If the paper title is too long for the running head, you can set an abbreviated paper title here

Sentiment Analysis Through Finite State Automata

Serena Pelosi[1], Alessandro Maisto[2], Lorenza Melillo[3], and Annibale Elia[4]

[1] University of Salerno, Department of Political and Communication Science
spelosi@unisa.it
[2] amaisto@unisa.it
[3] amelillo@unisa.it
[4] elia@unisa.it

Abstract. The present research aims to demonstrate how powerful Finite State Automata (FSA) can be, into a domain in which the vagueness of the human opinions and the subjectivity of the user generated contents make the automatic "understanding" of texts extremely hard. Assuming that the semantic orientation of sentences is based on the manipulation of sentiment words, we built from scratch, for the Italian language, a network of local grammars for the annotation of sentiment expressions and electronic dictionaries for the classification of more than 15,000 opinionated words. In the paper we explain in detail how we made use of FSA for both the automatic population of sentiment lexicons and the sentiment classification of real sentences.

Keywords: Finite State Automata · Sentiment Analysis · Contextual Valence Shifters · Sentiment Lexicon · Electronic Dictionary.

Introduction

The Web 2.0, as an interactive medium, offers to Internet users the opportunity to freely share thoughts and feelings with the web communities. This kind of information is extremely important under the consumers decision making process; we make particular reference to experience and search goods or to the e-commerce in general, if one think to what extent the evaluation of the products qualities is influenced by the past experiences of those customers that had already experienced the same goods and that had posted their opinions online. The automatic treatment of User Generated Contents becomes a relevant research problem when the huge volume of raw texts online makes their semantic content impossible to be managed by human operators. As a matter of fact, the largest amount of on-line data is semistructured or unstructured and, as a result, its monitoring requires sophisticated Natural Language Processing (NLP) tools, that must be able to pre-process them from their linguistic point of view and, then, automatically access their semantic content.

The Sentiment Analysis research field can have a large impact on many commercial, Government and Business Intelligence application. Examples are

ad-placement applications, Flame Detection Systems, Social Media Monitoring Systems, Recommendation Systems, Political Analysis, etc.
However, it would be difficult, indeed, for humans to read and summarize such a huge volume of data, but, in other respects, to introduce machines to the semantic dimension of the human language remains an open problem. The largest part of on-line data is semistructured or unstructured and, as a result, its monitoring requires sophisticated NLP strategies and tools, in order to pre-process them from their linguistic point of view and, then, automatically access their semantic content.

In the present work we present a method which exploits Finite State Automata (FSA) with the purpose of building high performance tools for the Sentiment Analysis. We computed the polarity of more than 15.000 Italian sentiment words, which have been semi-automatically listed into machine-readable electronic dictionaries, through a network of FSA, which consists of a syntactic network of grammars composed by 125 graphs.
We tested a strategy based on the systematic substitution of semantically oriented classes of (simple or compound) words into the same sentence patterns. The combined use of dictionaries and automata made it possible to apply our method on real text occurrences[5].

In Section 1 we will mention the most used techniques for the automatic propagation of sentiment lexicons and for the sentence annotation. Section 2 will delineate our method, carried out through finite state technologies. Then, in Section 3 and 4 we will go through our morphological and syntactic solutions to the mentioned challenges.

1 State of the Art

The core of this research consists of two distinguished Sentiment Analysis tasks: at the word level, the dictionary population and, at the sentence level, the annotation of complex expressions. In this paragraph we will summarize other methods used in literature to face those tasks.
Many techniques have been discussed in literature to perform the Sentiment Analysis. They can be classified into lexicon based methods, learning methods and hybrid methods.

In Sentiment Analysis tasks the most effective indicators used to discover subjective expressions are adjectives or adjective phrases [67], but recently it became really common the use of adverbs [6], nouns [72] and verbs as well [55].

Among the state of the art methods used to build and test dictionaries we mention the Latent Semantic Analysis (LSA) [41]; bootstrapping algorithms [65]; graph propagation algorithms [71, 33]; conjunctions and morphological relations

[5] We chose a rule-based method, among others, in order to verify the hypothesis that words can be classified together in accordance to both semantic and syntactic criteria.

between adjectives [29]; Context Coherency [35]; distributional similarity[6] [79]. Pointwise Mutual Information (PMI) using *Seed Words*[7] has been applied to sentiment lexicon propagation by [69, 70, 63, 71, 22].
It has been observed, indeed, that positive words occur often close to positive seed words, whereas negative words are likely to appear around negative seed words [69, 70].
Learning and statistical methods for Sentiment Analysis intent usually make use of Support Vector Machine [57, 52, 80] or Naïve Bayes classifiers [68, 37].
In the end, as regards the hybrid methods we must cite the works of [64, 42, 1, 9, 10, 24, 60] and [76].

