



## Building Dynamic Lexicons for Sentiment Analysis

Nicolás Mechulam, Damián Salvia, Aiala Rosá, Mathias Etcheverry

Instituto de Computación, Facultad de Ingeniería, Universidad de la República. Uruguay  
[nicomebu@gmail.com](mailto:nicomebu@gmail.com)

[damian.salvia@gmail.com](mailto:damian.salvia@gmail.com)

[aialar@fing.edu.uy](mailto:aialar@fing.edu.uy)

[mathiase@fing.edu.uy](mailto:mathiase@fing.edu.uy)

**Abstract:** Nowadays, many approaches for Sentiment Analysis (SA) rely on affective lexicons to identify emotions transmitted in opinions. However, most of these lexicons do not consider that a word can express different sentiments in different predication domains, introducing errors in the sentiment inference. Due to this problem, we present a model based on a context-graph which can be used for building domain specific sentiment lexicons (DL: Dynamic Lexicons) by propagating the valence of a few seed words. For different corpora, we compare the results of a simple rule-based sentiment classifier using the corresponding DL, with the results obtained using a general affective lexicon. For most corpora containing specific domain opinions, the DL reaches better results than the general lexicon.

**Keywords:** Lexicon Induction, Sentiment Analysis, Natural Language Processing.

### 1 Introduction

Sentiment Analysis identifies and extracts opinions in order to classify their sentiments. According to Taboada et al. (2011), sentiment can be induced by lexicon based approaches, or supervised text classification, eventually including lexicon based features. Generally, the sentiment is inferred by applying heuristics or machine learning methods on texts, eventually using sentiment information of words, such as *sentiment lexicon*.

The basic structure of sentiment lexicons is a reduced and selected set of words with their *valence* (sentiment or affection, e.g. ‘bad’ as negative or ‘good’ as positive). Many classification algorithms take this dictionary as input for identifying the sentiment of a whole text or for improving the results obtained by machine-learning techniques. Even though its use is necessary, usually it is not enough.

As Levin (1991) says, the lexicon usage “has often proved to be a bottleneck in the design of large-scale natural language systems”. This fact, despite technological advances from the last two decades, is still being a problem for creating and applying this kind of resources because of heterogeneous and ambiguous word’s meaning. For instance, Rosá et al. (2017) remarks that the use of sentiment lexicons does not produce significant improvements in sentiment prediction.

According to Pang and Lee (2008), when we face a lexicon construction problem, we must define as a first step which will be the lexicon’s units. There are many options to achieve this goal depending on text-preprocessing cost and sentiment dissipation from same-meaning lexical variations (i.e. by token, by

word, by lemma, by stem). For example, ‘*Educated*’ and ‘*educaated*’ tokens refers to ‘*educated*’ word, this word and ‘*educating*’ word refers to ‘*educate*’ lemma, and this lemma with ‘*educative*’ lemma refers to ‘*educ*’ stem. Note how sentiment is grouped when lexical unit is more general, but it requires more preprocessing on each lexical treatment.

Once lexical units are defined, we should define how sentiment is expressed and how lexical items will be retrieved. For the first problem, according to Jurafsky and Martin (2017), sentiments can be expressed in three ways: in a discrete domain (or categorical), in a continuous domain (a real number), or by vectors (where each component represents a sentiment feature). The same authors, for the second problem, say that there are two main approaches for retrieving relevant words and their sentiment: 1) by manual inference (human annotator/s), and 2) by machine learning strategies. The first family presents high accuracy with a considerable human effort, while the second one leads to a wider lexical coverage.

The process of building sentiment lexicons is a complex task because there is not a ‘universal receipt’ that indicates how sentiment is expressed. In consequence, these kinds of resources have been built by linguistics or experts, and it is hard to achieve high coverage in any context. For example, the word ‘long’ expresses opposite sentiments in the sentences “it is a long road to get there” and “a long weekend is coming”. In the first case, ‘long’ expresses a negative opinion over ‘the travel time’, and in the second example it expresses a positive opinion over ‘the weekend duration’. Also, on the same predication domain, there can be conflicting phrases (e.g. ‘low’ in “low price” and “low salary” for financial domain), or there can be phrases influenced by valence shifters (e.g. negation in ‘it’s not clean’ really means ‘it’s dirty’).

In this paper we present a method for the automatic construction of affective lexicons. For this, we estimate the valence of words according to their contexts and to the domain on which opinions are expressed. We call these lexicons Dynamic Lexicons. Despite the method was tested only for Spanish, its nature makes it language-independent.

The remaining document is structured in four main sections. Section 2 corresponds to related works that present some novelty about lexicon inference, Section 3 describes the bases of the generation model, Section 4 shows empirical results and Section 5 concludes the work made.

## 2 Related work

Liu (2012) exposes four problems when using lexicons for predicting sentiments: 1) Opposite affection depending on the opinion object (e.g. ‘big’ is positive for capacity but negative for electronic devices), 2) Opinions without affection words (e.g. “last night my phone got wet” is negative but it does not have any key word), 3) Neutral sentences with affection word (e.g. questions like “will my phone have a good signal there?”), and 4) Sarcasm/irony sentences (e.g. “what a good phone I bought! Two days it works...”). The problem becomes more complex when combining other linguistic phenomena, such as: intonation in written text (e.g. emojis or words as ‘EXCELLENT’ or ‘eeeexcellent’) (Muhammad et al., 2013), intensification/negation (e.g. ‘tiniest’) (Polanyi and Zaenen, 2006), or when treating global vs. local sentiment (e.g. “it’s a good phone with a bad design”) (Pang and Lee, 2008).

Hatzivassiloglou and McKeown (1997) propose one of the first methods for inducing a sentiment lexicon of adjectives. Their system starts from 1336 seed words with known valence assuming there are some conjunctions that preserve the valence (e.g. ‘and’) and others that invert it (e.g. ‘but’). They report a precision of 82.05%.

