

# Universal, Unsupervised, Uncovered Sentiment Analysis

David Vilares, Carlos Gómez-Rodríguez and Miguel A. Alonso

Grupo LyS, Departamento de Computación, Universidade da Coruña

Campus de A Coruña s/n, 15071, A Coruña, Spain

{david.vilares, carlos.gomez, miguel.alonso}@udc.es

## Abstract

We present a novel unsupervised approach for multilingual sentiment analysis driven by compositional syntax-based rules. On the one hand, we exploit some of the main advantages of unsupervised algorithms: (1) the interpretability of their output, in contrast with most supervised models, which behave as a black box and (2) their robustness across different corpora and domains. On the other hand, by introducing the concept of compositional operations and exploiting syntactic information in the form of universal dependencies, we tackle one of their main drawbacks: their rigidity on data that are differently structured depending on the language. Experiments show an improvement both over existing unsupervised methods, and over state-of-the-art supervised models when evaluating outside their corpus of origin. The system is freely available<sup>1</sup>.

## 1 Introduction

Semantic composition is a natural process for humans when understanding the sentiment of an opinion. In the sentence *‘He is not very handsome, but he has something that I really like’*, humans have the ability to infer that the word *‘very’* emphasizes *‘handsome’*, *‘not’* affects the whole expression *‘very handsome’*, and *‘but’* decreases the relevance of *‘He is not very handsome’* and increases the one of *‘he has something that I really like’*. Based on this, a human could justify a positive overall sentiment on that sentence.

Our main contribution is the introduction of the first universal and unsupervised model for compositional sentiment analysis (SA) driven by syntax-

based rules. We introduce a formalism for compositional operations, allowing the creation of arbitrarily complex rules to tackle relevant phenomena for SA, for any language and syntactic dependency annotation. A set of practical universal operations is evaluated on different corpora and languages. The model outperforms existing unsupervised approaches, and state-of-the-art compositional supervised models (Socher et al., 2013) on domain-transfer settings, and shows that the operations can be shared across languages, as they are defined using part-of-speech (PoS) tags and dependency types under the universal guidelines of (Petrov et al., 2011; McDonald et al., 2013).

## 2 Related work

A naïve approach to emulate the comprehension of the meaning of multiword phrases for SA consists in using  $n$ -grams with  $n > 1$  (Pang et al., 2002). The approach is limited by the curse of dimensionality, although crawling data from the target domain can help to reduce that problem (Kiritchenko et al., 2014). Joshi and Penstein-Rosé (2009) went one step forward and proposed generalized dependency triplets as features for subjectivity detection, capturing non-local relations. Socher et al. (2012) modeled a recursive neural network that learns compositional vector representations for phrases and sentences of arbitrary syntactic type and length. Socher et al. (2013) presented an improved recursive deep model for SA over dependency trees, and trained it on a sentiment treebank tagged using Amazon Mechanical Turk, pushing the state of the art up to 85.4% on the Pang and Lee (2005) dataset. Kalchbrenner et al. (2014) showed how convolutional neural networks (CNN) can be used for semantic modeling of sentences. The model implicitly captures local and non-local relations without the need of a parse

<sup>1</sup><http://grupolys.org/software/UUUSA/>

tree. It can be adapted for any language, as far as enough data is available. Severyn and Moschitti (2015) showed the effectiveness of a CNN in a SemEval SA shared task (Rosenthal et al., 2015), although crawling tens of millions of messages was first required to achieve state-of-the-art results.

In spite of being powerful and accurate, supervised approaches also present drawbacks. Firstly, they behave as a black box. Secondly, they do not perform so well on domain transfer applications (Aue and Gamon, 2005; Pang and Lee, 2008). Finally, feature and hyper-parameter engineering can be time and resource costly options.

When these limitations need to be addressed, unsupervised (rule-based) approaches are useful. In this line, Turney (2002) proposed an unsupervised learning algorithm to calculate the semantic orientation (SO) of a word. Taboada et al. (2011) presented a lexical rule-based approach to handle relevant linguistic phenomena such as intensification, negation, ‘*but*’ clauses and *irrealis*. Thelwall et al. (2012) released SentiStrength, a multilingual unsupervised system for micro-text SA that handles negation and intensification, among other web linguistic phenomena. Regarding syntax-based approaches, the few described in the literature are language-dependent. Jia et al. (2009) define a set of syntax-based rules for handling negation in English. Vilares et al. (2015a) propose a syntactic SA method, but limited to Spanish reviews and Ancora trees (Taulé et al., 2008).

In brief, most unsupervised approaches are language-dependent, and those that can manage multilinguality, such as SentiStrength, cannot apply semantic composition.

### 3 Unsupervised Compositional SA

#### 3.1 Dependency graphs

Let  $w=w_1, \dots, w_n$  be a sentence, where each word occurrence  $w_i \in W$  is assigned a PoS tag  $t_i \in T$ .

**Definition 1.** A **dependency tree** for  $w$  is an edge-labeled directed tree  $T = (V, E)$  where  $V = \{0, 1, 2, \dots, n\}$  is the set of nodes and  $E = V \times D \times V$  is the set of labeled arcs. Each arc, of the form  $(i, d, j)$ , corresponds to a syntactic **dependency** between the words  $w_i$  and  $w_j$ ; where  $i$  is the index of the **head** or **parent** word,  $j$  is the index of the **dependent** or **child** word and  $d$  is the **dependency type** representing the kind of syntactic relation between them. Following standard practice, we use node 0 as a dummy root node that acts

as the head of the syntactic root(s) of the sentence.