The largest part of the state of the art works on polarity lexicons for Sentiment Analysis purposes has been carried out on the English language. Italian lexical databases are mostly created by translating and adapting the English ones, SentiWordNet and WordNet-Affect, among others. The works on the Italian language that deserve to be mentioned are [5], merged the semantic information belonging to existing lexical resources in order to obtain an annotated lexicon of senses for Italian, Sentix (Sentiment Italian Lexicon)[8]. Basically, MultiWordNet [58], the Italian counterpart of WordNet [47, 20], has been used to transfer polarity information associated to English synsets in SentiWordNet [19] to Italian synsets, thanks to the multilingual ontology BabelNet [53].
Every Sentix's entry is described by information concerning its part of speech, its WordNet synset ID, a positive and a negative score from SentiWordNet, a polarity score (from -1 to 1) and an intensity score (from 0 to 1). [8] presented a lexical sentiment resource thant contains polarized simple words, multiwords and idioms which has been annotated with polarity, intensity, emotion and domain labels[9]. [12] built a lexicon for the EVALITA 2014 task by collecting adjectives, adverbs (extracted from the De Mauro - Paravia Italian dictionary [11]), nouns and verbs (from Sentix) and by classifying their polarity through the online Sentiment Analysis API provided by *Ai Applied*[10].
Another Italian Sentiment Lexicon is the one semi-automatically developed from ItalWordNet v.2 starting from a list of seed key-words classified manually[66]. It includes 24.293 neutral and polarized items distributed in XML-LMF format[11]. [30] achieved good results in the SentiPolC 2014 task by semi-automatically

[6] Word Similarity is a very frequently used method in the dictionary propagation over the thesaurus-based approaches. Examples are the Maryland dictionary, created thanks to a Roget-like thesaurus and a handful of affixe [48], and other lexicons based on WordNet, like SentiWordNet, built on the base of quantitative analysis of glosses associated to synsets [17, 18] or other lexicons based on the computing of the distance measure on WordNet [34, 17].
[7] *Seed words* are words which are strongly associated with a positive/negative meaning, such as *eccellente* ("excellent") or *orrendo* ("horrible"), by which it is possible to build a bigger lexicon, detecting other words that frequently occur alongside them.
[8] http://valeriobasile.github.io/twita/downloads.html
[9] https://www.celi.it/
[10] http://ai-applied.nl/sentiment-analysis-api
[11] http://hdl.handle.net/20.500.11752/ILC-73

translating in Italian different lexicons; namely, SentiWordNet, Hu-Liu Lexicon, AFINN Lexicon, Whissel Dictionary, among others.

As regards the works on the lexicon propagation, we mention three main research lines: the first one is grounded on the richness of the already existent thesauri, WordNet[12] [47] among others.
The second approach is based on the hypothesis that the words that convey the same polarity appear close in the same corpus, so the propagation can be performed on the base of co-occurrence algorithms [69, 78] and, [69, 4, 36, 61, 78]. In the end, the morphological approach, which is the one that employs morphological structures and relations for the assignment of the prior sentiment polarities to unknown words, on the base of the manipulation of the morphological structures of known lemmas[13] [50, 40, 50, 77].

However, it does not seem to be enough to just dispose of sentiment dictionaries. Actually, the syntactic structures in which the opinionated lemmas occur have a strong impact on the resulting polarity of the sentences. That is the case of negation, intensification, irrealis markers and conditional tenses.
Rule-based approaches, that take into account the syntactic dimension of the Sentiment Analysis, are [51, 49]. FSA have been used for the linguistic analysis of sentiment expressions by [27, 3, 43].

Rule-based approaches, that take into account the syntactic dimension of the Sentiment Analysis, are [51, 49, 81].

2 Methodology

The present research has been grounded on the richness, in term of lexical and grammatical resources, of the linguistic databases built in the Department of Political and Communication Science (DSPC) of the University of Salerno by the Computational Linguistic lab "Maurice Gross", which started its study on language formalization since the 1981 [16, 73].
These resources take the shape of lexicon-grammar tables, that cross-check the lexicon and the syntax of any given language, in this case Italian; domain independent machine-readable dictionaries and inflectional and derivational local grammars in the form of finite state automata.

Differently from other lexicon-based Sentiment Analysis methods, our approach has been grounded on the solidity of the Lexicon-Grammar resources

[12] Although WordNet does not include semantic orientation information for its lemmas; semantic relations, such as synonymy or antonymy, are commonly used in order to automatically propagate the polarity, starting from a manually annotated set of seed word. [34, 31, 39, 18, 2, 13, 28, 45, 31, 18, 45]. This approach presents some drawbacks, such as the lack of scalability, the unavailability of enough resources for many languages and the difficulty to handle newly coined words, which are not already contained in the thesauri.

[13] Morphemes allow not only the propagation of a given word polarity (e.g. *en-*, *-ous*, *-fy*), but also its switching (e.g. *dis-*, *-less*), its intensification (e.g. *super-*, *over-*) and its weakening (e.g. *semi-*) [54].

and classifications [16, 73], that provide fine-grained semantic but also syntactic descriptions of the lexical entries. Such lexically exhaustive grammars distance themselves from the tendency of other sentiment resources to classify together words that have nothing in common from the syntactic point of view.