Wu and Wen (2010) try to infer the polarity of ambiguous adjectives. They propose the concept of *sentiment expectation* that relates a noun (with known valence) with an adjective. For instance, if ‘salary’ has positive expectation, then ‘low’ should be negative-like in the phrase “the salary is low” due this phrase is negative. The authors report 80.06% of macro-F measure and 79.72% of accuracy.

Turney and Littman (2002) use PMI (Pointwise Mutual Information) and LSA (Latent Semantic Analysis) as a semi-supervised mechanism, reporting 74% of precision. Probably, the most important fact of their work is that they use only two seeds, one of each class (‘excellent’ and ‘poor’). Also, the nature of the applied method (PMI+LSA) encapsulates domain-specific information from the corpora used to build the lexicon.

Many works for sentiment lexicon construction rely on a graph based approach, in the following we comment the most relevant ones for our work.

Kamps et al. (2004) use WORDNET to find word synonyms from a set of seeds in order to propagate their valence, or invert them when they are antonyms. The valence assigned to new words is calculated by a function called EVA (short for ‘evaluation’) which measures the distance between ‘good’ and ‘bad’ adjectives. The authors report 68.19% of precision.

Esuli and Sebastiani (2007) also use WORDNET to build SENTIWORDNET, but they build a graph based on word’s glosses and then apply a modified version of PageRank algorithm over it. For evaluation they use *Kendall Correlation* measure, reporting 0.324 for positive words and 0.284 for negative words.

Another work inspired in WORDNET is Hamilton et al. (2016), who introduce SENTPROP. They build a graph using word embeddings where each word is a node connected with their  $k$ -nearest neighbors in the semantic word space model. Then, they apply a label-propagation mechanism over the valence from a few seed words following a probabilistic model defined on the weight of each edge. They report 86.7% of accuracy, 68.8 % of f-measure and 0.4 of Kendall correlation.

Chen and Skiena (2014) build a graph with semantic inter-language edges using a variety of linguistic resources, starting from an English sentiment lexicon as seeds. They cover 136 different languages and evaluate their results using existing non-English sentiment lexicons.

Our method is inspired by these graph based approaches. We take the idea of building a word’s graph in order to propagate the valence of a few seed words (with known valence) through it. However, our method differs from them in the fact that edges in our graph represent syntagmatic co-occurrence of words in texts from a specific domain, instead of representing semantic relations between words, provided by resources such as WORDNET or word embeddings.

### 3 Our Approach

Our work is based on the following hypothesis: “*the sentiment of a word depends on its context*”. Such context should be understood by 1) the neighbor words (or textual context), 2) the negation influence, and 3) the predication domain. In these terms, we create a mechanism for building a lexicon of  $\langle \text{word}, \text{valence} \rangle$  pairs based on the architecture described in Figure 1.

We started from a set of diverse corpus ( $C_1, \dots, C_n$ ) with sentiment-tagged opinions (e.g. reviews from movies, sports, restaurants, etc.), described in Tables 1 and 2 – refer to the author’s paper for further information (annotation method, valence meaning, etc). The complete corpus has 79.643 opinions. Each corpus was preprocessed applying word correction and lemmatization, identifying negated words and standardizing their heterogeneous valences.

To detect words within a negation scope we trained a bidirectional long-short term memory network using the SFU-NEG corpus (Martí Antonín et al., 2016). The inputs of the neural model are word embeddings built using skip-gram on a Spanish corpus of almost six billion words Azzinnari and Martínez (2016). More details on these experiments can be found in Salvia and Mechulam (2018).

After these preprocessing steps, we used the 10% of each corpus to build an *Independent Lexicon* (*IL*) where words sentiment does not depend on the context to which the words belong, they are always positive or negative.

From the remaining corpora, we used 72% of each corpus  $C_i$  to build each *Dynamic Lexicon* ( $DL_i$ ), using the words of the *IL* as seeds. The dynamic lexicon is built by generating a context-graph and applying a semi-supervised valence propagation mechanism over it.

We used the remaining 18% of each  $C_i$  as test corpus. To evaluate we compared the results obtained using the domain specific generated lexicon against a general purpose lexicon in sentiment analysis detection.

In the following, we describe the mechanisms applied to build the *IL* and the *DL*.

#### 3.1 Independent Lexicon

We built an *Independent Lexicon* (*IL*) taking the ideas of Ghag and Shah (2014). These authors propose Senti-TFIDF, a mechanism which treats sentiments as a Bag-of-Words (BoW). The result will be a lexicon where word’s sentiment is independent of the context, since it includes words that are frequently used

<sup>1</sup>The complete list is: cars, hotels, washing machines, books, cellphones, music, computes, movies.