We will write  $i \xrightarrow{d} j$  as shorthand for  $(i, d, j) \in E$  and we will omit the dependency types when they are not relevant. Given a dependency tree  $T = (V, E)$ , and a node  $i \in V$ , we define a set of functions to obtain the context of node  $i$ :

- $ancestor_T(i, \delta) = \{k \in V : \text{there is a path of length } \delta \text{ from } k \text{ to } i \text{ in } T\}$ , i.e., the singleton set containing the  $\delta$ th ancestor of  $i$  (or the empty set if there is no such node),
- $children_T(i) = \{k \in V \mid i \rightarrow k\}$ , i.e., the set of children of node  $i$ ,
- $lm\text{-}branch_T(i, d) = \min\{k \in V \mid i \xrightarrow{d} k\}$ , i.e., the set containing the leftmost among the children of  $i$  whose dependencies are labeled  $d$  (or the empty set if there is no such node).

#### 3.2 Operations for compositional SA

Our compositional SA system will associate an SO value  $\sigma_i$  to each node  $i$  in the dependency tree of a sentence, representing the SO of the subtree rooted at  $i$ . The system will use a set of compositional operations to propagate changes to the semantic orientations of the nodes in the tree. Once all the relevant operations have been executed, the SO of the sentence will be stored as  $\sigma_0$ , i.e., the semantic orientation of the root node.

A compositional operation is triggered when a node in the tree matches a given condition (related to its associated PoS tag, dependency type and/or word form); and it is applied to a scope of one or more nodes calculated from the trigger node by ascending a number of levels in the tree and then applying a scope function. More formally, we define our operations as follows:

**Definition 2.** Given a dependency tree  $T(V, E)$ , a **compositional operation** is a tuple  $o = (\tau, C, \delta, \pi, S)$  such that:

- $\tau : \mathbb{R} \rightarrow \mathbb{R}$  is a **transformation function** to apply on the SO ( $\sigma$ ) of nodes,
- $C : V \rightarrow \{\text{true}, \text{false}\}$  is a predicate that determines whether a node in the tree will **trigger** the operation,
- $\delta \in \mathbb{N}$  is a number of levels that we need to ascend in the tree to calculate the scope of  $o$ ,
- $\pi$  is a priority that will be used to break ties when several operations coincide on a given node, and

- $S$  is a scope calculation function that will be used to determine the nodes affected by the operation.

In practice, our system defines  $C(i)$  by means of sets of words, tags and/or dependency types such that the operation will be triggered if  $w_i$ ,  $t_i$  and/or the head dependency of  $i$  are in those sets. Compositional operations where  $C(i)$  is defined using universal tags and dependency types only are universal and can be used across languages.

We propose two options for the transformation function  $\tau$ :

- $shift_\alpha(x) = \begin{cases} x - \alpha & \text{if } x > 0 \\ x + \alpha & \text{if } x < 0 \end{cases}$  where  $\alpha$  is the shifting factor and  $\alpha, x \in \mathbb{R}$ .
- $weighting_\beta(x) = x \times (1 + \beta)$  where  $\beta$  is the weighting factor and  $\beta, x \in \mathbb{R}$ .

The scope calculation function,  $S$ , allows us to calculate the nodes of  $T$  whose SO is affected by the transformation  $\tau$ . For this purpose, if the operation was triggered by a node  $i$ , we apply  $S$  to  $ancestor_T(i, \delta)$ , i.e., the  $\delta$ th ancestor of  $i$  (if it exists), which we call the **destination node** of the operation. The proposed scopes are as follows (see also Figure 1):

- *dest* (destination node): The transformation  $\tau$  is applied directly to the SO of  $ancestor_T(i, \delta)$  (see Figure 1.a).
- *lm-branch<sup>d</sup>* (branch of  $d$ ): The affected nodes are  $lm-branch_T(ancestor_T(i, \delta), d)$  (see Figure 1.b).
- *rc<sup>n</sup>* ( $n$  right children):  $\tau$  affects the SO of the  $n$  smallest indexes of  $\{j \in children_T(ancestor_T(i, \delta)) \mid j > i\}$  (see Figure 1.c).
- *lc<sup>n</sup>* ( $n$  left children): The transformation affects the  $n$  largest elements of  $\{j \in children_T(ancestor_T(i, \delta)) \mid j < i\}$  (see Figure 1.d).
- *subjr* (first subjective right branch): The affected node is  $\min\{j \in children_T(ancestor_T(i, \delta)) \mid j > i \wedge \sigma_j \neq 0\}$  (see Figure 1.e).
- *subjl* (first subjective left branch): The affected node is  $\max\{j \in children_T(ancestor_T(i, \delta)) \mid j < i \wedge \sigma_j \neq 0\}$  (see Figure 1.f).

Compositional operations can be defined for any language or dependency annotation criterion. While it is possible to add rules for language-specific phenomena if needed (see § 3.3), in this paper we focus on universal rules to obtain a truly multilingual system.<sup>2</sup>

### 3.3 An algorithm for unsupervised SA

To execute the operations and calculate the SO of each node in the dependency tree of the sentence, we start by initializing the SO of each word using a subjective lexicon, as traditional unsupervised approaches do (Turney, 2002). Some options to obtain multilingual subjectivity lexica are SentiStrength (subjective data for up to 34 languages) or the Chen and Skiena (2014) approach, which introduced a method for building sentiment lexicons for 136 languages. Our implementation supports the lexicon format of SentiStrength, which can be plugged directly into the system. Additionally, we provide the option to create different dictionary entries depending on PoS tags to avoid conflicts between homonymous words (e.g. ‘*I’m fine*’ versus ‘*They gave me a fine*’).

Then, we traverse the parse tree in postorder, applying Algorithm 1 to update semantic orientations when visiting each node  $i$ . In this algorithm,  $O$  is the set of compositional operations defined in our system,  $A_i$  is a priority queue of the compositional operations to be applied at node  $i$  (because  $i$  is their destination node); and  $Q_i$  is another priority queue of compositional operations to be queued for upper levels at node  $i$  (as  $i$  is not yet their destination node).  $\oplus$  defines the operation to merge two priority queues, *push* inserts  $o$  in a priority queue and *pop* pulls the operation with the highest priority (ties are broken by giving preference to the operation that was queued earlier).

