



HAL
open science

Towards a Lexicon-grammar based Framework for NLP: an Opinion Mining Application

Annibale Elia, Serena Pelosi, Alessandro Maisto, Raffaele Guarasci

► **To cite this version:**

Annibale Elia, Serena Pelosi, Alessandro Maisto, Raffaele Guarasci. Towards a Lexicon-grammar based Framework for NLP: an Opinion Mining Application. RANLP, 2015, Hissar, Bulgaria. pp.160-167. hal-04401054

HAL Id: hal-04401054

<https://hal.science/hal-04401054v1>

Submitted on 17 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Towards a Lexicon-grammar based Framework for NLP an Opinion Mining Application

Annibale Elia, Serena Pelosi,

Alessandro Maisto and Raffaele Guarasci

Department of Political, Social and Communication Science

University of Salerno

{elia,spelosi,amaisto,rguarasci}@unisa.it

Abstract

The present research exploits the large amount of linguistic resources developed into the Lexicon-grammar paradigm in the domain of the Opinion Mining. Grounded on the Semantic Predicates theory, the proposed system is able to automatically match the syntactic structures selected by special classes of verbs, indicating positive or negative Sentiment, Opinion or Physical acts, with the semantic frames evoked by the same lexical items. This method has been tested on a large dataset composed of short texts, such as tweets and news headings.

1 Introduction

In our research we propose a computational use of the Lexicon-grammar (LG) theories in the domain of the Opinion Mining.

We take advantage of both the huge amount of linguistic facts, accurately formalized and described in the LG paradigm, and from the possibility to apply and test them on big data. The purpose is to build a fine grained Information Extraction tool able to locate meaningful information in raw texts and to characterize them with thorough semantic descriptions.

According with the Semantic Predicates theory (Gross, 1981), it has been possible to perform a matching between the definitional syntactic structures, attributed to each class of verbs, and the semantic information we attached in the database to every lexical entry.

This way we could create a strict connection between the arguments, selected by a given Predicate listed in our tables, and the actants involved into the same verb's Semantic Frame (Fillmore, 1976; Fillmore and Baker, 2001; Fillmore, 2006). Thanks to our LG-based linguistic rules, anchored

on the Semantic Predicates, we started with this research the development of an NLP framework that, on the base of sophisticated syntactic and semantic analyses, extracts real text occurrences and labels them with the semantic roles involved in every matched sentence.

This ambitious work, that in this preliminary stage focused just on the predicates indicating sentiments, opinions and physical acts, intends, in future works, to become a larger development of an LG-based Italian cross-platform open library for various kind of linguistic analyses.

We excluded from this work the Transfer, the Spatial and the also the Psychological Predicates, because they have been already tested on different kinds of raw data with satisfactory results (Vietri, 2014; Elia et al., 2010; Elia and Vietri, 2010; Elia et al., 2013; Maisto and Pelosi, 2014b).

2 Theoretical Background

The Lexicon-grammar (LG), the method and the practice of formal description of the natural language, introduced important changes in the way in which the relationship between lexicon and syntax was conceived. (Gross, 1971; Gross, 1975).

In the LG theoretical framework the minimum discourse units endowed with meaning are the whole nuclear sentences, generally anchored on the verbs, which hold together the relationships between the selected arguments.

That means that the sentence structure is already contained in the operator (Harris, 1971; Harris, 1976).

We chose this paradigm because of its compatibility with the purposes of the computational linguistics, that, in order to reach high performances in results, requires a large amount of linguistic data, as much as possible, exhaustive, reproducible and well organized.

The collection of the linguistic information is constantly registered into LG tables, binary matrices

that cross-check the lexical entries with transformational, distributional and structural properties (see Table 1).