In the present work, we started from the annotation of a small sized dictionary of opinionated words[14].

FSA are used in both the morphological expansion of the lexicon and in the syntactic modeling of the words in context. In this research we assume that words can be classified together not only on the base of their semantic content, but also according to syntactic criteria.

Thanks to finite state technologies, we computed the polarity of individual words by systematically replacing them with other items (endowed with the same and/or different individual polarity) into many sentence (or phrase) patterns. The hypothesis is that classes of words, characterized by the same individual annotation, can be generalized when considered into different syntactic contexts, because they undergo the same semantic shifting when occurring in similar patterns. Dictionaries and FSA used in tandem made it possible to verify these assumptions on real corpora.

2.1 Local grammars and Finite-state Automata

Sentiment words, multiwords and idioms used in this work are listed into Nooj electronic dictionaries while the local grammars[15] used to manipulate their polarities are formalized thanks to Finite State Automata.

Electronic dictionaries have then been exploited in order to list and to semantically and syntactically classify, into a machine readable format, the sentiment lexical resources. The computational power of Nooj graphs has, instead, been used to represent the acceptance/refuse of the semantic and syntactic properties through the use of constraint and restrictions.

Finite-State Automata (FSA) are abstract devices characterized by a finite set of nodes or "states" connected one another by transitions that allow us to determine sequences of symbols related to a particular path.

These graphs are read from left to right, or rather, from the initial state to the final state [26].

[14] While compiling the dictionary, the judgment on the words "prior polarity" is given without considering any textual context. The entries of the sentiment dictionary receive the same annotation and, then, are grouped together if they posses the same semantic orientation. The *Prior Polarity*[56] refers to the individual words Semantic Orientation (SO) and differs from the SO because it is always independent from the context.

[15] Local grammars are algorithms that, through grammatical, morphological and lexical instructions, are used to formalize linguistic phenomena and to -parse texts. They are defined "local" because, despite any generalization, they can be used only in the description and analysis of limited linguistic phenomena.

2.2 Sentita and its Manually-built resources

In this Paragraph we will briefly describe the Sentiment lexicon, available for the Italian language, which have been semi-automatically created on the base of the resources of the DSPC.

The tagset used for the Prior Polarity annotation of the resources is composed of four tags: POS *positive*; NEG *negative*; FORTE *intense* and DEB *weak*.
Such labels, if combined together, can generate an evaluation scale that goes from -3 to +3 and a strength scale that ranges from -1 to +1.
Neutral words (e.g. *nuovo* "new", with score 0 in the evaluation scale) have been excluded from the lexicon[16].

In our resources, adjectives and bad words have been manually extracted and evaluated starting from the Nooj Italian electronic dictionary of simple words, preserving their inflectional (FLX) and derivational (DRV) properties. Moreover, compound adverbs [15], idioms [74, 75], verbs [16, 14] have been weighted starting from the Italian Lexicon-Grammar tables[17], in order to maintain the syntactic, semantic and transformational properties connected to each one of them.

3 Morphology

In this paragraph we will describe how FSA have been exploited to enrich the sentiment lexical resources. The adjectives have been used as starting point for the expansion of the sentiment lexicon, on the base of the morphophonological relations that connect the words and their meanings[18].

Thanks to a morphological FSA it has been possible to enlarge the size of SentIta on the base of the morphological relations that connect the words and their meanings.
More than 5,000 labeled adjectives have been used to predict the orientation of the adverbs with which they were morphologically related. All the adverbs contained in the Italian dictionary of simple words have been used as an input and a mophological FSA has been used to quickly populate the new dictionary by extracting the words ending with the suffix *-mente*, "-ly", and by making such words inherit the adjectives' polarity. The Nooj annotations consisted in

[16] The main difference between the words listed in the two scales is the possibility to use them as indicators for the subjectivity detection: basically, the words belonging to the evaluation scale are "anchors" that begin the identification of polarized phrases or sentences, while the ones belonging to the strength scale are just used as intensity modifiers (see Paragraph 4.3).

[17] available for consultation at http://dsc.unisa.it/composti/tavole/combo/tavole.asp.

[18] The morphological method could be also applied to Italian verbs, but we chose to avoid this solution because of the complexity of their argument structures. We decided, instead, to manually evaluate all the verbs described in the Italian Lexicon-grammar binary tables, so we could preserve the different lexical, syntactic and transformational rules connected to each one of them [16].

a list of 3,200+ adverbs that, at a later stage, have been manually checked, in order to adjust the grammar's mistakes[19].
In detail, the Precision achieved in this task is 99% and the Recall is 88%.

