Lexical and affective models in early acquisition of semantics

Conference Paper · November 2017

DOI: 10.21437/WOCCI.2017-7

3 authors:

Athanasia Kolovou
National and Kapodistrian University of Athens

Elias Iosif
Technical University of Crete

Alexandros Potamianos
National Technical University of Athens

WOCCI 2017: 6th International Workshop on Child Computer Interaction
13 November 2017, Glasgow, Scotland, UK

Lexical and affective models in early acquisition of semantics

Athanasia Kolovou[1,2], *Elias Iosif*[1], *Alexandros Potamianos*[1]

[1]School of ECE, National Technical University of Athens, Zografou 15780, Athens, Greece
[2]Department of Informatics, National and Kapodistrian University of Athens, Athens, Greece
akolovou@di.uoa.gr, iosife@central.ntua.gr, potam@central.ntua.gr

Abstract

Motivated by theories of early language development in children we investigate the contribution of affective features to early acquisition of lexical semantics. For the task of semantic similarity between words, semantic and affective spaces are modeled using network-based distributed semantic models. We propose a method for constructing semantic activations from a combination of lexical and affective relations and show that affective information plays a prominent role in our lexical development model.

Index Terms: early lexical acquisition, distributional semantic models, semantic-affective models

1. Introduction

Computational models of lexical semantics can construct semantic representations for linguistic units ranging from words (word embeddings) to sentences and beyond. Distributional semantic models (DSMs) have revolutionized the way we represent meaning. Traditional DSMs deal with high-dimensional spaces exploiting patterns of word co-occurrences extracted from text corpora [1]. Such spaces can be further processed for creating embeddings with desired properties, e.g., low-dimensional dense representations [2, 3]. Despite recent progress in various semantic tasks (e.g., semantic classification, similarity computation) these models are fundamentally different from humans regarding the acquisition of word semantics [4]. Children are capable of inferring the meaning of new (unknown) words even from few examples [5, 6], while DSMs are not robust with respect to data sparsity. Inferring the semantics of words and their relations based on few or no in-domain data is an example of one-shot or zero-shot learning (e.g., see [7]), which is the focus of the present work. DSMs being robust to data sparsity enable the development of semantic modules in the framework of human-child interaction where the system has to be grounded in the contextual environment that is likely to include unknown objects.

DSMs have been criticized as "ungrounded", since they rely solely on linguistic information ignoring features from other modalities and experiential information that are related to the acquisition of semantic knowledge. This is also referred to as the *symbol grounding problem* [8]. Experimental findings indicate that real-world experiences also play a role for the acquisition of lexical semantics [9]. For example, [10] suggests that language acquisition is (also) grounded on communication episodes where partners exchange feelings.

The key hypothesis of this work is that the affective content of words can facilitate the acquisition of lexical semantics in the early stages of first language acquisition. The computational framework used here is a two-tier system, motivated by cognitive considerations such as lexical/affective activations and priming, where a target word facilitates the cognitive processing of another [11, 12]. In tier I, local network areas (subspaces) are activated, triggering a number of attributes that are (semantically/affectively) related with the target. We exploit both lexical and affective activations, as well as their fusion, for computing semantic similarity between words. In tier II, we operate on the semantic/affective activations of words to estimate semantic similarity. We set-up our experiments to simulate data sparsity conditions that roughly correspond to early acquisition by humans, namely, networks of small sizes, and availability of few examples for the target (unknown) words.

The rest of this paper is structured as follows. Related work is briefly discussed in Section 2. In Section 3, we present the lexical and affective features used for the creation of activations, while in Section 4, we present the activation-based computational model used in our analysis. The experimental data and settings are provided in Section 5, and the evaluation results are reported in Section 6. Section 7 concludes this work.

2. Related work

Vocabulary growth has been studied extensively in the linguistic and computational literature. The rate of lexical acquisition is not linear: there is a sudden surge in the rate of learning new words, especially nouns known as the "vocabulary spurt" or "naming explosion". Over the last 15 years a number of computational models have been used to model the vocabulary spurt pheonomenon on the basis of factors such as: associative learning, selective attention, rational inference ([13, 14, 15, 16]). There are few computational models based on the socio-pragmatic approach using intentions and feelings as input, e.g., [17, 18] in infant word learning.