The LG classification and description of the Italian verbs¹ (Elia et al., 1981; Elia, 1984; D’Agostino, 1992) is grounded on the differentiation of three different macro-classes: transitive verbs; intransitive verbs and verbs that select completive clauses as complement. Every LG class has its own definitional structure, that corresponds with the syntactic structure of the nuclear sentence selected by a given number of verbs². All the lexical entries are, then, differentiated from one another in each class, by taking into account all the transformational, distributional and structural properties accepted or rejected by every item.

The formal notation used in the LG framework can be summarized in the following way: *N*, that always indicates a nominal group, is followed by a number, which specifies its nature. (*NO* stands for the sentence formal subject, *N1* for the first complement and *N2* for the second complement); *V* stands for the verbs; *Prep* for the prepositions and *Che F* suggests the presence of completive or subjective clauses.

2.1 Semantic Predicates

The whole set of syntactical structure of a given language (*Sy*) is linkable to the entire collection of the semantic items of the same language (*Se*) by means of specific interpretation rules. This is the basic assumption on which has been build the Semantic Predicates theory into the LG framework, that postulates a parallelism between the *Sy* actants and the *Se* arguments (Gross, 1981). As an example, in [1]

[1] Quello slogan^[N0/h] *offende*^[V/O] le donne^[N1/t]
 “That slogan offends the women”

the verb *offendere* “to offend”, belonging to the LG class *20UM (NO V NIhum)*, will be associated to a Predicate with two variables, described by the function $O(h,t)$, through the following rules of interpretations (see Section 3 for the other semantic functions for the annotation of Semantic Predi-

¹freely available at the address <http://dsc.unisa.it/composti/tavole/combo/tavole.asp>

²e.g. *V* for *piovere* “to rain” and all the verbs of the class 1; *NO V* for *bruciare* “to burn” and the other verbs of the class 3; *NO V da NI* for *provenire* “to come from” and the verbs belonging to the class 6; etc...

cates of different nature):

1. the Opinion Holder (*h* in the *Se*) corresponds to the formal subject (*NO* in *Sy*);
2. the opinion Target (*t* in the *Se*) is the human complement (*NI* in *Sy*).

As shown in [2], the syntactic transformations in which the same Predicate is involved do not modify the role played by its arguments, that, in order to be semantically labeled in a correct way, must be always led back to their original form [3].

[2] Il fuoriclasse è stato *offeso* da un politico messicano
 “The champion has been offended by a Mexican politician”

[3] Un politico messicano^[N0/h] *ha offeso*^[V/O] Il fuoriclasse^[N1/t]
 “a Mexican politician offended the champion”

Special kinds of Semantic Predicates have been already used in NLP applications into a lexicon-grammar context; we mention (Vietri, 2014; Elia et al., 2010; Elia and Vietri, 2010) that formalized and tested the Transfer Predicates on the Italian Civil Code; (Elia et al., 2013) that focused on the Spatial Predicates and (Maisto and Pelosi, 2014b) that exploited the Psychological Semantic Predicates for Sentiment Analysis purposes.

2.2 Frame Semantics

“Some words exist in order to provide access to knowledge of such frames to the participants in the communication process, and simultaneously serve to perform a categorization which takes such framing for granted” (Fillmore, 2006). With these words it has been depicted the Frame Semantics, which describes the sentences on the base of *predicators* able to bring to mind the *semantic frames* (inference structures, linked through linguistic convention to the lexical items meaning) and the *frame elements* (participants and props in the frame) involved in these frames (Fillmore, 1976; Fillmore and Baker, 2001; Fillmore, 2006). A frame semantic description starts from the identification of the lexical items that carry out a given meaning and, then, explores the ways in which the frame elements and their constellations are realized around the structures that have such items as

head (Fillmore et al., 2002).

Based on these principles, the FrameNet research project produced a lexicon of English for both human use and NLP applications (Baker et al., 1998; Fillmore et al., 2002; Ruppenhofer et al., 2006). Its purpose is to provide a large amount of semantically and syntactically annotated sentences endowed with information about the valences (combinatorial possibilities) of the items derived from annotated contemporary English corpus. Among the semantic domains covered there are also *emotion* and *cognition* (Baker et al., 1998).