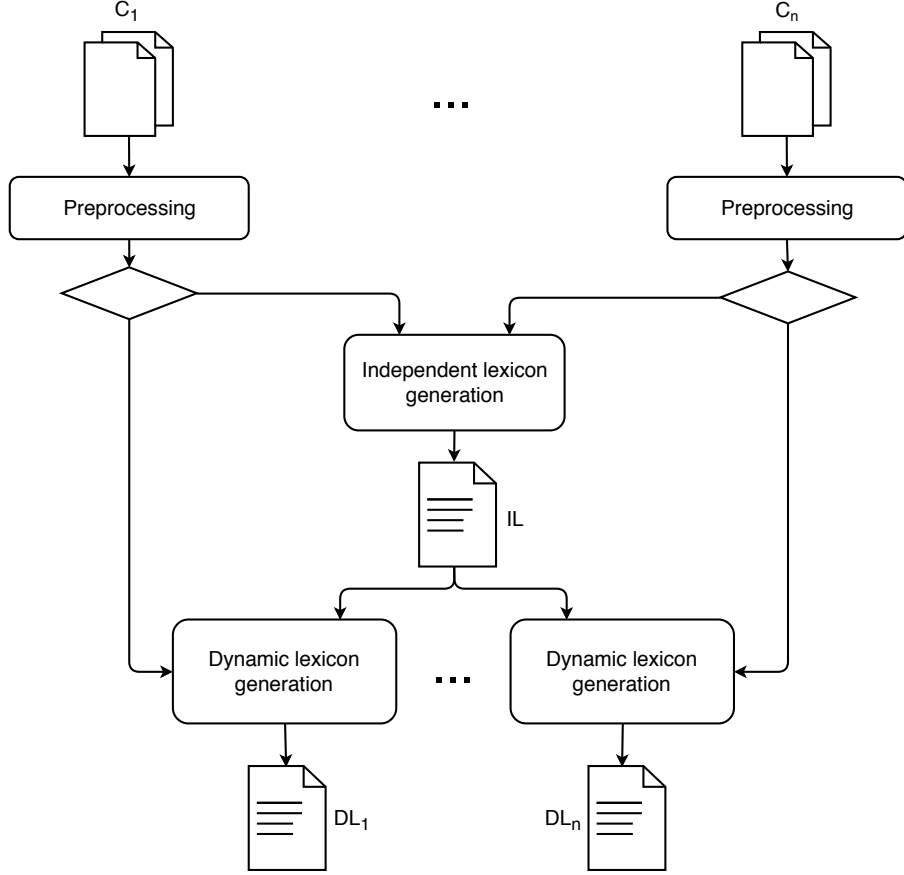


Figure 1: Architecture

with the same polarity in the different domains we studied. These words are used as seeds for sentiment propagation, in the construction of the dynamic lexicon. In this work, despite we know when a word is negated or not, we modify the original mechanism to include the negation in the calculus, resulting in the following equations:

$$\text{X-TF}_{t,d} = \frac{\begin{cases} f_{t^+,d} & \text{if } d \in D_X \\ f_{t^-,d} & \text{if } d \in D_{\bar{X}} \end{cases}}{\sum_{x \in D_X} f_{t^+,x} + \sum_{x \in D_{\bar{X}}} f_{t^-,x}}$$

$$\text{X-IDF}_t = \log \frac{|D_X \cup D_{\bar{X}}|}{|\{d \in D_X : t^+ \in d\} \cup \{d \in D_{\bar{X}} : t^- \in d\}|}$$

where  $X$  represents the *POS* or *NEG* class and  $\bar{X}$  the complementary class of  $X$ ; besides,  $t^-/t^+$  represent if term  $t$  is negated or not, respectively. The words in the *IL* will be the top 150 with higher absolute value of TF-IDF (product of both formulas).

Taking a portion of the resources described in Table 1 (as it was detailed in the introduction of this chapter), we built an *IL* limited to 150 words, including 75 positive words and 75 negative words. For instance, the most representative words on each category were *desastre* ('disaster'), as negative, and *excelente* ('excellent'), as positive.

### 3.2 Dynamic Lexicon

In this section we describe the construction of the context graph and the mechanism for valence propagation, which uses the words from the *IL* as seeds.

Reference	Valences	Domain
APPS by Dubian (2013b)	pos, neg	Mobile apps
MOV by Cruz Mata et al. (2008)	1,2,3,4,5	Movies
HOTEL by Molina-González et al. (2014)	1,2,3,4,5	Hotels
MED1 by Plaza-Del-Arco et al. (2016)	0,1,2,3,4,5	Medicine
MED2 by Jiménez-Zafra et al. (2017)	positive, negative, neutral	Medicine
MISC1 by Vilares et al. (2015)	pos 1-5, neg 1-5	General
MISC2 by Martí Antonín et al. (2016)	pos, neg	Cars, hotels,... <sup>1</sup>
MISC3 by Mori et al. (2016)	P, N, NEU, NONE	Politics
JOUR1 by Mordecki et al. (2015)	Pos, Neu, Neg	Journalism
JOUR2 by Boldrini et al. (2009)	positive, negative, $\varepsilon$	Journalism
REST1 by Jiménez-Zafra et al. (2009)	1,2,3,4,5	Restaurants
REST2 by Dubian (2013a)	pos, neg	Restaurants

Table 1: Collected corpus with sentiment annotation by domain

Corpus	Total	IL (10%)	DL (72%)	Eval (18%)
APPS	8.281	828	7.453	1.491
MOV	3.878	388	3.490	698
HOTEL	1.816	182	1.634	327
MED1	743	74	669	134
MED2	453	45	408	82
MISC1	3.600	360	3.240	648
MISC2	400	40	360	72
MISC3	2.466	247	2.219	444
JOUR1	2.260	226	2.034	407
JOUR2	974	97	877	175
REST1	2.202	220	1.982	396
REST2	52.570	5.257	47.313	9.463
<b>Total</b>	<b>79.643</b>	<b>7.964</b>	<b>71.679</b>	<b>14.337</b>

Table 2: Corpus sizes after splitting for IL, DL and Evaluation.

### 3.2.1 Context Graph

The idea behind the construction of the *Context Graph*, inspired by the semantic graphs used in (Kamps et al., 2004; Esuli and Sebastiani, 2007; Hamilton et al., 2016; Santos et al., 2012), is that nodes represent the vocabulary of the domain, and edges represent the textual contexts (contiguous words in the text). For instance, if the words ‘A’ and ‘B’ appear together in the text, then there is an edge  $\langle A, B \rangle$  in the graph, representing an *adjacency relation*. Each specific corpus will be associated to a graph from which a *DL* will be obtained. Thus, a context graph will be derived from each particular domain.

Normally, in sentences like “The sky is turning red”, the words ‘The’ and ‘is’ lack lexical content, and their occurrence avoids adjacency relations such as ‘sky’ - ‘beautiful’. To solve this issue we decided to discard stop words and auxiliary verbs.