Word-level representations constitute the core of DSMs typically constructed from co-occurrence statistics of word tuples. Word-level DSMs can be broadly categorized into unstructured and structured with respect to the extraction of contextual features. A comparison can be found in [19]. The bag-of-words model is a widely approach for the extraction of such contextual features (e.g., see [20]) based on word co-occurrence patterns. Recently, the computation of contextual features was posed in a learning-based framework where the goal is to estimate the context in which the words of interest are expected to occur [2, 3]. A comparison of this advancement with traditional DSMs is discussed in [21] for several tasks of lexical semantics. Word-level representations are the building blocks for phrase- and sentence-level models [22, 23] motivated by the principle of semantic compositionality [24]. A research direction that aims to alleviate the lack of grounding in DSMs deals with the incorporation of features from modalities other than text in order to augment the text-based DSMs, e.g., see [25] for image-derived features, and for audio-based features [26].

Based on the key hypothesis of the present work, in this paragraph a brief overview of affective text analysis is pre-

10.21437/WOCCI.2017-7

sented. Text can be analyzed for estimating affect and sentiment at different levels ranging for single words to entire sentences. Relevant applications include polarity recognition and opinion mining in domains such as product reviews [27] and tweets [28]. A mapping from semantic to affective spaces was proposed in [29], enabling the estimation of the continuous affective scores for unknown words. The mapping was based on the affective ratings of a given small set of words (seeds) and the semantic relatedness between the unknown and the seed words. An extension of this approach was proposed in [30] for computing sentence-level affective scores. The semantic-affective mapping was expanded in [31] for estimating scores for other dimensions, such as word familiarity and age acquisition. An example of the exploitation of affective spaces for semantic tasks can be found in [32] dealing with the detection of semantic opposition.

3. Lexical and affective features

The underlying assumption of this work is that the affective content of words can facilitate the acquisition of lexical semantics. This is investigated via the exploitation of lexico-semantic and affective spaces, which are constructed by employing the similarity metrics presented in Section 3.1 and 3.2, respectively.

3.1. Semantic similarity metrics

Numerous metrics have been proposed for the estimation of semantic similarity between words (a more detailed analysis can be found here [33]). In this work we utilize corpus co-occurrence statistics and, specifically, the Dice coefficient D metric defined as follows

$$D(w_i, w_j) = \frac{2 \cdot |M; w_i, w_j|}{|M; w_i| + |M; w_j|} \quad (1)$$

where $|M|$ is a set of all sentences in a corpus and $|M; w_i, ..., w_{i+n}|$ stands for the number of occurrences of words $w_i, ..., w_{i+n}$ within $|M|$. The use of the co-occurrence metrics is motivated by experimental findings suggesting that lexical co-occurrences stand as salient cues for the early acquisition of word semantics (e.g., see [34]).

3.2. Affective similarity metrics

A word w is characterized regarding its affective content in a continuous (within the $[-1, 1]$ interval) space consisting mainly of three dimensions (affective features), namely, valence, arousal, and dominance. This model that was first proposed by [30] and enhanced by [35] relies on the assumption that given some metric of similarity between two words, one may derive the similarity between their affective ratings. For each dimension, the affective content of w is estimated as a linear combination of its semantic similarities to a set of N words with known affective ratings, referred as seed words, as follows.

$$\hat{v}(w) = \alpha_0 + \sum_{i=1}^{N} \alpha_i v(t_i) S(t_i, w), \quad (2)$$

where $t_1 ... t_N$ are the seed words, $v(t_i)$ is the affective rating for seed word t_i with u denoting one of the aforementioned dimensions, α_i is a trainable weight corresponding to seed t_i and $S(\cdot)$ stands for the semantic similarity metric (see Section 5.3.1) between t_i and w.

4. Network-based model

Here, we provide a brief overview of the used computational framework [32], which consists of two layers. The first one is the *activation layer* that includes the words that are semantically/affectively related to target word w. It is computed according to the metrics defined in Section 3. The second layer is referred to as the *similarity layer*, and it is used for the computation of similarity between words based on their activations created in the previous layer. The proposed network can be represented as an undirected graph, whose set of vertices include the words under investigation and the set of edges contain links between the vertices. The links between words (nodes) in the network are determined and weighted according to their pairwise (lexical or affective) similarity metrics defined in Section 3.

