UNIVERSITY OF BERGAMO

School of Doctoral Studies

Doctoral Degree in Linguistic Sciences

XXX Cycle

SSD: L-LIN/01 GLOTTOLOGY AND LINGUISTICS

Lexical Opposition and Discourse Contrast: A Data-driven Investigation

Advisor

Prof. Elisabetta Ježek

Prof. Bernardo Magnini

Doctoral Thesis

Anna FELTRACCO

Student ID 1036289

Academic year 2016/17

This Ph.D. Thesis has been developed within the Ph.D. School in Linguistic Sciences of University of Bergamo and University of Pavia, in partnership with Fondazione Bruno Kessler (FBK), Trento, Italy. Specifically, the author carries on her Ph.D. research at the Human Language Technologies Group, in Fondazione Bruno Kessler within the FBK Ph.D. International Programm.

Abstract

This thesis focuses on the study of both lexical opposition between verbs (e.g. *to increase / to decrease*) and contrast at the textual level (e.g. 'The price of x *increased* by 5%, while the price of y *decreased* by 2.3%'). Intuitively, lexical opposition and discourse contrast have some overlap, as they involve the comparison of two elements that are similar in some aspects but different in others.

We adopt a data-driven approach and investigate opposition and contrast, as well as their co-occurrence, in the Italian language, observing the two linguistic phenomena in context. We aim at collecting both quantitative and qualitative data on opposition and contrast relations, useful for both linguistic research and computational purposes. We also aim at enriching and creating lexical resources and annotated corpora that encode the two relations.

In particular, towards the enrichment of a lexical resource with opposition, we define and test an annotation schema, and discuss methodologies for acquiring reliable data at the verb sense level, focusing on crowd-sourcing. We also provide evidence that different types of opposition are recognized by people.

Moreover, adopting a corpus-based approach, we focus on discourse contrast relations both when they are marked by textual connectives and when they are implicitly conveyed. We provide an analysis of the contrast relation in Italian in terms of frequency, explicitness/implicitness of the relation, and type of contrast, as a result of the annotation of a corpus of news (Contrast-Ita Bank). In addition to this, we present an investigation on the connectives of contrast in Italian for the creation of an exhaustive list of these lexical elements, starting from a lexicon of connectives we have created (LICO).

Finally, we analyze the connection between lexical opposition and discourse contrast. In particular, we investigate the contribution of opposition when in the context of a contrast relation, by observing their co-occurrence in a corpus. We also propose an exploratory research about how the co-occurrence of the two linguistic phenomena can be formally represented.

Acknowledgments

I would like to express my gratitude to Elisabetta Ježek and Bernardo Magnini for their continuous support of my study, for their motivation, and constant feedback during the entire time of researching and writing this thesis. They encouraged my initiatives, and always found time for discussing my ideas.

Besides my advisors, I would like to thank Professor Michele Prandi for his insightful comments on this thesis. I also acknowledge the anonymous reviewers of this thesis, whose indications will certainly be helpful for improving this contribution.

My sincere thanks also go to Martha Palmer who provided me the opportunity to join the CLEAR team at University of Colorado at Boulder, as a visiting student, for her stimulating suggestions on my research. I also thank the other members of the group, in particular Susan Brown, Timothy O'Gorman, Wei-Te Chen, and Julia Bonn.

I thank my PhD colleagues and friends at Fondazione Bruno Kessler: Lea, Erick, Lorenzo, Simone, Serra, Didem, Gözde, Ilia, Enrica, Shahab, Rajen, Mattia, Amin, Duygu, Josè, Prashant, Surafel, Nick, Rachele, Stefano, Giovanni, Alessio, and Giulio. I thank them for the stimulating discussions, for their suggestions, and for sharing joyful and stressful moments.

I also acknowledge all of the researchers in the HLT-NLP, HLT-MT and DH groups, in particular: Manuela Speranza, Anne-Lyse Minard, Mohammed R H Qwaider, Carlo Strapparava, and Luisa Bentivogli for sharing their experience with me; they have never turned down a request for help.

Finally I acknowledge Lorenzo and, especially, his patience, which I have challenged in many occasions during these last three years.

Contents

List of Tables

List of Figures

Introduction

1.1 Lexical Opposition and Discourse Contrast

This thesis focuses on the study of the opposition relation and of the discourse contrast relation. Intuitively, both these relations hold between contrasting elements: the first at the lexical level, the other at the discourse level. In this work we investigate these relations in the Italian language using a data-driven approach, and making use and developing lexical resources.

In order to clarify what we intend with *opposition*, we propose the following examples of opposition between verbs:

(1.1) a. *to pass - to fail* (an examination)

b. *to like - to dislike* (a person)

c. *to build - to destroy* (a building)

We define opposition as the relation between two lexical units that contrast with each other with respect to one key aspect of their meaning and that are similar for other aspects. In fact, both verbs in the pair *to pass - to fail* refer to the result of an examination, but they describe two possible opposite results. Similarly, *to like - to dislike* describe an attitude towards something, in one case this feeling is positive, in the other it is negative. In Example 1.1c, *to build - to destroy* represent a change of state between the "non existence" of the building and its "existence". In all the pairs, we can see that opposition combines similarity and dissimilarity.

The same examples also give us the chance to highlight two characteristics of opposition: i) opposition holds among the senses of the words: *to pass* in the sense of 'to pass an examination' is the opposite of *to fail* in 'to fail an examination', although this is not the case if we consider *to pass* in the sense of 'to pass the pen'; ii) there exist different types

of opposition: the opposite verbs in Example 1.1a indicate the two possible results of an examination and one excludes the other; the opposite verbs in Example 1.1b indicate two situations that do not exclude each other (e.g. you can not like a person, and also not dislike that person); and the opposite verbs in Example 1.1c describe a change of state in the opposite direction (these types of opposition are respectively called *complementaries, antonyms* and *reversives* by Cruse (1986), among other authors). The analysis in this thesis will consider both these characteristics of opposition.

The examples in 1.2 help us to introduce the concept of *discourse contrast*:

(1.2) a. Liza has blue eyes, while Alice has green eyes.

b. Although it was raining, we went to the beach.

c. Mary did the exam. John did not show up.

Discourse Contrast is a type of *discourse relation*: a relation between two parts of a coherent sequence of sentences, propositions or speeches (i.e., discourse)[1]. In this thesis, we consider as *discourse contrast*:

(i) cases in which one the two parts is similar to the other in many aspects but different in one aspect for which they are compared, as in Example 1.2a, where the two parts are 'Liza has blue eyes' and 'Alice has green eyes'; both refer to the topic 'color of the eyes', but with different values;

(ii) cases in which one part is denying an expectation that is triggered from the other part as in Example 1.2b, where 'going to the beach' denies the expectation that, since it is raining, one would stay home.

Contrast in text might be conveyed explicitly, by means of a lexical element, as *while* in 1.2a and *although* in 1.2b, or implicitly as in 1.2c. In this contribution, we will refer to lexical units that convey contrast explicitly as *discourse connectives*.

Intuitively, the concepts of opposition and discourse contrast presented here indeed have some overlap. Both relations involve the concept of comparing elements that have some aspects in common but are different under some respect.

Consider the following examples in which lexical opposition co-occurs with contrast:

[1] In the thesis we will use *discourse level* and *textual level* as synonymous, since we are studying the discourse contrast as a relation which holds among portions of text.

(1.3) a. The price of x *increased* by 5%, while the price of y *decreased* by 2.3.

b. I thought I *passed* the exam: I *failed* it!

Examples in 1.3 show that lexical opposition can be used in the context of a contrast at the textual level, both in the presence of discourse connectives that mark the discourse relation (such as *while* in Example 1.3a), and in the absence of such connectives (as in Example 1.3b).

In this thesis, we investigate the two linguistics phenomena and their co-occurrence in the Italian language, observing the two relations in context. In particular, starting from the definition of these linguistic phenomena (considering both theoretical contributions and annotation schemas in which the relations are defined), we aim at collecting oppositions and contrast relations providing quantitative and qualitative data useful for linguistic research and computational purposes. In this thesis, special attention is also paid to the enrichment and the creation of lexical resources.

1.2 Research Field and Methodology

This research project is being developed taking into consideration both theoretical linguistic studies and research in the field of Computational Linguistics[2].

We follow a corpus linguistics approach in the line of Sinclair (1991); Biber et al. (1998); Hanks (2004): the research is based on the analysis of linguistic data in their context of occurrence. This means that we analyze data in which the phenomena we are studying (opposition and contrast) are observable in context, which is of crucial importance for the acquisition of reliable data. This approach allow us to empirically study how these relations are realized in texts, discovering phenomena that otherwise would go unnoticed. Moreover, this data-focused approach allow us to verify linguistic classifications proposed in the literature for both opposition and contrast, and to evaluate whether examples of those proposed classifications are easy to identify and tag. Finally, the observation of the linguistic phenomena in corpora lead us to collect quantitative data over these relations (e.g., raw frequency or frequency in relation of other phenomena), which in turn will allow us to provide some interesting insights useful for both theoretical investigations

[2]In the thesis we will use the terms *Computational Linguistics* and *Natural Language Processing (NLP)* without distinction.

and applied purposes.

The field of Computational Linguistics in which this thesis has been developed, and the data-driven approach we decided to adopt, require methodologies and tools proper to the field, and the use of specific terminology.

We will see, for example, that one of the topics of this thesis concerns the *annotation of linguistic data*. For our purposes, we needed to use annotation tools which support (also with graphic strategies) the process of the annotation of data. Along with this, we learned to define task guidelines, so that other persons can contribute to the annotation following the same instructions. The annotated data, and linguistic data in general, then need to be organized and managed in order for the researcher to be able to retrieve the information that needs to be analyzed: this requires basic computational skills (e.g. the manipulation of TXT and XML files), which have not be taken for granted.

Other instruments that we used for the linguistic data analysis are statistical measures, for example for calculating how similar the annotations provided by two or more persons are: this is called *inter-annotator agreement* measure. There are several measures to calculate the *agreement* (such as, the Dice's coefficient and the Fleiss kappa) and some of them have been used in this thesis. In general, the *agreement* (which may include a reconciliation step in which the persons can explain their choices with the possibility of revising them) allows us to capture some insights into the annotation exercise concerning the clearness of the guidelines provided, or about how complicated the annotation is, and possibly, where these difficulties lie. The *agreement* also permits us to evaluate the expected performance of a system in doing the same operation: human judgments are in fact taken as the upper bound that systems can reach. Indeed, if a human agrees poorly with another human, then it is unlikely that an automatic system will agree more. With reference to a system, the group of statistical measures includes also those used for the evaluation of system performance; in other words, one needs some evaluation measures for understanding how good a system is in doing what it has been created for, and testing it against data for which the answer is known (a *gold standard*). Some of these measures, such as *precision* and *recall* will be mentioned and used in this thesis.

In this thesis we will also refer frequently to *linguistic resources*, used not only as a lexicographic product for linguistic analysis but also as knowledge bases for the training and testing of systems created to solve NLP tasks. For example, a lexical resource like

WordNet for English (which contains information for nouns, verbs, adjectives and adverbs) has been largely used for the task of automatically disambiguating the sense of a word. We will see in this thesis that the creation of a resource primarily involves the definition of a schema, which includes: the definition of the information that is going to be registered in the resource, the decision on how, practically, this information will be registered, and choosing how it will appear to users. The first aspect requires one to understand how to classify a linguistic phenomena and its characteristics, and define whether these characteristics are compatible with one another. The other aspect concerns the choices on how to label and organize the data in a structured manner. Many resources are written in XML files with a hierarchical structure (for instance, in a dictionary, the information for a lemma is contained in an entry that is at the same level as other entries; in an ontology of concepts, typically the broader concept is at the root and the more specific concepts are at the leaves).

Once the schema is defined, a resource needs to be populated with data. We will see that there are several strategies to collect data: some are more automatic than others. In general manual annotation is more reliable, but it requires a certain cost. Recently, crowdsourcing has also become very popular because it permits us to acquire a great amount of information in a limited period of time. Other strategies can rely on existing lexical resources, or on the use of statistical measures applied to large data (e.g. considering the frequency of a phenomenon, or calculating distributional information). The knowledge of this methodological background was necessary for the development of our work.

Linguistic data, annotated or collected with different strategies, can be further analyzed with tools, such as part-of-speech taggers (that analyze a word in its context and provide its part-of-speech) or parsers (typically for syntactic analysis and building syntactic trees). To learn which instruments for *text processing* are at our disposal, and to learn how to use them, is part of the expertise required for operating in the field of Computational Linguistics.

Given that this is the framework in which our work has been developed, we will put continual effort into clarifying the methodologies we are adopting for the annotation or the collection of the linguistic data, and their analysis, and into explaining the meaning of technical terms when introduced.

1.3 Research Goals

As mentioned before, in this work we investigate lexical opposition and discourse contrast in the Italian language. In this section, we define our goals, which are schematized in Figure 1.1.

1.3.1 Opposition

Opposition holds between the senses of two words: for instance, as we have seen for Example 1.1a, *to pass* is the opposite of *to fail* as long as it taken in its sense of 'to succeed (e.g., an examination)' and not in its sense of 'to give (e.g., a pen)'. At the same time, however, words sense disambiguation and lexical-semantic relation identification are not easy processes, not only for systems but also for human annotators. Moreover, we have also mentioned that there are different types of opposition: for instance, some opposites exclude each other (for instance, *to pass* / *to fail* an examination, Example 1.1a), while others do not (for instance, *to like* / *to dislike* a person, Example 1.1b). Taking into account these differences, various classifications of the opposition relation have been proposed in the literature (e.g. by Lyons (1977) and Cruse (1986)). Considering this background, we are interested in acquiring evidence about the different types of opposition by collecting multiple judgments on the opposition relations.

Another aspect we consider in this thesis concerns lexical resources for opposition. In the NLP field, lexical resources annotated with opposition are of crucial importance: they can be used as knowledge-bases in several tasks for providing reliable data. A survey on whether and how the opposition relation is represented in the most used lexical resources shows, however, that opposition is poorly encoded. Moreover, it shows that opposition is mainly registered using the term *antonymy*, and that this term indicates opposition in general: types of opposition are not distinguished. Considering this background, we aim at defining and testing a *schema* for enriching lexical resources with different types of opposition.

Concerning the acquisition of opposition relations among words (which can be used to create a dataset of opposite pairs or to populate already existing lexical resources), potential approaches include pattern-based methods (i.e., retrieving information from a corpus by first defining and then searching for textual patterns in which opposites often

occur) and distributional semantic approaches (based on the hypothesis that words that are semantically similar are expected to share many similar contexts). However, identifying opposition with pattern-based approaches does not guarantee full coverage, and to adopt distributional methods can result in obtaining noisy data (i.e., it is still a challenging task to discriminate between opposites and synonyms with this approach as they share many common features). On the other hand, crowd-sourcing has been used for the acquisition of reliable data in a relatively short period of time. In this thesis, we aim at evaluating whether the crowd-sourcing methodology is a valid approach for the acquisition of the opposition relation.

1.3.2 Contrast

Despite the importance that discourse contrast plays in textual understanding, we highlight a lack of empirical studies on discourse contrast in Italian. We are referring in particular to annotated corpora that can be used both for the investigation of the discourse relation based on empirical data (for instance, its realization and characteristics), and for developing methods and tools for the automatic identification and disambiguation of discourse relations (thus bringing benefits for NLP applications for which a discourse level analysis is required). For English, the Penn Discourse Treebank (PDTB, Prasad et al., 2007) is the largest and most used annotated corpora with discourse relations. Similar resources that take inspiration from the PDTB have also been created, for instance, for Czech, Chinese, Arabic, Turkish and Hindi. For Italian, the only attempt we know about is by Tonelli et al. (2010) that annotated discourse relations in a corpus of 60 conversational spoken dialogues in the domain of software/hardware troubleshooting. To the best of our knowledge, there is still no high coverage, non domain specific resource for discourse relations available for Italian. Given this situation, in the thesis we intend to examine how *contrast* is realized in Italian by collecting quantitative data from corpora.

A parallel investigation concerns the study of *discourse connectives* (e.g. *while* in 'Liza has blue eyes, *while* Alice has green eyes.') which indeed can help the automatic identification of contrast in texts. However, the automatic identification of a *connective* and the discrimination of the type of relation it conveys is not trivial. In fact, the same lexical unit can be a connective in some contexts while not playing that role in other contexts. Moreover, *connectives* can be polysemous and convey different relations in different con-

texts (e.g. *while* in 'Liza was singing, *while* she was walking.' conveys a *temporal* relation and not one of *contrast*). Furthermore, a survey of the panorama of resources for Italian shows that there is a lack of a lexicon (and no exhaustive repository) for connectives that collect quantitative data on connectives of contrast, or, more in general, that can be used for NLP purposes. Given these observations, we aim at collecting an exhaustive list of *discourse connectives of contrast* in Italian, and retrieve quantitative data on their use from corpora.

1.3.3 Co-occurrence of Opposition and Contrast

The third area of investigation concerns the co-occurrence of lexical opposition and contrast in corpora, and the representation of this co-occurrence. This analysis starts with the observation that these relations can appear together. As we have seen in the Examples in 1.3 ('The price of x *increased* by 5%, while the price of y *decreased* by 2.3.%' and 'I thought I *passed* the exam: I *failed* it!'), opposites in the arguments of a contrast relation can represent the element of "difference" that the relation of contrast requires. One of the goals of our work is to verify the co-occurrence of opposition in the arguments of a contrast relation (i.e., the two portions of text involved in a contrast relation). In the cases they co-occur, we intend to observe whether opposition actually participates to the discourse contrast phenomena, and we aim at investigating the formal representation of this co-occurrence.

In Figure 1.1 we summarize the goals we outlined in this section.

1.4 Structure of the Thesis

The thesis is structured as follows. Chapter 2 presents the background that introduces our data-driven contributions on the opposition relation at the lexical level. First, we discuss the concept of opposition; the definition we adopt is mainly based on the study of Lyons (1977) and Cruse (1986), among others. We present and define four types of opposition relations -*complementarity, antonymy, converseness*, and *reversivesness*- which have been discussed at length in that literature. In the same chapter, we propose an overview of contributions from the computational field for the study and the identification of opposition (e.g. pattern-based approaches and distributional models): these works provide

opposition

- ◦ acquiring evidence of the opposition classification proposed in the literature by collecting multiple judgments on the opposition relations
- ◦ defining and testing a schema for enriching lexical resources with types of opposition
- ◦ methodology overview and evaluation of the crowd-sourcing methodology for enriching a lexical resource with the opposition relation

contrast

- ◦ collecting quantitative data from corpora on contrast relations in Italian
- ◦ collecting an exhaustive list and quantitative data from corpora on the discourse connectives of contrast in Italian

co-occurrence of opposition and contrast

- ◦ verifying the co-occurrence of opposition in the arguments of a contrast relation
- ◦ investigating the contribution of opposition when in the context of a contrast relation
- ◦ formally represent the co-occurrence of opposition and contrast

Figure 1.1: Research goals: a summary.

interesting information on the relation itself and help in defining it. This is followed by a survey dedicated to investigate whether and how opposition is encoded in the most used lexical resources in the NLP field, e.g. WordNet 3.1 (Miller et al., 1990). We propose a detailed study also for Italian, since the focus of our investigation is the opposition among Italian verbs.

In Chapter 3, we describe our contribution for the identification of a methodology for enriching lexical resources with the opposition relation. In particular, we introduce the main motivations that brought us to investigate the annotation of opposites in lexical resources and we highlight the characteristics a lexical resource that encodes opposites should have. In the same section we present T-PAS (Ježek et al., 2014), an inventory of typed predicate argument structures for Italian verbs. T-PAS is the resource of reference for the investigations we will present.

In the same chapter, we define a *schema* for representing different types of opposition in a resource and its evaluation. Moreover, we present the collection of data we performed through a crowdsourcing experiment in order to populate T-PAS with opposition relations: first we discuss the reason that lead us to adopt a crowd-sourcing methodology and the characteristics of the methodology itself; then we present the acquisition of the data. Through this data acquisition, we indirectly collected judgments on four types of

opposition (*complementarity, antonymy, converseness*, and *reversivesness*).

In Chapter 4 we present some background that helps to understand the concept of *discourse contrast*. The first section is dedicated to *discourse relations* in general: we propose an overview of the two most influential theories on discourse relations (the Rhetorical Structure Theory (RST, Mann and Thompson, 1988), and Segmented Discourse Representation Theory (SDRT, Asher and Lascarides, 2003)). We also present the major annotated corpora that have discourse relations. In the section that follows, we focus on the contrast relation by comparing different definitions provided in the above mentioned theories and we describe how the contrast relation is defined in the Penn Discourse Treebank (Prasad et al., 2007), the major work of reference for discourse relations in the field of Computational Linguistics. The aim of this overview is to compare definitions of *contrast* in order to define a schema for the annotation of the relation in a corpus.

Chapter 5 is dedicated to the presentation of a data-driven investigation that focuses on: the contrast relation in Italian, the connectives of contrast, and the co-occurrence of contrast with opposition. More precisely, we present Contrast-Ita Bank: a corpus of 169 documents of news that we annotated with the relation of contrast. We present the schema of annotation and the choices we made in order to annotate contrast in the corpus (we mainly adopted the PDTB schema and guidelines), and we discuss the results of the annotation in terms of frequency and explicitness of the contrast relation. The analysis of the inter-annotator agreement we performed is also presented.

In the same chapter, we discuss the concept of *discourse connectives* and present the list of the connectives of contrast that we retrieved from Contrast-Ita Bank. This list represents an integration of LICO (Lexicon of Italian Connectives), the lexicon of connectives that we created, in which the list of connectives has been collected from other lexical resources. A deeper analysis of the most frequent connectives of contrast is also provided.

Chapter 6 investigates the co-occurrence of opposition and contrast. In particular, we show how frequently lexical opposites are found in the arguments of contrast relations in Contrast-Ita Bank, and we run a qualitative analysis of the data in order to investigate the contribution of opposition when in the context of contrast. We also propose a discussion about how their co-occurrence can be formally represented using the PDTB schema and the AMR format.