Note that the resulting graph contains many connected components, this issue hinders the propagation of the valences outside each component, reducing the graph coverage. As a way to reduce the generation of connected components, we decided to compact the different inflections of a word in a single node using

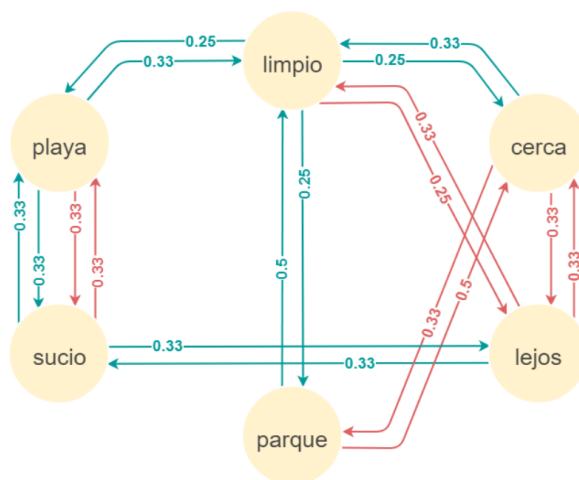


Figure 2: Context graph of a simple corpus.

its lemma and, in this way, decrease the sparsity without losing affective content.

The presence of negations in the text affects the valence of the words within its scope. Since the context graph seeks to model the contexts of words, negations are an additional aspect to be represented in the structure.

To model the negation, we decided to maintain two types of relationships: ((direct)) and ((inverse)). However, the affection transmitted by the relationships  $\langle \text{no-}X, Y \rangle$  and  $\langle X, \text{no-}Y \rangle$  is the same. That is, considering “The dog that does not bite is good” and “The dog that bites is not good”, the relations between ((bites)) and ((good)) is assumed to express the same sentiment, although the sentences are semantically different <sup>2</sup>.

Formally, direct relations ( $\langle x, y \rangle_{dir}$ ) are understood as links between two neighboring words with the same affective value (both are denied or none are). On the other hand, an inverse relation ( $\langle x, y \rangle_{inv}$ ) is one in which the neighboring words have a different affective value (one has the opposite value than the other one). In these conditions, for the first case, the valence is maintained, while the second is inverted.

Although for a set of texts the graph reflects the words contexts, it does not take into account the frequency with which they occur. That is, assuming that the words <dog, bites> co-occur more frequently than <dog, plays>, the former should take more importance in the representation, since ‘dog’ is closely related to ‘bites’ and probably has a similar affective value. For this reason, we decided to apply a weighting mechanism based on the number of co-occurrences in the texts, the weight ( $w$ ) of an edge is the probability that two words are adjacent (directly or indirectly).

It is important to note that the probability  $P(y|x)$  depends on the relation of the node “ $x$ ”. Therefore, since the weight of  $\langle x, y \rangle$  and  $\langle y, x \rangle$  could take different values, there is a need to express the origin and destination of the relation. Consequently, the context graph must be directed in order to adequately represent the weights. Given that there exist direct and inverse relations, the context graph is a directed graph with multiple weighted edges.

In figure 2 we show the context graph resulting from the following corpus, designed to explain the concept:

- La playa es sucia y está lejos. / The beach is dirty and far.
- El parque es limpio y no está lejos. / The park is clean and it is not far.
- La playa está limpia y es cerca. / The beach is clean and near.
- El parque no está cerca, está lejos. / The park is not near, it is far.

<sup>2</sup>Logically bites good does not express the same thing that bites good.

- La playa no está sucia. / The beach is not dirty.

As we can see, *sucia/sucio* (*dirty*) and *limpia/limpio* (*clean*) are represented by a unique node, labeled with the corresponding lemma (*sucio* and *limpio*). Stop words (*el, la*) and auxiliary verbs (*es, está*) are not included in the graph. Green edges indicate direct relations while red ones indicate inverse relations. Each edge weight is computed dividing the number of co-occurrences of its endpoints (distinguishing negated and non-negated contexts) by the number of occurrences of its origin. In order to illustrate this, we analyze the three occurrences of the lemma *playa*:

- the first one is in the non-negated context of *sucio*, generating a direct edge (green) with a weight of 0.33 (1/3) between the nodes *playa* and *sucio*,
- the second one is in the negated context of *sucio*, hence an inverse edge (red) with a weight of 0.33 (1/3) is added,
- the last one is in the non-negated context of *limpio*, resulting in a direct edge (green) with a weight of 0.33 (1/3).

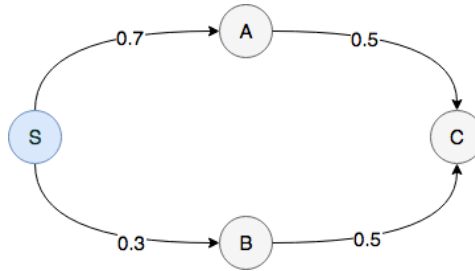
### 3.2.2 Valence Propagation

In this section, we describe the algorithm used for valence propagation over the *context graph*. This algorithm allows us to obtain the valence of each node of the *context graph*, starting from the seeds (*IL*), whose affective value is known. The result of this algorithm will determine the *DL*.