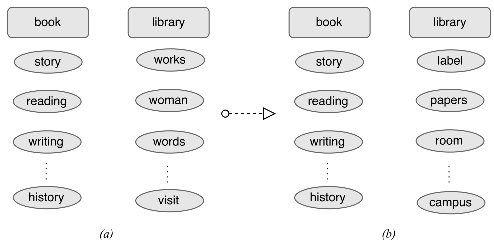


Figure 1: *Example of semantic activations representing words "book","library" when the number of examples encountered for the word "library" in the corpus is increased from (a) 5 to (b) 50.*

4.1. Layer 1: Activation models

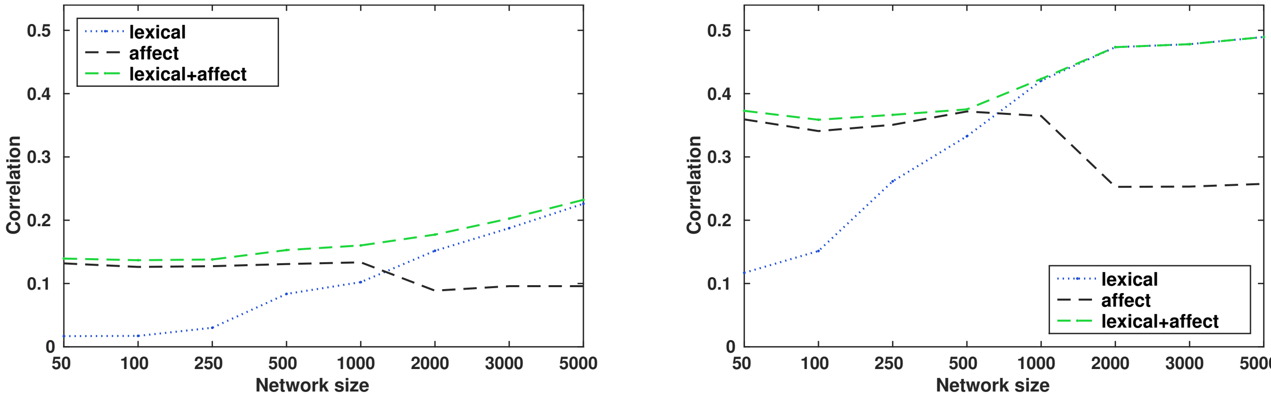
Lexical activations (L_w): The computation of the lexical activation model is motivated by semantic priming [11]. Given a target word w, the members of L_w are the n most similar words to w. Any metric of semantic similarity can be applied. In this work, we used the word co-occurrence metric defined in (1).

Affective activations (A_w): The affective activation of a target word w is motivated by affective priming [12]. The members of A_w are selected according to a metric of affective similarity (see Section 3.2) i.e., the n most similar words to w are selected.

Lexical and affective activations (F_w): We assume that both semantic and affective activations are triggered given lexical stimuli, e.g., the target words for which similarity is computed. Also, we further hypothesize that both activation types, L_w and A_w, can be fused rather being exploited independently. Here, we adopt a fusion scheme proposed in [36] for computing the activation $N_i(n)$ of a target word w_i: that is defined as follows:

$$F_w^{N_i(n)} = f(Q(w_i, L_i(|H|)), Q(w_i, A_i(|H|)); n) \quad (3)$$

where $Q(w_i, L_i(|H|)$ and $Q(w_i, A_i(|H|))$ stand for the vectors including the semantic and affective similarity scores between target w_i and the members of $L_i(|H|)$ and $A_i(|H|)$, respectively. H is the largest possible semantic/affective activation. $f(.)$ stands for a function that computes the maximum element-wise value, i.e., for each lexicon entry and the target w_i the respective maximum semantic or affective similarity score is selected. Before applying the maximum element-wise function the two feature vectors are aligned and normalized. The $f(.)$



(a) 5 instances *(b) 50 instances*

Figure 2: *Correlation for word similarity computation as a function of network size for different number of instances (for* $n = 40$*). The words of network were selected with respect to the estimated age of acquisition (from early to late).*

fusion function results in a single vector of size $|H|$ from which the n top-ranked values are considered as members of the activation $N_i(n)$.