The derivation of quality nouns from qualifier adjectives is another derivation phenomenon of which we took advantage for the automatic enlargement of SentIta. These kind of nouns allow to treat as entities the qualities expressed by the base adjectives.
A morphological FSA, following the same idea of the adverbs grammar, matches in a list of abstract nouns the stems that are in morpho-phonological relation with our list of hand-tagged adjectives. Because the nouns, differently from the adverbs, need to have specified the inflection information, we associated to each suffix entry, into an electronic dictionary dedicated to suffixes of quality nouns, the inflectional paradigm that they give to the words with which they occur.
In order to effortlessly build a noun dictionary of sentiment words we firstly exploit the hand-made list of nominalization of the psychological verbs [25, 27, 46].

Furthermore, we took advantage from other derivation phenomena connected to nouns: the derivation of quality nouns from qualifier adjectives. We built a morphological FSA that, following the same idea of the adverbs grammar, matches into a list of abstract nouns the stems that are in morphophonological relation with our list of hand-tagged adjectives.

Suffixes	Inflection	Correct	Precision
-ità	N602	666	98%
-mento	N5	514	90%
-(z)ione	N46	359	86%
-ezza	N41	305	99%
-enza	N41	148	94%
-ia	N41	145	98%
-ura	N41	142	88%
-aggine	N46	72	97%
-eria	N41	71	95%
-anza	N41	57	86%
TOT	-	2579	93%

Table 1. Analytical description of the most productive quality nouns suffixes.

[19] The meaning of the deadjectival adverbs in *-mente* is not always predictable starting from the base adjectives from which they are derived. Also the syntactic structures in which they occur influences their interpretation. Depending on their position in sentences, the deadjectival adverbs can be described as adjective modifiers (e.g. *altamente* "highly"), predicate modifiers (e.g. *perfettamente* "perfectly") or sentence modifiers (e.g. *ultimamente* "lately").

As regards the suffixes used to form the quality nouns (Table 1) [62], it must be said that they generally make the new words simply inherit the orientation of the derived adjectives. Exceptions are *-edine* and *-eria* that almost always shift the polarity of the quality nouns into the weakly negative one (-1), e.g. *facileneria* "slapdash attitude". Also the suffix *-mento* differs from the others, in so far it belongs to the derivational phenomenon of the deverbal nouns of action [21]. It has been possible to use it into our grammar for the deadjectival noun derivation by using the past participles of the verbs listed in the adjective dictionary of sentiment (e.g. V:*sfinire* "to wear out", A:*sfinito* "worn out", N:*sfinimento*; "weariness"). The Precision achieved in this task is 93%.

In this work we draw up a FSA which can also interact, at a morphological level, with a list of prefixes able to negate (e.g. *anti-*, *contra-*, *non-* among others) or to intensify/downtone (e.g. *arci-*, *semi-* among others) the orientation of the words in which they appear [32].

4 Syntax

Contextual Valence Shifters are linguistic devices able to change the prior polarity of words when co-occurring in the same context [38, 59]. In this work we handle the contextual shifting by generalizing all the polar words that posses the same prior polarity.

A network of local grammars as been designed on a set of rules that compute the words individual polarity scores, according to the contexts in which they occur.
In general, the sentence annotation is performed through an Enhanced Recursive Transition Network, by using six different metanodes[20], that, working as containers for the sentiment expressions, assign equal labels to the patterns embedded in the same graphs.
Among the most used Contextual Valence Shifters we took into account linguistic phenomena like Intensification, Negation, Modality and Comparison. Moreover, we formalized some classes of frozen sentences that modify the polarity of the sentiment words that occur in them.
Our network of 15,000 opinionated lemmas and 125 embedded FSA has been tested on a multi-domain corpus of customer reviews[21] achieving in the sentence-level sentiment classification task an average Precision of 75% and a Recall of 73%.

[20] Metanodes are labeled through the six corresponding values of the evaluation scale, which goes from -3 to +3.

[21] The dataset contains Italian opinionated texts in the form of users reviews and comments from e-commerce and opinion websites; it lists 600 texts units (50 positive and 50 negative for each product class) and refers to six different domains, for all of which different websites (such as www.ciao.it; www.amazon.it; www.mymovies.it; www.tripadvisor.it) have been exploited [44].

4.1 Opinionated Idioms

More than 500 Italian frozen sentences containing adjectives [74, 75] have been evaluated and then formalised with a pair of dictionary-grammar. Among the idioms considered there are the comparative frozen sentences of the type *N0 Agg come C1*, described by [74], that usually intensify the polarity of the adjective of sentiment they contain, as happens in (1).

(1) Mary è *bella*[+2] come il sole [+3]
"Mary is as beautiful as the sun"

Otherwise, it is also possible for an idiom of that sort to be polarised when the adjective (e.g. *bianco*, "white") contained in it is neutral (2), or even to reverse its polarity as happens in (3) (e.g. *agile*, "agile", is positive). In that regard, it is interesting to notice that the 84% of the idioms has a clear SO, while just the 36% of the adjectives they contain is polarised[22].