For the Italian language, it has been developed LexIt, a tool that, following the FrameNet approach, automatically explores syntactic and semantic properties of Italian predicates in terms of distributional profiles. It performs frame semantic analyses using both *La Repubblica* corpus and the *Wikipedia* taxonomy (Lenci et al., 2012).

2.3 Case Study: Opinion Mining and Sentiment Analysis

Sentiment Analysis, also called opinion mining, subjectivity analysis, or appraisal extraction, consists in the computational treatment of opinions, and emotions freely expressed in texts. It represents a really active NLP field that includes as specific research challenges the Sentiment and Subjectivity Classification, the Feature-based Sentiment Analysis, Sentiment analysis of comparative sentences, the Opinion search and retrieval, or the Opinion spam detecting and, in the end, the Opinion Holder and Target extraction. This research fields have a large impact on many commercial, Government and Business Intelligence application.

The most used approaches in the Sentiment Analysis include, among others, the lexicon-based methods, that always start from the following assumption: the text sentiment orientation comes from the semantic orientations of words and phrases contained in it.

The most commonly used SO indicators are adjectives or adjective phrases (Hatzivassiloglou and McKeown, 1997; Hu and Liu, 2004; Taboada et al., 2006), but recently became really common the use of adverbs (Benamara et al., 2007), nouns (Vermeij, 2005; Riloff et al., 2003) and verbs as well (Neviarouskaya et al., 2009).

Among the most popular lexicons for the Sentiment Analysis we account: the General In-

	N0=Nhum			N1= che F					
		N0= il fatto Ch F	Verb		N1= che Fcong	di N lhum	diN1-hum	di N1 contro N2	prep V-inf comp
+	-		profittare	+	+	+	+	+	-
+	-		ridersene	+	+	+	+	-	-
+	-		risentirsi	+	+	+	+	-	+
+	-		strafottersene	+	+	+	+	-	-
+	-		vergognarsi	+	+	+	+	-	+

Table 1: Extract of the Lexicon-grammar table of the verb Class 45.

quirer (Stone et al., 1966), the Hatzivassiloglou Lexicon (Hatzivassiloglou and McKeown, 1997), WordNet-Affect (Strapparava et al., 2004), the Wilson Lexicon (Wiebe et al., 2004), SentiWordNet (Esuli and Sebastiani, 2006), the Appraisal Lexicon (Argamon et al., 2009), the Maryland dictionary (Mohammad et al., 2009) SentiFul (Neviarouskaya et al., 2011), the SO-CAL dictionary (Taboada et al., 2011), Q-WordNet (Agerri and García-Serrano, 2010), Velikovich Web-generated lexicon (Velikovich et al., 2010), and the SentiSense (de Albornoz et al., 2012).

3 Methodology

The starting point of our research are 66 Lexicon-grammar tables of the Italian verbs, developed at the Department of Communication Science of the University of Salerno. Among the 3000 lexical entries listed in such matrices, we manually extracted about 1000 verbs endowed with a defined semantic orientation. Furthermore, on this base, we manually built a set of electronic dictionaries enriched with both the properties listed in the LG tables (Table 1) and the Semantic details associated with each lexical item (Table 2). In detail, 28 LG classes contained at least one opinionated item.

The examples in Tables 1 and 2 concern a small group of verbs belonging to the Lexicon-grammar class 45. This class includes all the verbs that can entry into a syntactic structure such as *N0 V di NI*, in which the “subject” (*N0*) selected by the verb (*V*) is generally a human noun (*Nhum*) and the complement (*NI*) is a completive (*Ch F*) or infinitive (*V-inf comp*) clause, usually introduced by the preposition “di” (see Table1).