4.2. Layer 2: Similarity layer

In this layer, the semantic similarity between pairs of words is computed based on their respective activations. We follow a similarity metric proposed in [32] that was motivated by attributional similarity suggesting that semantically similar words are expected to exhibit correlated similarities with respect to affective or lexical features. Given two words, w_i and w_j, which are represented by their activations, N_i and N_j, their semantic similarity is defined as follows:

$$R_n(w_i, w_j) = \max\{b_{ij}, b_{ji}\} \tag{4}$$

where $b_{ij}{=}\rho(C_i^{N_i}, C_j^{N_i})$, $b_{ji}{=}\rho(C_i^{N_j}, C_j^{N_j})$, $N_i{=}\{x_1, ..., x_n\}$ and $C_i^{N_i}{=}(S(w_i, x_1), S(w_i, x_2), ..., S(w_i, x_n))$. The ρ is the Pearson's correlation coefficient, N_i refers to the activation of w_i and $S(.)$ is a semantic similarity metric. The vectors $C_j^{N_i}$, $C_i^{N_j}$, and $C_j^{N_j}$ are computed similarly to $C_i^{N_i}$. The activations utilized in (4) are assumed to include semantic attributes (properties) of the words of interest w_i and w_j. Such attributes have been suggested to play an important role during the early acquisition of word meaning (for example, see [37, 38])

5. Experimental data and settings

In this section, we present the experimental setup for the proposed activation models.

5.1. Corpus and evaluation dataset

We defined a vocabulary, V, consisting of 8752 English nouns extracted from the SemCor3 corpus. For each entry of the vocabulary, a query was submitted to a web search engine and the snippets of the 1000 top documents were downloaded [39]. The retrieved snippets were aggregated for creating a text corpus.

The performance of the activation-based similarity metric defined in (4) was evaluated for the task of semantic similarity computation between nouns. For this purpose, we used a subset[1] of the MEN dataset [25] consisting of 684 pairs. The Pearson's correlation coefficient between the automatically estimated similarities and the human ratings (ground truth) was used as evaluation metric.

5.2. Sparsity

Unlike computational methods such as DSMs, the real-life acquisition of lexical semantics is performed by the exploitation of very few word examples in an incremental fashion [40]. The following scenario was adopted in order to simulate the aforementioned process. Consider a pair of words, w_i and w_j, for which the semantic similarity is to be computed. We assumed that for a member of the pair (randomly selected) few examples (i.e., sparsity) were available in the used corpus. For this purpose, we performed appropriate corpus decimation (also random). Under this condition, we created lexical and affective activations, as well as their fusion as described in Section 4.1. Based on these activations we computed the semantic similarity between w_i and w_j according to (4). We experimented with a varying number of instances. For each number of instances, the entire process (i.e., from corpus decimation to similarity computation) was repeated ten times. The performance is reported in terms of average correlation taking into account the correlation scores that were achieved for each of the ten runs.

In Figure 1, we present the fused (lexical and affective) activations of *book* and *library* which can be exploited according to (4) for computing their similarity. In this example, we assume that a child already knows the word *book* but has only heard the word *library* a limited number of times. For this example, the word *library* is encountered five (Figure 1(a)) and 50 times (Figure 1(b)) in our network-based model. In the case of Figure 1(a), the activation of *library* includes words that are moderately related with it. The semantic similarity of *library* and the members of its activation is enhanced for the case of Figure 1(b). For this case the similarity score between *book* and *library* is expected to be more accurate compared to the former case.

[1] We used those words included in the vocabulary of 8752 nouns.

5.3. Activations and network filtering

In this section, we briefly present the parameters of the experimental model.

5.3.1. Activations

The following are the parameters used for the construction of lexical/affective activations:

1. Network size: the number of words that constitute the network. We report results for $H = \{50, 100, \ldots, 5000\}$.
2. Number of instances used for investigating sparsity as mentioned in Section 5.2. In the reported experiments we used $k = \{3, 5, \ldots, 100\}$.
3. Activation size: number of words included in the activation layer defined in Section 4.1. We experimented with $n = \{10, 20, ..., 100\}$, however, here we report results only [2] for $n = 40$.