(2) Mary è *bianca*[0] come un cadavere [-2]
"Mary is as white as a dead body" (Mary is pale)

(3) Mary è *agile*[+2] come una gatta di piombo [-2]
"Mary is as agile as a lead cat" (Mary is not agile)

4.2 Negation

As regards negation, we included in our grammar negative operators (e.g. *non*, "not", *mica*, *per niente*, *affatto*, "not at all"), negative quantifiers (e.g. *nessuno*, "nobody" *niente*, *nulla*, "nothing") and lexical negation (e.g. *senza*, "without", *mancanza di*, *assenza di*, *carenza di*, "lack of") [7]. As exemplified in the following sentences, negation indicators not always change a sentence polarity in its positive or negative counterparts (4); they often have the effect of increasing or decreasing the sentence score (5). That is why we prefer to talk about valence "shifting" rather than "switching".

(4) Citroen *non*[neg] produce auto *valide*[+2] [-2]
"Citroen does not produce efficient cars"

(5) Grafica *non proprio*[neg] *spettacolare*[+3] [-2]
"The graphic not quite spectacular"

[22] Other idioms included in our resources are of the kind *N0 essere (Agg + Ppass) Prep C1* (e.g. *Max è matto da legare*, "Max is so crazy he should be locked up"); *N0 essere Agg e Agg* (e.g. *Max è bello e fritto*, "Max is cooked"); *C0 essere Agg (come C1 + E)* (e.g. *Mary ha la coscienza sporca* $\leftrightarrow$ *La coscienza è sporca*, "Mary has a guilty conscience" $\leftrightarrow$ "The conscience is guilty"), *N0 essere C1 Agg* (e.g. *Mary è una gatta morta*, "Mary is a cock tease").

4.3 Intensification

We included the Intensification rules into our grammar net, firstly, by combining in the words belonging to the strength scale (tags FORTE/DEB) with the sentiment words listed in the evaluation scale (tags POS/NEG)[23]. Besides, also the repetition of more than one negative or positive words, or the use of absolute superlative affixes have the effect of increasing the words' Prior Polarity.

In general, the adverbs intensify or attenuate adjectives, verbs and other adverbs, while the adjectives modify the intensity of nouns.
Intensification and negation can also appear together in the same sentence.

4.4 Modality

According to [7], modality can be used to express possibility, necessity, permission, obligation or desire, through grammatical cues, such as adverbial phrases (e.g. "maybe", "certainly"); conditional verbal moods; some verbs (e.g. "must", "can", "may"); some adjectives and nouns (e.g. "a probable cause").
When computing the Prior Polarities of the SentIta items into the textual context, we considered that modality can also have a significant impact on the SO of sentiment expressions.

According to the literature trends, but without specifically focusing on the [7] modality categories, we recalled in the FSA dedicated to modality the following linguistic cues and we made them interact with the SentIta expressions: sharpening and softening adverbs; modal verbs and conditional and imperfect tenses. Examples of the modality in our work are the following:

- ***"Potere" + Indicative Imperfect + Oriented Item***:
  (6) *Poteva* $^{\text{[Modal+IM]}}$ essere una trama *interessante* $^{[+2]}$ [-1]
  "It could be an interesting plot"

- ***"Potere" + Indicative Imperfect + Comparative + Oriented Items***:
  (7) *Poteva*$^{\text{[Modal+IM]}}$ andare *peggio*$^{\text{[I-OpW +2]}}$ [-1]
  "It might have gone worse"

- ***"Dovere" + Indicative Imperfect***:
  (70) Questo *doveva*$^{\text{[Modal+IM]}}$ essere un film *di sfumature*$^{[+1]}$ [-2]
  "This one was supposed to be a nuanced movie"

---

[23] Words that, at first glance, seem to be intensifiers but at a deeper analysis reveal a more complex behavior are *abbastanza* "enough" *troppo* "too much" and *poco* "not much".

In this research we noticed as well that the co-occurrence of *troppo*, *poco* and *abbastanza* with polar lexical items can provoke, in their semantic orientation, effects that can be associated to other contextual valence shifters. The *ad hoc* rules dedicated to these words (see Table **??**) are not actually new, but refer to other contextual valence shifting rules that have been discussed in this Paragraph.

– ***"Dovere" + "Potere" + Past Conditional***:
(71) Non$^{[\text{Negation}]}$ *avrei*$^{[\text{Aux+C}]}$ *dovuto*$^{[\text{Modal+PP}]}$ buttare via i miei soldi [-2]
"I should not have burnt my money"

4.5 Comparison

Sentences that express a comparison generally carry along with them opinions about two or more entities, with regard to their shared features or attributes [23].

As far as the comparative sentences are concerned, we considered in this work the already mentioned comparative frozen sentences of the type *N0 Agg come C1*; some simple comparative sentences that involve the expressions *meglio di*, *migliore di*, "better than", *peggio di*, *peggiore di*, "worse than", *superiore a*, "superior to" *inferiore a*, "less than"; and the comparative superlative.

The comparison with other products has been evaluated with the same measures of the other sentiment expression; so the polarity can range from -3 to +3.

4.6 Other Sentiment Expressions

In order to reach high levels of Recall, the lexicon-based patterns require also the support of lexicon independent expressions.