As shown in Table 2, our databases contain also semantic information concerning the nature, the semantic orientation and the strength of the Predi-

cates under consideration.

Differently from the most used Italian tagsets (Bosco et al., 2009), in order to avoid high computational costs, our lexical databases are provided with basic semantic description. In detail, the tagset used in this work is the following:

1. Type
 - (a) SENT, sentiment
 - (b) OP, opinion
 - (c) PHY physical act
2. Orientation
 - (a) POS, positive
 - (b) NEG, negative
3. Intensity
 - (a) STRONG, intense
 - (b) WEAK, feeble

Speaking in terms of Frame Semantics, we identified in the Opinion Mining and in the Emotion Detection field three Frames of interest, recalled by specific Predicates: *Sentiment*, *Opinion* and *Physical act*. The frame elements evoked by such frames are described below.

Sentiment. It refers to the expression of any given frame of mind or affective state. The “sentiment” words can be put in connection with some WordNet Affect categories (Strapparava et al., 2004), such as *emotion*, *mood*, *hedonic signal*. Examples are *sdegnarsi* “to be indignant” (class 10); *odiare* “to hate” (class 20); *affezionarsi* “to grow fond” (class 44B); *flirtare* “to flirt” (class 9); *disprezzare* “to despise” (class 20); *gioire* “to rejoice” (class 45).

Predicates of that kind evoke as frame elements an *experiencer* (e), that feels the emotion or other internal states, and a *causer* (c), an event or a person that instigates such states (Gildea and Jurafsky, 2002; Swier and Stevenson, 2004; Palmer et al., 2005). This semantic frame summarizes the FrameNet ones connected to emotions, such as *Cause_to_experience*, *Sensation Emotions*, *Cause_emotion*, *Emotions_of_mental_activity*, *Emotion_active*, etc...

In this work, they are described by a function of that sort: $S(e,c)$

Opinion. The type “Opinion”, instead, is the expression of positive or negative viewpoints, beliefs or judgments, that can be personal or shared by most people. It comprises, among the WN-affect categories, *trait*, *cognitive state*, *behavior*, *attitude*. OP examples are *ignorare* “to neglect” (class 20); *premiare* “to reward” (class 20); *difendere* “to defend” (class 27); *esaltare* “to exalt” (class 22); *dubitare* “to doubt” (class 45); *condannare* “to condemn” (class 49); *deridere* “to make fun of” (class 50).

The frame elements they evoke are an *opinion holder* (h), that states an opinion about an object or an event, and an *opinion target* (t), that represents the event or the object on which the opinion is expressed about (Kim and Hovy, 2006; Liu, 2012).

Into the FrameNet frame *Opinion* and *Judgment*, the *opinion holder* is called *Cognizer*, but we preferred to use a word which is more common in the Sentiment Analysis and in the Opinion Mining literature.

$O(h,t)$ is the function by which they are semantically described.

Physical act. The type “Physical act” comprises verbs like *baciare* “to kiss” (class 18); *suicidarsi* “to commit suicide” (class 2); *vomitare* “to vomit” (class 2A); *sparare* “to shoot” (class 4); *schiaffeggiare* “to slap” (class 20); *palpeggiare* “to grope” (class 18).

For this group of predicates the selected frame elements are a *patient* (z) that is the victim (for the negative actions) or the beneficiary (for the positive ones) of the physical act carried out by an *agent* (a) (Carreras and Màrquez, 2005; Màrquez et al., 2008).

It includes a large number of FrameNet frames, such as *Cause_bodily_experience*, *Cause_harm Killing*, *Rape*, *Sex*, *Shoot_projectiles*, *Violence*, etc...

The meaning of the sentences in which occur predicates of that kind is summarized in the function $P(z,a)$.

Semantic Orientation and Intensity. To perform the Orientation and the Intensity attribution, we manually explored the Italian LG tables of verbs and weighted the Prior Polarity (Osgood, 1952) of the words endowed with a positive or negative SO.