For the creation of affective spaces, (2) was applied relying on the ANEW lexicon [41] as proposed in [30]. Cosine similarity was applied for the computation of $S(\cdot)$ that appears in (2) using DSMs. The affective similarity between two words, w_i and w_j, was computed as the cosine similarity over the three-dimensional (valence, arousal, dominance) space. For the activation-based similarity metric shown in (4), as $S(\cdot)$ we used the co-occurrence-based metric defined by (1).

5.3.2. Network filtering

The largest size of the network equals to $|V|$ words (see Section 5.1). We experimented with varying network sizes, considering the following criteria for selecting a subset of vocabulary V of H size:

1. Corpus frequency.
2. *Familiarity*, i.e., the degree of exposure to and knowledge of the word.
3. *Age of acquisition*, i.e., the expected age at which one acquires the word.

For each word of the vocabulary the aforementioned scores were computed. The computation of familiarity and age of acquisition scores was performed according to the approach proposed in [31].

6. Evaluation results

In Figure 2, we present Pearson correlation as a function of the network size for activation size $n = 40$. This is shown for different number of instances, namely, five (Figure 2(a)) and 50 instances (Figure 2(b)). The words that were included in the network were selected with respect to the earlier age of acquisition. The performance is shown for both lexical, L_w, and affective activations, A_w, as well as for their fusion F_w denoted as lexical+affective. The main observation is that the model based on affective activations outperforms the lexical model for small network sizes (< 1000) regardless of the number of instances used. Also, the proposed fusion of lexical and affective activations yields performance that is higher (or equal) to the best individual model. Specifically, approximately 0.50 correlation[3] is achieved for networks consisting of 2000 words when 50 instances are used (see Figure 2(b)).

[2] Due to space limitations. The results are consistent for other values of n.

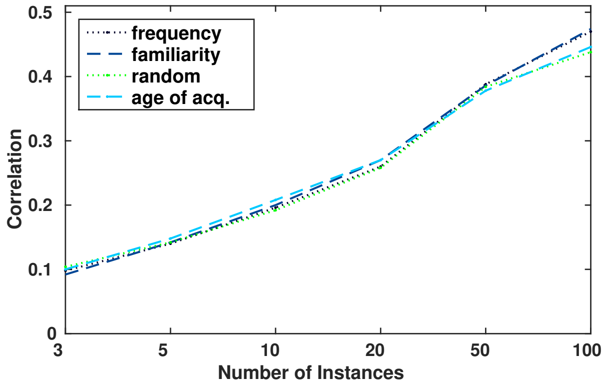


Figure 3: *Correlation of affect+lexical (fusion) as a function of number of instances shown various network filtering criteria (for $n = 40$ and network size equal to 250).*

In addition, we investigated the role of the network selection method as a function of the number of instances used. This is shown in Figure 3 for the fusion-based model using a network of 250 words and activation sizes $n = 40$. The performance is shown for three selection methods according to word scores dealing with corpus frequency, familiarity, and the earlier (younger) age of acquisition. In the same plot, we also present the performance for the random selection of network words. Slightly higher correlation is achieved by the method based on age of acquisition compared to the rest methods for up to 20 instances. As the number of instances increases, familiarity and corpus frequency tend to yield higher correlation than the other methods.

7. Conclusions

We investigated a computational framework motivated by cognitive considerations, namely, network activations and semantic/affective priming. This was applied to the computation of semantic similarity between words. We focused on sparsity conditions that roughly model to the early acquisition of lexical semantics namely, networks of small sizes, and few examples for the words of interest. The key finding is that the exploitation of affective activations facilitates the acquisition of lexical semantics for small networks. In addition, we found that the fused lexical and affective activations outperform the respective individual models.

Future work will deal with the investigation of more fusion schemes. Also, we plan to fuse the affective activations with activations created by other modalities other than text, e.g., using visual and acoustic features and similarity metrics. Last but not least, we aim to further verify the universality of the presented models using datasets in other languages.