1.5 Main Contributions

This thesis contains several contributions concerning:

- empirical evidence about characteristics of opposition:
  - We provide evidence that different types of opposition are recognized: some types of opposition are mutually exclusive (i.e., there is a distinction between *complementarity* and *antonymy*), while *reversiveness* seem not to be an exclusive relation.
  - We provide evidence that opposition does not co-occur very often when contrast is explicitly marked by a connective, but it is more frequent when no connective is present. Limited to the cases we observed, we show that if opposites appear in the context of a contrast relation they frequently participate in it.
- a methodology for the acquisition of opposition relations:
  - We experiment with crowd-sourcing for enrichment of the T-PAS resource (Ježek et al., 2014) with different types of opposition relation: *complementarity, antonymy, converseness*, and *reversivesness*, demonstrating it is a valid option for the acquisition of the data.
- a corpus analysis about the contrast relation and the connectives that conveys it:
  - We provide quantitative data about frequency and explicitness/implicitness of the contrast relation, resulting from the annotation of the relation in a corpus of news (Contrast-Ita Bank).
  - We collect information on the connectives of contrast. Through the annotation of Contrast-Ita Bank we gather information on the frequency of these connectives, the type of contrast they convey, and their syntactic behaviour. We also show that some of them do not always play the role of connectives of contrast.
- the creation of freely available lexical resources:
  - We created Contrast-Ita Bank: a corpus of 169 documents of news annotated with the relation of contrast and the relation of opposition when occurring in

the portion of text involved in a contrast relation. Contrast-Ita Bank also represents an attempt to create a corpus annotated with discourse relations for Italian, compatible with the PDTB project (Prasad et al., 2007).

  - We created LICO, Lexicon of Italian Connectives, a lexicon in which connectives are listed together with orthographic, syntactic and semantic information, and examples of their usage.

- a proposal of the representation of the co-occurrence of opposition and contrast

  - We discuss whether the co-occurrence of opposition and contrast can be formally represented integrating the annotation of lexical opposition in the PDTB schema, or adopting the AMR format.

Some of the results presented in this thesis have already been published or have been submitted for publication in:

Anna Feltracco, Elisabetta Jezek and Bernardo Magnini. Enriching LICO with Corpus-based Data. Submitted to *The Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 7-12, 2018

Anna Feltracco, Elisabetta Jezek and Bernardo Magnini. Contrast-Ita Bank: A corpus for Italian Annotated with Discourse Contrast Relations. To appear in *Proceedings of the Fourth Italian Conference on Computational Linguistics (CLiC-it 2017)*, Rome, December, 2017

Anna Feltracco. T-PAS: costruire una risorsa per l'italiano basata sull'analisi di un corpus in *RiCognizioni*, 2016, Vol 3, No 6

Anna Feltracco, Elisabetta Jezek, Bernardo Magnini and Manfred Stede. LICO: A Lexicon of Italian Connectives, *Proceedings of the Third Italian Conference on Computational Linguistics (CLiC-it 2016), 2016 and Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian.* Final Workshop (EVALITA 2016), Napoli, December 5-7, 2016

Anna Feltracco, Simone Magnolini, Elisabetta Jezek and Bernardo Magnini. Acquiring Opposition Relations among Italian Verb Senses using Crowdsourcing, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pp. 2138-2144, Portorož, Slovenia, May 23-28, 2016

Anna Feltracco, Elisabetta Jezek, Bernardo Magnini and Simone Magnolini. Annotating opposition among verb senses: a crowdsourcing experiment, *Proceedings of the Second Italian Conference on Computational Linguistics (CLiC-it 2015)*, 2015, Trento, Italy, December, 3-4, 2015

Anna Feltracco, Elisabetta Jezek and Bernardo Magnini. Opposition Relations among verb Frames, *Proceedings of the Third Workshop on EVENTS: Definition, Detection, Coreference, and Representation, 2015,* pp. 16-24, Denver, Colorado, June, 2015

Chapter 2

Lexical Opposition: Background

This chapter presents the concept of opposition at the lexical level. An overview of the most important study on the topic in lexical semantics and in the field of Natural Language Processing is provided, followed by a survey dedicated to investigating whether and how opposition is encoded in the most-used lexical resources in Computational Linguistics.

We define *lexical opposition* as a kind of contrast relation between two lexical units that contrast each other with respect to one key aspect of their meaning and are similar for other aspects. Examples of opposition include the following pairs for English (2.1a, 2.1b) and the equivalents for Italian (2.1c, 2.1d):

(2.1) a. *to increase / to decrease*
b. *long / short*
c. *aumentare / diminuire*
d. *lungo / corto*

The members of the pairs in 2.1 share an element, i.e., movement along a scale (examples 2.1a and 2.1c) or being a description of a degree of length (examples 2.1b and 2.1d). They also have a key point of differentiation: directionality of the movement in one case, length above or below an average in the other.

In this chapter, we first present an overview of the most significant definitions for the relation of opposition which have been proposed in lexical semantic studies (Section 2.1.1). By doing this, we are defining the concept of *opposition* that will be at the centre of the next sections. We will then present and discuss the various types of opposition that have been identified in the literature (Section 2.1.2): we have empirically validated

the existence of these types through a crowd-sourcing experiment (presented in following chapter, Section 3.4). This study also bring us to define a scheme of annotation for types of opposition that can be added to a lexical resource (Section 3.2).

In Section 2.2 we propose an overview of contributions from the computational field for the study and the identification of *opposition*; these contributions provide interesting information on the relation itself and help us in defining it. Finally, Section 2.3 is dedicated to analyzing whether and how *opposition* is encoded in the most-used lexical resources in the area of Computational Linguistics. The survey shows that opposition is poorly registered in the resources examined.

2.1 Opposition in Lexical Semantics

This section is dedicated to clarifying the definition of opposition in lexical semantics. As we will explain, our notion of opposition is mainly based on the studies of Lyons (1977) and Cruse (1986, 2011), and is presented in Section 2.1.1. In Section 2.1.2, we also focus on describing the different types of opposition that have been proposed in the literature. A significant part of the investigation that we will present in Chapter 3 is dedicated to empirically validating whether these categories are perceived by speakers.

2.1.1 Defining the Opposition Relation

As mentioned in Introduction, we define *lexical opposition* as the relation between two lexical units that contrast each other with respect to one key aspect of their meaning and are similar for other aspects. According to Cruse (1986), two opposites are two lexical units that "typically differ along only one dimension of meaning: in respect to all other features they are identical" (Cruse, 1986, p.197). Similarly, Murphy and Andrew (1993) indicate that "words have opposites only if one dimension of their meaning is particularly salient, and if there is a word that is equivalent except for its value on that dimension" (Murphy and Andrew, 1993, p. 302). Thus, opposites combine similarity and dissimilarity in their meanings. In Lyons's words: "when we compare and contrast two objects with respect to their possession or lack of one or more properties, we generally do so on the basis of their similarity in other respects. [..] Oppositions are drawn along some dimension of similarity" (Lyons, 1977, p.286). In line with these definitions, Ježek

(2016) suggests that "paradoxically, the first process for identifying an opposition often consists in identifying something that the meanings of the words under examination have in common. A second step is to identify a key aspect in which the two meanings oppose each other". Notice also that opposition is a symmetrical relation, that is *to increase* is the opposite of *to decrease*, and viceversa.

Cruse (1986) defines opposition also as a case of *incompatibility*. In fact, opposites cannot be true simultaneously for the same entity at the same time; for example, a *price* cannot be said *to increase* and *to decrease* at exactly the same point in time. A basic test to identify an opposition is "It is both X and Y". Based on this test, "*The price is both increasing and decreasing" is ruled out as odd because *to increase* and *to decrease* are opposites in the sense of being mutually exclusive (Ježek, 2016). Notice, however, *incompatibility* is a property shared also by other pairs that are not opposites, as, for example, by the pair *dog - cat*. According to Lyons (1977), *incompatibility* (for which he adopts the general term of *contrast*[3]) involves, as *opposition*, the concepts of similarity; for instance, the incompatible terms *dog* and *cat* are similar in that they both refer to animals.

So, what distinguishes *incompatibility* from *opposition*? The peculiarity of opposition with respect to other forms of incompatibility is that opposition is a "dichotomous, or binary, contrast" (Lyons, 1977, p.279). In Cruse's terms, opposition is a form of *inherently binary* (and not be accidentally binary) *incompatibility* (Cruse, 1986).[4] The author explains the difference between *inheritance* and *accidentality* providing the example of *tea* and *coffee* (Cruse, 2011, p.144): these drinks can accidentally (that is, contextually) be the only two drinks offered in a meeting but they are not the only *logical* choices for drinks, so they do not have to be considered opposites. On the contrary, *to increase / to decrease* are opposites as they are logically -and not accidentally- the only two possible 'movements along a linear axis'.

According to Prandi (2017), an important characteristic of the pairs of lexical units that can be in a relation of opposition (e.g. *to increase / to decrease*) -contrary to others

[3]In this thesis, we are reserving the term *contrast* to indicate contrast as a *discourse relation*, thus at the discourse level and not at the lexical one. We will discuss *contrast* in chapters 4 and 5.

[4]According to Cruse (1986) this *binarity* is "notion independent", since (binary) opposites can also be found in the same notional area in which non-opposites are. To explain this idea, he provides the example of the opposite *black* and *white* for the area of colours which includes also *blue*, *red*. Lyons (1977) further specifies that these binary sets do not belong to some lexical domain or to area of the vocabulary.

that are not (e.g. *dog - cat*)- is related to their belonging (or not) to *(close) paradigms*. Members of *(close) paradigms* depend on their correlation "with the competing values organizing the same conceptual area" (Prandi, 2017, p.46): for example, the meaning of *to increase* is "competing" with *to decrease* in the same area they belong to (let's say, 'movement along an axis') and 'depends' on it since its *value* is recognized only if compared *to decrease*. On the contrary, members of *(open) series* (e.g. *dog* and *cat*) do not "depend" on the other member (i.e., *dog* does not affect the *value* of *cat*).[5] However, *close paradigms* do not include only opposition, but also *differential paradigms*, which "organize a homogeneous root meaning into a network of differential value through a set of differential dimentions" (Prandi, 2017, p. 46). Examples of differential paradigms include the paradigm composed by 'kill', 'stab', 'strangle', 'shoot', and 'poison' which organizes the root 'take somebody else's life' according to different a dimension such as 'use of an instrument' or 'required force' or 'type of injury'. Conversely, according to Prandi (2017), "opposition organizes a whole conceptual area around two polar values" and "no root meaning can be isolated independently of the correlation" (Prandi, 2017, p. 47).[6] Polarity is, for the author, evident in some cases (i.e., in cases which the author calls *binary paradigms*, e.g. or *good / bad*) but less apparent (although still present) when there are more possible values for the same paradigm (for example, *hot / cold* are the opposite poles of a scale that also includes *lukewarm* and *cool*). Notice that this idea is not dependent on the notion of *exclusivity* of the opposites that, as we will see in the next section, characterizes a certain type of opposition. For instance, as we just said, in the pair *good / bad*, polarity is more evident given the un-availability of other lexical values in the paradigms; conversely, polarity is less evident for *hot / cold*), for which, we find e.g. *lukewarm*. However, both the pairs *good / bad* and *hot / cold* are *non-exclusive*: something can be not good and also not bad, something can be not hot, and also not cold. Therefore, we can say that the characteristic of "occupying opposite poles" (Cruse, 1986) or "organizing a conceptual area in two poles" (Prandi, 2004) is, according to the authors, shared by all opposites, independently of the type of opposition to which they belong.[7]

[5]Prandi (2017) provides the example of the class of all flowers as *open series*.

[6]In order not to be confused, the author also specifies that the existence of a label for identifying the conceptual area (for example, *temperature* for *hot / cold*) "does not imply that the root meaning is accessible beneath the opposition and independently from it" (Prandi, 2017, p. 47).

[7]We believe that a clarification is required in order for the reader not to get confused by the term *polar* in the next sections. Even if *polarity* is a characteristic of all opposites, as we just mentioned, we will see in Section 2.1.2 that

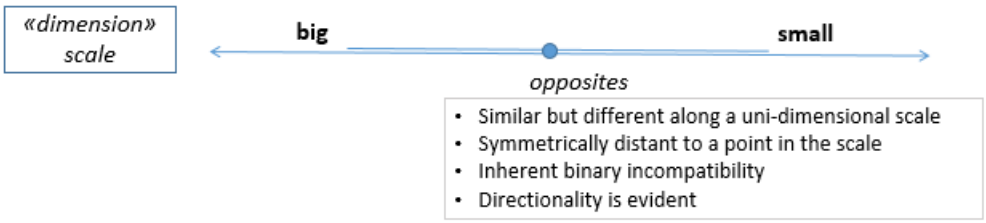

Figure 2.1: Characteristics of opposition in (Cruse, 1986).

Even if *binarity* is a property of opposition, it is not sufficient to define it: opposition is also characterized by *directionality* (Cruse, 1986). According to the author, opposite pairs "can be seen to embody a directional opposition" (Cruse, 1986, p. 261), such as a change in the opposite direction or a movement in opposite direction. For some types of opposition[8] this characteristic is more evident (e.g. in *to enter / to leave*), while it is more veiled for others (e.g. *present / absence*, for which, according to the author, the only possibility of change for either one is in the *direction* of the other). According to Cruse, in the case of opposites, *binary directionality* that forms part of the meaning of a pair of opposites also has to be evident, i.e., clear (to some degree). For example *Friday* and *Sunday* are the two elements in the set of 'one day removed from Saturday' (binarity) and they are in opposite direction with reference to *Saturday* (i.e., before and after) (directionality), but "their meaning in no way highlights a mutual orientation towards *Saturday*" (Cruse, 1986, p. 261). This binary directionality is definitely more evident, for example, in *to increase / to decrease* (as well as in the other examples in 2.1): according to Cruse (1986, p. 262), "the more patent a contained opposition is, the better examples of the category of opposites a pair of lexical items will be ".

Figure 2.1 schematizes the characteristics of opposites according to (Cruse, 1986).

According to Cruse (1986), however, "Good" opposition also depends on:

i how easy it is to identify a uni-dimensional scale in which opposites can be symmetrically disposed. For example, the scale of *long / short* is easy to identify: they are symmetrically disposed on the scale of *length*.

ii the portion of meanings that is exhausted by the opposition: the greater the portion of meaning that is exhausted in the opposition, the "better" is the opposition. Cruse

Pustejovsky (2000) uses *polar* opposites for indicating a type of opposites, i.e., *antonymy*, while Cruse (1986) himself uses the term *polar* for defining a subtype of *antonyms* in a more fine grained classification.

[8] In Section 2.1.2 we will present a classification of types of opposition.

(1986) proposes a comparison between *father / mother* and *man / woman*: the first pair is a weaker case of opposition, since the two terms in the pair capture the meaning of 'be a parent', which is not contrasted in the opposition;

iii whether the two terms match with respect to the "non-propositional" meaning. In this sense, *tubby / skinny* is a better pair of opposites with respect to *tubby / emaciated*, since these two belong to different registers.

The "goodness" of the opposition (also called "prototypicality") has also been discussed together with the concept of *canonicity* of opposition, that is, in a more cognitive perspective, how "good" two opposites are perceived to be[9]. Murphy (2003) for example distinguishes between canonical pairs (e.g. *hot / cold*) and non canonical pairs (e.g. *steamy / frigid*) being the second less common and more context dependent with respect to the former.

The concepts of "good opposites" and of "canonicity" are also brought into the discussion about the semantic or lexical nature of the opposition relation. For example, Jones et al. (2007) use the term "canonical antonym[10] pairs" to indicate when two words have been "learned as pairings of lexical units [...], not just derived by semantic rules". For supporting the hypothesis that opposition is "more related to the word form instead to word meaning", Gross et al. (1989) propose the example *big / little* in comparison with the pair *big / small*, the first being called "direct opposites" (*direct antonyms* in their terminology) and the second being called "indirect opposites" (*indirect antonyms* in their terminology). A pair of"indirect opposites" is composed by a member of a direct opposite pair, and by a synonyms of the other. Similarly, Miller et al. (1990) differentiate lexical opposition (*antonymy* in their terms) from conceptual opposition, discussing the pairs *fall / rise* and *ascend / descend*: the two pairs "seem identical in meaning, yet are distinguished by the way their members pick out their direct antonyms: rise/descend and ascend/fall are conceptually opposed, but are not direct antonyms" (Fellbaum, 1998, p.82), (Miller et al., 1990, p.50). In order to explain these examples, Murphy and Andrew (1993, p.305) use the terms "true antonyms" and "quite good antonyms, even if they are not perfect

[9]The distinction between protipicality and canonicity is discussed in Benotto (2014). For an overview of the discussion about *canonicity* of opposition and further investigation on this topic, see also Jones et al. (2007); Paradis et al. (2009); Pastena and Lenci (2016).

[10]The term *antonymy* has been used in literature as a synonym of opposites. However, in this thesis, we adopt the terminology of (Cruse, 1986) for which *antonymy* is a type of opposition; we will discuss this in Section 2.1.2.

ones".[11] Figure 2.2 schematizes the differences between "good opposites" and "less good opposites", according to the authors we have cited.

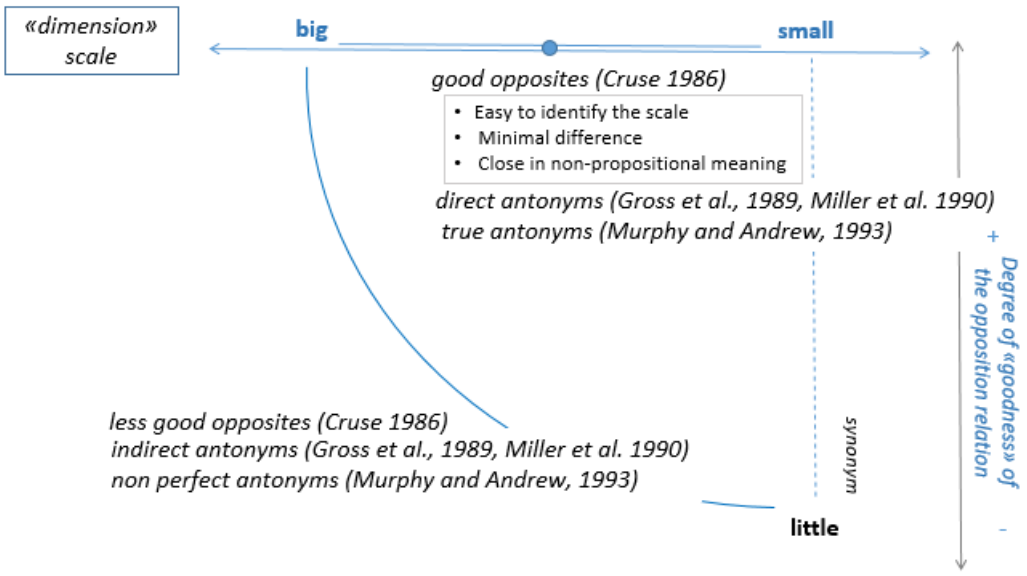

Figure 2.2: "Good" vs "Less good" opposites: a graphic representation.

It is thus an open discussion whether opposition is a lexical or semantic relation. What is clear is that the lexical unit that is considered opposite of another does not activate this relation for all of its senses. For example, the verb *to lose* can be considered opposite of *to win* as far as the former is considered in its sense of 'not to come first (in a competition)' and the latter in its sense of 'to come first (in a competition)' (Example 2.2a). The opposition relation does not hold if *to lose* is considered in its sense of 'to fail to keep (a wallet)' (Example 2.2b). Notice that in this case, the opposite would be *to find* (2.2c), which, similarly, is not a valid opposite for *to lose* in all its senses (2.2d): this means that it can happen that a lexical unit has an opposite if considered in one of its senses, and another opposite if considered in another sense[12].

(2.2) a. *to lose* (a competition) / *to win* (a competition) → opposition

b. *to lose* (a wallet) / *to win* (a competition) → no-opposition

c. *to lose* (a wallet) / *to find* (a wallet) → opposition

d. *to lose* (a competition) / *to find* (a wallet) → no-opposition

[11] Murphy and Andrew (1993) raise some objections against this model of *antonymy as a lexical relation between word forms and not a semantic relation between word meanings*. The reader can refer to their contribution for further discussion.

[12] Prandi (2017, p.48-49) discusses similarly how polysemy can affect opposition.

This also happens for Italian verbs. For instance, *abbattere* is the opposite of *costruire* as far as the former is considered in its sense of 'to destroy (a building)', and the latter in its sense of 'to build (a building)' (Example 2.3a). The opposition relation does not hold if *abbattere* is considered in its sense of 'to kill (an animal)' (Example 2.3b), for which there is no ontological opposite (Example 2.3c). This highlights that not every word (or every sense of a word) has an opposite.

(2.3) a. *abbattere* (un muro) - *costruire* (un muro) → opposition
b. *abbattere* (un animale) - *costruire* (un muro) → no-opposition
c. *abbattere* (un muro) - ?? (un animale)

Notice that opposites that are not actually used in their opposite senses can also be found in a unique sentence, as in Example 2.4 in which *to win* and *to lose* are not used in their opposite senses.

(2.4) Something always happens to him: last week he *won* the competition, today he *lost* his wallet in the gym.

It is clear that this poses interesting challenges for an automatic system that would, for instance, identify if two verbs in a portion of text actually hold a semantic opposition or not.

Moreover, opposition can be found among lexical units that are not opposite if considered in isolation (i.e., that "are not directly in opposition from an immanent lexical point of view"), but that instead are opposite in a specific context. According to Prandi (2017) this can be due to a variation in the meaning of a word in the context, not to the polysemy of the word.[13] Prandi (2017) proposes as an example the pair *happy / frightened*: the opposite of *happy* is typically *sad* but in cases such as "[she] was both frightened and happy", *happy / frightened* are perceived as contrasting[14]. In Section 6.2, we will present the annotation of some of these cases in a corpus of Italian documents.

One last remark that is worth mentioning concerns pairs of contrasting units that are not considered opposites. We have already seen that there is a certain agreement in recognizing that some opposites are "better" examples than others (e.g. "better vs less good",

[13]The author recalls Cruse (1986) and the notion of *modulation*: 'a single sense can be modified in an unlimited number of ways by different contexts, each context emphasizing certain semantic traits, and obscuring or suppressing others [...] This effect of a context on an included lexical unit will be termed *modulation*" (Cruse 1986: 52).