The algorithm is based on the concept of influence (*inf*), a positive real value that represents how similar is the valence of a node respect to another. This value is given by the path that maximizes the product of the edges. It is calculated recursively according to Equation 1, where *in* returns the incident nodes to the target. Notice how the function, when looking for the incidents of the destination node (*in(y)*), implicitly calculates the maximum path from *x* to *y*. To clarify this idea, the Example 3.1 is proposed (for simplicity the bidirectionality of the edges is omitted).

$$inf(x, y) = \begin{cases} 1 & : \text{if } x = y \\ \max_{z \in in(y)} \{ inf(x, z) * \omega(< z, y >) \} & : \text{else} \end{cases} \quad (1)$$

**Example 3.1.** Given the following simplified context graph, it is shown the calculation (top-down) of the influence of node *S* (seed) towards node *C*.



$$\begin{aligned} inf(S, C) &= \max \left\{ \begin{array}{l} inf(S, A) * \omega(< A, C >) \\ inf(S, B) * \omega(< B, C >) \end{array} \right\} \\ &= \max \left\{ \begin{array}{l} inf(S, S) * \omega(< S, A >) * \omega(< A, C >) \\ inf(S, S) * \omega(< S, B >) * \omega(< B, C >) \end{array} \right\} \\ &= \max \left\{ \begin{array}{l} 1 * 0.7 * 0.5 = 0.35 \\ 1 * 0.3 * 0.5 = 0.15 \end{array} \right\} \\ &= 0.35 \end{aligned}$$

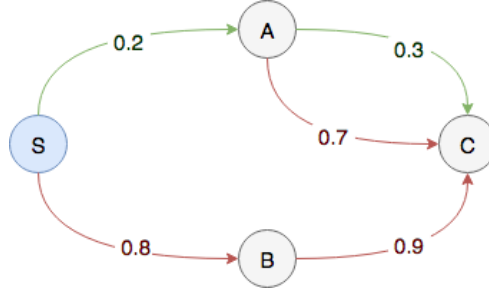
Note that Example 3.1 does not include inverse edges. This simplification is intentional since the Equation 1 only describes the influence of direct relations. Remember that a direct relation represents two neighboring words with the same sentiment, while an inverse relation represents words with opposite sentiments. In this sense, it is necessary to define **direct influence** ( $inf_{dir}$ ) and **inverse influence** ( $inf_{inv}$ ) in a mutually recursive way according to the Equation 2

$$inf_{dir}(x, y) = \begin{cases} 1 & : \text{if } x = y \\ \max_{z \in in(y)} \left\{ \begin{array}{l} inf_{dir}(x, z) * \omega(< z, y >_{dir}) \\ inf_{inv}(x, z) * \omega(< z, y >_{inv}) \end{array} \right\} & : \text{else} \end{cases} \quad (2)$$

$$inf_{inv}(x, y) = \begin{cases} 0 & : \text{if } x = y \\ \max_{z \in in(y)} \left\{ \begin{array}{l} inf_{dir}(x, z) * \omega(< z, y >_{inv}) \\ inf_{inv}(x, z) * \omega(< z, y >_{dir}) \end{array} \right\} & : \text{else} \end{cases}$$

It is important to note that the *direct influence* of a node on itself is 1 since a word always has the same valence as itself. Analogously, the *inverse influence* of a node on itself is 0 since a word never has a valence opposite to itself. As a way to analyze the influence calculations with this new formula, Example 3.2 is proposed. A real context graph is much more complex, for this reason an algorithm inspired by Dijkstra was designed to optimize the calculations.

**Example 3.2.** Given the following context graph, the calculation (bottom-up) of the direct and inverse influences from the node *S* (seed) to node *C* is shown.



**Node S (base step)**

$$inf_{dir}(S, S) = 1$$

$$inf_{inv}(S, S) = 0$$

**Node A**

$$inf_{dir}(S, A) = \max \left\{ \begin{array}{l} inf_{dir}(S, S) * \omega(< S, A >_{dir}) \\ inf_{inv}(S, S) * \omega(< S, A >_{inv}) \end{array} \right\} = \max \left\{ \begin{array}{l} 1 * 0.2 = 0.20 \\ 0 * 0.0 = 0.00 \end{array} \right\} = 0.20$$

$$inf_{inv}(S, A) = \max \left\{ \begin{array}{l} inf_{dir}(S, S) * \omega(< S, A >_{inv}) \\ inf_{inv}(S, S) * \omega(< S, A >_{dir}) \end{array} \right\} = \max \left\{ \begin{array}{l} 1 * 0.0 = 0.00 \\ 0 * 0.2 = 0.00 \end{array} \right\} = 0.00$$

**Node B**

$$inf_{dir}(S, B) = \max \left\{ \begin{array}{l} inf_{dir}(S, S) * \omega(< S, B >_{dir}) \\ inf_{inv}(S, S) * \omega(< S, B >_{inv}) \end{array} \right\} = \max \left\{ \begin{array}{l} 1 * 0.0 = 0.00 \\ 0 * 0.8 = 0.00 \end{array} \right\} = 0.00$$

$$inf_{inv}(S, B) = \max \left\{ \begin{array}{l} inf_{dir}(S, S) * \omega(< S, B >_{inv}) \\ inf_{inv}(S, S) * \omega(< S, B >_{dir}) \end{array} \right\} = \max \left\{ \begin{array}{l} 1 * 0.8 = 0.80 \\ 0 * 0.0 = 0.00 \end{array} \right\} = 0.80$$

**Node C**

$$inf_{dir}(S, C) = \max \left\{ \begin{array}{l} inf_{dir}(S, A) * \omega(< A, C >_{dir}) \\ inf_{inv}(S, A) * \omega(< A, C >_{inv}) \\ inf_{dir}(S, B) * \omega(< B, C >_{dir}) \\ inf_{inv}(S, B) * \omega(< B, C >_{inv}) \end{array} \right\} = \max \left\{ \begin{array}{l} 0.2 * 0.3 = 0.06 \\ 0 * 0.7 = 0.00 \\ 0 * 0.0 = 0.00 \\ 0.8 * 0.9 = 0.72 \end{array} \right\} = 0.72$$

$$inf_{inv}(S, C) = \max \left\{ \begin{array}{l} inf_{dir}(S, A) * \omega(< A, C >_{inv}) \\ inf_{inv}(S, A) * \omega(< A, C >_{dir}) \\ inf_{dir}(S, B) * \omega(< B, C >_{inv}) \\ inf_{inv}(S, B) * \omega(< B, C >_{dir}) \end{array} \right\} = \max \left\{ \begin{array}{l} 0.2 * 0.7 = 0.14 \\ 0 * 0.3 = 0.00 \\ 0 * 0.9 = 0.00 \\ 0.8 * 0.0 = 0.00 \end{array} \right\} = 0.14$$