8. Acknowledgements

This work has been partially funded by the BabyRobot project supported by the EU Horizon 2020 Programme grant num-

[3] Note that the used network-based similarity metrics achieve state-of-the-art results (0.80 correlation) for used evaluation dataset when large corpora are used (for more details see [42]).

ber 687831, and by the project "BabyAffect" which was implemented under the "ARISTEIA II" Action of the OPERATIONAL PROGRAMME "EDUCATION AND LIFELONG LEARNING" and was co-funded by the European Social Fund (ESF) and National Resources. The authors wish to thank G. Athanasopoulou for useful comments and contribution on an early version of this work.

9. References

[1] M. Baroni and A. Lenci, "Distributional memory: A general framework for corpus-based semantics," *Computational Linguistics*, vol. 36, no. 4, pp. 673–721, 2010.

[2] Y. Bengio, R. Ducharme, P. Vincent, and C. Janvin, "A neural probabilistic language model," *Journal of Machine Learning Research*, vol. 3, pp. 1137–1155, 2003.

[3] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *http://arxiv.org/abs/1301.3781/*, 2013.

[4] P. Bloom and L. Markson, "Capacities underlying word learning," *Trends in Cognitive Sciences*, vol. 2, no. 2, pp. 67–73, 1998.

[5] J. Trueswell, T. Medina, A. Hafri, and L. Gleitman, "Propose but verify: Fast mapping meets cross-situational word learning," *Cognitive Psychology*, vol. 66, no. 1, pp. 126–156, 2013.

[6] P. Bloom, *How children learn the meanings of words*, 2000, vol. 377.

[7] M. Palatucci, D. Pomerleau, G. Hinton, and T. Mitchell, "Zero-shot learning with semantic output codes," in *Advances in neural information processing systems*, 2009, pp. 1410–1418.

[8] S. Harnad, "The symbol grounding problem," *Physica D: Nonlinear Phenomena*, vol. 42, pp. 335–346, 1990.

[9] B. Landau, L. Smith, and S. Jones, "Object Perception and Object Naming in Early Development," *Trends in Cognitive Science*, vol. 27, pp. 19–24, 1998.

[10] M. Tomasello, M. Carpenter, J. Call, T. Behne, and H. Moll, "Understanding and Sharing Intentions: The Origins of Cultural Cognition," *Behavioral and Brain Sciences*, vol. 28, pp. 675–691, 2005.

[11] T. McNamara, *Semantic priming: Perspectives from memory and word recognition*. Psychology Press, 2005.

[12] P. Ferré and R. Sánchez-Casas, "Affective priming in a lexical decision task: Is there an effect of words' concreteness?" *Psicológica*, vol. 35, pp. 117–138, 2014.

[13] P. Li, X. Zhao, and B. Whinney, "Dynamic Self-organization and Early Lexical Development in Children," *Cognitive science*, vol. 31, no. 4, pp. 581–612, 2007.

[14] K.Plunkett, C. Sinha, M. Møller, and O. Strandsby, "Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net," *Connection Science*, vol. 4, no. 3-4, pp. 293–312, 1992.

[15] T. Regier, "Emergent constraints on word-learning: A computational perspective," *Trends in Cognitive Sciences*, vol. 7, no. 6, pp. 263–268, 2003.

[16] J. Siskind, "A computational study of cross-situational techniques for learning word-to-meaning mappings," *Cognition*, vol. 61, no. 1, pp. 39–91, 1996.

[17] G. Caza, "Computational model of cognitive development: Filtering ambient speech to facilitate word learning," *Proceedings of the New Zealand Computer Science Research Student Conference*.

[18] C.Yu, D. Ballard, and R. Aslin, "The role of embodied intention in early lexical acquisition," *Cognitive science*, vol. 29, no. 6, pp. 961–1005, 2005.

[19] E. Agirre, E. Alfonseca, K. Hall, J. Kravalova, M. Pasca, and A. Soroa, "A Study on Similarity and Relatedness Using Distributional and Wordnet-based Approaches," in *Proc. of the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2009, pp. 19–27.

[20] G. Grefenstette, *Explorations in Automatic Thesaurus Discovery*. Kluwer Academic Publishers, 1994.

[21] M. Baroni, G. Dinu, and G. Kruszewski, "Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors," in *Proc. of ACL*, 2014, pp. 238–247.