[14]The example "was both frightened and made happy thereby" is from Anna Karenina (Prandi, 2017, 49) .

"true vs non perfect", "direct vs indirect", see Figure 2.2). However, in the literature it is possible to find a definition for such pairs that do not satisfy the requirement of being symmetrically disposed on a uni-dimensional scale (see Cruse, 1986). Murphy and Andrew (1993) use the name *near-antonyms* to refer to pairs of adjectives such as *lukewarm / cold*, in which one adjective is the pole in the scale of temperature (*cold*) and the other is not (the other pole will be *hot*).

In their investigation, Mohammad et al. (2013) use the term "opposite" to indicate "word pairs with a strong binary incompatibility relation and /or that are saliently different across a dimension of meaning" (e.g. *hot / cold*). However, they also introduce the concept of *contrasting words* that are not opposite. They define as *contrasting* two words in a pair A:B for which there is a pair of opposites C:D such that A and C, and B and D respectively are strongly *related*. As an example, they propose the contrasting pair *tropical / freezing* related to the pair of opposites *hot /cold*, given that *tropical* is related to *hot* and *freezing* is related to *cold*. In more detail, they differentiate contrasting pairs from opposite pairs according to the degree of *binary incompatibility* and *difference across a dimension of meaning* they display. In the case of opposites (e.g. *hot / cold*), the degree of binary incompatibility is strong and the difference across a dimension of meaning is salient. On the other hand, word pairs that are contrastive have some "non-zero" degree of binary incompatibility and of difference across a dimension of meaning. For example, the binary incompatibility of *hot:cold* is stronger than the one of *tropical:freezing*, and their difference in the dimension of *temperature* is more salient.

In Figure 2.3, in which the opposites *big / small* are graphically represented on the scale of 'dimension', we also propose an example of 'contrasting words' (Mohammad et al., 2013); the same figure shows a case of *near-antonyms* (Murphy and Andrew, 1993). In the figure, we interpret the two definitions by identifying *modest* is as a *near-antonym* of *big* (i.e., *modest* is in the scale, but not symmetrically disposed with respect to *big*), and *large / little* as contrasting words being the first term related to *big* and the second to *little*.

2.1.2 Types of Opposition

Various classifications of types of oppositions have been discussed in the literature, even with some points of divergence. In this section, we will focus on four types which appear

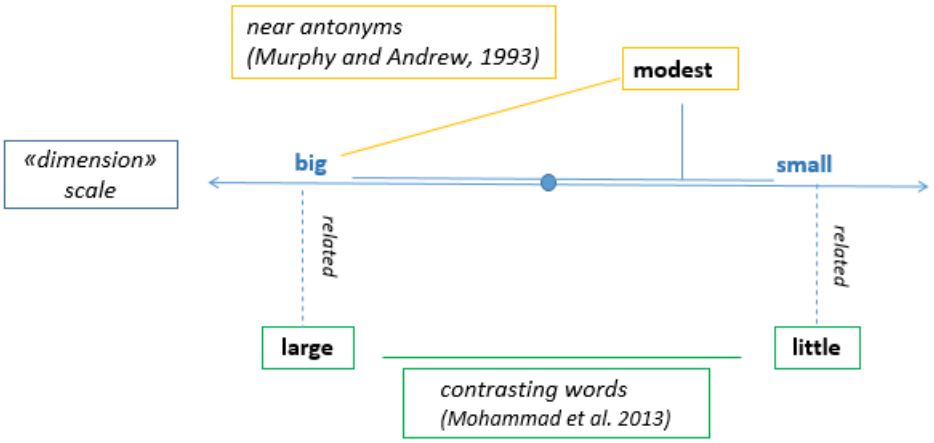

Figure 2.3: Near-antonyms (Murphy and Andrew, 1993) and contrasting pairs (Mohammad et al., 2013): our examples and graphic representation.

to recur frequently across the vocabulary and are also the most cited ones (Lyons, 1977; Cruse, 1986; Prandi, 2004, 2017; Ježek, 2016; Fellbaum, 1998; Murphy, 2010). The four types are: *complementarity, antonymy, converseness* and *reversiveness*; Table 2.2 (at the end of the section) summarizes their main characteristics. The examples of opposites that we provide in order to discuss these types of opposition mainly belong to the verb category, since opposition between verbs is the focus of the investigation we are going to present in the following chapter.

2.1.2.1 Complementaries

The first category we discuss is the category of *complementaries* (Cruse, 1986; Lyons, 1977), also called *ungradable opposites* (Lyons, 1977), *binary opposites* (Pustejovsky, 2000) or *exclusive opposites* (Prandi, 2017). Examples of *complementaries* are reported in 2.5:

(2.5) a. *to stay / to leave* ((in) a place)
b. *to succeed / to fail* (an examination)

Complementaries are opposites that "divide some conceptual domain in two mutual exclusive counterparts, so that what does not fall into one of the compartments must necessarily fall into the other" (Cruse, 1986, p.198). To understand this definition consider Example 2.5a: if a person is not leaving a place, necessarily it means that s/he is staying. This is because for *complementaries* "tertium non datur" (Prandi, 2004, 2017): there is

no intermediate degree between them. In fact, they are not poles of a continuum or a scale (as, for example, *big / small*).

Complementaries respond positively to the relationship "X is equivalent to non-Y", but also to "non-X is equivalent to Y": in the examples in 2.5, *to stay* is equivalent of not-*to leave*, *to fail* is equivalent to not-*to succeed*; at the same time, not-*to stay* is equivalent of *to leave*, not-*to fail* is equivalent to *to succeed*. For this reason, *complementaries* are also called *contradictory*: the assertion of one member of the pair implies and is implied by the negation of the other (i.e., *to stay* implies and is implied by not-*to leave*) (Cruse, 1986, 2011; Lyons, 1977). Similarly, Pustejovsky (2000) affirms that these opposites (that he calls *binary* opposites) refer to properties that are "inherently contradictory when applied to its natural predicate type" (Pustejovsky, 2000, p.455) (e.g. *male* and *female* when applied to "animate-gender") as "they exhaustively partition the property that they are sorts of"(Pustejovsky, 2000, p.456). From this, it follows that the simultaneous negation of both is a contradiction (i.e., not- *to stay* and not-*to leave* = contradiction) and the correlation between complementaries is a tautology (i.e., *to stay* or *to leave* = tautology) (Cruse, 2011).

Complementaries are generally not gradable with degree modifiers (e.g. **to succeed* a bit an examination). There are however some cases in which the members seem gradable. Cruse (1986, 2011) mentions the pair *clean / dirty*, discussing that *clean* entails *not dirty* and viceversa, but that it is possible to describe something as "slightly dirty" or "cleaner than before". According to the author, this is due to the fact that the two words are used in different senses: in the first, the adjectives refer to something that can be either clean or dirty (and even if it is not "that dirty", it is necessarily not clean); in the latter, *clean / dirty* are members of a gradable scale of "cleanliness" in which the two elements capture a certain degree of "cleanliness" (and in which *dirty* is more distant to *clean* than *slightly dirty*). Other cases in which *complementaries* are graded are the ones of the kind "more dead then alive" (in Italian: "più morto che vivo"): (Lyons, 1977) points out that in these cases we are not putting under discussion the ungradability of *dead* and *alive* but we are grading "secondary implications, or connotation" of the opposites.

Notice also that in some cases the relation holds perfectly (as the one in 2.5a: if one does not stay, s/he leaves; if one does not leave, s/he stays), in other cases, it can be more questionable. Example 2.5b for instance is an example of *complementarity* as far

as one can succeed or fail an examination. If, hypothetically, the exam is divided in two parts, and those who fail one part can integrate with an exercise after few days, then one could say that there is third option in which it is possible neither to succeed nor to fail the exam. Cruse (1986) proposes the example of *alive / dead* that are complementaries as far as the domain of "animate being" is concerned: if we include, for instance, ghosts and vampires, these can maybe considered nor alive, nor dead; in this case the pair *alive / dead* will succeed the "simultaneous negation" test (i.e., to be non *alive* and non *dead*) that *complementaries* usually fail. Thus, it is important to highlight that *complementarity* holds within a consistent domain (that however is not always easy to distinguish).

Cruse also notices that in case of verbs, there might exist a lexical unit that "contracts distinctive relation of oppositeness with one of the member of the *complementary* pairs" and that "express the necessary presupposition for the pairs of *complementaries*" (Cruse, 1986, p.201). For example, for the pair *to stay / to leave* (a place) in 2.5a, there exists *to arrive* that is an opposite (of the kind *reversiveness* that we will discuss in Section 2.1.2.4) of *to leave* and that expresses the presupposition for the pair (i.e., one can stay in a place or leave a place only if before s/he arrived in the place); for the pair *to succeed / to fail* (an examination) in 2.5b, the lexical unit is *to try*[15]. A similar consideration is made also by Fellbaum (1998) who speaks about a third "entailed verb".

2.1.2.2 Antonyms

Antonyms (Cruse, 1986; Lyons, 1977) are also called *gradable opposites* (Lyons, 1977). Examples of *antonyms* are reported in 2.6:

(2.6) a. *to increase / to decrease* (a price)
b. *to lengthen / to shorten* (some trousers)

Two antonyms oppose each other in relation to a scale of values for a given property. Instead of being exclusive counterparts and bisect a domain, as *complementaries*, they are two poles in the scale symmetrically distant from a neutral centre. This means that in the scale (which zero value is not reached by none of the two antonyms (Cruse, 1986)) there is range of values of the property between the two antonyms (Cruse, 1986), a residual space (Prandi, 2017). To understand this, Cruse (1986) suggests to think that there is a

[15]For a detailed study of which are the relations that can be involved in these "triplets" together with *complementarity* (such as *reversiveness*) see (Cruse, 1986, p.201).

neutral region of the scale (*pivotal region* in his terminology) around which the antonyms are symmetrical disposed (e.g. for example a neutral region in which you do not increase, nor decrease a price).

Antonyms are also said to "manifest the property of polarity more strikingly than other opposites" (Lyons, 1977, p.279). Pustejovsky (2000) refers to them using the term *polar* opposites[16]. According to the author, the polar adjectives "are measured as positive or negative values of placement on [the] scale [they are relative to]".

Antonyms are fully gradable. Consider Example 2.6a: increasing and decreasing a price denote a change of property that is gradual from a conceptual point of view (you can increase the price a bit or you can increase it a lot). The same is not possible, for instance, for the pair *to stay / to leave* in 2.5a: in fact you cannot "leave a bit"[17] and there is not intermediate value between *to stay* and *to leave*. This property of *being gradable* can be expressed by mean of a series of terms that identify very refined gradations of the specific property, for example with temperature: *freeze*, *cool*, *warm up*, *boil*. Potentially, along the scale *antonyms* are relative to, we could have many terms lexicalizing different degrees. In reality, the use of degree modifiers (such as *a bit* or *slightly*) to refine the concept is more common (Ježek, 2016); for example, we say "The weather warmed slightly". Thus, a standard test to identify antonyms given that they express a scalable dimension is "it is quite / moderately / [..] x, y" (Ježek, 2016, p.177).

According to Lyons (1977) gradable antonyms involve implicit or explicit grading; for instance, in 'our house is big' there is an implicit grading; the house is bigger than the average dimension for a house (i.e., for the element we are considering). Cruse (1986) specifies that the second implicit element in the comparison is "some sort of average value".

From a logical point of view, antonyms are said to be *contrary*, not contradictory (Lyons, 1977; Cruse, 1986): the negation of one term is not equivalent to the opposite term. For example, not-*increased* does not necessarily mean *decreased*. The joint negation of both is thus possible (Prandi, 2017; Lyons, 1977; Cruse, 1986; Ježek, 2016) and

[16]The author distinguishes *polar opposites* from *binary opposites* (which we called *complementaries*, see the previous related section). He underlines this distinction in a broader project that aims at incorporating what he calls "opposition structure" (that is a representation of the opposition) into the "event structure of the predicate's semantics" (Pustejovsky, 2000, p.457). See also Footnote 7.

[17]This has not to be confused with the duration of the action. It is possible to say "I stay a bit longer" but the action *to stay* is not gradable.

not paradoxical: e.g. not-*to increase* and not- *to decrease* is not a contradiction.

According to Cruse (1986), the relation is very frequent among adjectives and less frequent among verbs. Two verbs are antonyms when they are opposite and denote a change in property that has the characteristic of being gradual from a conceptual point of view.

Sub-types of antonyms. Cruse (1986) propose a more granular distinction of antonyms.[18] The author identifies three sub-types of antonyms, whose characteristics are summarized in Table 2.1 below: *polar, overlapping, equipollent*. *Polar antonyms* (e.g. *heavy/light*, *long/short*) are usually "evaluatively neutral and objectively descriptive" (Cruse, 1986). Both members have a *pseudo-comparative* correspondent; this means that both members x and y have a comparative form (x-er, y-er) that means "of greater x / of great y", or "more units of x / y"; for example in "This book is light, but it is heavier that the other", *heavier* means "of great weight" (*pseudo-comparative*). Their comparative forms are also in a converse relation: for instance, if A is heavier that B, then B is lighter that A, and viceversa. Only one member in the pair can be used in the question *how x is it?* and this question will be *impartial*, that is, it will not carry any presupposition, for example, "how long is it?", "how heavy is it?" carry no presuppositions about the length or the weight of an object.

Overlapping antonyms (e.g. *good/bad*, *safe/dangerous*) are not equally interpreted, i.e., one term is commendatory (e.g. *good, safe*), the other deprecatory (e.g. *bad, dangerous*) (Cruse, 1986). Just one member in the pair has a *pseudo-comparative*; the other will be a *true comparative*, that is x-er means "x to greater degree" with respect to something. For example, in the pair *safe/dangerous*, *safe* has a *pseudo-comparative* (e.g. "This path is dangerous, but it is safer that the other", safer = "of great safety"), and *dangerous* has a *true-comparative* (e.g. in "This path is more dangerous that the other", more dangerous = "more dangerous than.." but it is not possible to say *"This path is safe, but it is more dangerous that the other", that is more dangerous = *"of greater danger"). Both members can appear in the question *how x is it?* but one of them leads to an impartial question, the other to a non-impartial (*committed*) question: to ask "how bad is the movie?" is non-impartial since it presupposes that the movie is not good.

[18]This distinction is presented for the sake of completeness; we will not refer further to these subtypes in the thesis.

Antonyms		
polar (e.g. *long/short*)	**overlapping (e.g. *good/bad*)**	**equipollent (e.g. *hot/cold*)**
○ usually neutral and descriptive	○ one is commendatory, one is deprecatory	○ usually subjective sensations or emotions
○ both have a pseudo-comparative	○ one has a pseudo-comparative	○ both have true comparatives
○ both are impartial	○ one is impartial, one is committed	○ both are committed
○ one ok in "how x is it?"	○ both ok in "how x is it?"	○ both ok in "how x is it?"
○ "how x is it?" is *impartial*	○ one "how x is it?" is *committed*	○ both "how x is it?" are committed

Table 2.1: Scheme of the antonymy subtypes according to Cruse (1986).

Equipollent antonyms (e.g. *hot/cold*) usually refer to subjective sensations or emotions or psychological states (Cruse, 1986). Both members in the pair have *true comparatives* (e.g. "Today is hotter that yesterday" but not *"Today is cold, but it is hotter that yesterday"). Both terms can be used in *how x is it?* questions, and these questions are committed, since they carry a presupposition. According to Cruse (1986), some pairs of stative verbs enter in this subtype, such as *like/dislike*. They refer to psychological states and they are both committed in the "How do you x?" question (e.g. "How do you like him?", "How do you dislike him?").

Hay et al. (1999) propose a distinction among gradable adjectives considering the structure of the scale to which they refer.[19] The authors define a *scale* as a graphic representation of the dimension of meaning to which the opposites refer, and is considered a set of order points, in which it is possible to identify a set of positive and negative degrees.[20] Briefly, they distinguish between *open-range adjectives* (e.g. *long, wide*) for which a single scale is involved (similarly to Cruse's *polar antonyms*) and "it is not possible to identify maximal values on the scale" (Hay et al., 1999, p.135) (you cannot say "it is *completely* long"), and *closed-range adjectives* (e.g *straight, empty, dry*) for which it is possible to identify maximal values on the scale, "where maximality is relative to the adjective's polarity." (Hay et al., 1999, p.135) (you can say "it is *completely* empty").[21]

[19]Cruse (1986, 2011) distinguishes the subtypes also in terms of the different scales that underlie them. For a more detailed analysis of the subtypes of antonyms see (Cruse, 1986, p. 210).

[20]The authors adopt this model in order to support their analysis of *degree achievements* (e.g. *widen, lengthen, cool*), that are verbs which "describe an event consisting of a sequence of stages in which one participant exhibits different degree of the property expressed by the verb" (Ježek, 2016). They propose an analysis that is partially based on the inherent scalar structure of these verbs associated with their base adjectives (e.g. *wide, long, cool*).

[21]Adjectives in this latter group can have *true comparative* in Cruse's terms (e.g. emptier can just mean "emptier

The notion of "scale" has also been formalized in other dedicated studies. For instance, the characteristics for a scale of having or not having a bound is a distinction recalled by Hovav (2008) in her definition of *multi-point* scale. The author defines the scales that can be specified by verbs (often a scale corresponds to non-derived adjective), first distinguishing *two-point* scales from *multi-point* scales. In *two-point* scales a property can be associated or not be associated; the author provides the example of *to die*, that means to pass from having the property of "not being not death" to the property of "being death". On the other hand, in *multi-point* scale there are many values for the attribute, for example *widen* (e.g. *The river widen*) "is true if there is any increase in the value of wide"(Hovav, 2008, p.19). *Multi-point* scales are further distinguished according to whether they are bounded or not, leading to distinguishing between *flatten* (bounded scale) and *lengthen* (non bounded scale).

2.1.2.3 Converses

Converses (Cruse, 1986; Lyons, 1977) or *relational opposites* (Cruse, 1986; Prandi, 2017) are opposites that express a relation between two entities by "specifying the direction of one relative to the other along an axis" (Cruse, 1986). Examples of *converses* verbs are reported in 2.7.

(2.7) a. *to lend / to borrow* (an object)
b. *to precede / to follow* (something)

Converses are logically inseparable and interdependent (Prandi, 2017): each member in a pair of converses suggests the existence of the other. For example, consider the pair *to lend / to borrow*: if a person lends an object to another person, this entails that the second person (the receiver) is borrowing the object from the first person. This is why converse are said to be inherently *relational*: the meaning of converses involves necessarily a relation between at least two elements.

Another characteristic of two converse terms is that each one expresses the same underlying relation in the opposite way from the other, thus we can say that the relation is seen from two points of view (Ježek, 2016), and that, in case of verbs, they refer to the same action (Fellbaum, 1998). For example, consider Example 2.8: if John *lends* some-

than another" and not "of great emptiness").

thing to Mary (2.8a) and consequently "Mary *borrows* something from John" (2.8b), then John can say "I *lend* something to Mary" and Mary can say "I *borrow* something from John".

(2.8) a. John *lends* something to Mary;

b. Mary *borrows* something from John.

From Example 2.8, it is also possible to see that, given the appropriate grammatical changes (e.g. interchange of the roles of the participants), when one member of a pair is substituted for the other in a sentence, the new sentence can be made logically equivalent to the original.[22] If in 2.8a *to lend* is substituted for *to borrow* and the roles of the participants *Mary* and *John* are interchanged, the new sentence that will be obtained (that is 2.8b) is logically equivalent to the first. This is the same for two-place predicates, such as *to precede / to follow*: 'A precedes B' is equivalent to 'B follows A'. Cases in which the interchangeable component is an obligatory argument of the verb (as for *to precede / to follow*, in which subject and object are interchanged) have been also called *direct converses* (Cruse, 1986); on the other hand, cases in which the relation involves a central participant and a "peripheral" (Cruse, 1986, p. 234) one (e.g. John *lend* something [to Mary], where [to Mary] is "peripheral"), are called *indirect converses*.

According to Lyons (1977), *converses* are frequent in the social relation area (e.g. *doctor / patient*), in kinship relations (e.g. *father / son*) and in temporal and spacial relations (e.g. *above / below*). Comparative forms of antonyms are also converses (e.g. *bigger / smaller*) and so active and passive forms of the verbs (e.g. *to call / to be called*) (Lyons, 1977; Cruse, 1986). These examples follow the pattern described above: e.g. for *father / daughter*, "John is the *father* of Mary" is equivalent to "Mary is the *daughter* of John".

Cruse (1986) points out that *indirect converses* are sometimes felt to be relatively weakly opposed. He notices that for those pairs of converses, which on the contrary, seem to be good examples of opposites, a "movement in opposite direction" is perceived as a salient characteristic of the pair[23]. For example, for *to lend / to borrow* (that is considered

[22]Using other words, Fellbaum (1998) says that converses "refer to the same activity, but the thematic roles associated with them (SOURCE and GOAL) are mapped differently in the surface structure of the sentences in which they occur"(Fellbaum, 1998, p.81).

[23]For Cruse (1986, 2011) converse opposition is a subtype of what he calls *directional opposition*, together with *reversiveness* that is discussed here after converseness. This position is however not shared by e.g. Lyons (1977) which

a good example) the thing lent moves away from a participant towards the other. On the contrary, in *to buy / to pay* no movement in opposite direction is perceived and the opposition is felt more weak. To discern this "directional nature" is however pretty hard (Cruse, 2011).

2.1.2.4 Reversives

The last category we are discussing in this overview is the one of *reversives*, which denote motion or change in opposite direction (Cruse, 1986), and which, according to Cruse (2011), belong to the verbal category. Examples of *reversives* are provided in 2.9:

(2.9) a. *to tie / to untied* (a lace)
b. *to appear / to disappear* (of a ghost)

Reversive opposition is considered by Cruse (1986) a subtype of the more broad category of *directional opposites* (in which he includes also *converses*). Directional opposites are opposites that "owe their oppositeness to the fact that they encapsulate a movement in opposite directions." (Cruse, 1986, p.226). Examples denoting pure opposites directions are according to the author only adverbs or preposition such as *up / down*, *north / south*. Notice that these are defined with respect to a point of reference, given a base point (for instance, *north* direction is referred to the point "north pole", and *south* direction is referred to the point "south pole") or with respect to the orientation or motion of another object (*towards / away from* are opposite in relation to an object). In general, however, the notion of opposite direction needs to be interpreted abstractly (e.g. in case of the reversives *to tie / to untie* in which the change is not in a spacial direction). Also Lyons (1977) identifies a class of directional opposites, in which involves a "motion in one of the two opposed directions with respect to a place"; in the class the author includes *up / down*, *come / go*. According to the author, it is difficult to distinguished directional from the other three already mentioned types (Lyons, 1977, p. 281), since, as also Benotto (2014) suggests, members of these categories could be interpreted as directionally opposed as well.

