'_j)$  is the corresponding affective rating estimated using the compositional model described in (4).

**3D compositional model (com3D):** Here we assume that three affective dimensions (valence, arousal, dominance) contribute to the affective content of the word pairs. The 3D compositional model is a special case of the general compositional model of (4), where  $W \in \mathbb{R}^{3 \times 3}$  and contains the weight coefficients for the three affective dimensions and the bias vector  $\vec{\beta} \in \mathbb{R}^{3 \times 1}$  and contains the bias of each affective dimension. The coefficient matrix  $W$  and the bias vector  $\vec{\beta}$  are estimated using LSE. Specifically, for each test pair  $p$  with  $K$  training pairs  $p'$   $K$  linear equations with 4 unknown variables (3 for the affective dimensions and 1 for the bias) have to be solved. The linear system is shown in (3). The columns of the weight matrix correspond to valence ( $w_{1*}$ ), arousal ( $w_{2*}$ ) and dominance ( $w_{3*}$ ), while the first row is the bias. With  $\hat{v}(\cdot)$ ,  $\hat{a}(\cdot)$ ,  $\hat{d}(\cdot)$  we denote the valence, arousal and dominance affective ratings respectively.

**1D compositional model (com1D):** The 1D compositional model is a special case of the 3D compositional model and follows the assumption that each affective dimension is independent of the other affective dimensions.

The dimensionality of the model is reduced and thus the behaviour of the modifier, i.e., the coefficient matrix  $W$  and the bias vector  $\vec{\beta}$  are substituted by two scalars. The general compositional model of (4) is transformed to the 1D compositional model using scalar variables. These two scalar coefficients are estimated similarly to the 3D model.

#### 4 Fusion of affective models

The same word may have different contribution to the affective content of a phrase depending on the context. For example, when the word “dog” appears in the word pair “happy dog” the conveyed affect is positive, which it is reversed when it appears in the word pair “dead dog”. The accuracy of the proposed affective models depend on the compositionality degree that characterize the pairs of interest. By combining different models we aim to achieve more accurate affective scores.

The first fusion scheme combines the affective ratings estimated by the compositional and semantic-affective models. The underlying assumption is that all models contribute equally to the affective meaning of a word pair, that is a word pair exhibits both compositional and non-compositional aspects. This scheme is based on the averaging of affective ratings defined as follows:

$$\Phi_{avg}(p) = \frac{1}{M} \sum_{i=1}^M \hat{v}_i(p), \quad (6)$$

where  $\Phi_{avg}(p)$  is the fused affective rating of the word pair  $p$ ,  $M$  denotes the number of fused models and  $\hat{v}_i(p)$  stands for the estimated affective rating of word pair  $p$ , i.e.,  $\hat{v}_1(p) = \hat{v}_U(p)$ ,  $\hat{v}_2(p) = \hat{v}_B(p)$ ,  $\hat{v}_3(p) = \hat{v}_{c1}(p)$ ,  $\hat{v}_4(p) = \hat{v}_{c3}(p)$ . A weighted variant of (6) was also investigated, as follows:

$$\Phi_{avg}^w(p) = \frac{1}{\sum_{i=1}^M w_i} \sum_{i=1}^M w_i \hat{v}_i(p), \quad (7)$$

where  $\Phi_{avg}^w(p)$  is the fused affective rating of the word pair  $p$ ,  $\hat{v}(p)$  stands for the affective rating of word pair  $p$  estimated by the affective models, as explained in (6) and  $w_i$  are weights estimated via linear regression. Motivated by the fusion scheme proposed in (Georgiladakis et al., 2015), we propose the use of MSE weight appropriately the compositional and the semantic-affective models. MSE is estimated during training phase as shown in (5) and the weight parameter is defined as  $\lambda'(p) = \frac{0.5}{1+e^{-MSE(p)}}$ , where  $MSE(p)$  is the MSE measured for each word pair  $p$  and  $\lambda'(p)$  is estimated for each word pair  $p$  based on the compositional model. (4) is applied both on 1D and 3D compositional models and the derived  $\lambda'(p)$  are averaged in order to formulate the parameter

$\lambda(p)$  that is used in the fusion scheme as follows:

$$\Phi_{avg}^{MSE}(p) = \lambda(p)(w_1 \hat{v}_B(p) + w_2 \hat{v}_U(p) + (1 - \lambda(p))(w_3 \hat{v}_{c1}(p) + w_4 \hat{v}_{c3}(p)), \quad (8)$$

where  $w_1, \dots, w_4$  are the weights that correspond to each affective model and estimated through LSE and  $\sum_{i=1}^4 w_i = 1$ ,  $\hat{v}_*(p)$  is the affective rating of each word pair derived from each affective model, and  $\lambda(p)$  is meant for weighting the contribution of compositional and non-compositional models.