In our work, we listed and computed many cases in which expressions that do not imply the words contained in our dictionaries are sentiment indicators as well. This is the case in which one can see the importance of the Finite-state automata. Without them it would be really difficult and uneconomical for a programmer to provide the machine with concise instructions to correctly recognise and evaluate some kind of opinionated sentences that can often reach high levels of variability. Examples of patterns of this kind are *valerne la pena*$^{[+2]}$, "to be worthwhile"; *essere (dotato + fornito + provvisto) di*$^{[+2]}$, "to be equipped with"; *grazie a*$^{[+2]}$, "thanks to"; *essere un (aspetto + nota + cosa + lato) negativo*$^{[-2]}$, "to be a negative side"; *non essere niente di che*$^{[-1]}$, "to be nothing special"; *tradire le (aspettative + attese + promesse)*$^{[-2]}$, "not live up to one's expectations"; etc. For simplicity, in the present work we put in this node of the grammar the sentences that imply the use of frozen, semi-frozen expression and words that, for the moment, are not part of the dictionaries.

5 Conclusion

In this paper we gave our contribution to the most two challenging tasks of the Sentiment Analysis field: the lexicon propagation and the sentence semantic annotation.
The necessity to quickly monitor huge quantity of semistructured and unstructured data from the web, poses several challenges to Natural Language Processing, that must provide strategies and tools to analyze their structures from lexical, syntactical and semantic point of views.

Unlike many other Italian and English sentiment lexicons, SentIta, made up on the interaction of electronic dictionaries and lexicon dependent local grammars, is able to manage simple and multiword structures, that can take the shape of distributionally free structures, distributionally restricted structures and frozen structures.

According with the major contribution in the Sentiment Analysis literature, we did not consider polar words in isolation. We computed they elementary sentence contexts, with the allowed transformations and, then, their interaction with contextual valence shifters, the linguistic devices that are able to modify the prior polarity of the words from SentIta, when occurring with them in the same sentences. In order to do so, we took advantage of the computational power of the finite-state technology. We formalized a set of rules that work for the intensification, downtoning and negation modeling, the modality detection and the analysis of comparative forms. Here, the difference with other state-of-the-art strategies consists in the elimination of complex mathematical calculation in favor of the easier use of embedded graphs as containers for the expressions designed to receive the same annotations into a compositional framework.

References

1. Andreevskaia, A., Bergler, S.: When specialists and generalists work together: Overcoming domain dependence in sentiment tagging. In: ACL. pp. 290–298 (2008)
2. Argamon, S., Bloom, K., Esuli, A., Sebastiani, F.: Automatically determining attitude type and force for sentiment analysis pp. 218–231 (2009)
3. Balibar-Mrabti, A.: Une étude de la combinatoire des noms de sentiment dans une grammaire locale. Langue française pp. 88–97 (1995)
4. Baroni, M., Vegnaduzzo, S.: Identifying subjective adjectives through web-based mutual information **4**, 17–24 (2004)
5. Basile, V., Nissim, M.: Sentiment analysis on italian tweets. In: Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. pp. 100–107 (2013)
6. Benamara, F., Cesarano, C., Picariello, A., Recupero, D.R., Subrahmanian, V.S.: Sentiment analysis: Adjectives and adverbs are better than adjectives alone. In: ICWSM (2007)
7. Benamara, F., Chardon, B., Mathieu, Y., Popescu, V., Asher, N.: How do negation and modality impact on opinions? pp. 10–18 (2012)
8. Bolioli, A., Salamino, F., Porzionato, V.: Social media monitoring in real life with blogmeter platform. ESSEM@ AI* IA **1096**, 156–163 (2013)
9. Dang, Y., Zhang, Y., Chen, H.: A lexicon-enhanced method for sentiment classification: An experiment on online product reviews. In: Intelligent Systems, IEEE, vol. 25, pp. 46–53. IEEE (2010)
10. Dasgupta, S., Ng, V.: Mine the easy, classify the hard: a semi-supervised approach to automatic sentiment classification. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2. pp. 701–709. Association for Computational Linguistics (2009)