In the class of directional opposites (Cruse, 1986) includes *antipodals* when one term represents an extreme in one direction along some salient axis, while the other denotes

does not include converses in *directional opposition*, but considers this as a fourth different type.

the corresponding extreme in the other direction (e.g. *top / bottom*, *always / never*); *counterparts* when one term represents the counterpart of the other in which the "defining direction are reversed" (Cruse, 1986) (e.g. *concave / convex*, *hill / valley*); and finally *reversives* that denote movement in opposite direction and are discussed here in detail since, as already mentioned, they are all verbs (Cruse, 1986).[24]

As just said, *reversives* denote motion or change in opposite directions. We can also say that one term indicates a change from a state to another and *viceversa*, for instance, *to tie* means to pass from 'being tied' to 'being untied', *to untie* means to pass from 'being tied' to 'being untied'. The process or the action can also be not precisely reversed for the two verbs (e.g. they possibly require different movements or different instruments); according to Cruse (1986, 2011) the manner of the process and details of the path that lead from a state to the other do not count: what is crucial is the initial and resulting states involved. The same author also highlights that these states are relative and not absolute, for example, *to accelerate* means to pass from a lower speed to a higher speed and *to decelerate* to pass from a higher speed to a lower speed, no matter the actual speed.

Cruse (1986) further divides the class of reversives into *restitutives* and *independent reversives*. *Restitutives* involve the return of a former state for one of the member, e.g in *to damage / to repair*, *to repair* denotes a change in a former state of 'being fixed, unbroken'. It is interesting to highlight that for pairs in this group, Fellbaum (1998) notices that the relation of opposition interacts with the one of lexical entailment in a systematic way (Fellbaum, 1998, p. 82); for example one can only *repair* something which has been previously *damaged*, thus *repair* something entails that this something has been *damaged*. Chklovski and Pantel (2004) suggest that in these cases opposition interacts with the *happens-before* relation, that is, a temporal relation.

The second group of *independent reversives* includes pairs as *to accelerate / to decelerate* that denote a change in opposite direction but the resulting state is not necessarily a former one. For example, I can accelerate and decelerate independently, there is no necessity for the final state of one of the two verbs to be a recurrence of a former state. Cruse (1986), however, specifies that there can be some expectations; the author provides the example of *lock / unlock*: even if logically independent, we tend to imagine that something that we unlock (e.g. a door) was actually locked in its former state.

[24] As mentioned, in Chapter 3, we will present our investigation on opposition among verbs in Italian.

Another characteristic of reversives is that they are positive to the test of "first-time interpretation of *again*" (Cruse, 1986). The "again-test" verifies the possibility of using unstressed *again* without the process denoted by the verb having happened before, but instead occurring for the first time. For example, it is possible to say "the engine started, then stopped again", even if the "stop" occurs for the first time.

As introduced, reversives are all verbs: mainly intransitive verbs whose subject undergoes the change of states (e.g. *the ghost appeared / the ghost disappeared*) or causative verbs whose subject is the agent that operates the change and the object undergoes the change (e.g. *to build a house / to destroy a house*). Cruse also suggests that while reversives verbs can derive from opposite adjectives (as to *lengthen / to shorten* derived from *long / short*), "the essential opposition underlying all reversives is not derived from any other opposition type:[..] it is an extension of the elementary notion of movement in contrary direction" Cruse (1986, p. 231).

Table 2.2 summarizes the characteristics of the four examined types of opposition.

Types and description	Test and example
COMPLEMENTARIES: ◦ divide a conceptual domain into mutually exclusive parts ◦ no intermediate degree between them ◦ the denial of X implies the assertion of Y	Negative to: "neither X nor Y"; Negative to: "It X moderately / lightly / a bit" Positive to: "X entails and is entailed by non Y" e.g. "*He was neither accepted nor rejected"
ANTONYMS: ◦ oppose each other in relation to a scale of values for a given property, of which they may specify the two poles ◦ are gradual	Positive to: "neither X nor Y"; Positive to: "It X moderately / lightly / a bit" e.g. "The water did not cool nor warm (up)"; e.g. "The weather has warmed moderately".
CONVERSES: ◦ describe the same action from an opposite perspective ◦ involve a relation between at least two elements ◦ can be two-place predicates (two elements involved) or three/four-place predicates (more elements involved)	If syntactical changes are adopted, they can be substituted without affecting the sentence meaning: e.g. a. Mary lends something to John b. John borrows something from Mary
REVERSES: ◦ denote reversive actions or events ◦ interact with lexical entailment (Fellbaum, 1998)	If x and then y, the situation returns the initial one. Positive to the 'again-test" e.g. "to build / to destroy", "to wrap / to unwrap"

Table 2.2: Types of Oppositions (Cruse, 1986; Lyons, 1977; Ježek, 2016; Murphy, 2010).

To distinguish different types of opposition is relevant for several reasons, and crucial for systems that aim at recognizing relations among larger portions of text. Consider for instance the following pairs of sentences:

(2.10) a. John did not remain at home
b. John did not leave home
c. The price did not increase
d. The price did not decrease

In these examples, 2.10a and 2.10b contradict each other but this is not the case for 2.10c and 2.10d. In fact, while 2.10a implicitly denies the truth of 2.10b (i.e., John actually left his home as he did not remain in it), 2.10c does not coincide with the negation of 2.10d and the two sentences can be jointly stated (i.e., The price did neither increase nor decrease). According to our definition *to remain / to leave* are *complementaries* (they are by definition exhaustive and the negation of one implies the affirmation of the second), while *to increase / to decrease* can be said to be *antonyms* and do not answer positively to this test.

Similarly, we have seen that *reversives* (e.g. *damage / repair*) can possibly capture also entailment (e.g. *repair* entails *damage*) and the information can be useful, for instance, for automatic systems in order to detect the entailment relation among parts of a text.

In conclusion, we believe that to know whether a pair of verb senses is holding an opposition relation and to know which type of opposition is held is relevant, also for automatic systems to capture the relation between the events the two verbs describe.

2.2 Opposition in Computational Linguistics

In the computational field, a number of works have been dedicated to the opposition relation using different approaches. In this section we will present, in particular, works dedicated to the verification of the co-occurrence hypothesis (i.e., opposites co-occur in windows of words significantly more often than chance) and the substitutability hypothesis (i.e., given a word, its opposite is frequently selected as sustitutable for the word in the same context), to the automatic identification of opposition (even discriminating opposition from synonymy), to the distinction of different types of opposition. These works highlight characteristics of the relation (Section 2.2.4) which we believe are important, and represent part of the background on the topic.

2.2.1 Verifying the Co-occurrence and the Substitutability of Opposition

The key idea of the *Substitutability Hypothesis* is that "the context in which one word occurs in a sentence may call to mind other words syntactically substitutable [..] Under

such hypothesis the greater the syntactic and semantic appropriateness of a substitutable word in a context, the greater the mental activation" (Justeson and Katz, 1991, p.2), that is, the more likely it is for this word to be selected as a substitute of the initial word. We believe that this is related to the characteristic of opposites of combining dissimilarity with similarity; as we have seen, opposites "typically differ along only one dimension of meaning: in respect to all other features they are identical" (Cruse, 1986, p.197).

The idea was tested for adjectives by Charles and Miller (1989) who extract sentences that contain one of the adjectives belonging to a pair of opposites, remove the adjective from the sentences leaving a blank space and ask people to fill in the missing adjective. The *Substitutability Hypothesis* predicts that participants will have no preferences in choosing the adjective or its opposites. However, the results show that in some contexts, only one of the adjectives is appropriate.

Charles and Miller (1989) also suggest that opposite words tend to occur together in the same context more frequently than chance (*co-occurrence hypothesis*) and they provided quantitative evidences that sentences including pair of opposites are "higher than expected".[25]

Many studies have been be conducted considering this *co-occurrence hypothesis* for opposition. For example, Justeson and Katz (1991) provide further support for the hypothesis focusing on adjectives of the English language, in three different dataset. They show that these adjectives are usually syntactically paired: they have been found, for example, in coordinate structures, such as "she felt *cold* and *hot*"(Justeson and Katz, 1991, p.11), or in contrasting identical noun phrases, such as "one of *low* anionic binding capacity and one of *high* anionic binding capacity" (Justeson and Katz, 1991, p.11).

Fellbaum (1995) tests the *co-occurrence hypothesis* for four parts of speech for a total of 47 noun, verbs, adverb and adjectives pairs, demonstrating that antonyms frequently co-occur.

Also Mohammad and colleagues (Mohammad et al., 2013) study the opposition relation in different parts of speech demonstrating that opposite words present a high tendency to co-occur in the same span of context when compared with randomly selected pairs. In particular, they calculate the *pointwise mutual information* (PMI, Church and

[25]For instance, they provide evidence that *big / little* appear in sentences together more frequently than expected. (Charles and Miller, 1989).

Hanks, 1990)[26] for a number of opposite pairs in a set (they call this set *opposite set*), average it, and compare this result with the one obtained for pairs in a *random set*, i.e., a set in which words are matched randomly (except for a frequency criteria). PMI for *opposites* results to be higher.

Also Jones et al. (2007) demonstrate that (*canonical* or *good*) opposites have a tendency to appear together in constructions such as "from X to Y " and "either X or Y", claiming that "the ways in which antonyms co-occur in text go beyond collocation"(Jones et al., 2007, p.150).

Notice that several of these works treat co-occurrence or substitutability as keys for indicating canonicity of the opposites (Charles and Miller, 1989; Jones et al., 2007) or for measuring the strength of their relation (Mohammad et al., 2013).

2.2.2 Automatically Identifying Opposition

Detecting opposition is a fundamental requirement for any approach that aims at a deep language understanding. A number works in the field of Computational Linguistics have been dedicated to the automatic identification of lexical relations, including opposition, using different methods. In this section we proposed an overview.

Pattern-Based Methods. Works that adopt a pattern-based approach to extract pairs of opposites from text include the work of Chklovski and Pantel (2004), who use this approach to retrieve pairs of opposite verbs from the Web. More in concrete, they define recurrent patterns (e.g. "either *X* or *Y*" for *antonymy*) specific for the following relations: *similarity* (e.g. produce-create), *strength* (e.g. permit-authorize), *antonymy* (e.g. buy-sell, live-die), *enablement* (e.g., assess-review), *happens-before* (e.g. enroll-graduate). Then, they extract verbs pairs from the web using these patterns and test the accuracy of the system against a gold standard. They obtain that in 43.8% of the cases the system was able to assign correctly a relation *antonymy* to a pair of verbs in a pattern.

In the previous section, we already mentioned the work by Jones et al. (2007). They demonstrate that opposites have a tendency for appearing together in specific construc-

[26]Briefly, PMI is a measure that indicates how significant is the occurrence of the two words, by considering their individual frequency and the frequency of their co-occurrence (occurrence in the same windows of words). For example, consider two words that occur very rarely in a document but the few times they occur they appear frequently close (e.g. a rare adjectives followed by a noun). In this case, their co-occurrence is significant, potentially more significant than the co-occurrence of two very frequent words that appear frequently with other words (e.g. an article followed by a noun).

tions such as "from *X* to *Y* " and "either *X* or *Y*" showing that pattern-based research is a useful methodology for retrieving pairs of opposites.

Lobanova (2012)[27] uses a limited set of opposites and identifies patterns of consecutive words containing these opposites (e.g. given *rich* and *poor*, she retrieved the span of text "is a mix of *rich* and *poor*", identifying the pattern "is a mix of *X* and *Y*"); then, she searches for all occurrences of the patterns in a corpus, in order to extract other opposites that fill the pattern. The retrieved pairs are further evaluated by judges.[28] She applies this methodology to the Dutch language.

Distributional Semantic Methods. A number of works have been dedicated to the automatic identification of opposition using distributional methods. Distributional methods are based on the *Distributional Hypothesis*(Harris, 1954) for which words that are semantically similar are expected to share many similar contexts: the more similar the contexts in which we found two words, the more semantically similar they will be.[29] Distributional semantic models (DSM, Turney and Pantel (2010)) are computational methods that use this hypothesis for semantic analysis. In turn, DSM have also been used for identifying semantically similar lexemes, and, since opposites are semantically similar (as discussed in Section 2.1), they have also been applied to identify opposition. Specifically, in DSM, words or phrases from a vocabulary are represented as *vectors* of real numbers. These are calculated on the base of their co-occurrences with other words in corpora; the result of the computation of all the words in the vocabulary is a *vector space*. The more similar the words, the closer they will be in the *vector space* (and, in turn, the higher their cosine similarity).

Adel and Schütze (2014) use vectors computed on co-reference chain (i.e., group of words that refer to the same entity/event, for example: 'John-president-man') to improve the performance of a system for the discrimination of opposite noun pairs from non-

[27]This study provides also a more complete overview on this approach and of examples of experiments on the identification of opposition with pattern-based methods.

[28]Specifically, patterns that were found more than once were automatically scored (according to the probability that the pattern contains a pair of opposite seeds: patterns that extract a lot of seeds receive higher scores). Based on the scoring of patterns, candidate word pairs were also automatically scored (the scoring was based on the number of times a pair was found in each pattern). Pairs with high score were evaluated by judges.

[29](Harris, 1954):" [..] if we consider words or morphemes A and B to be more different than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C. In other words, difference in meaning correlates with difference in distribution." The underlying idea, from a corpus linguistics point of view, is that the meaning of a word is always contextual (cf. Firth, 1957).

opposite noun pairs (i.e., a system that has to classify a pair as "opposites" or "non-opposites"). In particular, they improve the performance of a system that takes as input *text-based* vectors (i.e., vector representations trained on a large corpora) by adding also *co-reference* vectors (i.e., vector representations trained on coreference chain). They show that this helps in increasing the distance among opposite words in the representational vector space (their cosine similarity differ more strongly), improving the results. The authors suggest that this information is discrimitative for nouns, and less significant for adjectives: in fact, opposite adjectives with opposite meanings can occur in the same co-reference chain, but this is unlikely for opposite nouns to belong to the same co-reference chain (Adel and Schütze, 2014).

In general, however, contributions adopting distributional methods confirm that it is difficult to discriminate between opposites and synonyms (Mohammad et al., 2013; Santus et al., 2014b; Ono et al., 2015; Pham et al., 2015; Nguyen et al., 2016) as they share many common characteristics and, for both relations, the members of the pairs in the relation are distributionally similar. This mean that to identify opposites also requires a system to be able to distinguish them from synonyms. Works in this area include Mohammad et al. (2013); in their attempt to distinguish between synonymy, opposites and pairs of randomly-matched words, they observe that distributional similarity is not sufficient alone to determine whether two words are contrasting or synonymous.

For the discrimination of synonyms and opposites, Turney (2008) proposes a classifier, i.e., a system that has to classify a pair of word as "opposites" or "synonyms", trained on a set of labelled word pairs (i.e., pairs for which the relation is known). These are represented in the classifier as vectors. The vectors are created considering the frequency of the two words (in the examined pair) in textual patterns extracted from a large corpus. The system is tested on a set of 136 pairs of synonyms and antonyms, and achieves an accuracy[30] of 75%; however, always if the system would always classify a pair just considering the most common class (i.e., opposition or synonymy), the results in an accuracy of 65.4%.

Also Schulte Im Walde and Köper (2013) propose a method to distinguish between synonymy, opposition and hypernymy that takes advantage of the different lexico-syntactic patterns in which pairs of words in these relations are found in a corpus. First, they col-

[30]Accuracy is the number of of correctly predicted answers divided by the total number of answers.

lect pairs of synonyms, opposites or hypernyms via crowd-sourcing, and automatically acquire patterns in which they appear. Then, they generate a vector space model from these patterns. Using the training data, for each word class (noun, verb, adjective) they calculate three mean vectors, one for each lexical relations. Finally, they predict the relation for each pair in the test set, by calculating its vector representation and measuring which of the three main 'relation vectors' is more close to it (calculating the cosine similarity). They achieve an F-score[31] of 70.75% in distinguishing opposites from synonyms and of 61.25% in distinguishing opposites from hypernyms. They also show that the relations are more difficult to be distinguished in case of verbs, and in general, of polysemous words.

Other works combine DSM with information from lexical resources in order to distinguish opposites from synonyms. Ono et al. (2015) propose two models for the identification of opposites. The first model uses word vectors trained on information on opposites (*antonyms* in their terms) and synonyms from lexical resources (i.e., WordNet (Fellbaum, 1998) and Roget (Kipfer, 2005).): they tune the model in order for synonyms to have higher similarity, and antonyms and indirect antonyms (i.e., opposite-of-x - synonym-of-x, e.g. *garner* - *scatter*, being the first opposite of *disperse* and the second a synonym of *disperse*) to have lower similarity. The second model uses distributional information from a corpus. Results show the limitation of the second model in distinguishing antonyms, but evidence the crucial importance of adding information from lexical resources. The highest result is obtained by combining the two models (89% F-score versus 80% of the first model and only 7% of the second model).

Pham et al. (2015) introduce information from WordNet into a skip-gram model[32] in order to optimize the vector representation of words in a space and make the words to be closer to their WordNet synonyms than to their antonyms (same effect obtained by the above cited worked by Adel and Schütze (2014)). They prove the model to be successful (with respect to use the model without this information) in lexical tasks, including opposites and synonyms discrimination.

[31]The F score (F1 score or F measure) is a measure of a test's accuracy. It is defined as the harmonic mean of the *precision* and *recall*: where *precision* is the number of correct positive results divided by the number of all positive results, and *recall* is the number of correct positive results divided by the number of positive results that should have been returned.

[32]Briefly, a skip-gram model optimizes word vectors predicting the context of a given a word, that is approximating the probability of other words to occur in the context of a target word. For more details see (Mikolov et al., 2013).

Nguyen et al. (2016) add a weighted feature[33] to word vectors. This feature is calculated by taking into consideration that, given a target word *t*, some words that co-occur with *t*, co-occur also with synonyms of *t* but not with antonyms of *t*.[34] The idea behind this is that if a word *w* occurs frequently with a target word *t* but not with its opposites, then *t* is considered a significant feature for *w* (the feature would be weight with a high value); if *w* occurs both with *t* and with its opposite the feature score will be near zero (minimal difference); if *w* occurs only with the opposite of *t* and not with *t*, the feature will be evaluated negatively. Results show that including this weight helps in increasing the distance between opposites and synonyms in the representational space with respect to non considering this weight.

Santus et al. (2014a,b) propose a measure for the unsupervised discrimination of opposites from synonyms (i.e., they call the measure *APAnt*). The starting point of their works is to acknowledge, as recognized by other authors that we have mentioned in this section, that both opposite words and synonym words share salient dimensions of meaning but in opposites there is one dimension in which they contrast. For example, *size* is the dimension of meaning for which the opposites *giant* and *dwarf* contrast. Accordingly, authors hypothesize that there is a salient dimension of meaning for which opposites have a different distributional behaviour, and this dimension can help in distinguishing synonyms (that, instead, are expected to have very similar distributional behaviour) from opposites. In their works authors predict that synonyms share a number of salient contexts that is significantly higher than the ones shared by antonyms. They calculate the salience by computing the *Local Mutual Information*(Evert, 2005) [35] of the contexts for a word. Then they rank these salient contexts and compare this rank with the one of the other word in the pair, and calculate the intersection. The boarder and the more salient the intersection, the higher the probability of the two words to be synonyms. This measure proves to better in distinguishing opposites from synonyms with respect to both a vector

[33]Through a weighted feature it is possible to emphasize some data with respect to others. Technically, a value (a weight) can be assigned to each element in a group of elements depending whether or how it reflects a characteristic

[34]They derive information about which pairs are synonyms or antonyms from lexical resources. Specifically, WordNet (Fellbaum, 1998) and Wordnik `https://www.wordnik.com/`.

[35]The Mutual Information (MI) is a measure for the estimation of the association among two elements (*x* and *y)*, typically lexical units: it considers the observed frequency of the co-occurrence of two lexical units in a defined windows of units in a context ($f_{x,y}$), and it compares this frequency with the expected ones considering the individual frequency of the two units f_x and f_y. Here the formula: MI = $\log_2$ * [$f_{x,y}$ / (f_x * f_y)] The Local Mutual Information multiples MI for the frequency of the co-occurrence lexical units, LMI = ($f_{x,y}$) * MI.

cosine similarity measure and a co-occurrence measure (based on the already mentioned *co-occurrence hypothesis* by Charles and Miller (1989)).

Exploiting the co-occurrence of opposition with the *discourse contrast relation*. Other works have approached the identification of lexical opposition by exploiting the co-occurrence of opposition together with the *discourse contrast relation*. For example, in "John is tall but Bill is short" the antonyms adjectives *tall / short* appear in a sentence in which a relation of *contrast* is explicitly conveyed by the connectives *but*[36].

Re-implementing the model of Schulte Im Walde and Köper (2013), Roth and Im Walde (2014) take into account the frequency of the two words in co-occurring with some specific lexical elements (e.g. *but*) in order to distinguish between the relations of synonymy, opposition and hypernymy in German and in English. These specific elements are taken from a list of discourse connectives that convey discourse relations, e.g. *but* is typically a discourse connectives that indicates a *contrast* relation. The underlying hypothesis is that if two words frequently co-occur with a specific discourse marker, then the discourse relation expressed by the corresponding marker (e.g. *but* is typically a discourse connectives that indicates a *contrast*) should also indicate the relation between the words in the affected portions of text (authors related *contrast* to *opposition*). More concretely, they build a space model in which pairs of words are represented using as features the discourse markers[37] that appear between the two words. The incorporation of this information in the pattern-based model suggested by Schulte Im Walde and Köper (2013) increases the results both for the opposites-synonyms and for the opposites-hypernyms distinction[38].