At a practical level, the set of compositional operations are specified using a simple XML file:

- **<forms>**: Indicates the tokens to be taken into account for the condition  $C$  that triggers the operation. Regular expressions are supported.

<sup>2</sup>Apart from universal dependencies and PoS tags, the only extra information used by our rules is a short list of negation words, intensifiers, adversative conjunctions and words introducing conditionals (like the English “if” or “would”). While this information is language-specific, it is standardly included in multilingual sentiment lexica which are available for many languages (§ 3.3), so it does not prevent our system from working on a wide set of languages without any adaptation.

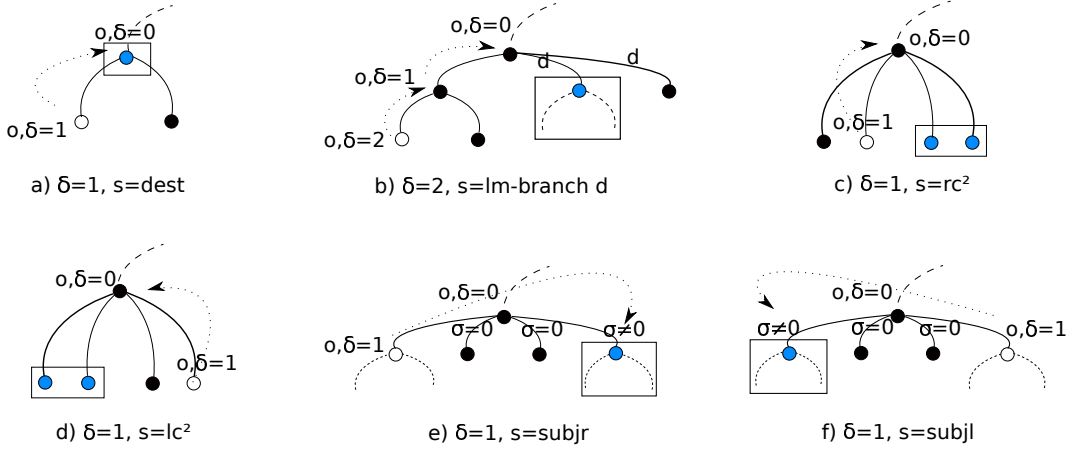


Figure 1: Graphical representation of the proposed set of influence scopes  $S$ .  $\bigcirc$  indicates the node that triggers an operation  $o$ ,  $\square$  the nodes to which it is applied.

#### Algorithm 1 Compute SO of a node

```

1: procedure COMPUTE( $i, O, T$ )
2:    $A_i \leftarrow \square$ 
3:    $Q_i \leftarrow \square$ 
4:    $\triangleright$  Enqueue operations triggered by node  $i$ :
5:   for  $o = (\tau, C, \delta, \pi, S)$  in  $O$  do
6:     if  $C(i)$  then
7:       if  $\delta > 0$  then
8:          $push((\tau, C, \delta, \pi, S), Q_i)$ 
9:       else
10:         $push((\tau, C, \delta, \pi, S), A_i)$ 
11:    $\triangleright$  Enqueue operations coming from child nodes:
12:   for  $c$  in  $children_T(i)$  do
13:     for  $o = (\tau, C, \delta, \pi, S)$  in  $Q_c$  do
14:       if  $\delta - 1 = 0$  then
15:          $push((\tau, C, \delta - 1, \pi, S), A_i)$ 
16:       else
17:          $push((\tau, C, \delta - 1, \pi, S), Q_i)$ 
18:    $\triangleright$  Execute operations that have reached their destination
19:   node:
20:   while  $A_i$  is not empty do
21:      $o = (\tau, C, \delta, \pi, S) \leftarrow pop(A_i)$ 
22:     for  $j$  in  $S(i)$  do
23:        $\sigma_j \leftarrow \tau(\sigma_j)$ 
24:    $\triangleright$  Join the SOs for node  $i$  and its children:
25:    $\sigma_i \leftarrow \sigma_i + \sum_{c \in children_T(i)} \sigma_c$ 

```

- **<dependency>**: Indicates the dependency types taken into account for  $C$ .
- **<postags>**: Indicates the PoS tags that must match to trigger the rule.
- **<rule>**: Defines the operation to be executed when the rule is triggered.
- **<levelsup>**: Defines the number of levels from  $i$  to spread before applying  $o$ .
- **<priority>**: Defines the priority of  $o$  in case than more than one operation needs to be applied over  $i$  (a larger number implies a bigger priority).

## 4 NLP tools for universal unsupervised SA

The following resources serve us as the starting point to carry out state-of-the-art universal, unsupervised and syntactic sentiment analysis.

The system by Gimpel et al. (2011) is used for tokenizing. Although initially intended for English tweets, we have observed that it also performs robustly for many other language families (Romance, Slavic, etc.).

For part-of-speech tagging we rely on the free distribution of the Toutanova and Manning (2000) tagger. Dependency parsers are built using Malt-Parser (Nivre et al., 2007) and MaltOptimizer (Ballesteros and Nivre, 2012). We trained a set of taggers and parsers for different languages using the universal tag and dependency sets (Petrov et al., 2011; McDonald et al., 2013). Table 1 shows their performance under standard metrics. The tagging<sup>3</sup> and parsing<sup>4</sup> models are also available.