11. De Mauro, T.: Dizionario italiano. Torino: Paravia (2000)
12. Di Gennaro, P., Rossi, A., Tamburini, F.: The ficlit+ cs@ unibo system at the evalita 2014 sentiment polarity classification task. In: Proceedings of the Fourth International Workshop EVALITA 2014 (2014)
13. Dragut, E.C., Yu, C., Sistla, P., Meng, W.: Construction of a sentimental word dictionary pp. 1761–1764 (2010)
14. Elia, A.: Le verbe italien. Les complétives dans les phrases àa un complément (1984)
15. Elia, A.: Chiaro e tondo: Lessico-Grammatica degli avverbi composti in italiano. Segno Associati (1990)
16. Elia, A., Martinelli, M., D'Agostino, E.: Lessico e Strutture sintattiche. Introduzione alla sintassi del verbo italiano. Napoli: Liguori (1981)
17. Esuli, A., Sebastiani, F.: Determining the semantic orientation of terms through gloss classification. In: Proceedings of the 14th ACM international conference on Information and knowledge management. pp. 617–624. ACM (2005)
18. Esuli, A., Sebastiani, F.: Determining term subjectivity and term orientation for opinion mining. **6**, 2006 (2006)
19. Esuli, A., Sebastiani, F.: SentiWordNet: A publicly available lexical resource for opinion mining. In: Proceedings of LREC. vol. 6, pp. 417–422 (2006)
20. Fellbaum, C.: WordNet. Wiley Online Library (1998)
21. Gaeta, L.: Nomi d'azione. La formazione d elle parole in italiano. Tübingen: Max Niemeyer Verlag pp. 314–51 (2004)
22. Gamon, M., Aue, A.: Automatic identification of sentiment vocabulary: exploiting low association with known sentiment terms pp. 57–64 (2005)
23. Ganapathibhotla, M., Liu, B.: Mining opinions in comparative sentences. In: Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1. pp. 241–248. Association for Computational Linguistics (2008)
24. Goldberg, A.B., Zhu, X.: Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization. In: Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing. pp. 45–52. Association for Computational Linguistics (2006)
25. Gross, M.: Les bases empiriques de la notion de prédicat sémantique. Langages pp. 7–52 (1981)
26. Gross, M.: Les phrases figées en français. In: L'information grammaticale, vol. 59, pp. 36–41. Peeters (1993)
27. Gross, M.: Une grammaire locale de l'expression des sentiments. Langue française pp. 70–87 (1995)
28. Hassan, A., Radev, D.: Identifying text polarity using random walks pp. 395–403 (2010)
29. Hatzivassiloglou, V., McKeown, K.R.: 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. pp. 174–181. Association for Computational Linguistics (1997)
30. Hernandez-Farias, I., Buscaldi, D., Priego-Sánchez, B.: Iradabe: Adapting english lexicons to the italian sentiment polarity classification task. In: First Italian Conference on Computational Linguistics (CLiC-it 2014) and the fourth International Workshop EVALITA2014. pp. 75–81 (2014)
31. Hu, M., Liu, B.: Mining and summarizing customer reviews pp. 168–177 (2004)
32. Iacobini, C.: Prefissazione. La formazione delle parole in italiano. Tübingen: Max Niemeyer Verlag pp. 97–161 (2004)

33. Kaji, N., Kitsuregawa, M.: Building lexicon for sentiment analysis from massive collection of html documents. In: EMNLP-CoNLL. pp. 1075–1083 (2007)
34. Kamps, J., Marx, M., Mokken, R.J., De Rijke, M.: Using wordnet to measure semantic orientations of adjectives (2004)
35. Kanayama, H., Nasukawa, T.: Fully automatic lexicon expansion for domain-oriented sentiment analysis. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. pp. 355–363. Association for Computational Linguistics (2006)
36. Kanayama, H., Nasukawa, T.: Fully automatic lexicon expansion for domain-oriented sentiment analysis p. 355 (2006)
37. Kang, H., Yoo, S.J., Han, D.: Senti-lexicon and improved naïve bayes algorithms for sentiment analysis of restaurant reviews. In: Expert Systems with Applications, vol. 39, pp. 6000–6010. Elsevier (2012)
38. Kennedy, A., Inkpen, D.: Sentiment classification of movie reviews using contextual valence shifters. Computational intelligence **22**(2), 110–125 (2006)
39. Kim, S.M., Hovy, E.: Determining the sentiment of opinions p. 1367 (2004)
40. Ku, L.W., Huang, T.H., Chen, H.H.: Using morphological and syntactic structures for chinese opinion analysis pp. 1260–1269 (2009)
41. Landauer, T.K., Dumais, S.T.: A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. In: Psychological review, vol. 104, p. 211. American Psychological Association (1997)
42. Li, F., Huang, M., Zhu, X.: Sentiment analysis with global topics and local dependency. In: AAAI (2010)
43. Maisto, A., Pelosi, S.: Feature-based customer review summarization. In: On the Move to Meaningful Internet Systems: OTM 2014 Workshops, pp. 299–308. Springer (2014)
44. Maisto, A., Pelosi, S.: A lexicon-based approach to sentiment analysis. the italian module for nooj. In: Proceedings of the International Nooj 2014 Conference, University of Sassari, Italy. Cambridge Scholar Publishing (2014)
45. Maks, I., Vossen, P.: Different approaches to automatic polarity annotation at synset level pp. 62–69 (2011)
46. Mathieu, Y.Y.: Les prédicats de sentiment. Langages pp. 41–52 (1999)
47. Miller, G.A.: Wordnet: a lexical database for english. Communications of the ACM **38**(11), 39–41 (1995)
48. Mohammad, S., Dunne, C., Dorr, B.: Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2. pp. 599–608. Association for Computational Linguistics (2009)
49. Moilanen, K., Pulman, S.: Sentiment composition pp. 378–382 (2007)
50. Moilanen, K., Pulman, S.: The good, the bad, and the unknown: morphosyllabic sentiment tagging of unseen words pp. 109–112 (2008)
51. Mulder, M., Nijholt, A., Den Uyl, M., Terpstra, P.: A lexical grammatical implementation of affect pp. 171–177 (2004)
52. Mullen, T., Collier, N.: Sentiment analysis using support vector machines with diverse information sources. In: EMNLP. vol. 4, pp. 412–418 (2004)
53. Navigli, R., Ponzetto, S.P.: BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. vol. 193, pp. 217–250 (2012)
54. Neviarouskaya, A.: Compositional approach for automatic recognition of fine-grained affect, judgment, and appreciation in text (2010)