On the other hand, other works highlight that lexical opposition is itself a cue for identifying discourse contrast and contradiction. These works will be presented in Section 6.1 when we will discuss more in detail how opposition is related to discourse contrast.

[36] As mentioned in the Introduction, the relation between the lexical opposition and the discourse contrast is a crucial point of our work and it will be brought in several times through the thesis. *Discourse contrast relation* will be defined in Chapter 4, the definition of *connectives* is provided in Section 5.2.1, and the co-occurrence of opposition and contrast is discussed in Section 6.

[37] The reader should not be confused by the terminology *discourse connectives* and *discourse markers*. We reserve the term *discourse connectives* for those markers that convey a discourse relation. We believe that the more general term *discourse markers* is properly used by Roth and Im Walde (2014), since when searching in a large corpus for these elements there is no guarantee that they always play the role of discourse connectives (e.g. *but* can be a noun phrase conjunction). More details on the definition of *connectives* and the distinction between *markers* and *connectives* can be found in Section 5.2.1.

[38] Results are not directly comparable to the ones by Schulte Im Walde and Köper (2013), as authors re-implemented the experiment. See Roth and Im Walde (2014).

Other methods. As we saw in Section 2.2.1, Mohammad et al. (2013) demonstrate that opposite words present a high tendency to co-occur; however, they advise that this is not a sufficient condition for identifying opposites. This is because other lexically related words co-occur together significantly more often than chance (for instance, as we mentioned, synonyms).

To distinguish opposites from synonyms, the authors take advantage of how information is organized in thesauri, which, indeed, " [..] are organized such that categories corresponding to opposing concept are placed adjacent to each other" Mohammad et al. (2013, p. 572). They report that, for instance, in the Macquarie Thesaurus for example *hiding* is in category 360 and *revealing* is in category 361. Relying on this and predicting that words pairs which are in the same thesaurus category are synonyms, while words which are in adjacent categories or paragraphs are opposites, they define some rules for distinguishing opposite from synonyms (e.g. "If the words in a pairs occur in adjacent categories, they are opposite"). They test this method on the same test set of 136 pairs of synonyms and antonyms used by (Turney, 2008) (see above), using the Macquarie Thesaurus, and obtain a F-score of about 90%.

2.2.3 Distinguishing Types of Opposition

The studies presented so far do not address the issue of distinguishing different types of opposites; an exception is represented by the above cited contribution by Mohammad et al. (2013). The authors take advantage of the crowd-sourcing methodology to collect information for classifying pairs of opposites into one of the different types of opposition the authors distinguish. Inspired by the work of Cruse (1986), they distinguish: *antipodals*, *complementaries*, *disjoint*, *gradable opposites*, *reversibles*. Observing the collected data, the authors find that there is a variation in the agreement among the participants when answering the questions; for example in the question "does X imply not to Y" the agreement is of 87.4%, while in the "are X and Y at two ends of a dimension?" question, the agreement is of 72.1%. They also verify that contrasting pairs can be classified into more than one type. In the same investigation, Mohammad and his colleagues use crowd-sourcing to determine the level of agreement on considering terms in a pair (adjectives, adverbs, nouns and verbs) as *contrasting* or *opposites*[39]. Annotators agree more in this

[39] As we explain at the end of Section 2.1, Mohammad et al. (2013) distinguish between *opposites* (i.e., word pairs that have a strong binary incompatibility relation with each other and/or are saliently different across a dimension of

task, with respect to type of opposition identification task.

2.2.4 Characteristics of Opposition Highlighted by Works in Computational Linguistics

In the previous sections, we presented contributions in the field of Computational Linguistics concerning lexical opposition. These works propose different approaches and certainly provide interesting results, for instance, for the identification of opposition or its discrimination from other relations. They also give some insight on the relation itself.

We have seen that many works have proved (and have taken advantage of) the characteristic of opposites of co-occurring in the same context (that is, a pair of opposites, e.g. *increase / decrease*, appear in a window of words more frequently than a randomly combined pair of words, e.g. *increase / eat*) and in certain patterns (e.g. either *increase* or *decrease*). Moreover, these studies highlight the characteristic of opposites of sharing similar textual contexts.

Some of the referred works also highlight how *opposites* and *synonyms* are distributionally similar: they have both been proven to share similar contexts, and this in turn poses a challenge for systems that rely on distributional methods that aim at automatically discriminating the two relations.

The study conducted by Mohammad et al. (2013) in order to discriminate among types of opposition is the only one of its kind in the field of Computational Linguistics we have knowledge of. By gathering data using crowd-sourcing, the authors show that there is some degree of disagreement among people when they are asked questions about the types of opposition; they also demonstrate that contrasting pairs can be classified into more than one type, suggesting that there is not a clear-cut distinction among these types.

Finally, we mentioned some works that use discourse connectives of contrast as a feature for identifying opposition, highlighting the relation between contrast at the lexical level (opposition) and contrast at the discourse level (discourse contrast). We will return to this in Section 6.1.

meaning) and *contrasting pairs* (i.e., word pairs that have some non-zero degree of binary incompatibility and/or have some non-zero difference across a dimension of meaning).

2.3 Opposition and Lexical Resources: a survey

In this section, we examine whether and how the opposition relation is recorded in lexical resources frequently used in Computational Linguistics. First, we consider the most important resources for English, such as WordNet (Miller et al., 1990; Fellbaum, 1998) and FrameNet (Ruppenhofer et al., 2010). Then, we will propose a detailed study for Italian, focusing on how and to which extent opposition between verbs senses is registered in resources for this language. We are particularly interested in this analysis, since the focus of the data-driven investigation we propose in Chapter 3 is the study of opposition among Italian verbs. The overview shows that opposition is poorly encoded.

WordNet. The most widely resource used in the field of Natural Language Process is WordNet[40] (Miller et al., 1990; Fellbaum, 1998). WordNet is a large lexical database of English nouns, verbs, adjectives and adverbs, which are grouped into sets of synonym terms called *synsets* (the version 3.1 counts 117,000 synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations; *semantic relations* (e.g., *hyponymy*) hold between synsets (e.g., a relation of *hyponymy* holds between the two synsets $\{dog\}$ and $\{puppy\}$), while lexical relations (such as *pertains-to*) connect words (e.g. the relation *pertains-to* holds between *criminal* and *crime*). In WordNet, the label *antonymy* is used to identify the opposition relation, which is considered "a lexical relation between word forms, not a semantic relation between word meanings". As referenced in Section 2.1, Miller et al. (1990) and Fellbaum (1998) recognize that two words, such *rise* and *descend* can be conceptual opposites, but they are not considered (direct) antonyms since they lexically pick other words as their antonyms (i.e., *rise / fall* and *ascend / descend*). Notice that in WordNet, an adjective that is "semantically similar" (i.e., is related by the *similar to* relation) to one of the two members of a *direct antonym* adjective pair (e.g. *arid* is "similar to" *dry* that is a member of the opposite pairs *wet / dry*) is an "indirect antonym" of the other member (*arid* is "indirect antonym" of *wet*). Figures 2.4 and 2.5 show the results of the online research in WordNet 3.1. for *dry* and *arid* and their antonyms. In the figures, it can be also noticed that the relation is considered a *lexical relation*, since it holds between words (i.e., in the figure *W:*) and not among *synset* (*S:*).

[40] `http://wordnet.princeton.edu/`

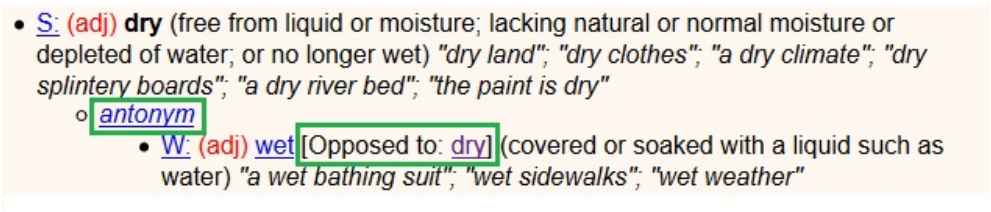

Figure 2.4: Example of opposition in WordNet 3.1 (Miller et al., 1990; Fellbaum, 1998): *dry / wet*.

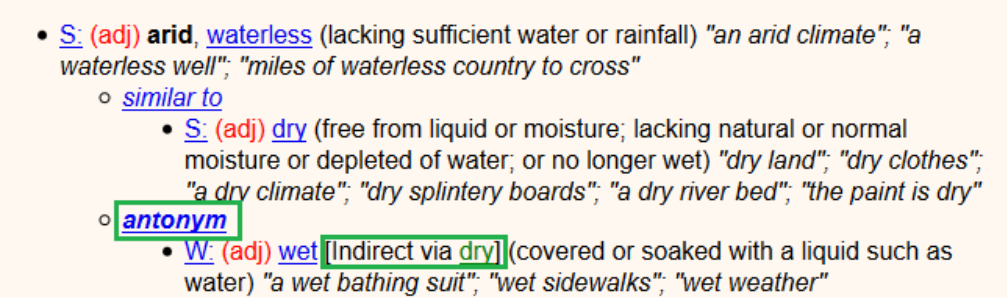

Figure 2.5: Example of indirect opposition in WordNet 3.1 (Miller et al., 1990; Fellbaum, 1998): *arid / wet* via *dry / wet*.

As we said, in WordNet opposition is referred as *antonymy*, but this word subsumes also the categories discussed in Section 2.1.2, i.e., complementaries (e.g., *succeed / fail*), converses (e.g., *buy / sell*) and reversives (e.g., *tie / untie*).

For what concerns the coverage of *antonymy* in WordNet, Lobanova (2012, p. 33) reports that "while previous studies agree that antonymy is not well covered in WordNet and many useful opposites remain uncovered in this resource, it remains to be the main source used in the experiments for identification and validation of good opposites".

FrameNet. FrameNet[41] (Baker et al., 1998; Ruppenhofer et al., 2010) is the most comprehensive corpus-based repository of semantic frames for the English language, developed according to the "Frame Semantics Theory" proposed by Charles J. Fillmore (c.f. Fillmore (1976)). Each frame is composed by frame elements that are similar to semantic roles participating in the situation described in the frame. For example, for the frame 'building', the frame elements are, for instance, 'Agent', 'Components', and 'Created_entity'; for the frame 'commerce_good-transfer', the frame elements are e.g. 'Buyer', and 'Seller'. Notice that frame elements are frame specific. The FrameNet lexical database contains over 1,200 semantic frames in 202,000 example sentences.

[41] `https://framenet.icsi.berkeley.edu/fndrupal/home`

In FrameNet, no relation of opposition is considered. However, the *perspective on* relation relates two frames describing the same basic situation from a different perspective; it seems to relate pairs of *converses* opposites (see Section 2.1.2). Figure 2.6 provides an example in which 'commerce_buy' and 'commerce_sell' are both related by the *perspective on* relation (via an arrow) with the 'commerce_good-transfer' frame. Indeed, they are two different possible points-of-view of this last frame since 'buying' is an event that only occurs when the situation (the "commerce_good-transfer") can also be described as 'selling'.[42]

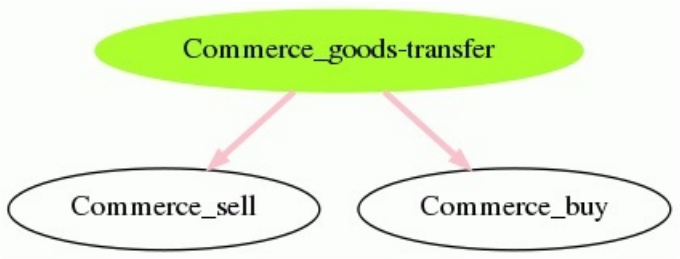

Figure 2.6: Example of *perspective on* relation in FrameNet (Ruppenhofer et al., 2010).

PropBank. The Proposition Bank (PropBank)[43](Kingsbury and Palmer, 2002; Palmer et al., 2005) is a large corpus in which the arguments of each predicate are annotated with their semantic roles. In particular, the predicate-argument relations are added to the syntactic structures of the Penn Treebank (Marcus et al., 1993). For each predicate in PropBank, it is also possible to explore its set of syntactic frames, each one associated with its set of roles. For example, for 'eat.01 (to consume) the identified roles are ARG0 (consumer, eater) and ARG1 (meal); for 'build.01' (to construct) the roles are ARG0 (builder), ARG1 (construction), ARG2 (material, start state), ARG4 (end state). A polysemous verb may have more than one frameset when the differences in meaning are distinct enough to require a different set of roles (Palmer et al., 2005).[44] There is no relation annotated between different predicates.

[42]The example is taken from (Ruppenhofer et al., 2010, p. 82).

[43]`http://propbank.github.io/`. PropBank has also been mapped to VerbNet and FrameNet as part of *SemLink* `http://verbs.colorado.edu/verb-index/`.

[44]The list of *frameset* is published here: `https://verbs.colorado.edu/propbank/framesets-english/`

VerbNet. VerbNet[45] (Kipper-Schuler, 2005) is a hierarchical domain-independent, broad coverage verb lexicon. VerbNet is organized into verb classes extending Levin (1993) classes that has been refined to achieve syntactic and semantic coherence among members of each class. Each verb class in VerbNet is described by thematic roles, selectional restrictions on the arguments, and frames consisting of a syntactic description and semantic predicates with a temporal function (Kipper-Schuler, 2005, p.4). As an example, we report the information associated to one of the specified frames for the class build-26.1.1 (that has 7 members, including *build*):

class:	build-26.1.1
roles::	Agent [+animate | +machine], Material [+concrete], Product, Beneficiary [+animate | +organization], Asset [+currency]
frame:	*NP V NP P.asset*
example:	The contractor builds houses for $100,000.
syntax:	Agent V Product {for} Asset
semantics	not(exist(start(E), Product)) exist(result(E), Product) made_of(result(E), Product, ?Material) cause(Agent, E) cost(E, Asset)

VerbNet does not account for lexical or semantic relations between the listed verbs.

VerbOcean VerbOcean[46] is the result of the investigation by Chklovski and Pantel (2004) that we presented in the Section 2.2.2. It is a broad coverage lexical resource that provides semantic relationships between verb pairs (i.e., it is also called "a semantic network of verbs"), in which the pairs are retrieved from the web using lexico-syntactic patterns (e.g. either X or Y). These patterns capture different relations (e.g. *similarity*, *happens-before*) including opposition (*antonyms* in their terms), for example *win / lose*.

MultiWordNet: Italian. As regards Italian, MultiWordNet[47] (MWN) (Pianta et al., 2002) is a multilingual lexical database in which the Italian WordNet is strictly aligned with WordNet (version 1.6). The Italian synsets are created in correspondence with the WordNet synsets for English, whenever possible, and semantic relations are imported from the corresponding English synsets. Thus, since, as mentioned before, *antonymy* is conceived as lexical relation in WordNet, it has not been imported: in MultiWordNet opposition is not encoded.

[45] `https://verbs.colorado.edu/~mpalmer/projects/verbnet.html`

[46] `http://demo.patrickpantel.com/demos/verbocean/`

[47] `http://multiwordnet.fbk.eu/english/home.php`

ItalWordNet. ItalWordNet (IWN)[48] (Roventini et al., 2000) is a lexical-semantic database structured in the same way as WordNet, namely around the notion of synset or set of synonymous word meanings. In ItalWordNet, relations are encoded between Italian synsets and the closest concepts in an Inter-Lingual Index, a separate module containing all WordNet1.5 synsets (but not the relations among them). Among the internal language relations that are listed in the annotation scheme also *antonymy* is considered, and has further been distinguished in COMPLEMENTARY_ANTONYMY and GRADABLE_ANTONYMY (however, the use of the more general ANTONYMY label is also allowed, according to the scheme). According to the authors, the relation is defined at the *synset* level: it is thus considered a semantic relation, differently from WordNet (in which, as we clarified in the dedicated paragraph, it is considered a relation among word forms, and not among word meanings). We have, however, no information about the coverage of this relation, and, as far as the on-line version is consulted, it seems not to be instantiated.

Senso Comune. Senso Comune[49] (Oltramari et al., 2013) is an open knowledge base available through a Web-based collaborative platform. The project has as stating point the De Mauro Paravia dictionary, that collected very frequent Italian words. The resource integrates data coming from both users and other resources. In *Senso Comune*, each sense of the lemmas is mapped to ontological categories. The annotation of the opposition relation appears not to be instantiated, even if the tag for the relation (*antonimia*) is present in the scheme.

T-PAS. T-PAS[50](Ježek et al., 2014), Typed Predicative Argument Structure, is a repository of verb patterns acquired from corpora by the manual clustering of distributional information about Italian verbs. For each structure, the specification of the expected *semantic type* (e.g. [[Human]], [[Event]]) for each argument slot in the structure is provided. Each typed structure represents a sense of the verb and it is associated with instances from the corpus from which it is manually retrieved. These structures are not linked by any relation. [51]

[48] `http://www.ilc.cnr.it/iwndb_php/`

[49] `http://www.sensocomune.it/`

[50] `http://tpas.fbk.eu`

[51] In Section 3.1.1 we will present in detail this resource that has been largely used in our investigation.

We provide an example of *t-pas* for the verb *costruire*. In the example, the *semantic type* [[Human]], [[Institution]] are specified for the position subject ([[Human]] and [[Institution]] are alternatives) and [[Building]] and [[Route]] for the object position of the *t-pas*#1 of the verb.

t-pas#1 of the verb *costruire*:
t-pas: [[Human | Institution]-subj] *costruire* [[Building | Route]-obj]
example: "Il muratore costruisce la parete."
(Eng. 'The bricklayer buildss the wall.').

LexIT. LexIT[52] (Lenci et al., 2012) is a large-scale resource providing automatically acquired distributional information about verbs, adjectives and nouns. LexIt provides statistical information, syntactic information registering the typical subcategorization frames in which the lemma occurs, and semantic information providing the selectional preferences (specified via semantic classes) of the lemma. For example, for the lemma *costruire* the resource registers ten subcategorization frames, being the most frequent the *subj#obj* frame for which 'Group' is the most frequent class for the subject position, and 'Artifact' the most frequent class for the object position. LexIT does not account for relation between these syntactic frames, which are not distinct on a semantic base.

The SIMPLE project. The SIMPLE project (Lenci et al., 2000) aims at adding a semantic layer to a subset of the existing morphological and syntactic layers for each of its entries. SIMPLE encodes semantic information for nouns, verbs, and adjectives, such as domains and semantic classes from an ontology. For 'costruire' (to build), they provide the following example (that we simplify):

unit: <build> //verb//
Predicative Representation: pred_ BUILD (<arg0>, <arg1>,<arg2>)
Selectional Restrictions: <arg0> = [Human], <arg1> = [Artifact], <arg2>:default = [Substance]

SIMPLE also encodes relations between word senses. Among these, the opposition relation (*antonymy*) is limited to adjectives and it has been defined by using two labels: *AntonymComp*, if the adjectives are complementary (e.g., *dead/alive*) and *AntonymGrad* if they are gradual (e.g., *hot/cold*).

[52] http://lexit.fileli.unipi.it/

IMAGACT. IMAGACT[53] (Moneglia et al., 2014; Panunzi et al., 2014) is a visual ontology of action that provides a translation and disambiguation framework for action verbs. Similar actions are grouped in concepts and related to the verbs that can be used to refer to them: so each verb is connected to a number of different action concepts (according to its semantic variation) and, at the same time, each concept is connected to a number of verbs. In Figure 2.7 we report the results of our search in ImagAct for the verb 'to cook', setting Italian as output language. As can be seen, the same concept is related to different verbs both in English and in Italian.

Figure 2.7: An example from IMAGACT (Moneglia et al., 2014; Panunzi et al., 2014).

The resource does not register when two verbs express "opposite" actions.

In this chapter we have defined the concept of opposition at the lexical level, presenting various contributions from the lexical semantics literature (in 2.1), and proposing an overview of research in Computational Linguistics which points out some characteristics of the relation itself. We also highlight that there are different types of opposition that have, however, received less study in the area of NLP, and are only partially registered in highly used lexical resources, both for English and for Italian; we have also shown (in 2.3) that the relation of opposition in general is poorly encoded.

The investigation that we will present in Chapter 3 will consider the definition of opposition we provided and the classification of opposites in the types we presented, in order to define and test an annotation scheme for tagging these types of opposition between opposite-sense verbs in a lexical resource. We will also present our methodology for the acquisition of pairs of opposites, including identifying the type to which they belong.

[53] http://www.imagact.it/

Chapter 3

Encoding Opposition in Lexical Resources through a Data-driven Approach

In this chapter, we describe our contribution for the identification of a methodology for enriching lexical resources with the opposition relation. First, we define a schema for representing opposition; then we present the collection of data we performed through a crowd-sourcing experiment in order to populate an Italian resource with opposition relations for verbs.

In this chapter, we describe our contribution for the identification of a methodology to enrich lexical resources with the opposition relation. In particular, we want the final resource annotated with oppositions to be reliable and to respond to specific criteria considering that i) opposition is among senses of words and that ii) there exist different types of opposition.

To do this we first define in this chapter a *schema* for representing different types of opposition in a resource (i.e., an abstract definition of the relations that we intend to represent in the resource), and then we propose a methodology to collect reliable data on opposition at large scale, with the goal of populating a lexical resource.

We chose to adopt an empirically grounded approach: the schema is in fact not only defined but also tested on available data, and the opposition relations were acquired by considering terms (*verbs* in our case) in context, and not just relying on abstract judg-

ments. More specifically, we gathered this data through crowd-sourcing collection, which relies on the identification of opposite verbs and of different types of opposition by observing pairs of opposites in context.

Through the same project, we also indirectly conduct a linguistic investigation on opposition based on empirical data. We provide empirical support that there exist different types of opposition relations with different characteristics, and that some of them do not seem to be exclusive. We also provide results concerning the agreement among people in identifying opposition between verbs in context.

After describing the data acquisition process, we discuss whether crowd-sourcing is a valid methodology for acquiring reliable data on opposition for Italian.