## 5 Defining compositional operations

We presented above a formalism to define arbitrarily complex compositional operations for unsupervised SA over a dependency tree. In this section, we show the definition of the most important rules that we used to evaluate our system. In practical terms, this implies studying how syntactic constructions that modify the sentiment of an ex-

<sup>3</sup><http://grupolys.org/software/TAGGERS/universal-tag-sets/monolingual/>

<sup>4</sup><http://www.grupolys.org/software/PARSERS/universal-tag-sets/monolingual/>



Language	Tagger		Parser	
	Acc.	Acc. unk	LAS	UAS
English	98.12	91.81	89.36	91.02
Spanish	96.05	84.15	80.60	84.75
German	95.13	91.89	78.27	84.03
Italian	97.83	92.07	84.40	87.54
French	95.89	87.18	79.29	84.47
Portuguese	97.72	93.44	84.88	87.39
Indonesian	95.67	93.19	77.69	84.47

Table 1: Performance of the taggers and parsers provided together with our system on the corresponding McDonald et al. (2013) test set.

pression are represented in the annotation formalism used for the training of the dependency parser, in this case, Universal Dependencies. We are using examples following those universal guidelines, since they are available for more than 40 languages and, as shown in § 6, the same rules can be competitive across different languages.

## 5.1 Intensification

Intensification amplifies or diminishes the sentiment of a word or phrase. Simple cases of this phenomenon can be ‘*I have **huge** problems*’ or ‘*This is **a bit** dissappointing*’. Traditional lexicon-based methods handle most of these cases with simple heuristics (e.g. amplifying or diminishing the sentiment of the word following an intensifier). However, ambiguous cases might appear where such lexical heuristics are not sufficient. For example, ‘*huge*’ can be a subjective adjective introducing its own SO (e.g. ‘*The house is huge*’), but also an amplifier when it modifies a subjective noun or adjective (e.g. ‘*I have huge problems*’, where it makes ‘*problems*’ more negative).

Universal compositional operations overcome this problem without the need of any heuristic. A dependency tree already shows the behavior of a word within a sentence thanks to its dependency type, and it shows the role of a word independently of the language. Figure 2 shows graphically how universal dependencies represent the cases discussed above these lines. Formally, the operation for these forms of intensification is:  $(weighting_{\beta}, w \in \text{intensifiers} \wedge t \in \{\text{ADV}, \text{ADJ}\} \wedge d \in \{\text{advmod}, \text{amod}, \text{nmod}\}, 1, 3, dest \cup lm\text{-branch}^{acomp})$ , with the value of  $\beta$  depending on the strength of the intensifier as given by the sentiment lexicon.

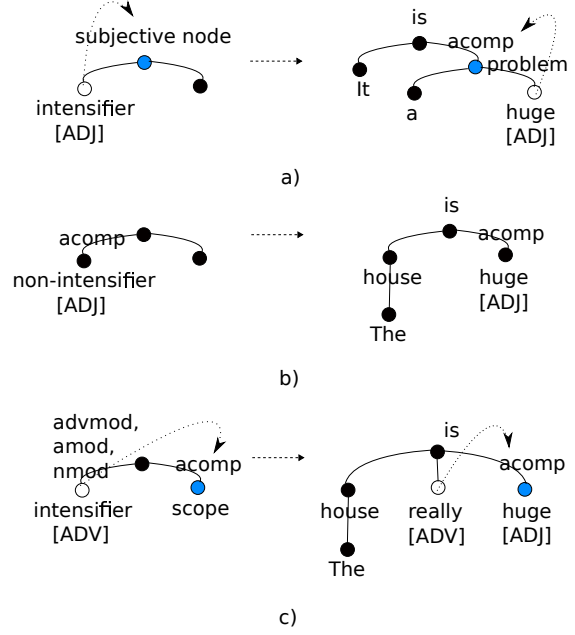


Figure 2: Skeleton for intensification compositional operations (2.a, 2.c) and one case without intensification (2.b), together with examples annotated with universal dependencies.

### 5.1.1 ‘But’ clauses

Compositional operations can also be defined to manage more challenging cases, such as clauses introduced by ‘*but*’, considered as a special case of intensification by authors such as Brooke et al. (2009) or Vilares et al. (2015a). It is assumed that the main clause connected by ‘*but*’ becomes less relevant for the reader (e.g. ‘*It is expensive, **but** I love it*’). Figure 3 shows our proposed composition operation for this clause, formally:  $(weighting_{\beta}, w \in \{\text{but}\} \wedge t \in \{\text{CONJ}\} \wedge d \in \{\text{cc}, 1, 1, \text{subj}\})$  with  $\beta = -0.25$ . Note that the priority of this operation ( $\pi = 1$ ) is smaller than that of intensification ( $\pi = 3$ ), since we first need to process intensifiers, which are local phenomena, before resolving adversatives, which have a larger scope.

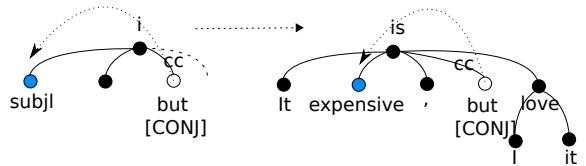


Figure 3: Skeleton for ‘*but*’ compositional operation illustrated with one example according to universal dependencies.

## 5.2 Negation

Negation is one of the most challenging phenomena to handle in SA, since its semantic scope can be non-local (e.g. ‘*I do not plan to make you suffer*’). Existing unsupervised lexical approaches are limited to consider a snippet to guess the scope of negation. Thus, it is likely that they consider as a part of the scope terms that should not be negated from a semantic point of view. Dependency types help us to determine which nodes should act as negation and which should be its scope of influence. For brevity, we only illustrate some relevant negation cases and instructional examples in Figure 4. Formally, the proposed compositional operation to tackle most forms of negation under universal guidelines is:  $(shift_{\alpha}, w \in \text{negations} \wedge t \in U \wedge d \in \{\text{neg}\}, 1, 2, \text{dest} \cup \text{lm-branch}^{\text{attr}} \cup \text{lm-branch}^{\text{acomp}} \cup \text{subj})$ , where  $U$  represents the universal tag set. The priority of negation ( $\pi = 2$ ) is between those of intensification and ‘*but*’ clauses because its scope can be non-local, but it does not go beyond an adversative conjunction.