Once defined the influence between nodes (direct and inverse), the valence of a word is calculated according to the influence it receives from the seeds (Equation 3).

$$val(x) = \frac{\sum_{s \in S} (inf_{dir}(s, x) * val(s) - inf_{inv}(s, x) * val(s))}{\sum_{s \in S} (inf_{dir}(s, x) + inf_{inv}(s, x))} \quad (3)$$

In order to summarize the strategy hereby described, Algorithm 1 is presented considering all the aspects discussed.

```

Input: context graph (G), seeds (S)
Output: lexicon (L)

influences_nodes  $\leftarrow \emptyset$ 
foreach (seed, valence) in seeds do
    influences  $\leftarrow$  get_influences(graph, seed)
    foreach (node, inf_dir, inf_inv) in influences do
        | influences_nodes[node]  $\leftarrow$  (seed, inf_dir, inf_inv)
    end
end
L  $\leftarrow \emptyset$ 
foreach inf_node in influences_nodes do
    sum_val  $\leftarrow$  0
    sum_inf  $\leftarrow$  0
    foreach (s, inf_dir, inf_inv) in inf_node do
        | sum_val  $\leftarrow$  sum_val + seeds[s] * (inf_dir - inf_inv)
        | sum_inf  $\leftarrow$  sum_inf + inf_dir + inf_inv
    end
    L[node]  $\leftarrow$  sum_val / sum_inf
end
return L

```

**Algorithm 1:** Propagation algorithm by seeds influence

Another propagation mechanism has been tested in Salvia and Mechulam (2018) original work, inspired in BFS algorithm (*Breath First Search*) with a slightly poorer different result.

## 4 Results

In order to evaluate the Dynamic Lexicons obtained for each domain, we separated 18% of each corpus (14.337 opinions, see Table 2) before building the respective DL. We apply a simple lexicon based heuristic for sentiment detection: we compute the sentiment of an opinion through the valence average of the affective words it contains (inverting the valence of negated words). For each domain, we compare the results obtained using the corresponding DL with the results obtained using a *General Lexicon* (GL) with word polarities for Spanish ((Saralegi and San Vicente, 2013), version *ElhPolar-esV1.lex* has 5.200 words). In Table 3 we show the Macro-F measure for each corpus and for each lexicon (DL vs GL). This evaluation was carried out on the portion of the corpus reserved for testing.

As can be seen, for most specific domain corpora (movies, hotels, restaurants, medicine) the classifier based on the Dynamic Lexicon reaches slightly better results than the one based on the General Lexicon, except for apps domain. On the other hand, for corpora with opinions covering different domains (miscellaneous and journalistic texts) the General Lexicon has a better performance, as expected. The best performance was obtained using the domain specific lexicon on the largest corpus (REST2), reaching a Macro-F of 87.5

Notice that in most cases the improvement obtained is not statistically significant, excepting the REST2 corpus, which is remarkably larger than the others. These results show the relevance of the corpus size in the Dynamic Lexicon generation. Moreover, as a positive aspect, it is important to note that similar (even better) results are obtained using the Dynamic Lexicons, which are much smaller than the general one (20 times smaller approximately).

Corpus	Test Size	Dynamic Lex. Size	Dynamic Lexicon	General Lexicon
APPS	1.491	177	71.1 $\pm$ 2.3	<b>73.7</b> $\pm$ 2.2
MOV	698	214	<b>68.9</b> $\pm$ 3.4	63.5 $\pm$ 3.5
HOTEL	327	246	<b>85.1</b> $\pm$ 3.8	84.6 $\pm$ 3.9
MED1	134	155	<b>69.4</b> $\pm$ 7.8	60.1 $\pm$ 8.3
MED2	82	79	<b>49.0</b> $\pm$ 10.8	40.7 $\pm$ 10.6
MISC1	648	211	48.8 $\pm$ 3.8	<b>63.5</b> $\pm$ 3.7
MISC2	72	137	59.0 $\pm$ 11.4	<b>68.8</b> $\pm$ 10.7
MISC3	444	123	61.2 $\pm$ 4.5	<b>63.1</b> $\pm$ 4.5
JOUR1	407	49	59.9 $\pm$ 4.8	<b>75.6</b> $\pm$ 4.2
JOUR2	175	113	51.1 $\pm$ 7.4	<b>54.7</b> $\pm$ 7.4
REST1	396	266	<b>84.8</b> $\pm$ 3.5	84.0 $\pm$ 4.5
REST2	9.463	210	<b>87.5</b> $\pm$ 0.6	84.9 $\pm$ 0.7

Table 3: Macro-F<sub>1</sub> for each corpus using the Dynamic Lexicons and the 5.200 words General Lexicon.



Figure 3: Word clouds results.

In Figure 3 we show some word clouds for restaurant and medicine domains. Bigger words represent higher valences, while green/red colors represent positive/negative words respectively. This graphical representation of the result helps to intuitively understand how well the Dynamic Lexicon behaves in the domain where it had been inferred. Note that words like ‘cerdo’ (‘pig’), that have a positive valence in the Restaurants domain, could have a negative valence in other cases.

## 5 Conclusions

We have presented a semi-supervised mechanism for building a sentiment lexicon from unlabeled corpora and a set of seed words (named *IL*). The resources we have generated prove that it is possible to build a lexicon considering the words contexts as long as the corpus belongs to a specific domain. The generated lexicons achieve a competitive score in comparison to external general-purpose lexicons, even having much fewer words. Also, as it was expected, our sentiment lexicons induced from specific domains produced better results than those induced from general domain corpora, in which case a general-purpose lexicon works slightly better.

The context-graph proposal has value in itself. We have mentioned only one valence propagation mechanism but other algorithms over it could be applied. Also, the context-graph can be used for other problems in which the words context must be taken into account.