The chapter is structured as follows. In Section 3.1 we introduce the main motivations for the annotation of opposites in lexical resources and we highlight the characteristics a lexical resource that encodes opposites should have. In the same section, we present the T-PAS resource (Ježek et al., 2014), which we used in our data-driven investigations. In Section 3.2, we present the schema we defined and tested for annotating oppositions in lexical resources. We evaluated the schema by calculating the degree of agreement between two annotators on its application. In Section 3.3, we discuss the reasons that led us to adopt a crowd-sourcing methodology which we then adopted for collecting data. This acquisition is presented in Section 3.4 in terms of settings and results; a discussion over the methodology itself is provided at the end of that section as an additional contribution to this thesis.

3.1 Towards the Creation of a Resource for Opposites: Motivations and Desiderata

Lexical resources are largely used in the NLP community for the resolution of many tasks, such as the task of automatically disambiguating the sense of a word, or Word Sense Disambiguation (WSD) (Agirre and Edmonds, 2007); the task of automatically identifying a relation of entailment between two portions of text, or Recognizing Textual Entailment (RTE) (Dagan et al., 2009); the task of measuring the similarity between two portions of text, called Semantic Textual Similarity (STS) (Agirre et al., 2015), or the task of identifying which semantic relation (e.g. opposition) exists between aligned text chunks, known as Interpretable Semantic Textual Similarity (iSTS), (Agirre et al., 2016).

In Section 2.3, we have seen that the most used resources in the NLP community are scarcely populated with lexical opposition, and in some cases they do not encode the relation at all. If they do, opposites are not distinguished by the types of opposition they belong to; instead, they are grouped into one category, frequently under the name of *antonyms* (e.g. in WordNet (Fellbaum, 1998)). Other lexical resources, such as synonym and antonym dictionaries, list semantic opposition using the term *contraries* (e.g. in Dizionario dei Sinonimi e dei Contrari - Rizzoli Editore[54]). This highlights that the types of opposition are in general not distinguished.

In Computational Linguistics, a number of datasets have also been created as extremely useful knowledge-bases, e.g. for training systems for solving NLP tasks. Datasets that account for the opposition relation include, for example, the one created by Schulte Im Walde and Köper (2013). As seen in Section 2.2, the authors collect synonyms, opposites and hypernyms for building the dataset they use for distinguishing paradigmatic relations in German. Inspired by this work, Benotto (2014) collects information for different relations using crowd-sourcing; the people that collaborated were asked to propose synonyms, opposites and hypernyms for some target terms across word categories (noun, adjective, verb). To generate pairs of opposites or "contrasting word pairs"[55], Mohammad et al. (2013) propose a system that relies on existing lexical resources[56] and affix patterns (e.g. *X*:*unX* $\rightarrow$ *tie*:*untie*). Notice that with this method, the opposition relations are not collected at the sense level of the lexical units, and the type of the opposition relation is not considered.

Additionally, pattern based approaches (see Section 2.2.2), in which it can be claimed that the verb is disambiguated in the pattern, have been used for extracting opposition, for example, by Lobanova (2012)[57]. Once again, these datasets do not provide information about the senses of the terms involved in the opposition: still, one would need to run a WSD system in order to understand, for instance, that the sense of *to lose* when contrasting with *to win* is not the same as that contrasting with *to find*.

[54] `http://dizionari.corriere.it/dizionario\_sinonimi\_contrari`

[55] As mentioned in Section 2.1.1, according to Mohammad et al. (2013), *contrasting words* are two words in a pair A:B for which there is a pair of *opposites* C:D such that A and C, and B and D are strongly related. For example, *hot* and *cold* are opposite and *tropical* and *freezing* are *contrasting words* since *tropical* is related to *hot*, and *freezing* is related to *cold*.

[56] These are: the Macquarie Thesaurus and WordNet (Fellbaum, 1998).

[57] Examples of patterns for detecting opposition are "either X or Y" (Chklovski and Pantel, 2004) or "is a mix of *X* and *Y*" (Lobanova, 2012). See Section 2.2.2.

For what concerns the types of opposition, as we said in the Section 2.2.2, the only contribution we have knowledge about is Mohammad et al. (2013). The authors retrieve pairs of opposites in English from WordNet and collect information via a questionnaire for classifying these pairs into one of the different types of opposition they distinguish: *antipodals*, *complementaries*, *disjoint*, *gradable opposites*, and *reversibles*. In their questionnaire, however, the two terms under examination were shown to the people isolated, out of context, and are not linked to any definition; that is, with no information about which sense of the term was intended. In other words, they obtained a dataset in which the types of opposition are distinct but the terms are not disambiguated.

Unlike these datasets, we intend to tag opposition at the sense level among terms for which a sense repository is available and, most importantly, linked. That is why we propose to enrich a pre-existing lexical resource in which the registered terms are already disambiguated. We believe that to annotate opposition at the sense level is important since the same term can (or cannot) hold this relation with a number of different terms, depending on which of its senses is considered (see Section 2.1.1). Moreover, we would like the types of opposition to be distinguished in the resource. As already clarified at the end of Section 2.1, systems that aim to recognize relations among portions of text can benefit from this specification.

In the next sections we will investigate a methodology for reliably enriching lexical resources with opposites taking into account that opposition is at the sense level, and thus needs to be encoded at the sense level, and that there exist different types of opposition. We will concentrate on the Italian language and, as already announced, we will focus on verbs, a very polysemous class in the lexicon. We apply our methodology to the T-PAS resource (Ježek et al., 2014), an inventory of typed predicate argument structures for Italian verbs acquired from corpora, conceived for both linguistic research and computational tasks. In this resource, each structure corresponds to a distinct verb sense. In other words, in T-PAS verbs senses are distinguished, and this makes it possible to annotate opposition at the sense level. We can also take advantage of the corpus instances associated with each structure (i.e., each sense of the verb) in the resource. These instances represent a rich set of information and benefit the interpretation of the different senses of the verbs. A detailed presentation of the resource is provided in the next section.

3.1.1 T-PAS: a Corpus-based Lexicographic Resource for Italian

In this section, we present the T-PAS resource (Ježek et al., 2014). This resource, which the author of this thesis contributed to develop, is the reference resource for the investigation that is presented along this chapter. Specifically, first, we describe the resource itself providing some data about its current version; then, we draw some comparisons with other resources; and finally we illustrate recent applications of the resource.

3.1.1.1 An Overview of T-PAS

T-PAS[58], Typed Predicate Argument Structures, is a repository of verb patterns acquired from corpora by manual clustering distributional information about Italian verbs (Ježek et al., 2014). For each typed structure (henceforth *t-pas*s) the specification of the expected *semantic type* (ST) for each argument slot is provided.

An example of *t-pas* for the verb *abbattere* is given in 3.1. In the example, the STs [[Human]], [[Event]] and [[Building]][59] are specified for the position subject ([[Human]] and [[Event]] are alternatives for this position) and object ([[Building]]) of the *t-pas*#2 of the verb.

(3.1) *t-pas*#2 of the verb *abbattere*: (Eng. to destroy, to knock down)
[[Human | Event] - subj] *abbattere* [[Building]-obj]
example: "Il muratore abbatte la parete." (Eng. 'The bricklayer knocks down the wall.')

The resource has been developed following the lexicographic procedure called Corpus Pattern Analysis (CPA, Hanks, 2004; Hanks and Pustejovsky, 2005; Hanks, 2013), that is reported in detail in Figure 3.1. The procedure requires that the *t-pas*s are identified and defined by analyzing examples found in a corpus of sentences. After identifying the *t-pas*, the instances in the corpus are associated to the corresponding *t-pas*, which turns to be paradigmatic for the usage of that particular sense of the verb (Popescu, 2012). The *t-pas* acquisition is thus totally corpus-driven.

The STs in the argument positions (e.g. [[Human]] and [[Building]] in Example 3.1) are drawn from a list of 224 corpus-derived categories; these are semantic classes obtained by generalizing over sets of lexical items and are induced by the analysis of the

[58] `https://tpas.fbk.eu`

[59] Semantic types are indicated with uppercase and between square brackets.

T-PAS acquisition and ST tagging involves the following steps:

1 creation a sample of 250 concordances in the corpus for the verb

2 identification of the relevant syntagmatic structures while browsing the corpus lines, corresponding to the minimal contexts where all words are disambiguated;

3 identification of the ST for each argument slot of the structure by inspecting the lexical set of fillers. Each ST corresponds to a category from a provided inventory. If no ST captures the selectional properties of the predicate, a new ST can be proposed or list a lexical set, if no generalization can be done;

4 registration of the *t-pas* with a unique identification number and a description of its sense in the form of an implicature linked to the ST of the *t-pas*. The corpus lines are assigned the same number associated with the *t-pas*.

5 Concordances containing errors are tagged as *x*. Verb uses that do not come close to any of the normal *t-pas* are tagged as *u* (that stands for *unclassifiable*.

Figure 3.1: Step of the Corpus Pattern Analysis (CPA Hanks, 2004) for the *t-pas*s acquisition.

selectional properties of verbs.[60] For instance, in Example 3.1, [[Building]] generalizes over the lexical items: *parete, immobili, casa* (Eng. 'wall', 'buildings', 'house'). In the list, the STs are organized in a hierarchy in which the elements are linked by a "IS-A" relation; e.g. [[Plane]] IS-A [[Vehicle]], [[Vehicle]] IS-A [[Machine]] and so on, as can be seen in Figure 3.2 that reports a section of the hierarchy (Ježek et al., 2016).[61] In case no generalization is possible for a certain argument position in the structure, the lexicographer lists the set of lexical items found in the corpus for that position.

[[Inanimate]]
▷ [[Artifact]]
▷ [[Machine]]
▷ [[Vehicle]]
▷ [[Plane]]
▷ [[Road Vehicle]]
▷ [[Water Vehicle]]
▷ [[Train]]

Figure 3.2: Section of the T-PAS STs hierarchy.

T-PAS accounts for the following argument positions: *subject, object, indirect object, complement, adverbial* and *clausal*[62].

[60]The list has been created by manual clustering and generalization over sets of lexical items found in the argument positions in the corpus for about 1.500 verbs in English, Italian and Spanish. They are thus not derived from categories of entities defined on the basis of ontological axioms but are language-driven. (See Ježek et al., 2014). The same list has been used for the English resource PDEV (Hanks and Pustejovsky, 2005), `http://pdev.org.uk`.

[61]The hierarchy can be found in `http://pdev.org.uk/#onto`.

[62]For further details about the argument positions, see Cinková and Hanks (2010) and Maarouf (2013).

In the resource the *t-pas*s are semantically motivated: i.e., each *t-pas* corresponds to a distinct sense of the verb. For each *t-pas* a free-text description of its sense is provided (called *implicature*) anchored to the STs of the *t-pas* as far as this is possible (Cinková and Hanks, 2010). For example, the *implicature* for the *t-pas*#2 of the verb *abbatere* is following:

t-pas: [[Human | Event]] abbatte [[Building]]
implicature: [[Human | Event]] *demolisce, distrugge, butta giù* [[Building]]
(Eng. *destroys, knocks down*)

Figure 3.3 reports a more complete example of three *t-pas*s for the verb *abbattere*, the corresponding *implicatures* and some examples from the corpus. Each *t-pas* of the verb *abbattere* is a sense of the verb (*t-pas*#2: to demolish (a building), *t-pas*#3: to kill (an living being), *t-pas*#8: (referred to a weather event) to hit (a location)). We report in Appendix A the XML version of one of these *t-pas*s as it is registered in the resource.

Verb: *abbattere*

▷ ***abbattere*#2:**
- ▷ T-PAS:
  [[Human | Event]] *abbattere* [[Building]]
- ▷ IMPLICATURE:
  [[Human | Event]] *demolisce, distrugge, butta giù* [[Building]]
- ▷ **Annotated Corpus:**
  dovranno essere abbattuti in tutto altri tre immobili
  ..muratori occupati ad abbattere muraglie..

▷ ***abbattere*# 3:**
- ▷ T-PAS:
  [[Human]] *abbattere* [[Animate]]
- ▷ IMPLICATURE:
  [[Human]] *uccide, toglie la vita a* [[Animate]]
- ▷ **Annotated Corpus:**
  ..Kenai, il più giovane, *abbatte* l' orso...
  ...un bracconiere *abbatteva* un coniglio...

▷ ***abbattere*#8:**
- ▷ T-PAS:
  [[Weather Event]] *abbattersi* [[NO OBJ]] su [[Location]]
- ▷ IMPLICATURE:
  [[Weather Event]] *si rovescia* [[NO OBJ]] su [[Location]]
- ▷ **Annotated Corpus:**
  ...un acquazzone si abbatte sull'isola...
  Una pioggia insistente si abbatte sulla città.

Figure 3.3: *t-pas* 2, 3 e 8 for the verb *abbattere*.

The three components of the T-PAS resources are shown in Figure 3.4 and are: the list of the *t-pas*s; the inventory of the 224 corpus-derived semantic types (STs); the list of cor-

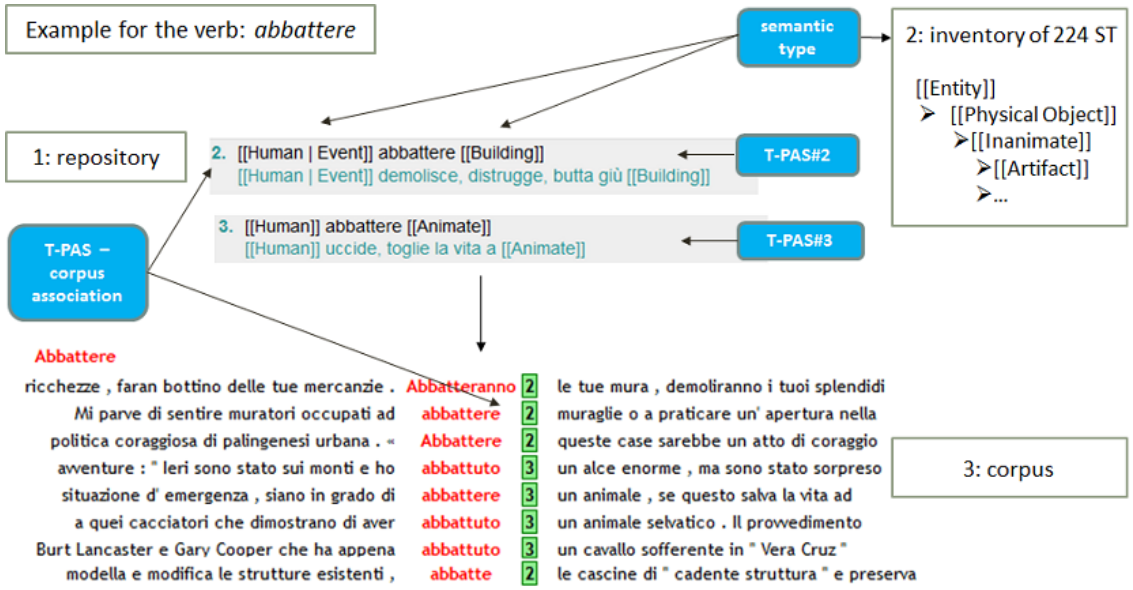

Figure 3.4: The components of the T-PAS resources.

pus sentences, whose observation is used for the *t-pas* definition, and that are associated to each *t-pas*.

3.1.1.2 The Current Version of T-PAS

T-PAS has been developed at the Department of Humanities of the University of Pavia, in collaboration with the Human Language Technology group of Fondazione Bruno Kessler, Trento, Italy and the technical support of the Faculty of Informatics at Masaryk University in Brno (CZ). The resource is distributed under a Creative Commons Attribution 3.0 license[63]. The current release of T-PAS (Ježek et al., 2014) contains 1000 verbs. These verbs have been retrieved from the total set of fundamental lemmas of Sabatini Coletti 2008 (Sabatini and Coletti, 2007), according their polysemy and with the following proportions: 10% 2 sense verbs, 60% 3-5 sense verbs, 30% 6-11 sense verbs.

The total number of *t-pas* identified is 4241; the minimum per verb being 1 and the maximum 20, as reported in Table 3.1. The reference corpus is a reduced version of ItWAC (Baroni and Kilgarriff, 2006).

At present, *t-pas* are stored in the resource as a list and they are not linked by any relation. However, in the following chapter, we will describe a first step for extending the resource by adding opposition relations between the *t-pas*s.

[63] `https://tpas.fbk.eu`

Total number of verbs	1000
Total number of T-PASs	4241
Minimum number of T-PASs per verb	1
Maximum number of T-PASs per verb	20

Table 3.1: Quantitative data on T-PAS.

3.1.1.3 T-PAS and Other Resources

The main resource of reference for the T-PAS project is the Pattern Dictionary of English Verbs[64] (PDEV) described in Hanks and Pustejovsky (2005), which is the main product of the CPA procedure applied to English. PDEV served as a main inspiration for T-PAS. The two resources also use the same hierarchy of STs.

Important resources to which T-PAS can be compared to are PropBank (Kingsbury and Palmer, 2002; Palmer et al., 2005), FrameNet (Baker et al., 1998; Ruppenhofer et al., 2010) and VerbNet (Kipper-Schuler, 2005), however they differ from T-PAS in several ways. PropBank is similar to T-PAS in being a verb-oriented resource that is based on the analysis and the tagging of a corpus. The list of frames for each verb in PropBank has also some similarity with the list of *t-pas* in T-PAS: e.g., they both provide semantic structure information and examples of the frame/structure. They, however, differ because PropBank lists semantic roles in syntactic frames, while T-PAS uses Semantic Types that generalize over the lexical items found in a corpus for that argument position. Also in VerbNet thematic roles are used in combination with a syntactic description of the verbs, with (similarly to T-PAS) a specification of the selectional restrictions on the arguments (e.g. *buy*, syntactic frame: AGENT verb THEME, where the role AGENT is [+animate | +organization]). However, in VerbNet frames are not verb-sense distinct, while in T-PAS each structure corresponds to a sense of a verb. Moreover, differently from T-PAS, frames are not acquired from corpora following a systematic procedure. This happen also in FrameNet in which example sentences are extracted from a large corpus and then assigned to the frames, while in T-PAS the structures are defined by observing sentences in the corpus.

As for Italian, a complementary project is LexIt (Lenci et al., 2012), that provides automatically acquired distributional information about verbs, and also information about semantic classes. Differently from T-PAS, in LexIt patterns are not semantically distinct,

[64] `http://pdev.org.uk`

do not address the argument-adjunct distinction, and the categories used for classifying the semantics of arguments are not corpus-driven. Finally, as affirmed by Ježek et al. (2014), inventory of senses such as MultiWordNet (Pianta et al., 2002) and Senso Comune (Oltramari et al., 2013) are resources to which T-PAS can be successfully linked with the goal of populating the former with corpus-driven pattern-based sense distinctions for verbs.[65]

3.1.1.4 Works with T-PAS

Several works have been proposed so far using T-PAS[66]. Popescu et al. (2014) present a probabilistic model for the disambiguation of verbs using T-PAS, and, in particular, the STs identified for each argument position in the *t-pas*s. Results show good precision.

Ježek et al. (2014) propose an experiment that evaluates the agreement between two expert annotators in grouping the same sample of corpus instances (250) of the verb into *t-pas*s that they create. The calculated agreement intends to measure the similarity among the two distribution. Results show a consistent overlap among the two distributions[67].

The author of this thesis and other collaborators (Feltracco et al., 2016a) attempt also at automatically annotating the lexical items in the argument positions (i.e., the items over which the ST generalizes) with the ST provided by the *t-pas* (for instance, assigning [[Building]] to *parete* in the sentence *Il muratore abbatte la parete* associated to the *t-pas* [[Human | Event]] *abbattere* [[Building]]). The experiment uses a mapping between the list of the STs and the lemma entries in MultiWordNet (Pianta et al., 2002) and exploits the structure of this latter resource.[68] The system shows good results in terms of precision.[69]

Della Moretta et al. (2017) contribute to expand the resource by manually tagging the examples from the corpus associated to each *t-pas* with the STs that generalize over the argument positions, thus replicating manually the attempt proposed in (Feltracco et al., 2016a). They report both quantitative data about the tagging and a qualitative analysis of

[65] All these mentioned resources have been presented in Section 2.3.

[66] For an overview, check the page: `http://tpas.fbk.eu/publications`

[67] For this project and the previously mentioned one, the author of this thesis collaborated as annotator of the T-PAS resource.

[68] The mapping is available `http://tpas.fbk.eu/publications` and has been further analyzed in Ježek et al. (2016).

[69] For this project, the author of this thesis collaborated in the definition of the system, she prepared the gold standard for the evaluation, and helped in checking the automatic mapping between T-PAS and MultiWordNet.

cases of disagreement between two annotators.[70]

Ravelli et al. (2017) evaluate a system for the mapping of T-PAS with IMAGACT (Moneglia et al., 2014; Panunzi et al., 2014) through a rule-based algorithm which converts argument structures in thematic roles. Results are good in terms of recall, while precision values are low.[71]

In the next sections, we present the investigations on opposition relation we conducted by making extensive use of the information in the T-PAS resource.

3.2 A Schema for Annotating Opposition among Verb Frames

In this section we propose a schema for annotating opposition relations among verb senses in lexical resources.[72] The motivations that lead us to run this investigation are multiple. First, as we saw in Section 2.3, lexical resources are largely used in the NLP community for the resolution of many tasks, but we verified that opposition is poorly encoded in many of the most used ones. Second, we believe that it is important to annotate opposition at the sense level and it is useful to distinguishing between different types of opposition.

In the computational field, several initiatives have proposed schemes for the annotation of the relations among events, including temporal relations as proposed in the TimeML scheme reported in (Pustejovsky et al., 2003), but less works have systematically addressed the annotation of the relation of opposition and its types in existing lexicons or sense repertoires.

Our goal is to define a schema of annotation that distinguishes between different types of opposition (Section 3.2.1) and test if it is reliable for the annotation of opposition between verb senses (Section 3.2.2). As we said, our resource of reference is the T-PAS resource (Ježek et al., 2014), described in Section 3.1.1. In the resources, each *t-pas* of a verb corresponds to a sense the verb, so the annotation of opposition between verbs at sense level is formalized as the annotation of opposition between the *t-pas*s of the verbs. To evaluate the schema, we calculated the degree of agreement between two annotators on

[70]For this project, the author of this thesis collaborated in the definition of the guidelines and in the analysis of the inter-annotator agreement.

[71]For this project, the author of this thesis collaborated in the creation of the gold standard, in the definition of the mapping rules, and in the evaluation and analysis of the results.