We created two separate scales for the evaluation of the strength (intense/weak) and of the polar-

Verb	LG Class	Type	Orientation	Intensity	Influence
profittare	45	op	neg	-	N0
ridersene	45	op	neg	weak	N1
risentirsi	45	sent	neg	-	N0
strafottersene	45	op	neg	strong	N1
vergognarsi	45	sent	neg	strong	N0

Table 2: Extract of the semantic description of the opinionated verbs belonging to the LG class 45.

ity (positive/negative) through the combination of four tags: POS, NEG, STRONG and WEAK, creating, this way, an evaluation scale that goes from -3 to +3 and a strength scale that ranges from -1 to +1.

Semantic Role Labeling. Thanks to lexical resources of this kind, it is possible to automatically extract and semantically describe real occurrences of sentences, like [4]

[4] Renzi^[N0/e] *si vergogna*^[V/S] di parlare di energia in Europa^[N1/c]
 “Renzi feels ashamed of talking about energy in Europe”

in which the syntactic structure of the verb, *vergognarsi* “to feel ashamed”, *N0 V (*di) Ch F*, is matched, by means of interpretation rules to the semantic function *S(e,c)*, that put in relation an *experiencer* (e) and a *causer* (c) thanks to a *Sentiment Semantic Predicate* (S).

Moreover, we provided our LG databases with the specification of the arguments (N0, N1, N2, etc...) that are semantically influenced by the semantic orientation of the verbs. The purpose is to correctly identify them as *features* of the opinionated sentences and to work on their base also into feature-based sentiment analysis tasks.

4 Experiment

4.1 Datasets

The reliability of the LG method on the Semantic Role Labeling in the Opinion Mining and the Emotion Detection tasks has been tested on three different datasets, two of which have been extracted from social network or web resources.

In detail, the first two datasets came from Twitter, the third was a free web news headings dataset provided by DataMediaHub (www.

datamediahub.it) and Human Highway (www.humanhighway.it).

The tweets have been downloaded using the two hashtags #Mattarellapresidente, that groups together the user comments on the election of the Italian President Mattarella and #Masterchefit, that collects the comments on the homonymous Italian TV show.

1. Tweets (46.393 tweets)
 - (a) #Mattarellapresidente (10.000 tweets)
 - (b) #Masterchefit (36.393 tweets)
2. News Headings (80.651 titles)

4.2 System and Tools

The LG based approach includes the following basic steps:

1. a preprocessing pipeline, that includes two phases:
 - (a) a cleaning up phase, carried out with Python routines, that aims to distinguish in the datasets linguistic elements from structural elements (e.g. markup informations, web specific elements);
 - (b) an automatic linguistic analysis phase, with the goal to linguistically standardize relevant elements obtained from the cleaned datasets; in this phase texts are tokenized, lemmatized and POS tagged using TreeTagger (Schmid, 1994; Schmid et al., 2007) and, then, parsed using DeSR, a dependency-based parser (Attardi et al., 2009);
2. a Lexicon-grammar based automatic analysis, in which the raw data are semantically labeled according with the syntactic/semantic rules of interpretation connected with each LG verb class;

Figure 1 presents three headlines examples processed both with the dependency syntactic parser and the semantic LG-based semantic analyzer.

Notice that the elements of the traditional grammar automatically identified by DeSR, such as subjects and complements, have been renamed according with the lexico-grammar tradition.