## 5 Experimental Procedure and Results

### 5.1 Dataset

For evaluating the proposed semantic-affective models we use two word pairs datasets, one consisting of AN and one consisting of NN. The word pairs of the evaluation datasets were extracted from movie reviews by (Socher et al., 2013) as follows. Each movie review was first split into the constituent sentences and then into the constituent phrases. The derived sentences and phrases were annotated with respect to their polarity using crowdsourcing. We kept only the word pairs that have an adjective or noun as their first word, and their second word is a noun. The created dataset consists of 1009 AN and 357 NN pairs.

### 5.2 Semantic-affective models

The proposed models estimate the affective ratings for each affective dimension in a continuous scale in  $[-1, 1]$ , however we only report results for valence. The semantic-affective model shown in (1) was applied for the unigram (U) and bigram (B) models as defined in Section 2.1. The parameters of the semantic-affective model are set as follows:  $N = 600$  seeds, for  $S(\cdot)$  a context-based metric of semantic similarity was applied with window size equal to one, while the extracted features were weighted according to positive pointwise mutual information (Church and Hanks, 1990). LSE was applied for estimating the weights  $\alpha$  of (1). The parameters of the model are detailed in (Palogiannidi et al., 2015). The compositional model requires a large corpus<sup>2</sup> for extracting the training pairs of each modifier (Iosif et al., 2016). For each modifier all word pairs with the same modifier are extracted creating hundreds of training pairs.

### 5.3 Fusion

We investigate both weighted and unweighted average schemes while a compositionality criterion based on the compositional models’ MSE is also incorporated. In (6)

<sup>2</sup>We use a web harvested corpus that was created posing one query that was formulated for each vocabulary word on search engines and downloading and aggregating the snippets of 1K top-ranked document.

the average of all affective models is estimated, while a weighted version of this scheme is shown in (7). LSE was adopted in order to estimate the weights that capture each model’s contribution.  $\Phi_{avg}^{MSE}$  was implemented in a two-fold cross validation scenario. Moreover, we introduce a compositionality criterion based on the MSE that was measured during the compositional training process. Then the weighted average of the compositional and the non-compositional semantic-affective models were estimated as shown in (8). The weights  $w_i$  were estimated through LSE and two-fold cross validation. The parameter  $\lambda(p)$  used in (14) was computed as the average of the corresponding parameters that are estimated for the 1D and the 3D compositional models.

## 5.4 Results

We compare the valence ratings that were automatically estimated against the human valence ratings that were collected via crowdsourcing. We report evaluation results based on three evaluation metrics, namely, Pearson correlation coefficient between the estimated and the human valence ratings, binary classification accuracy (positive vs. negative valence ratings) and F-measure.

Aff. Model	Correlation		Acc. (%)		F-measure	
	NN	AN	NN	AN	NN	AN
Chance	-	-	76.4	74.1	-	-
$U$	0.581	0.573	84.3	80.0	0.903	0.874
$B$	0.507	0.451	76.8	74.6	0.841	0.824
$com1D$	0.523	0.552	79.2	77.2	0.880	0.866
$com3D$	0.538	0.552	79.8	78.1	0.883	0.870
$\Phi_{avg}$	0.624	0.609	<b>86.2</b>	80.9	<b>0.916</b>	<b>0.882</b>
$\Phi_{avg}^w$	<b>0.630</b>	0.608	85.7	<b>81.3</b>	<b>0.911</b>	<b>0.882</b>
$\Phi_{avg}^{MSE}$	0.624	<b>0.613</b>	85.5	80.9	<b>0.912</b>	<b>0.883</b>

**Table 1:** Performance of affective models for valence estimation.

The evaluation results are reported in Table 1 for several semantic-affective models. Regarding individual models, the highest performance is achieved by the unigram model for both AN and NN. This may be attributed to the very good performance at the unigram model as reported in (Malandrakis et al., 2013; Palogiannidi et al., 2015). However the fact that when moving from words to word pairs the performance drops by about 11% is a strong indicator of the need a compositional modeling. As expected, the accuracy of compositional models is between the accuracy of two semantic-affective models. Similar results have been also obtained for compositional models in the semantic space (Georgiladakis et al., 2015). Using fusion schemes that combine the compositional with the semantic-affective models we can achieve the best performance exceeding the performance of all individual models. Simple fusion schemes such as average of the affective ratings derived from the individual models

can increase the performance of the best model up to 5% in terms of correlation. Similar performance increase is observed for the rest of the evaluation metrics as well.

## 6 Conclusions and Future Work

We proposed a compositional model for estimating continuous affective ratings of AN and NN structures consisting of words that formulate modifier-head pairs. The composition was motivated by the affective interaction of modifier and head words, while it was implemented as a affine operation in the continuous affective space. The compositional models were compared and fused with two semantic-affective models defined at the unigram and bigram level. The best performance overall was achieved by the fusion-based approach suggesting that there is not a single model that works for all word pairs, i.e., the degree and type of compositionality is different for each word pair.

Currently, we are working on the improvement of the fusion schemes with focus on the identifying the parameters that control the degree of compositionality , i.e., the  $\lambda(p)$  parameter. Also, we are investigating the generalization of the proposed model across semantically/affectively more complex structures. Our long-term goal is to formulate a generic framework for integrating the compositional and non-compositional aspects of semantic and affective spaces bringing together theories from the areas of cognitive science psycholinguistics and data-driven computational models.