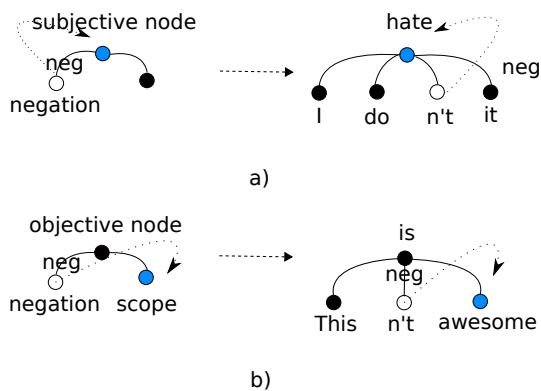


Figure 4: Skeleton for negation compositional operations illustrated together with one example.

## 5.3 Irrealis

Irrealis denotes linguistic phenomena used to refer to non-factual actions, such as conditional, subjunctive or desiderative sentences (e.g. ‘*He would have died if he hadn’t gone to the doctor*’). It is a very complex phenomenon to deal with, and systems are usually unable to tackle this issue or they simply define rules to ignore sentences containing a list of irrealis stop-words (Taboada et al., 2011). We do not address this phenomenon in detail in this study, but only propose a rule to deal with ‘*if*’ constructions (e.g. ‘*if I die [...]*’ or ‘*if you are*

*happy [...]*’, considering that the phrase that contains it should be ignored from the final computation. Formally:  $(weighting_{\beta}, w \in \{\text{if}\} \wedge t \in U \wedge d \in \{\text{mark}\}, 2, 3, \text{dest} \cup \text{subj})$ . Its graphical representation would be very similar to intensification (see Figures 1 a) and e)).

## 5.4 Discussion

Figure 5 represents an analysis of our introductory sentence ‘*He is not very handsome, but he has something that I really like*’, showing how compositional operations accurately capture semantic composition.

It is hard to measure the coverage of our rules and the potential of these universal compositional operations, since it is possible to define arbitrarily complex operations for as many relevant linguistic phenomena as wished. In this line, Poria et al. (2014) define a set of English sentic patterns to determine how sentiment flows from concept to concept in a variety of situations (e.g. relations of complementation, direct nominal objects, relative clauses, ...) over a dependency tree following the De Marneffe and Manning (2008) guidelines.

## 6 Experimental results

We compare our algorithm with respect to existing approaches on three languages: English, Spanish and German. The availability of corpora and other unsupervised SA systems for English and Spanish enables us to perform a richer comparison than in the case of German, where we only have an *ad-hoc* corpus.

We compare our algorithm with respect to two of the most popular and widely used unsupervised systems: (1) SO-CAL (Taboada et al., 2011), a language-dependent system available for English and Spanish guided by lexical rules at the morphological level, and (2) SentiStrength, a multilingual system that does not apply any PoS tagging or parsing step in order to be able to do multilingual analysis, relying instead on a set of subjectivity lexica, snippet-based rules and treatment of non-grammatical phenomena (e.g. character replication). Additionally, for the Spanish evaluation, we also took into account the system developed by Vilares et al. (2015a), an unsupervised syntax-based approach available for Spanish but, in contrast to ours, heavily language-dependent.

For comparison against state-of-the-art supervised approaches, we consider the deep recursive