As future work, concerning the generation of the specific domain lexicons, we propose to include aspect information using word’s local context, in order to better solve the problem of valence ambiguity.

## References

- Azzinnari, A. and Martínez, A. (2016). Representación de palabras en espacios de vectores. Proyecto de grado, UdelAR (Universidad de la República).
- Boldrini, E., Balahur, A., Martinez-Barco, P., and Montoyo, A. (2009). Emotiblog: an annotation scheme for emotion detection and analysis in non-traditional textual genres. *Proceedings of the 1st Workshop on Opinion Mining and Sentiment Analysis, WOMSA09*, pages 491–497.
- Chen, Y. and Skiena, S. (2014). Building sentiment lexicons for all major languages. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 383–389.
- Cruz Mata, F., Troyano Jiménez, J. A., Enríquez de Salamanca Ros, F., and Ortega Rodríguez, F. J. (2008). Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español. *Procesamiento del lenguaje natural. N. 41*.
- Dubian, L. (2013a). Analisis de sentimientos sobre un corpus en español: Experimentación con un caso de estudio. *Jornadas Argentinas de Informática*.
- Dubian, L. (2013b). Procesamiento de lenguaje natural en sistemas de análisis de sentimientos. Tesis de grado, UBA (Universida de Buenos Aires).
- Esuli, A. and Sebastiani, F. (2007). Pageranking wordnet synsets: An application to opinion mining. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pages 424–431. Association for Computational Linguistics.
- Ghag, K. and Shah, K. (2014). SENTITFIDF sentiment classification using relative term frequency inverse document frequency. *International Journal of Advanced Computer Science and Applications(IJACSA)*, 5(2).
- Hamilton, W. L., Clark, K., Leskovec, J., and Jurafsky, D. (2016). Inducing domain-specific sentiment lexicons from unlabeled corpora. *CoRR*, abs/1606.02820.
- Hatzivassiloglou, V. and McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, pages 174–181, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Jiménez-Zafra, S. M., Martín-Valdivia, M. T., Molina-González, M. D., and na López, L. A. U. (2009). Corpus COAR, with opinions about restaurants. Available in: <http://sinai.ujaen.es/coar/> (last access 02/11/17).
- Jiménez-Zafra, S. M., Martín-Valdivia, M. T., Molina-González, M. D., and Ureña López, L. A. (2017). Corpus annotation for aspect based sentiment analysis in medical domain. *International Artificial Intelligence in Medicine Conference (AIME)*.
- Jurafsky, D. and Martin, J. H. (2017). *Speech and Language Processing*. (Book Draft), 3 edition.
- Kamps, J., Marx, M., Mokken, R., and De Rijke, M. (2004). Using wordnet to measure semantic orientations of adjectives. *Institute for Logic, Language and Computation (ILLC), FNWI*.
- Levin, B. (1991). Building a lexicon: The contribution of linguistics. *International Journal of Lexicography*, 4(3):205.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan-Claypool.
- Martí Antonín, M. A., Taulé Delor, M., Nofre, M., Marsó, L., Martín Valdivia, M. T., and Jiménez Zafra, S. M. (2016). La negación en español: análisis y tipología de patrones de negación. *Procesamiento del Lenguaje Natural, Revista n 57*.

- Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T., and Urena-López, L. A. (2014). Cross-domain sentiment analysis using spanish opinionated words. In *International Conference on Applications of Natural Language to Data Bases/Information Systems*, pages 214–219. Springer.
- Mordecki, G., Kremer, F., and Dufort y Alvarez, G. (2015). Estudio de reputación a partir de comentarios extraídos de redes sociales. Proyecto de grado, Udelar (Universidad de la República).
- Mori, M., Tambucho, M., and Cardozo, D. (2016). Estudio de reputación a partir de comentarios extraídos de redes sociales. Proyecto de grado, Udelar (Universidad de la República).
- Muhammad, A., Wiratunga, N., Lothian, R., and Glassey, R. (2013). Contextual sentiment analysis in social media using high-coverage lexicon. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pages 79–93. Springer.
- Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2:1–158.
- Plaza-Del-Arco, F., Martín-Valdivia, M., Zafra, S. M., González, M. D., and Martínez-Cámara, E. (2016). Copos: Corpus of patient opinions in spanish. application of sentiment analysis techniques. *SINAI Research Group, Universidad de Jaen*, 57:83–90.
- Polanyi, L. and Zaenen, A. (2006). Contextual valence shifters, chapter 1. computing attitude and affect in text: Theory and applications.
- Rosá, A., Chiruzzo, L., Etcheverry, M., and Castro, S. (2017). RETUYT in TASS 2017: Sentiment analysis for spanish tweets using SVM and CNN. In *Proceedings of TASS 2017: Workshop on Sentiment Analysis at SEPLN co-located with 33rd SEPLN Conference*, volume 1896. CEUR Workshop Proceedings.
- Salvia, D. and Mechulam, N. (2018). Construcción de léxicos dinámicos para el análisis de sentimientos. Proyecto de grado, Udelar (Universidad de la República).
- Santos, A. P., Oliveira, H. G., Ramos, C., and Marques, N. C. (2012). A bootstrapping algorithm for learning the polarity of words. In *International Conference on Computational Processing of the Portuguese Language*, pages 229–234. Springer.
- Saralegi, X. and San Vicente, I. n. (2013). Elhuyar at tass 2013. In *Proceedings of XXIX Congreso de la Sociedad Española de Procesamiento de Lenguaje Natural, Workshop on Sentiment Analysis at SEPLN (TASS2013)*.
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2):267–307.
- Turney, P. D. and Littman, M. L. (2002). Unsupervised learning of semantic orientation from a hundred-billion-word corpus. *CoRR*, cs.LG/0212012.
- Vilares, D., Thelwall, M., and Alonso, M. A. (2015). The megaphone of the people? Spanish SentiStrength for real-time analysis of political tweets. *J. Inf. Sci.*, 41(6):799–813.
- Wu, Y. and Wen, M. (2010). Disambiguating dynamic sentiment ambiguous adjectives. In *Proceedings of the 23rd International Conference on Computational Linguistics, COLING '10*, pages 1191–1199, Stroudsburg, PA, USA. Association for Computational Linguistics.