**Acknowledgements:** The authors were partially funded by the SpeDial project supported by the EU Seventh Framework Programme (FP7), grant number 611396 and the BabyRobot project supported by the EU Horizon 2020 Programme, grant number: 687831. Many thanks to Spiros Georgiladakis for the useful discussions on compositional models.

## References

- Ion Androutsopoulos and P. Malakasiotis. 2010. A survey of paraphrasing and textual entailment methods. *Journal of Artificial Intelligence Research*, pages 135–187.
- M. Baroni and R. Zamparelli. 2010. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In *Proc. of Empirical Methods in Natural Language Processing (EMNLP)*.
- M. Bradley and P. Lang. 1999. Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings, technical report c-1. Technical report, The Center for Research in Psychophysiology, University of Florida.

- F. Celli. 2012. Unsupervised personality recognition for social network sites. In *Proc. of International Conference on Digital Society (ICDS)*, pages 59–62.
- K. W. Church and P. Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1):22–29.
- S. Clark and S. Pulman. 2007. Combining symbolic and distributional models of meaning. In *AAAI Spring Symposium: Quantum Interaction*, pages 52–55.
- A. Esuli and F. Sebastiani. 2006. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proc. of LREC*, pages 417–422.
- S. Georgiladakis, Iosif E., and Potamianos A. 2015. Fusion of compositional network-based and lexical function distributional semantic models. In *Proceedings of CMCL*, pages 39–47.
- B. Hammer. 2003. Compositionality in neural systems. *The handbook of brain theory and neural networks*, pages 244–248.
- Z. Harris. 1954. Distributional structure. *Word*, 10(23):146–162.
- E. Iosif, S. Georgiladakis, and A. Potamianos. 2016. Cognitively Motivated Distributional Representations of Meaning. In *Proc. of 10th Language Resources and Evaluation Conference (LREC)*.
- N. Malandrakis, A. Potamianos, E. Iosif, and S. S. Narayanan. 2013. Distributional semantic models for affective text analysis. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(11):2379–2392.
- J. Mitchell and M. Lapata. 2008. Vector-based models of semantic composition. In *Proc. of (ACL)*, pages 236–244.
- J. Mitchell and M. Lapata. 2010. Composition in distributional models of semantics. *Cognitive Science*, 34(8):1388–1429.
- E. Palogiannidi, E. Iosif, P. Koutsakis, and A. Potamianos. 2015. Valence, Arousal and Dominance Estimation for English, German, Greek, Portuguese and Spanish Lexica using Semantic Models. In *Proc. of Interspeech*, pages 1527–1531.
- B. Pang and L. Lee. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135.
- B. Pang, L. Lee, and S. Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proc. of the ACL-02 conference on Empirical methods in natural language processing*, pages 79–86. Association for Computational Linguistics.
- B. Partee. 1995. Lexical semantics and compositionality. *An invitation to cognitive science: Language*, 1:311–360.
- Francis J. Pelletier. 1994. The principle of semantic compositionality. *Journal of Topoi*, 13(1):11–24.
- D. Quercia, J. Ellis, L. Capra, and J. Crowcroft. 2011. In the mood for being influential on twitter. In *Proc. of Privacy, Security, Risk and Trust (PASSAT) and IEEE Conference on Social Computing (SocialCom)*, pages 307–314.
- S. Rosenthal, A. Ritter, P. Nakov, and V. Stoyanov. 2014. Semeval-2014 Task 9: Sentiment Analysis in Twitter. In *Proc. of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 73–80. Association for Computational Linguistics and Dublin City University.
- S. Rosenthal, P. Nakov, S. Kiritchenko, S. M. Mohammad, A. Ritter, and V. Stoyanov. 2015. Semeval-2015 Task 10: Sentiment Analysis in Twitter. In *Proc. of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 451–463. Association for Computational Linguistics.
- R. Socher, B. Huval, C. D Manning, and A.Y Ng. 2012. Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1201–1211. AC:.
- R. Socher, A. Perelygin, J.W., J. Chuang, C. Manning, A.N., and C. Potts. 2013. Parsing With Compositional Vector Grammars. In *EMNLP*.
- C. Strapparava and R. Mihalcea. 2007. SemEval-2007 Task 14: Affective text. In *Proc. of SemEval*, pages 70–74.
- C. Strapparava and A. Valitutti. 2004. WordNetAffect: an affective extension of WordNet. In *Proc. of LREC*, pages 1083–1086.
- P. Turney and M. Littman. 2002. Unsupervised learning of semantic orientation from a hundred-billion-word corpus, technical report ERC-1094 (NRC 44929). Technical report, National Research Council of Canada.

- P. D. Turney. 2012. Domain and function: A dual-space model of semantic relations and compositions. *Journal of Artificial Intelligence Research*, 44:533–585.
- T. Wilson, J. Wiebe, and P. Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proc. of Human Languages Technologies and Empirical Methods in Natural Language Processing*, pages 347–354.