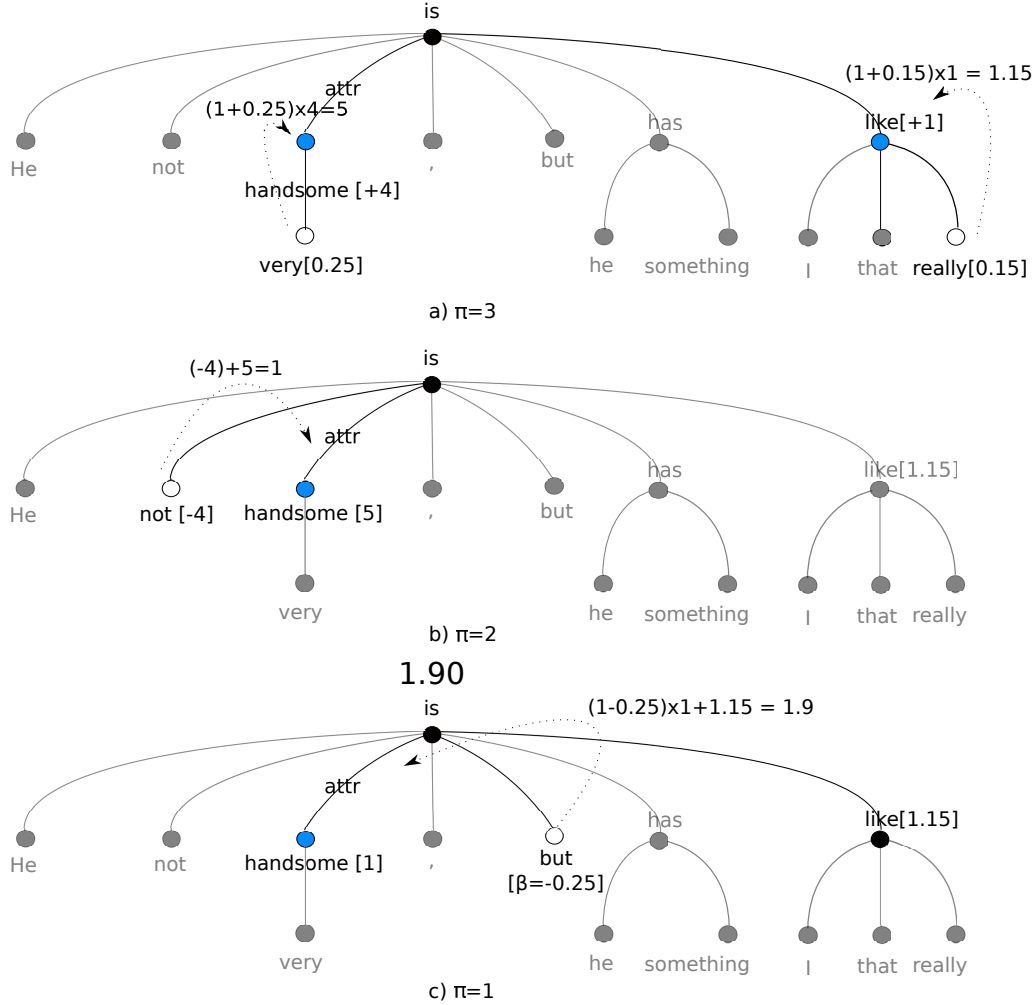


Figure 5: Analysis of a sentence applying universal unsupervised prediction. For the sake of clarity, the real post-order traversal is not illustrated. Instead we show an (in this case) equivalent computation by applying all operations with a given priority,  $\pi$ , at the same time, irrespective of the node. Semantic orientation, intensification and negation values are extracted from the dictionaries of Taboada et al. (2011). Phase a) shows how the intensification is computed on the branches rooted at 'handsome' and 'like'. Phase b) shows how the negation shifts the semantic orientation of the attribute (again, the branch rooted at 'handsome'). Phase c) illustrates how the clause 'but' diminishes the semantic orientation of the main sentence, in particular the semantic orientation of the attribute, the first left subjective branch of its head. Elements that are not playing a role in a specific phase appear dimmed. One of the interesting points in this example comes from illustrating how three different phenomena involving the same branch (the attribute 'handsome') are addressed properly thanks to the assigned  $\pi$ .

neural network presented by Socher et al. (2013), trained on a movie sentiment treebank (English). To the best of our knowledge, there are no semantic compositional supervised methods for Spanish and German.

Accuracy is used as the evaluation metric for two reasons: (1) it is adequate for measuring the performance of classifiers when the chosen corpora are balanced and (2) the selected systems for comparison also report their results using this metric.

## 6.1 Resources

We selected the following standard English corpora for evaluation:

- Taboada and Grieve (2004) corpus: A general-domain collection of 400 long reviews (200 positive, 200 negative) about hotels, movies, computers or music among other topics, extracted from epinions.com.
- Pang and Lee (2004) corpus: A corpus of 2 000 long movie reviews (1 000 positive, 1 000 negative).
- Pang and Lee (2005) corpus: A corpus of short movie reviews (sentences). In particular, we used the test split used by Socher et al. (2013), removing the neutral ones, as they did, for the binary classification task (total: 1 821 subjective sentences).

To show the universal capabilities of our system we include an evaluation for Spanish using the corpus presented by Brooke et al. (2009) (200 positive and 200 negative long reviews from ciao.es). For German, we rely on a dataset of 2 000 reviews (1 000 positive and 1 000 negative reviews) extracted from Amazon.

As subjectivity lexica, we use the same dictionaries used by SO-CAL for both English and Spanish. For German, we use the German SentiStrength dictionaries (Momtazi, 2012) instead, as Brooke et al. (2009) dictionaries are not available for languages other than Spanish or English.

## 6.2 Comparison to unsupervised approaches

Table 2 compares the performance of our model with respect to SentiStrength<sup>5</sup> and SO-CAL on

<sup>5</sup>We used the default configuration, which already applies many optimizations. We set the length of the snippet between a negator and its scope to 3, based on empirical evaluation, and applied the configuration to compute sentiment on long reviews.

the Taboada and Grieve (2004) corpus. With respect to SO-CAL, results show that our handling of negation and intensification provides better results (outperforming SO-CAL by 3.25 percent points overall). With respect to SentiStrength, our system achieves better performance on long reviews.

Table 3 compares these three unsupervised systems on the Pang and Lee (2004) corpus, showing the robustness of our approach across different domains. Our system again performs better than SO-CAL for negation and intensification (although it does not behave as well when dealing with irrealis, probably due to the need of more complex compositional operations to handle this phenomenon), and also better than SentiStrength on long movie reviews.

Rules	SentiStrength	SO-CAL	Our system
Baseline	N/A	65.50	65.00
+negation	N/A	67.75	<b>71.75</b>
+intensification	66.00	69.25	<b>74.25</b>
+irrealis	N/A	71.00	<b>73.75</b>

Table 2: Accuracy (%) on the Taboada and Grieve (2004) corpus. We only provide one column for SentiStrength since we are using the standard configuration for English (which already includes negation and intensification functionalities).

Rules	SentiStrength	SO-CAL	Our system
Baseline	N/A	<b>68.05</b>	67.77
+negation	N/A	70.10	<b>71.85</b>
+intensification	56.90	73.47	<b>74.00</b>
+irrealis	N/A	<b>74.95</b>	74.10

Table 3: Accuracy (%) on Pang and Lee (2004) test set.

Table 4 compares the performance of our universal approach on a different language (Spanish) with respect to: Spanish SentiStrength (Vilares et al., 2015b), the Spanish SO-CAL (Brooke et al., 2009) and a syntactic language-dependent system inspired on the latter (Vilares et al., 2015a). We used exactly the same set of compositional operations as used for English (only changing the list of word forms for negation, intensification and ‘but’ clauses, as explained in §3.2). Our universal system again outperforms SentiStrength and SO-CAL in its Spanish version. The system also obtains results very similar to the ones reported by Vilares et al. (2015a), even though their system is language-dependent and the set of rules is fixed and written specifically for Spanish.