[72]Some of the results presented in this section are already published in (Feltracco et al., 2015a).

its application in the annotation of opposition between a number of verbal *t-pas*s (Section 3.2.2.1).

3.2.1 Defining the Schema for Annotating Types of Opposition

In the lexical resources in which opposition is encoded the terms *antonymys* or *contraries* are mainly adopted to indicate opposites in general (see Section 2.3). Exceptions in resources for the Italian language include the ItalWordNet schema (Roventini et al., 2000) in which the *antonymy* relation has further been distinguished in COMPLEMENTARY_ANTONYMY and GRADABLE_ANTONYMY (yet the two relations seem not to be instantiated), and the SIMPLE project schema (Lenci et al., 2000) in which, similarly, two labels are defined for adjectives: *AntonymComp*, if they are complementary (e.g., *dead/alive*) and *AntonymGrad* if they are gradual (e.g., *hot/cold*).

We expand this two-types classification. Among the various types of opposition that can be said to exist among verbs, we focus on the types described in Section 2.1: *complementarity*, *antonymy*, *converseness* and *reversiveness*; this four-type distinction is mainly based on the linguistic studies of Lyons (1977) and Cruse (1986, 2011).

Following this classification, we propose guidelines for the annotation of these different types of opposition, whose characteristics are summarized in Table 2.2, here reported as Table 3.2. The linguistic tests that help us in discriminating the types are the following:

- Complementarity (tag: COMPL). The standard tests to determine whether two words are *complementaries* are the following: “neither X nor Y”; “It X moderately / lightly / a bit”. The “neither X nor Y” test verifies whether it is possible to negate both terms simultaneously, and whether there is a neutral interval with respect to the two terms: complementary terms fail the tests because the opposition they encode is exclusive, in the sense that the assertion of one term entails the negation of the other (and *viceversa*); there are no intermediate cases (“*he was neither accepted nor rejected”; “*He neither failed nor succeeded”). It is not possible to negate both terms simultaneously. The second test verifies whether the terms of the opposition express a scalable dimension: complementary terms fail also this test. An example of COMPL is the pair *to accept / to reject* (a request).

- Antonymy (tag: ANT). The same tests just described can be used for *antonyms*.

Antonyms are positive to both tests. For example: "The water did not cool nor warm (up)"; "The weather has warmed moderately". An example of ANT is the pair *to cool / to warm* (of the water).

- Converseness (tag: CONV). *Converses* describe the same action from an opposite perspective with regard to the participant roles. If syntactical changes are adopted, converses can be substituted without affecting the meaning of the sentence. Converses can be two-place predicates, where two elements are involved or three/four-place predicates, where more than two elements are involved (in this case, one of the arguments can be omitted). An example of CONV is the pair *to lend / to borrow* (an object).

- Reversiveness (tag: REV). *Reversives* denote reversive actions; a test for reversible pairs is the "again-test", which verifies the possibility of using unstressed *again* without the process denoted by the verb having happened before (Cruse, 1986). An example of REV is the pair *to wrap / to unwrap* (a box).

Types and description	Test and example
COMPLEMENTARIES: ○ divide a conceptual domain into mutually exclusive parts ○ no intermediate degree between them ○ the denial of X implies the assertion of Y	Negative to: "neither X nor Y"; Negative to: "It X moderately / lightly / a bit" Positive to: "X entails and is entailed by non Y" e.g. "*He was neither accepted nor rejected"
ANTONYMS: ○ oppose each other in relation to a scale of values for a given property, of which they may specify the two poles ○ are gradual	Positive to: "neither X nor Y"; Positive to: "It X moderately / lightly / a bit" e.g. "The water did not cool nor warm (up)"; e.g. "The weather has warmed moderately".
CONVERSES: ○ describe the same action from an opposite perspective ○ involve a relation between at least two elements ○ can be two-place predicates (two elements involved) or three/four-place predicates (more elements involved)	If syntactical changes are adopted, they can be substituted without affecting the sentence meaning: e.g. a. Mary lends something to John b. John borrows something from Mary
REVERSES: ○ denote reversive actions or events ○ interact with lexical entailment (Fellbaum, 1998)	If x and then y, the situation returns the initial one. Positive to the 'again-test" e.g. "to build / to destroy", "to wrap / to unwrap"

Table 3.2: Types of Oppositions (Cruse, 1986; Lyons, 1977; Ježek, 2016; Murphy, 2010).

Figure 3.5 shows an example of the schema applied to the *t-pas*#2 of the verb *abbattere* which is considered to hold an opposition relation of type REV with a sense of *erigere* (*erigere*#1).

Verb: *abbattere*

▷ ***abbattere*#2:**

▷ T-PAS:

[[Human | Event]] *abbattere* [[Building]]

▷ IMPLICATURE:

[[Human | Event]] *demolisce, distrugge, butta giù* [[Building]]

▷ Annotated Corpus:

dovranno essere abbattuti in tutto altri tre immobili

..muratori occupati ad abbattere muraglie..

▷ **Opposition Relation:**

REV: *erigere#1*

Figure 3.5: Example of *t-pas*#2 for the verb *abbattere* annotated with the opposition schema.

3.2.2 Testing the Schema

In order to test the schema in the T-PAS resource, we designed and performed a pilot annotation; the annotation consisted in applying the schema over a selected set of 25 verbs pairs (the verbs in the pairs being annotated in T-PAS). Specifically, these 25 pairs had been selected because, according to human judgment, verbs in the pairs hold a relation of opposition for at least one of their *t-pas* (i.e., one of their senses). Consequently, such verb pairs were expected to present a high frequency of the relations the schema is designed for. The 47 verbs[73] in the 25 pairs (for a total of 216 *t-pas*s) are different in terms of polysemy and argument structure. In particular, with respect to the average polysemy of verbs in the T-PAS resource (equals to 4.25), 11 verbs have a high degree of polysemy in T-PAS (more than 6 senses), 28 have a medium degree of polysemy (between 3 and 5 senses), 8 have a low degree of polysemy (two senses). This proportion is similar to the one adopted for selecting the 1000 verbs that are analyzed in T-PAS. The average is 4.6 senses per verb, similar to the average in the entire T-PAS. 10 verbs require a direct object argument for all their uses, while six are used in absence of a direct object, 31 permit both structures.

We provided the annotators with the list of pairs of verbs and their respective *t-pas*s and *implicatures*. Moreover, annotated corpus-derived examples in T-PAS could be consulted.

As an example, Figure 3.6 reports the *t-pas*s and *implicature* of the 25 pairs selected

[73]Three of these verbs are repeated in two pairs.

pair 1

Verb: *abbattere*	**Verb:** *erigere*
▷ ***abbattere*#1:** [[Human ǀ Event]] abbattere [[Inanimate ǀ Plant]] *Implicature:* [[Human ǀ Event]] *fa cadere, manda a terra* [[Inanimate ǀ Plant]] *colpendolo* ▷ ***abbattere*#2:** [[Human ǀ Event]] abbattere [[Building]] *Implicature:* [[Human ǀ Event]] *demolisce, distrugge, butta giù* [[Building]] ▷ ***abbattere*#3:** [[Human]] abbattere [[Animate]] *Implicature:* [[Human]] *uccide, toglie la vita a* [[Animate]] ▷ ...	▷ ***erigere*#1:** [[Human]] erigere [[Building ǀ Artwork =Sculpture]] *Implicature:* [[Human]] *costruisce, crea* [[Building ǀ Artwork =Sculpture]] ▷ ***erigere*#2:** [[Human ǀ Rule ǀ Institution1]] erigere [[Concept1 ǀ Institution2]] {a ǀ in [[Concept2]]} ǀ { a ǀ in [[Institution3]]} *Implicature:* [[Human ǀ Rule ǀ Institution1]] *porta* [[Concept1 ǀ Institution2]] {a ǀ in [[Concept2]] } ǀ { a ǀ in [[Institution3]]} *di un grado superiore* ▷ ***erigere*#3:** [[Human]] erigersi [NO OBJ] {a [[Human Role]]} *Implicature:* [[Human]] *si innalza, si eleva* {a [[Human Role]]}

Figure 3.6: Example of one of the 25 pairs selected to test the schema.

to test the schema: *abbattere* and *erigere*.[74]

The two annotators, both familiar with verbal *t-pas* structures and their acquisition, were asked to identify and classify opposition relations between *t-pas*s following the annotation schema proposed in Section 3.2.1. For each given pair of verbs, the annotation task consists in two main steps:

1. for each pair of *t-pas*s created by coupling all the *t-pas* of a verb with all the *t-pas*s of the other verb (e.g. *abbattere*#1 - *erigere*#1; *abbattere*#2 - *erigere*#1, etc..), to identify the presence of an opposition relation;

2. if the opposition relation is present, to recognize which type of opposition occurs (ANT, COMP, REV, CONV).

Figure 3.7 displays two examples: one in which *REV* was recognized between *abbattere*#2 and *erigere*#1; the other in which no opposition was identified between *abbattere*#3 and *erigere*#1.

[74]In the Figure, Semantic Types are in square brackets with the capital letters and further specification of the ST are signalled via equals sign (e.g. [[Artwork = Sculpture]]). The vertical bar (ǀ) indicate that the element are alternative (e.g. [[Human]] is alternative to [[Event]] for subject position of *abbattere#1*). If the same ST is expected in multiple positions, it is numbered progressively (e.g. [[Concept1]] and [[Concept2]] in *erigere#2*) Preposition are grouped with the relative ST inside curly brackets (e.g. {a ǀ in [[Concept2]]} in *erigere#2*).

Verb: *abbattere*	**Verb:** *erigere*
▷ ***abbattere*#2:**	▷ ***erigere*#1:**
[[Human | Event]] *abbattere* [[Building]]	[[Human]] erige [[Building | Artwork=Sculpture]]
Implicature:	*Implicature:*
[[Human | Event]] *demolisce, distrugge, butta giù* [[Building]]	[[Human]] *costruisce* [[Building | Artwork=Sculpture]]
(Eng.: demolishes, destroys)	(Eng.: rises, erects)
task 1: OPPOSITION	
task 2: REV	

Verb: *abbattere*	**Verb:** *erigere*
▷ ***abbattere*#3:**	▷ ***erigere*#1:**
[[Human]] *abbattere* [[Animate]]	[[Human]] erige [[Building | Artwork=Sculpture]]
Implicature:	*Implicature:*
[[Human]] *uccide, toglie la vita a* [[Animate]]	[[Human]] *costruisce* [[Building | Artwork=Sculpture]]
(Eng.: kills, suppresses)	(Eng.: rises, erects)
task 1: NO-OPPOSITION	
task 2: -	

Figure 3.7: Examples of the annotation of types of opposition among *t-pas*s. REV is annotated between *abbattere*#2 and *erigere*#1. No opposition is annotated between *abbattere*#3 and *erigere*#1.

In both steps of the task, annotators made use of the semantic types (STs) expressed in the verb *t-pas*. In particular, the STs in the argument positions (e.g. [[Human]], [[Event]]) helped the annotators in interpreting the sense of the *t-pas* and consequently in identifying which are the senses of the verbs in an opposition relation (if an opposition relation is realized). As an example, consider the *t-pas*s reported in Figure 3.7 for the pair *abbattere / erigere*. In this example, the STs (in particular [[Human]], [[Building]] and [[Animate]]) helped the annotators in understanding which senses of the two verbs s/he is comparing, and, possibly, to establish an opposition relation between *abbattere*#2 and *erigere*#1, but not between *abbattere*#3 and *erigere*#1.

In the case of multiple STs for the same argument slot, annotators were allowed to mark opposition relations between *t-pas*s even if they are realized only by a subset of the set of ST for each argument position. For instance, *abbattere*#2 and *erigere*#1 are opposites only as far as [[Human]] is considered as the subject of the two predicates (i.e., *erigere*#1 does not select [[Event]] as subject).

Finally, annotators could match the same *t-pas* of a verb to more than one *t-pas*s of

the other verb. In total each annotator had to judge 595 *t-pas* pairs.[75]

3.2.2.1 Inter Annotator Agreement

To calculate the agreement between the two annotators, we adopted the Dice's coefficient (Rijsbergen, 1979), which measures how similar two sets of elements are; it is calculated by dividing the number of shared elements of the two sets, by the total number of elements they are composed of. This produces a value from 1, if both sets share all elements, to 0, if they have no element in common[76].

We calculate the Dice's coefficient for two configurations. In the first configuration, *opposition recognition*, we consider 1 point of agreement if both annotators agree on recognizing opposition or non-opposition between two *t-pas*s, 0 if they do not agree. In the second configuration, we calculate the agreement considering *opposition type*, i.e., we consider as agreement if both annotators identify exactly the same opposition relation. Finally, for each type of opposition, we calculate the *per type disagreement* as the proportion of pairs in which the two annotators disagree over the total pairs in which the type has been recognized.

Out of 595 pairs of *t-pas*s used in the experiment, the two annotators agreed in recognizing a pair as displaying or not displaying an opposition relation in 588 cases (44 are marked as opposites by both annotators, 544 as non-opposites): the Dice value for opposition recognition is 0.98. This result suggests that identifying opposition relations between *t-pas*s is not to a controversial decision among annotators. Moreover, annotators identified the same type of opposition or agreed in recognizing non-opposition in 582 cases, thus Dice value for type of opposition is 0.97 showing that the annotation by the two annotators has a very high degree of overlap.

On the other hand, considering disagreement for each opposition type (see left part of Table 3.3), results show that most cases concern the COMPL type (annotators identified this type in 16 pairs but disagreed on 6 of them) and the REV type (disagreement on 9/21 pairs); by contrast, annotators agreed more consistently on recognizing CONV pairs (only one case of disagreement).

In order to understand the motivations of these discrepancies, we have adopted a rec-

[75]To complete the task annotators took approximately two days, including corpus examples consultation.

[76]Given two sets of annotation (by two different annotators) A and B, the formula is the following: Dice's value = [2 x (A ∩ B) / (A + B)]).

pre-reconciliation		
Type	**#disagreement / #total**	**%**
COMPL	6 / 16	37.5
ANT	3 / 13	23
CONV	1 / 9	11.1
REV	9 / 21	42.8
NON-OPP	7 / 551	1.2

post-reconciliation		
Type	**#disagreement / #total**	**%**
COMPL	3 / 15	20
ANT	2 / 12	16.6
CONV	0 / 9	0
REV	5 / 18	27.7
NON-OPP	5 / 550	0.9

Table 3.3: Per type disagreement (pre and post-reconciliation).

onciliation strategy among annotators. In particular, we asked annotators to motivate their choices with the possibility to revise their selections. After the reconciliation discussion, Dice values increased to 0.99 (considering only opposition recognition) and to 0.98 (considering opposition type) and the *per type* disagreement decreased for every type (see right part Table 3.3).

We observe three cases of disagreement among annotators.

A first case concerns disagreement when the semantic types specified in the *t-pas* include elements with different characteristics. This, in some cases, has conditioned annotators to consider a *t-pas* as opposite (or not) of another *t-pas*. As an example, consider *mettere*#1 in Example 3.2 paired with *togliere*#2 in Example 3.3[77]:

(3.2) *t-pas*#1 of the verb *mettere (to place)*:
[[Human]] mettere [[Artifact | Body Part]] {in [[Location]] | in [[Container]]}
(*eng.: [[Human]] place [[Artifact | Body Part]] {in [[Location]] | in [[Container]]}*)

(3.3) *t-pas*#2 of the verb *togliere (to remove)*:
[[Human]] togliere [[Inanimate]]
(*eng.: [[Human]] remove [[Inanimate]]*)

In this example, one annotator recognized the two *t-pas*s as REV. On the contrary, the other annotator, considering some examples in the corpus for *togliere*, *t-pas*#2, decided not to mark the opposition, as most of the lexical items over which the ST [[Inanimate]] generalises identify elements that cannot be placed or re-placed in a certain [[Location]] or [[Container]] (e.g. to remove a tooth).

The second case we discuss concerns disagreement between *opposition type* selection, as observable in *caricare*#1, in Example 3.4 paired with *scaricare*#3, in Example 3.5.

[77]The English version is intended only for readability purposes and it is not meant to represent a corresponding English *t-pas* of the Italian *t-pas*.

(3.4) *t-pas*#1 of the verb *caricare (to load)*
[[Human]] caricare [[Animate | Inanimate]] (su | in [[Vehicle]] | su | in {spalle|schiena})
(*eng.: [[Human]] load [[Animate | Inanimate]] (into [[Vehicle]]) or carry [[Animate | Inanimate]] on {his | her shoulders}*)

(3.5) *t-pas*#3 of the verb *scaricare (to unload)*:
[[Human | Machine]] *scaricare* [[Inanimate]]
(*eng.: [[Human | Machine]] unload [[Inanimate]]*)

For this pair, one annotator recognized the two *t-pas*s as REV (*reversives*), as the two events describe a change in opposite direction, and display a temporal relation; in contrast, the other annotator selected ANT (*antonyms*), considering that, the objects of both predicates, as observed in the corpus samples, are quantifiable, and thus the actions are in a certain way measurable.

The third case we discuss highlights disagreement due to the semantic interpretation of the verbal *t-pas*s as it is represented by *sorgere*#1, in Example 3.6, paired with *tramontare*#3, in Example 3.7. In this case, it seems that one annotator focused on the entailment relation exiting among the *t-pas*s and marked the pair as REV (as detailed in Section 2.1.2.4, some reversives hold an entailment relation: a dimension that is not captured by the other opposition types in the schema). On the other hand, the other annotator interpreted that the two *t-pas*s "divide in two a conceptual domain", thus selecting COMPL (*complementaries*).

(3.6) *t-pas*#1 of the verb *sorgere (to rise)*
{*sole*|*luna*|...} sorgere [NO OBJ]
(*Eng.:* {*sun*|*moon*|...} *rise [NO OBJ]*)

(3.7) *t-pas*#1 of the verb *tramontare (to set)*
{*sole*|*luna*|...} tramontare [NO OBJ]
(*Eng.:* {*sun*|*moon*|...} *set [NO OBJ]*)

In conclusion, we can say that the inter annotator agreement is very high in the identification of opposite *t-pas* pairs, and fair in distinguishing among types, even if some cases of disagreement could have not been solved. In particular, we have noticed that the type of reversives is frequently tagged together with other types of oppositions.

3.2.3 Contributions, Shortcomings and Further Work

In this section we have presented an annotation schema for the annotation of oppositions among verbal frames. In our schema, opposition relations have been classified in four categories: *complementaries, antonyms, converses, reversives*. For Italian, to the best of our knowledge, there is no annotation schema that identifies these different types for the annotation of opposition.

We have conducted a pilot annotation, selecting 25 verb pairs from the T-PAS resource to access the reliability of the scheme. Results show also that the inter annotator agreement is very high in the identification of opposite *t-pas* pairs, and fair in distinguishing among types. The exercises confirms that the annotation is doable and can be extended to all verbs in the resource, thus enriching it with opposition relations among frames. In Section 3.4, we will present our strategy for populating the T-PAS resource with oppositions; in particular, we adopt the crowd-sourcing methodology with a more systematic use of the corpus samples associated to each *t-pas*.

We also found that the *t-pas* pairs can have properties pertaining to more than one type of opposition. In particular, in some cases, the type of reversives appears to be a cross relation that co-exists with other types. We consider this as a first indication that the types of opposition we considered seem not to be exclusive.

3.3 Methodological Choices for Acquiring Opposition

In this section we present our methodological choices for the acquisition of the opposition relation to enrich a lexical resource. In particular, in Section 3.3.1 we discuss various methodologies and in Section 3.3.2 we provide a more detailed presentation of the crowd-sourcing methodology that we decided to adopt.

3.3.1 Using Automatic Methods to Identify Opposition

As we saw in Section 2.2, much research has been dedicated to the acquisition of lexical opposition in Computational Linguistics. Some of them rely on pattern-based approaches (for instance, Lobanova (2012)) which use textual patterns (e.g. "either *X* or *Y*") to distinguish antonymy related pairs from other pairs. As we already clarified, however (see Section 2.2), these methods do not permit the disambiguation of the terms under exami-

nation: even if the sense of the terms can possibly be disambiguated in the pattern itself, there is still no way to track the sense of the terms in the pattern in order, for example, to compare different patterns in which a term is used (for instance, one cannot tell that *to lose* has one sense when contrasting with *to win* and another sense when contrasting with *to find*). To account for this, we need to acquire information on opposite terms for which a repository of senses is available, and for which it is known which is the sense of the term in the particular context we are examining. The pattern-based approach also does not permit identifying the type of opposition a pair of opposites belongs to. Moreover, in terms of coverage, there is a good chance that only the most frequent pairs will be identified by these methods.[78]

Other work on the identification of opposition among words pairs has been conducted considering the distributional hypothesis (e.g. works by Adel and Schütze (2014) and Santus et al. (2014b))[79]. All these have the great advantage of acquiring information automatically or semi-automatically, but they often do not consider word polysemy: for example, they aim at identifying that *to lose* and *to win* are opposites, that *to lose* and *to find* are also opposites, and that *to lose* and *to mislay* are, on the contrary, synonyms. [80]

As already noted, our aim is instead to gather information for enriching a resource with opposition at the sense level, thus disambiguating the sense of the terms we are considering. Word Sense Disambiguation (WSD) is still a challenging task to solve with automatic methods[81], and it is particularly difficult in the case of verbs (the focus of our

[78] According to Santus et al. (2014a) these methods "suffer from low recall, because they can be applied only to frequent words, which are the only ones likely to occur with the selected pattern".

[79] See Section 2.2 for other works on the identification of opposition with distributional methods, and a brief introduction to that area of investigation.

[80] As we highlighted in Section 2.2, the most challenging aspect to use distributional methods for the identification of opposites is to distinguish them from other similar terms, e.g. synonyms.