## 6 Appendix

In this section we show some examples extracted from the corpora described in Table 1.

- **Corpus APPS (Dubian, 2013b)** - Rank: NEGATIVO - Compre un galaxy s3 ... Una estafa. Nunca responde el servicio de atención al cliente. Compre un galaxy s3 que prometía 50gb durante 2 años y ha pasado ya un año y pese a que reclame mil veces la respuesta nunca llego. Sansung y dropbox estafadores.
- **Corpus MOV (Cruz Mata et al., 2008)** - Rank: 1 - La alianza del mal. Resumen: Silicona, esteroides, pactos demoníacos y otras basuras habituales son la base que sustentan esta aberración. De vergüenza. Comentario: Una fiesta llena de excesos, rubias despampanantes, musculitos por doquier, algún que otro muerto. nada nuevo. La alianza del mal es el nombre de este Thriller sobrenatural que narra ... Porque la única realidad es que La alianza del mal es un desecho cinematográfico, una insufrible sucesión de imágenes cuya única virtud es la de no durar demasiado. De vergüenza.
- **Corpus HOTEL (Molina-González et al., 2014)** - Rank: 5 - Un hotel digno de mención! Como bien les comenté a los propietarios a la hora de abandonar el hotel, no dudaré un momento en recomendar una y otra vez el Hotel Albero de Granada. Su situación respecto del centro de Granada no es la mejor, pero para nuestros propósitos era perfecto (escapada de fin de semana con visita a la Alhambra). Se encuentra en la carretera de paso a Sierra Nevada y muy cercano a la Alhambra. Por la zona se puede encontrar aparcamiento y este se encuentra en una zona segura y tranquila. Los parkings del centro de Granada que nos recomendaron en el hotel fueron lo que nos dijeron (nada caros) y pudimos movernos por el centro perfectamente desde allí. Las habitaciones muy limpias y las camas confortables. El desayuno fue espectacular. Ya teníamos buenas referencias de este maravilloso hotel de una estrella (que para mí que viajo constantemente son más) pero ha superado con creces nuestras expectativas. Si vuelvo a Granada no dudaré en hospedarme en el mismo hotel. Muchas gracias por todo!
- **Corpus MED1 (Plaza-Del-Arco et al., 2016)** - Rank: 5 - Todo un acierto. Encantado de ser su paciente. Siempre se toma su tiempo para saber tu caso y resolver tus dudas.
- **Corpus MED2 (Jiménez-Zafra et al., 2017)** - Rank: NEGATIVE - En lugar de calmar el dolor por una infección dental, me da efectos secundarios desagradables y una horrible migraña al día siguiente. El Tramadol 5 es mucho mejor.
- **Corpus MISC1 (Vilares et al., 2015)** - Rank: 4 - @Chunjiram @LjoeRam Son tan adorables, ajkijshdkjhgeftws.
- **Corpus MISC2 (Martí Antonín et al., 2016)** - Rank: NEGATIVO - No lo he utilizado mucho. Actualmente (agosto 2002) tiene 22.800 km. Con solo 22.000, avería del embrague 70.000 ptas. Que dice Seat al cliente? ha pasado la garantía no se hacen cargo, a pesar, que el mecánico me dijo que NO se debía a mal uso por mi parte. Lo arreglo. Resulta que el pedal, una vez arreglado, no sube del todo. Me lo vuelven a arreglar y actualmente sigue mal. Tiene un ruido en los bajos que parece un concierto. No me atrevo a viajar.
- **Corpus MISC3 (Mori et al., 2016)** - Rank: POSITIVO - Uruguay se floreció en casa y está un poco más cerca de Rusia 2018 <https://t.co/mu5UJL4P> SOYCELESTE @SoyCeleste <https://t.co/tt8bpbEP2y>.
- **Corpus JOUR1 (Mordecki et al., 2015)** - Rank: NEGATIVO - anuncio de Batlle sobre juicio a ex socios del Comercial. Batlle dijo que cuando recuerda la conversación con Mulford, del Credit Suisse, siente "ganas de matar".
- **Corpus JOUR2 (Boldrini et al., 2009)** - Rank: POSITIVE- Ahora vamos a lo nuestro, el tema del calentamiento global es uno de los más interesantes y conocer sobre el protocolo de Kyoto se hace sumamente necesario. Es por eso que en esta oportunidad les traigo un informe que publicaron los 147 amigos de la BBC Mundo que espero te sirva para conocer un poco más sobre esta situación que a todos nos involucra.
- **Corpus REST1 (Jiménez-Zafra et al., 2009)** - Rank: 1,3 - Caro para el conjunto ofrecido. Aunque los alimentos estaban bien cocinados no me pareció coherente el precio que se paga con lo ofrecido en general. Desde mi punto de vista, cuando pagas 45-50 persona hay muchas cosas que valorar y que deben ser coherentes con el precio: la forma de servir los platos, la cubertería, la mantelería, etc. Todo debe ser coherente con esos precios y en este restaurante no lo es: manteles personales de plástico, vasos de tubo para refrescos, presentación del producto anticuado, etc. Creo que pagar esos precios para el conjunto que ofrecen no es apropiado sino más bien caro.
- **Corpus REST2 (Dubian, 2013a)** - Rank: NEGATIVO - Cada día esta peor, precios para turistas, la entrada casi incomible. El mozo trajo primero la ensalada, después las empanadas y la carne cuando se le ocurrió. Ya no hay tanta gente por algo será. Difícilmente regresemos fuimos 5 personas y todos coincidimos.