Rules	SentiStrength	SO-CAL	Our system	Vilares et al. (2015a)
Baseline	N/A	N/A	<b>63.00</b>	61.80
+negation	N/A	N/A	<b>71.00</b>	N/A
+intensification	73.00	N/A	74.25	<b>75.75</b>
+irrealis	N/A	74.50	<b>75.75</b>	N/A

Table 4: Accuracy (%) on the Spanish Brooke et al. (2009) test set.

In order to check the validity of our approach for languages other than English and Spanish, we have considered the case of German. It is worth noting that the authors of this article have no notions of German at all. In spite of this, we have been able to create a state-of-the-art unsupervised SA system by integrating an existing sentiment lexicon into the framework that we propose in this article.

We use the German SentiStrength system (Momtazi, 2012) for comparison. The use of the German SentiStrength dictionary allows us to show how our system is robust when using different lexica. Experimental results show an accuracy of 72.75% on the Amazon review dataset when all rules are included, while SentiStrength reports 69.95%. Again, adding first negation (72.05%) and then intensification (72.85%) as compositional operations produced relevant improvements over our baseline (69.85%). The results are comparable to those obtained for other languages, using a dataset of comparable size, reinforcing the robustness of our approach across different domains, languages, and base dictionaries.

### 6.3 Comparison to supervised approaches

Supervised systems are usually unbeatable on the test portion of the corpus with which they have been trained. However, in real applications, a sufficiently large training corpus matching the target texts in terms of genre, style, length, etc. is often not available; and the performance of supervised systems has proven controversial on domain transfer applications (Aue and Gamon, 2005).

Table 5 compares our universal unsupervised system to Socher et al. (2013) on a number of corpora: (1) the collection used in the evaluation of the Socher et al. system (Pang and Lee, 2005), (2) a corpus of the same domain, i.e., movies (Pang and Lee, 2004), and (3) the Taboada and Grieve (2004) collection. Socher et al.’s system provides sentence-level polarity classification with five possible outputs: *very positive*, *positive*, *neutral*, *negative*, *very negative*. Since the Pang and

Corpora	Socher et al. (2013)	Our system
<i>Origin corpus of Socher et al. (2013) model</i>		
Pang and Lee (2005)	<b>85.40</b>	75.01
<i>Other corpora</i>		
Taboada and Grieve (2004)	62.00	<b>73.75</b>
Pang and Lee (2004)	63.80	<b>74.10</b>

Table 5: Accuracy (%) on different corpora for Socher et al. (2013) and our system. On the Pang and Lee 2005 (Pang and Lee, 2005) collection, our detailed results taking into account different compositional operations were: 73.75 (baseline), 74.13 (+negation), 74.68 (+intensification) and 75.01 (+irrealis)

Lee (2004) and Taboada and Grieve (2004) corpora are collections of long reviews, we needed to collect the global sentiment of the text. We count the number of outputs of each class (*very positive* and *very negative* count double, *positive* and *negative* count one and *neutral* counts zero). We take the majority class, and in the case of a tie, it is classified as negative.<sup>6</sup>

The experimental results show that our approach obtains better results on corpora (2) and (3). It is worth mentioning that our unsupervised compositional approach outperformed the supervised model not only on an out-of-domain corpus, but also on another dataset of the same domain (movies) as the one where the neural network was trained and evaluated. This reinforces the usefulness of an unsupervised approach for applications that need to analyze a number of texts coming from different domains, styles or dates, but there is a lack of labeled data to train supervised classifiers for all of them. As expected, Socher et al. (2013) is unbeatable for an unsupervised approach on the test set of the corpus where it was trained. However, our unsupervised algorithm also performs very robustly on this dataset.

<sup>6</sup>These criteria were selected empirically. Assigning the positive class in the case of a tie was also tested, as well as not doubling the *very positive* and *very negative* output, but these settings produced similar or worse results with the (Socher et al., 2013) system.

## 7 Conclusions and future work

In this article, we have described, implemented and evaluated a novel model for universal and unsupervised sentiment analysis driven by a set of syntactic rules for semantic composition. Existing unsupervised approaches are purely lexical, their rules are heavily dependent on the language concerned or they do not consider any kind of natural language processing step in order to be able to handle different languages, using shallow rules instead.

To overcome these limitations, we introduce from a theoretical and practical point of view the concept of compositional operations, to define arbitrarily complex semantic relations between different nodes of a dependency tree. Universal part-of-speech tagging and dependency parsing guidelines make it feasible to create multilingual sentiment analysis compositional operations that effectively address semantic composition over natural language sentences. The system is not restricted to any corpus or language, and by simply adapting or defining new operations it can be adapted to any other PoS tag or dependency annotation criteria.

We have compared our universal unsupervised model with state-of-the-art unsupervised and supervised approaches. Experimental results show: (1) that our algorithm outperforms two of the most commonly used unsupervised systems, (2) the universality of the model's compositional operations across different languages and (3) the usefulness of our approach on domain-transfer applications, especially with respect to supervised models.

As future work, we plan to design algorithms for the automatic extraction of compositional operations that capture the semantic relations between tree nodes. We would also like to collect corpora to extend our evaluation to more languages, since collections that are directly available on the web are scarcer than expected. Additionally, the concept of compositional operations is not limited to generic SA and could be adapted for other tasks such as universal aspect extraction. Finally, we plan to adapt the Poria et al. (2014) sentic patterns as compositional operations, so they can be handled universally.

## Acknowledgments

This research is supported by the Ministerio de Economía y Competitividad (FFI2014-51978-C2). David Vilares is funded by the Ministerio

de Educación, Cultura y Deporte (FPU13/01180). Carlos Gómez-Rodríguez is funded by an Oportunus program grant (Xunta de Galicia). We thank Roman Klinger for his help in obtaining the German data.