[22] M. Baroni and R. Zamparelli, "Nouns are vectors, adjectives are matrices: Representing Adjective-noun constructions in semantic space," in *Proc. of EMNLP*, 2010, pp. 1183–1193.

[23] R. Socher, B. Huval, C. D. Manning, and A. Y. Ng, "Semantic Compositionality through Recursive Matrix-vector Spaces," in *Proc. of EMNLP-CoNLL*, 2012, pp. 1201–1211.

[24] F. J. Pelletier, "The Principle of Semantic Compositionality," *Topoi*, vol. 13, pp. 11–24, 1994.

[25] E. Bruni, N. Tran, and M. Baroni, "Multimodal Distributional Semantics," *J. Artif. Intell. Res.(JAIR)*, vol. 49, no. 2014, pp. 1–47, 2014.

[26] D. Kiela and S. Clark, "Multi- and Cross-Modal Semantics Beyond Vision: Grounding in Auditory Perception," in *Proc. of EMNLP*, 2015.

[27] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. of SIGKDD*, 2004, pp. 168–177.

[28] P. Nakov, S. Rosenthal, Z. Kozareva, V. Stoyanov, A. Ritter, and T. Wilson, "Semeval 2013 task 2: Sentiment analysis in twitter," in *Proc. of Second Joint Conference on Lexical and Computational Semantics (*SEM), Seventh International Workshop on Semantic Evaluation*, 2013, pp. 312–320.

[29] P. Turney and M. Littman, "Measuring praise and criticism: Inference of semantic orientation from association," *ACM Transactions on Information Systems (TOIS)*, vol. 21, no. 4, pp. 315–346, 2003.

[30] N. Malandrakis, A. Potamianos, E. Iosif, and S. Narayanan, "Distributional semantic models for affective text analysis," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 21, no. 11, pp. 2379–2392, 2013.

[31] N. Malandrakis and S. Narayanan, "Therapy language analysis using automatically generated psycholinguistic norms," *Companion submission to Interspeech*, 2015.

[32] E. Iosif and A. Potamianos, "Similarity computation using semantic networks created from web-harvested data," *Natural Language Engineering*, vol. 21, no. 01, pp. 49–79, 2015.

[33] E. Iosif, "Network-based distributional semantic models," Ph.D. dissertation, Technical University of Crete, 2013.

[34] P. Li, C. Burgess, and K. Lund, "The acquisition of word meaning through global lexical co-occurrences," in *Proceedings of the thirtieth annual child language research forum*, 2000, pp. 166–178.

[35] E. Palogiannidi, E. Iosif, P. Koutsakis, and A. Potamianos, "Valence, Arousal and Dominance Estimation for English, German, Greek, Portuguese and Spanish Lexica using Semantic Models," in *Proc. of Interspeech*, 2015, pp. 1527–1531.

[36] E. Iosif and A. Potamianos, "Feeling is Understanding: From Affective to Semantic Spaces," *IWCS 2015*, p. 162, 2015.

[37] L. B. Smith and C. Nelson, "Self-organizing processes in learning to learn words: Development is not induction," in *Basic and applied perspectives on learning, cognition, and development. The Minnesota Symposia on Child Psychology*, vol. 28. Lawrence Erlbaum Associates Mahwah, NJ, 1995, pp. 1–32.

[38] B. Landau, L. Smith, and S. Jones, "Object shape, object function, and object name," *Journal of Memory and Language*, vol. 38, no. 1, pp. 1–27, 1998.

[39] E. Iosif and A. Potamianos, "Semsim:Resources for Normalized Semantic Similarity Computation Using Lexical Networks," in *LREC*, 2012, pp. 3499–3504.

[40] C. Kit, "How does lexical acquisition begin? A cognitive perspective," *Cognitive Science*, vol. 1, no. 1, pp. 1–50, 2003.

[41] M. Bradley and P. Lang, "Affective norms for English words (ANEW): Instruction Manual and Affective Ratings," The Center for Research in Psychophysiology, Univ.of Florida, Tech. Rep., 1999.

[42] E.Iosif, S. Georgiladakis, and A. Potamianos, "Cognitively motivated distributional representations of meaning," in *10th Language Resources and Evaluation Conference (LREC)*, 2016.