55. Neviarouskaya, A., Prendinger, H., Ishizuka, M.: Compositionality principle in recognition of fine-grained emotions from text. In: ICWSM (2009)
56. Osgood, C.E.: The nature and measurement of meaning. Psychological bulletin **49**(3), 197 (1952)
57. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. pp. 79–86. Association for Computational Linguistics (2002)
58. Pianta, E., Bentivogli, L., Girardi, C.: MultiWordNet: developing an aligned multilingual database. In: Proceedings of the first international conference on global WordNet. vol. 152, pp. 55–63 (2002)
59. Polanyi, L., Zaenen, A.: Contextual valence shifters pp. 1–10 (2006)
60. Prabowo, R., Thelwall, M.: Sentiment analysis: A combined approach. In: Journal of Informetrics, vol. 3, pp. 143–157. Elsevier (2009)
61. Qiu, G., Liu, B., Bu, J., Chen, C.: Expanding domain sentiment lexicon through double propagation. **9**, 1199–1204 (2009)
62. Rainer, F.: Derivazione nominale deaggettivale. La formazione delle parole in italiano pp. 293–314 (2004)
63. Rao, D., Ravichandran, D.: Semi-supervised polarity lexicon induction. In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics. pp. 675–682. Association for Computational Linguistics (2009)
64. Read, J., Carroll, J.: Weakly supervised techniques for domain-independent sentiment classification. In: Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion. pp. 45–52. ACM (2009)
65. Riloff, E., Wiebe, J., Wilson, T.: Learning subjective nouns using extraction pattern bootstrapping. In: Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4. pp. 25–32. Association for Computational Linguistics (2003)
66. Russo, I., Frontini, F., Quochi, V.: OpeNER sentiment lexicon italian - LMF (2016), http://hdl.handle.net/20.500.11752/ILC-73, digital Repository for the CLARIN Research Infrastructure provided by ILC-CNR
67. Taboada, M., Anthony, C., Voll, K.: Methods for creating semantic orientation dictionaries. In: Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC), Genova, Italy. pp. 427–432 (2006)
68. Tan, S., Cheng, X., Wang, Y., Xu, H.: Adapting naive bayes to domain adaptation for sentiment analysis. In: Advances in Information Retrieval, pp. 337–349. Springer (2009)
69. Turney, P.D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews pp. 417–424 (2002)
70. Turney, P.D., Littman, M.L.: Measuring praise and criticism: Inference of semantic orientation from association. In: ACM Transactions on Information Systems (TOIS), vol. 21, pp. 315–346 (2003)
71. Velikovich, L., Blair-Goldensohn, S., Hannan, K., McDonald, R.: The viability of web-derived polarity lexicons. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. pp. 777–785. Association for Computational Linguistics (2010)
72. Vermeij, M.: The orientation of user opinions through adverbs, verbs and nouns. In: 3rd Twente Student Conference on IT, Enschede June (2005)
73. Vietri, S.: The italian module for nooj. In: In Proceedings of the First Italian Conference on Computational Linguistics, CLiC-it 2014. Pisa University Press (2014)

74. Vietri, S.: On some comparative frozen sentences in italian. Lingvisticæ Investigationes **14**(1), 149–174 (1990)
75. Vietri, S.: On a class of italian frozen sentences. Lingvisticæ Investigationes **34**(2), 228–267 (2011)
76. Wan, X.: Co-training for cross-lingual sentiment classification. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1. pp. 235–243. Association for Computational Linguistics (2009)
77. Wang, X., Zhao, Y., Fu, G.: A morpheme-based method to chinese sentence-level sentiment classification. Int. J. of Asian Lang. Proc. **21**(3), 95–106 (2011)
78. Wawer, A.: Extracting emotive patterns for languages with rich morphology. International Journal of Computational Linguistics and Applications **3**(1), 11–24 (2012)
79. Wiebe, J.: Learning subjective adjectives from corpora. In: AAAI/IAAI. pp. 735–740 (2000)
80. Ye, Q., Zhang, Z., Law, R.: Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. In: Expert Systems with Applications, vol. 36, pp. 6527–6535. Elsevier (2009)
81. Yi, J., Nasukawa, T., Bunescu, R., Niblack, W.: Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques pp. 427–434 (2003)