## References

- A. Aue and M. Gamon. 2005. Customizing sentiment classifiers to new domains: A case study. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP)*.
- M. Ballesteros and J. Nivre. 2012. MaltOptimizer: an optimization tool for MaltParser. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 58–62. Association for Computational Linguistics.
- J. Brooke, M. Tofiloski, and M. Taboada. 2009. Cross-Linguistic Sentiment Analysis: From English to Spanish. In *Proceedings of RANLP 2009, Recent Advances in Natural Language Processing*, pages 50–54, Bovorets, Bulgaria, September.
- Y. Chen and S. Skiena. 2014. Building Sentiment Lexicons for All Major Languages. In *The 52nd Annual Meeting of the Association for Computational Linguistics. Proceedings of the Conference. Volume 2: Short Papers. ACL 2014*, pages 383–389, Baltimore, June. ACL.
- M. De Marneffe and C. D. Manning. 2008. The Stanford typed dependencies representation. In *Coling 2008: Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation*, pages 1–8. Association for Computational Linguistics.
- K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanagan, and N. A. Smith. 2011. Part-of-speech tagging for twitter: Annotation, features, and experiments. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 42–47. Association for Computational Linguistics.
- L. Jia, C. Yu, and W. Meng. 2009. The effect of negation on Sentiment Analysis and Retrieval Effectiveness. In *CIKM'09 Proceeding of the 18th ACM conference on Information and knowledge management*, pages 1827–1830, Hong Kong, November. ACM, ACM Press.
- M. Joshi and C. Penstein-Rosé. 2009. Generalizing dependency features for opinion mining. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, ACLShort '09, pages 313–316, Stroudsburg, PA, USA. Association for Computational Linguistics.

- N. Kalchbrenner, E. Grefenstette, and P. Blunsom. 2014. A Convolutional Neural Network for Modelling Sentences. In *The 52nd Annual Meeting of the Association for Computational Linguistics. Proceedings of the Conference. Volume 1: Long Papers*, pages 655–665, Baltimore, Maryland, USA. ACL.
- S. Kiritchenko, X. Zhu, and S. M. Mohammad. 2014. Sentiment Analysis of Short Informal Texts. *Journal of Artificial Intelligence Research*, 50(1):723–762, may.
- R. McDonald, J. Nivre, Y. Quirmbach-brundage, Y. Goldberg, D. Das, K. Ganchev, K. Hall, S. Petrov, H. Zhang, O. Täckström, C. Bedini, N. Castelló, and J. Lee. 2013. Universal Dependency Annotation for Multilingual Parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 92–97. Association for Computational Linguistics.
- S. Momtazi. 2012. Fine-grained german sentiment analysis on social media. In *LREC*, pages 1215–1220. Citeseer.
- J. Nivre, J. Hall, J. Nilsson, A. Chanev, G. Eryigit, S. Kübler, S. Marinov, and E. Marsi. 2007. Malt-Parser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(2):95–135.
- B. Pang and L. Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*, pages 271–278. Association for Computational Linguistics.
- B. Pang and L. Lee. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 115–124. Association for Computational Linguistics.
- B. Pang and L. Lee. 2008. *Opinion Mining and Sentiment Analysis*. now Publishers Inc., Hanover, MA, USA.
- B. Pang, L. Lee, and S. Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of EMNLP*, pages 79–86.
- S. Petrov, D. Das, and R. McDonald. 2011. A universal part-of-speech tagset. *arXiv preprint arXiv:1104.2086*.
- S. Poria, E. Cambria, G. Winterstein, and G. Huang. 2014. Sentic patterns: Dependency-based rules for concept-level sentiment analysis. *Knowledge-Based Systems*, 69:45–63, October.
- S. Rosenthal, P. Nakov, S. Kiritchenko, S. M. Mohammad, A. Ritter, and V. Stoyanov. 2015. Semeval-2015 task 10: Sentiment analysis in Twitter. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*.
- A. Severyn and A. Moschitti. 2015. UNITN: Training Deep Convolutional Neural Network for Twitter Sentiment Classification. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 464–469, Denver, Colorado. 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. Association for Computational Linguistics.
- R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts. 2013. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In *EMNLP 2013. 2013 Conference on Empirical Methods in Natural Language Processing. Proceedings of the Conference*, pages 1631–1642, Seattle, Washington, USA, oct. ACL.
- M. Taboada and J. Grieve. 2004. Analyzing appraisal automatically. In *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text (AAAI Technical Report SS# 04# 07)*, Stanford University, CA, pp. 158q161. AAAI Press.
- M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2):267–307.
- M. Taulé, M. A. Martí, and M. Recasens. 2008. Ancora: Multilevel Annotated Corpora for Catalan and Spanish. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odjik, Stelios Piperidis, and Daniel Tapias, editors, *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC’08)*, pages 96–101, Marrakech, Morocco.
- M. Thelwall, K. Buckley, and G. Paltoglou. 2012. Sentiment strength detection for the social web. *J. Am. Soc. Inf. Sci. Technol.*, 63(1):163–173.
- K. Toutanova and C. D. Manning. 2000. Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. In *Proceedings of the 2000 Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics-Volume 13*, pages 63–70.
- P. D. Turney. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th An-*

*nual Meeting on Association for Computational Linguistics, ACL '02*, pages 417–424, Stroudsburg, PA, USA. ACL.

- D. Vilares, M. A. Alonso, and C. Gómez-Rodríguez. 2015a. A syntactic approach for opinion mining on Spanish reviews. *Natural Language Engineering*, 21(1):139–163.
- D. Vilares, M. Thelwall, and M. A. Alonso. 2015b. The megaphone of the people? Spanish SentiStrength for real-time analysis of political tweets. *Journal of Information Science*, 41(6):799–813.