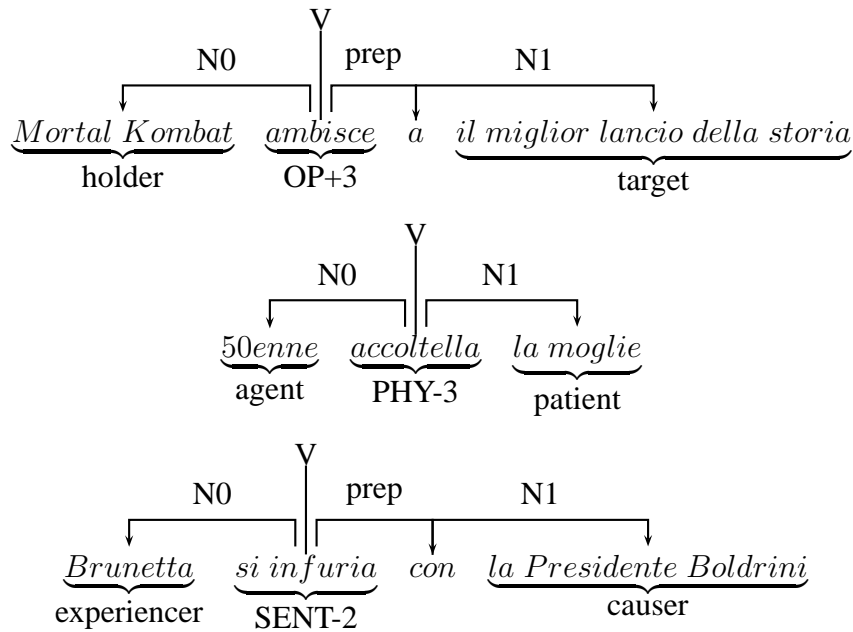


Figure 1: Examples of syntactically and semantically annotated sentences

4.3 Results and Open Issues

The corpus described in section 4.1, that counts 127,044 short texts, has been analyzed and semantically and syntactically annotated.

The representative sample on which the human evaluation has been performed, instead, has 42,348 texts.

The evaluation of the performances of our tool proved the effectiveness of the Lexicon-grammar approach. The average F-scores achieved in the different datasets are 0.71 in the Twitter and 0.76 in the Heading corpus.

Although such results, in this preliminary stage of the research, can be considered satisfactory, they shown that applying a lexicon-grammar method through a dependency parser is not the greatest solution for our purposes. The main goal of this work was, in fact, to demonstrate the validity and the reliability a LG based framework for NLP, but, in order to improve our performances, in future works we aim to build from scratch a syntactic parser completely inspired on the Lexicon-grammar theories, able to take into account not only the definitional syntactic structures of the LG verb classes, but also capable to handle every lemma's idiosyncrasies and any one of the properties systematically recorded into the LG tables.

In the end, it must be pointed out that this research represents just an aspect of a broader Sentiment Analysis framework, which involves not

only the verbs in its lexicon, but also other, simple and compound, parts of speech, including special kinds of opinionated idioms (Maisto and Pelosi, 2014b; Maisto and Pelosi, 2014a). The novel aspect introduced in this work concerns, above all, the lexicon-grammar idea that in the lexicon are already contained syntactic clues.

5 Conclusion

This paper has introduced the possibility to apply and test the Lexicon-grammar theories and lexical resources on large corpora for different kinds of information extraction and content analysis purposes.

In detail, this research focused on the automatic extraction from raw data of sentences regarding Sentiments, Opinions and Physical Acts and on the semantic annotation of the roles involved in each one of the mentioned frames. Both the extraction and the analysis are anchored on a lexicon of Semantic Predicates, able to evoke, at the same time, the syntactic structures of their arguments in real text occurrences and the nature of the roles that those arguments play into specific semantic frames.

Furthermore, thanks to the tags which the Predicates are provided with, it has been possible to annotate the same sentences with information regarding their semantic orientation and intensity.

The aim of the research was to demonstrate the re-

liability of a Lexicon-grammar based framework for many kinds of NLP purposes. We started the experimentation on a corpus of tweets and news headlines, with satisfactory results.

References

- Rodrigo Agerri and Ana García-Serrano. 2010. Q-wordnet: Extracting polarity from wordnet senses. In *LREC*.
- Shlomo Argamon, Kenneth Bloom, Andrea Esuli, and Fabrizio Sebastiani. 2009. Automatically determining attitude type and force for sentiment analysis. pages 218–231.
- Giuseppe Attardi, Felice Dell’Orletta, Maria Simi, and Joseph Turian. 2009. Accurate dependency parsing with a stacked multilayer perceptron. *Proceedings of EVALITA*, 9.
- Collin F Baker, Charles J Fillmore, and John B Lowe. 1998. The berkeley framenet project. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, pages 86–90. Association for Computational Linguistics.
- Farah Benamara, Carmine Cesarano, Antonio Picariello, Diego Reforgiato Recupero, and Venkatramana S Subrahmanian. 2007. Sentiment analysis: Adjectives and adverbs are better than adjectives alone. In *ICWSM*.
- Cristina Bosco, Simonetta Montemagni, Alessandro Mazzei, Vincenzo Lombardo, Felice Dell’Orletta, and Alessandro Lenci. 2009. Evalita09 parsing task: comparing dependency parsers and treebanks. *Proceedings of EVALITA*, 9.
- Xavier Carreras and Lluís Màrquez. 2005. Introduction to the conll-2005 shared task: Semantic role labeling. In *Proceedings of the Ninth Conference on Computational Natural Language Learning*, pages 152–164. Association for Computational Linguistics.
- Emilio D’Agostino. 1992. *Analisi del discorso: metodi descrittivi dell’italiano d’uso*. Loffredo.
- Jorge Carrillo de Albornoz, Laura Plaza, and Pablo Gervás. 2012. Sentsense: An easily scalable concept-based affective lexicon for sentiment analysis. In *LREC*, pages 3562–3567.
- Annibale Elia and Simonetta Vietri. 2010. Lexis-grammar & semantic web. *INFOtheca*, 11:15a–38a.
- Annibale Elia, Maurizio Martinelli, and Emilio D’Agostino. 1981. *Lessico e Strutture sintattiche. Introduzione alla sintassi del verbo italiano*. Napoli: Liguori.
- Annibale Elia, Simonetta Vietri, Alberto Postiglione, Mario Monteleone, and Federica Marano. 2010. Data mining modular software system. In *SWWS*, pages 127–133.
- Annibale Elia, Daniela Guglielmo, Alessandro Maisto, and Serena Pelosi. 2013. A linguistic-based method for automatically extracting spatial relations from large non-structured data. In *Algorithms and Architectures for Parallel Processing*, pages 193–200. Springer.
- Annibale Elia. 1984. Le verbe italien. *Les complétives dans les phrases à un complément*.
- Andrea Esuli and Fabrizio Sebastiani. 2006. Determining term subjectivity and term orientation for opinion mining. 6:2006.
- Charles J Fillmore and Collin F Baker. 2001. Frame semantics for text understanding. In *Proceedings of WordNet and Other Lexical Resources Workshop, NAACL*.
- Charles J Fillmore, Collin F Baker, and Hiroaki Sato. 2002. The framenet database and software tools. In *LREC*.
- Charles J Fillmore. 1976. Frame semantics and the nature of language*. *Annals of the New York Academy of Sciences*, 280(1):20–32.
- Charles J Fillmore. 2006. Frame semantics. *Cognitive linguistics: Basic readings*, 34:373–400.
- Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational linguistics*, 28(3):245–288.
- Maurice Gross. 1971. *Transformational Analysis of French Verbal Constructions*. University of Pennsylvania.
- Maurice Gross. 1975. *Méthodes en syntaxe*. Hermann.
- Maurice Gross. 1981. Les bases empiriques de la notion de prédicat sémantique. *Langages*, pages 7–52.
- Zellig Sabbetai Harris. 1971. *Structures mathématiques du langage*, volume 3. Dunod.
- Zellig Sabbetai Harris. 1976. Notes du cours de syntaxe, traduction française par maurice gross. *Paris: Le Seuil*.
- Vasileios Hatzivassiloglou and Kathleen R McKeown. 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. Association for Computational Linguistics.
- Minqing Hu and Bing Liu. 2004. Mining opinion features in customer reviews. In *AAAI*, volume 4, pages 755–760.

- Soo-Min Kim and Eduard Hovy. 2006. Extracting opinions, opinion holders, and topics expressed in online news media text. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, pages 1–8. Association for Computational Linguistics.
- Alessandro Lenci, Gabriella Lapesa, and Giulia Bonansinga. 2012. Lexit: A computational resource on italian argument structure. In *LREC*, pages 3712–3718.
- Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167.
- Alessandro Maisto and Serena Pelosi. 2014a. Feature-based customer review summarization. In *On the Move to Meaningful Internet Systems: OTM 2014 Workshops*, pages 299–308. Springer.
- Alessandro Maisto and Serena Pelosi. 2014b. 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.
- Lluís Màrquez, Xavier Carreras, Kenneth C Litkowski, and Suzanne Stevenson. 2008. Semantic role labeling: an introduction to the special issue. *Computational linguistics*, 34(2):145–159.
- Saif Mohammad, Cody Dunne, and Bonnie Dorr. 2009. 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*, pages 599–608. Association for Computational Linguistics.
- Alena Neviarouskaya, Helmut Prendinger, and Mitsuru Ishizuka. 2009. Compositionality principle in recognition of fine-grained emotions from text. In *ICWSM*.
- Alena Neviarouskaya, Helmut Prendinger, and Mitsuru Ishizuka. 2011. Sentiful: A lexicon for sentiment analysis. *Affective Computing, IEEE Transactions on*, 2(1):22–36.
- Charles E Osgood. 1952. The nature and measurement of meaning. *Psychological bulletin*, 49(3):197.
- Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1):71–106.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135.
- Ellen Riloff, Janyce Wiebe, and Theresa Wilson. 2003. Learning subjective nouns using extraction pattern bootstrapping. In *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4*, pages 25–32. Association for Computational Linguistics.
- Josef Ruppenhofer, Michael Ellsworth, Miriam RL Petruck, Christopher R Johnson, and Jan Schefczyk. 2006. Framenet ii: Extended theory and practice.
- H Schmid, M Baroni, E Zanchetta, and A Stein. 2007. The enriched treetagger system. In *proceedings of the EVALITA 2007 workshop*.
- Helmut Schmid. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the international conference on new methods in language processing*, volume 12, pages 44–49. Cite-seer.
- Philip J Stone, Dexter C Dunphy, and Marshall S Smith. 1966. The general inquirer: A computer approach to content analysis.
- Carlo Strapparava, Alessandro Valitutti, et al. 2004. Wordnet affect: an affective extension of wordnet. In *LREC*, volume 4, pages 1083–1086.
- Robert S Swier and Suzanne Stevenson. 2004. Unsupervised semantic role labelling. In *Proceedings of EMNLP*, volume 95, page 102.
- Maite Taboada, Caroline Anthony, and Kimberly Voll. 2006. Methods for creating semantic orientation dictionaries. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC), Genova, Italy*, pages 427–432.
- Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307.
- Leonid Velikovich, Sasha Blair-Goldensohn, Kerry Hannan, and Ryan McDonald. 2010. 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*, pages 777–785. Association for Computational Linguistics.
- MJM Vermeij. 2005. The orientation of user opinions through adverbs, verbs and nouns. In *3rd Twente Student Conference on IT, Enschede June*. Citeseer.
- Simona Vietri. 2014. The italian module for nooj. In *In Proceedings of the First Italian Conference on Computational Linguistics, CLiC-it 2014*. Pisa University Press.
- Janyce Wiebe, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin. 2004. Learning subjective language. *Computational linguistics*, 30(3):277–308.