[81] Much work is dedicated to discriminate the senses of a word even with no reference to a pre-defined repository of senses (i.e., Word Sense Discrimination). For instance, Schütze (1998) apply the following methodology. First, he creates a vector space model from a large corpus, and calculates the *word vector*, a vectorial representation of the single words. Then for each target word he calculates its *context vectors*, i.e., vectors representing each occurrence of the word in the corpus. These *context vectors* are then clustered into groups, each one representing a sense of the word. Finally, for each cluster, a *sense vector* is computed by calculating its centroid. Given a word in a new *context vector*, this can be disambiguated by calculating the distance with *sense vectors*: the *context vector* is assigned to the most similar *sense vectors*. This method has inspired many contributions such as the one by Reisinger and Mooney (2010) in which the authors create a vector-space model in order to determine the semantic similarity of both isolated words and words in context. Works in this sense also include those by Erk and Padó. In their contribution of 2008, the authors take into account the representation of the selectional preferences for words' argument positions, integrating this information into the computation of a word meaning in context (Erk and Padó, 2008).
However, we believe that in order to annotate a resource with opposition, word sense discrimination is not sufficient, since the information of opposition needs to be assigned to a specific sense of the words, as listed in the resource itself.

work) which are notoriously the most polysemous class of words[82]. Verb senses can in fact present subtle differences that are not easy to catch automatically. In an interesting overview presented by Raganato et al. (2017), the authors compare the performance of eleven systems on the WSD task: in the case of verbs, no system reported an F-score[83] above 58% (in their data set, the average polysemy of verbs is 10.4, more than double that of the other lexical categories). Human annotation provides better results: the Inter Annotator Agreement for the Senseval-2 Verb Lexical Sample by expert annotators is 71% (Palmer et al., 2006, p.90) (in their data set of 29 verbs, the average polysemy of verbs is 16.28, more than double that of the other lexical categories).

This discourages the use of automatic methods for the acquisition of precise, reliable data on opposition considering that we aim at enriching a resource that can be used as a reference for both linguistic investigation and NLP task resolution.

In order to gather high quality data, we decided to rely on human judgments. The collection of data by expert linguistic annotators is in fact more reliable but expensive in terms of time and cost. The solution we adopted was to use crowd-sourcing, which allows us to acquire a great amount of information in a limited period of time. In Section 3.3.2, we highlight the main characteristics of this methodology, the best practise to be followed when using this methodology, and we present an overview of contributions which use crowd-sourcing for collecting information on opposition.

3.3.2 Crowd-sourcing: Characteristics, Best Practice and Use

Crowd-sourcing consists in asking people in the "crowd" (here on: *workers*), that have no specific preparation and whose skills and background are unknown, to collaborate in small jobs (here on: *tasks*). In this way, it is possible to acquire a great amount of information from different persons, in a limited period of time; in our case, this is faster and cheaper than, for instance, to hire expert workers.

This methodology is used in many disciplines in order to acquire information on a large scale, however, there are some best practices that need to be followed, no matter which task workers are asked to solve. For example, the task should be understandable and feasible to solve, and, in order to benefit from this methodology, the task needs to be

[82] In WordNet 3.0 (Fellbaum, 1998), verbs are about 70% more polysemous than nouns, adjectives, adverbs. Source: `https://wordnet.princeton.edu/wordnet/man/wnstats.7WN.html`.

[83] For an explanation of the F-score, see footnote 31 in Section 2.2.2.

designed carefully and account for some quality requirements. The general best practices that we followed are detailed below.

Best Practices. In order to design the task properly in our investigation, we have tried to follow the best practices suggested by Sabou et al. (2014)[84]; the authors provide a number of examples in which paid crowd-sourcing has been used to create corpora that support a broad range of NLP problems. In particular, we followed these suggestions for what concerns:

- project definition: main task needs to be decomposed in simple tasks, suitable task setting has to be selected, rewarding need to be determined, tasks should be simple and intuitive;
- data preparation: the interface and the instruction should be clear, the task needs to be designed in order to prevent and reduce cheating (crowd-sourcing platforms offer some functionality that can help in this sense);
- project execution: it is important to attract and retain contributors and to filter cheating workers in order to improve quality. For example, it is possible to embed gold standard questions (that are questions for which authors of the task already know the answers) to determine the general quality of data provided by each worker.

The use of crowd-sourcing for collecting information on opposition. Works which use crowd-sourcing to collect information on opposition relation include the contribution of Mohammad et al. (2013). As mentioned in Section 2.2, the authors use crowd-sourcing to determine the level of human agreement on considering terms in a pair (adjectives, adverbs, nouns and verbs) as *contrasting words* pairs from *opposites*[85], and on classifying these pairs according to their type of opposition. Specifically, the authors find that workers agree markedly on identifying contrasting word pairs while they agree less in assigning the pairs to the different types of opposition. As we saw in Section 2.2, Schulte Im Walde and Köper (2013) collect synonyms, antonyms and hypernyms using crowd-sourcing for

[84]The essay suggests a set of best practice guidelines for paid crowd-sourcing for what concerns e.g. project definition - task decomposition, suitable task setting choice, rewarding, etc..-, data and gold data preparation, contributor recruitment, project execution.

[85]For a definition, see Section 2.1.1 and footnote 55 in 3.1.

building the dataset for their experiment aiming at distinguishing paradigmatic relations in German, and similarly Benotto (2014) defines a task in which the workers are asked to propose synonyms, antonyms, and hypernyms for a target term. More recently, Takabatake et al. (2015) use crowd-sourcing in order to collect 'contradictory event pairs' (e.g. 'the dish is expensive' / 'the dish is cheap') and create a large-scale database of Japanese contradictory event pairs by asking the crowd to write and evaluate in-domain contradictory sentences. In their study, the authors classify contradictory event pairs in a taxonomy which includes, among other, *binary event pairs*, i.e., events that contradict each other for including "mutually exclusive antonyms" (e.g. 'being single' and 'being married') and *continuous event pairs*, i.e., events that contradict each other for including "event pair consists of antonym predicates that represent continuous states" (Takabatake et al., 2015, p. 102) (e.g. 'being expensive' / 'being cheap').

Similarly to these works, we expect to acquire reliable information on opposition relations among verb senses at large scale, inexpensively and over a relative short period of time. As said, this information will possibly be used to enrich lexical resources with the opposition relation.

3.4 Acquiring Different Types of Opposition between Verbs Senses

In this section, we describe the large scale acquisition of the opposition relation among verbs we performed through the use of crowd-sourcing.

The schema we defined (Section 3.2) and the crowd-sourcing methodology we decided to use for the population of the resource can be applied to gather information for enriching many lexical resources; however, in our investigations we chose the T-PAS resource (Ježek et al., 2014) as the "target resource" we aim at enriching, and defined the crowd-sourcing task accordingly.

Moreover, we defined the task in order to be based as much as possible on empirical observation of the data. We did not confine our analysis to the observation of the verb in isolation (in which no information about which sense of the verb under examination is noticeable), neither we relied on definition which can explain the senses of a verb abstractly. In our task, we considered the verbs in context. More specifically, given a pair of verbs in context, we were interested in knowing whether these verbs (used in those specific senses) are opposite, and if so, to which types of opposition they belong.

The task for the crowd was defined as follows. We showed the workers in the crowd the pairs of sentences which differ just for one verb: the first sentence is extracted by a corpus and it contains a particular sense of the verb (according to the T-PAS resource); the second sentence is created by automatically substituting a verb in the first sentence with one of its opposites verb. We asked workers to judge if the sentences in which the two verbs appear have opposite meanings. Then, we asked them to disambiguate the sense of the second verb. Finally, if an opposition was identified, we asked workers to judge the type of opposition existing among these two verbs (as they appear in the sentences). To improve the reliability of the process, for each pair we collected the judgments of more than one worker.

Through the same crowd-sourcing experiment, we also collected information about the types of opposition between verbs. We would like to understand whether the types of opposition we distinguish (i.e., *complementarity*, *antonymy*, *converseness* and *reversiveness*, Section 2.1) are actually recognized by the crowd. We also intend to figure out whether it is possible for a pair of verb senses to have characteristics that belong to more than one type of opposition. To do this, we clustered the human judgments we collected into four classes. The specification of the type of opposition can be potentially added in the resource.

This Section is organized as follows. In Section 3.4.1 we present the three step task we defined to collect data. This is followed by the description of the task preparation in Section 3.4.2. In Section 3.4.3 the results of our crowd-sourcing experience are presented. Finally, Section 3.4.4 provides a discussion about the methodology we adopted and Section 3.4.5 presents some conclusions and directions for future work.

3.4.1 Define the Data Collection

As we previously mentioned, the investigation that we carried out is manifold. On the one hand, we aimed at collecting information about opposition among pairs of verbs at the sense level and about their types of opposition, in order to enrich lexical resources; on the other hand, we wanted to provide some insights of the different types of opposition: *complementarity*, *antonymy*, *converseness* and *reversiveness*. To reach this goal, we organized the investigation in three steps that represent also three parts of a unique, broader, crowd-sourcing task.

In the first step, *opposition identification* (Task A, Section 3.4.1.1), we wanted to determine whether there is an opposition relation between a certain sense of a verb (the *source verb*) and another verb (the *target verb*): thus, we asked the workers in the crowd if they recognize this relation. In the second step, *sense disambiguation* (Task B, Section 3.4.1.2) we focused on understanding which senses of the target verb is involved in the opposition -if previously identified. In the third step, *type-of-opposition identification* (Task C, Section 3.4.1.3), we asked the workers to give us information over the types of opposition that the *source verb* and the *target verb* hold. The three tasks are further detailed here below. Results are also presented *per task* in Section 3.4.3.1, 3.4.3.2, 3.4.3.3 respectively.

3.4.1.1 Task A: Opposition Identification

Aim: to determine whether there is an opposition relation between a certain sense of a verb (the source verb) and another verb (the target verb)

For the *opposition identification* step (Task A), we showed workers a pair of sentences, S1 and S2: S1 is a sentence that contains a *source verb*, while S2 is identical to S1, with the exception of the *source verb*, which is substituted with the *target verb*. The *target verb* is an opposite of the *source verb*. S1 is extracted from a corpus, while S2 is automatically produced via the substitution of the *source verb* with the *target verb*. Figure 3.8 reports an example in which the source verb *respingere* in S1 has been substituted with the target verb *approvare* generating S2.

S1: L' appello va, pertanto, **respinto**. (Eng.: The appeal must be rejected.)
→ *source verb* = **respingere**
S2: L' appello va, pertanto, **approvato**. (Eng.: The appeal must be approved.)
→ *target verb* = **approvare**

- S2 = S1, but *source verb* substituted with *target verb*
- S1 is retrieved from a corpus, S2 is created via substitution
- the *target verb* is an opposite of the *source verb*

Figure 3.8: Example and description of S1 and S2.

This substitution may, however, generate a S2 sentence that does not make sense. For instance, in Figure 3.9, *sorgere* (Eng. *to rise* of building or of a star) is the *source verb* and *tramontare* (Eng. *to set* of a star) is the *target verb*, and the substitution of *sorgere* with

tramontare produces a "no sense" sentence, since the sun or the moon can *tramontare* (Eng. *to set*) but a church cannot. Considering that the substitution of the verb can lead to this possibility, workers were asked also to judge whether S2 makes sense.

S1: La chiesa **sorge** su una collina. (Eng.: The church rises on a hill)
→ *source verb* = **sorgere**
S2: *La chiesa **tramonta** su una collina. (Eng.: *The church sets on a hill)
→ *target verb* = **tramontare**

Figure 3.9: Example of no sense S2.

Workers thus were asked to compare the two sentences and choose if: (i) S2 makes sense and holds an opposition relation with S1, or (ii) S2 makes sense but does not hold an opposition relation with S1, or (iii) S2 does not make sense. Figure 3.10 shows an English version of the task and Appendix B the original Italian version.

If a relation of opposition was identified, we asked workers to continue with the other steps and complete Task B and Task C, otherwise a new pair S1-S2 was shown.

Consider the following sentences:
S1: The appeal must be **rejected**. (*source verb* = *to reject*)
S2: The appeal must be **approved**. (*target verb* = *to approve*)

Task A: "Would you say that S2 S2 makes sense?If yes, would you say that there is an opposition relation between the sentences?
A1: S2 makes sense and holds an opposition relation with S1 → continue with Task B and C
A2: S2 makes sense but it does not hold an opposition relation with S1
A3: S2 does not make sense

Figure 3.10: Example for the *opposition identification* task (English version).

Notice that in this task (and in the followings), we showed the *source verb* and the *target verb* in context. This is motivated by the fact that we would like to collect information on opposition relations at verb senses level; thus showing verbs out of context will not help in clarifying which sense of the verbs we are referring to. Moreover, we think that the choice of showing the verbs in context helps in keeping the task simpler, more intuitive and less ambiguous for workers than to provide a definition of the senses of the verbs. Nonetheless, we think that showing the same sentence in which the verb is substituted makes the task faster, since the understanding of which sense of the two verbs is under observation is immediately possible; on the contrary, we believe that using different sentences in which the two verbs are shown would make the task more difficult to solve.

This setting is also in line with our general approach of gathering evidence of the existence of an opposition, that is, to retrieve data which prove that opposition really holds, and not just rely on intuitions.

3.4.1.2 Task B: Verb Sense Disambiguation

Aim: to understand which senses of the target verb are involved in the opposition.

In the *sense disambiguation* step (Task B), we asked workers to disambiguate the sense of the *target verb* in S2, in order to understand which are the senses of the *target verb* that hold the opposition relation with the *source verb*. We had no need to disambiguate also the *source verb* since, as we will see in Section 3.4.2, we retrieved S1 from the sentences associated to verb senses in the T-PAS resource.

To perform this disambiguation, we showed workers a list of sentences containing the *target verb* in one of its meanings.[86] We asked workers to mark the sentences in which the *target verb* has the "same" meaning as in S2. An example (in English) is reported in Figure 3.11.

Consider the following sentences:
S1: The appeal must be **rejected**. (*source verb = to reject*)
S2: The appeal must be **approved**. (*target verb = to approve*)

Task B: "Read the following sentences. In which of them *to approve* has the same meaning as in S2?"
B1: The Commission **approves** the law 2.15 of the speaker.
B2: The astronauts have *approved* the use of the TVIS in this configuration.
B3: In any case, candidates that registered absences for more than one third of the total number of the total amount of hours of class will not be *approved*.
B4: None of the previous.

Figure 3.11: Example for the *sense disambiguation* task (English version).

3.4.1.3 Task C: Type-of-Opposition Identification

Aim: to identify which is the type of opposition the pair of opposite verbs belongs to, and verify whether a pair can have characteristics that belong to more than one type.

In the *type-of-opposition identification* step (Task C), we asked workers to judge the type of opposition existing among the *source verb* and the *target verb* as they appear in the sentences S1 an S2, previously judged as opposite.

[86] This is different from other Word Sense Disambiguation (WSD) tasks, (e.g. (Mihalcea et al., 2004)), in which workers were asked to select among a sense inventory.

Recall that among the various types of oppositions that can be said to exist among verbs, we focus on the four described in Section 2.1 and in Section 3.2.1: *complementarity*, *antonymy*, *converseness* and *reversiveness* (Lyons, 1977; Cruse, 1986, 2011).

We represented the characteristics of each type as testing statements and asked workers to evaluate which of these statements can correctly apply to the pairs of verbs (*source verb* and *target verb*) they were considering each time. More in detail, workers were asked to consider S1 (with the *source verb*) and S2 (with the *target verb*) and answer the the following question: "which (of the following) statements are true?". Then, the list of statements was provided.

The statements, reported in Figure 3.12, are created as follows:

- statement 1 has been introduced as a control statement. It is meant not to be a valid option in the case of an opposition relation, since it concerns the characteristic of opposite of "not being true simultaneusly";
- statement 2 characterizes *complementarity* and refers to their "mutual exclusivity";
- statements 3 and 4 characterize *antonymy* and refer to the "existence of a third alternative" and to their characteristics of "being gradual";
- statements 5, 6, 7 characterize *reversiveness*, referring to their "dynamic dimension" (i.e., they express a change), their "interaction with entailment relation", their "being an expression of a change in opposite direction";
- statements 8 characterizes *converseness* and refers to their distinctive characteristics of "describing the same situation fro two point of view".

These statements were meant to express single characteristics of the four types of opposition as they are defined in the literature; at the same time, they were formulated by taking into account that the task was proposed to non expert workers "in the crowd", with no preparation or instruction on opposition relations and whose background was not known by the authors. Therefore, in order for the task to be more intuitive, we tried to create simple statements, clear to the crowd, synthetic, also avoiding definitions and providing statements with an example.

Notice that for *antonyms* and *reversives* we adopted more than one statement: we considered that the combination of these statements is more representative for the type than

a single one. Moreover, there was no restriction on the maximum number of statements a worker could select. One of the aim of this exercise was to see if characteristics that typically refer to different types are shared by the same pair of verb senses.

Consider the following sentences:
S1: The appeal must be **rejected**.
S2: The appeal must be **approved**.

Task C: Consider the two verbs **to reject** and **to approve** in S1 and S2. Which of the following statements are true?
1: A situation in which both events take place simultaneously and with the same participants is possible e.g. I breath/I write a letter
2: Given a situation, if one of the two events does not take place, the other occurs, e.g. to fail/to pass (an examination)
3: Given a situation, a "neutral" alternative in which none of the events occurs is possible, e.g. to increase/to decrease (the temperature)
4: It is possible to use gradable modifiers as "a bit", "moderately", "a lot", etc.. e.g. to improve/to get worse
5: It is possible that the two events occur repeatedly several times one after the other, e.g. to open/to close (a door)
6: One event is possible only if the other has occurred, e.g. to destroy/to build (a building)
7: If one event occurs, and then the other takes place, the situation returns to the initial state, e.g. to button/to unbutton (a shirt)
8: The two verbs describe the same situation by two different points of view, e.g. to give/to receive (a present)

Figure 3.12: Example for the *type-of-opposition identification* task (English version).

The final design of our task was tuned in an off-line pilot version we proposed to five workers of different background, age and experience in the NLP field [87].

3.4.2 Preparation of the Crowd-Sourcing Job

As previously described, we asked workers to provide judgments on the opposition existing between a *source verb* and a *target verb* by considering two sentences S1 and S2. The tasks implementation thus required: (i) the selection of the *source verb* and *target verb*, (ii) the extraction of sentences in which the *source verb* appears, i.e., S1, and (iii) the substitution of the verb in the sentences in order to create S2. This section will explain

[87] We are grateful to Lorenzo Gatti, Manuela Speranza, and Rachele Sprugnoli from Fondazione Bruno Kessler (Trento) for their contribution in testing and setting the final design of the task, and to Lorenzo Campostrini and Silvia Meli for testing the task.

the verb selection, the sentences extraction and the verb substitution processes, and finally will provide a description of the crowd-sourcing platform setting.

Verbs Pair Selection. We selected the *source verb* for S1 and the *target verb* for S2 according to three conditions:

1. both verbs are in the T-PAS resource (Ježek et al., 2014);
2. both verbs appear in the Dizionario dei Sinonimi e dei Contrari - Rizzoli Editore[88] as lemmas;
3. the *target verb* is annotated as "contrary" for the *source verb* and viceversa in the Dizionario dei Sinonimi e dei Contrari; thus, for each pair in which Verb A is the *source verb* and Verb B is the *target verb*, also the pair Verb B as *source verb* - Verb A as *target verb* has been considered. Notice that we retrieved also pairs such as *abbottonare / aprire* (Eng. 'to button up' / 'to open'). The opposition in these cases hold between a verb and the hyperonym/hyponym of its "best opposite"[89], i.e., *abbottonare / sbottonare* and *chiudere / aprire* (Eng. to button up: to unbutton / to close : to open). This condition however simply aimed at defining an initial list of possible opposites that the crowd is asked to confirm, so we kept the list as retrieved, including these pairs of "less good opposites".

The total number of verb pairs extracted according to these criteria is 420 (i.e., 210 symmetrical pairs)[90]. Since our aim is to annotate opposition among verb senses, we implemented Task A for each of the senses of the *source verb*s. As mentioned, we used the T-PAS resource (Ježek et al., 2014) as a sense inventory for verbs. This means that we implemented the task for each of the *t-pas*s of the verbs, e.g., for *t-pas*#1 of the source verb *respingere*, for its *t-pas*#2, for its *t-pas*#3, etc.., for a total of 1326 different *t-pas*s involved, and a total of 2166 *t-pas*s for which we needed to implemented the task (i.e., counting the *t-pas* for a *source verb* every time it was paired with a *target verb*) .

Sentences Extraction. Once the verb pairs were selected, we needed to find a number of sentences for each sense of the verbs in the pair. For example, for the pair *respingere*

[88] `http://dizionari.corriere.it//dizionario/_sinonimi/_contrari`

[89] See Section 2.1.1 for the characteristics of the "good opposition" proposed by Cruse (1986).

[90] These data have been corrected with respect to the number of 436 pairs published in (Feltracco et al., 2015b).

/ approvare, we needed to have sentences for *respingere#1*, *respingere#2*, etc., and for *approvare#1*, *approvare#2*, etc. These sentences represented the S1 sentences in Task A (when the verb is the *source verb*), and the answers proposed in Task B (when the verb is the *target verb*). We needed to retrieve sentences for both verbs in a pair because each of the two verbs plays alternatively the role of *source* or *target*.

We extracted these sentences from the T-PAS resource, in which each sense of the verb is associated with instances in the corpus of reference for the resource (Section 3.1.1). In order to increase the reliability of the annotation, we extracted up to three examples for each sense (*t-pas*) of the verbs from the resource, according to their availability (i.e., we extracted up to three examples for *respingere#1*, up to three examples for *respingere#2*, etc..). We discarded examples annotated as "non regular" such as metonymical uses and, to simplify the task, we selected the shortest examples, composed by at least 5 tokens.

Verb Substitution. We generated S2 from S1 substituting the *source verb* with the *target verb* automatically conjugated accordingly, using the library: italian-nlp-library[91]. The library analyzes only the verb and not the whole sentence and does not manage all the suffixes; to solve this we added some simple rules. This system grants a quick implementation avoiding parsing or deeper analysis of the sentence.

Crowdflower Platform Settings. We used the Crowdflower Platform[92], with the following parameter setting. We initially set the payment to 0.04 USD, then to 0.05 USD for each page and the number of sentence pairs for page to 5. In order to reduce the likelihood of unreliable users to participate in the task, we included in each page a Test Question (TQ): a question for which we already knew the answer.[93] If a worker missed many TQs s/he was not permitted to continue the annotation and his/her judgments were rejected: we set the threshold of this accuracy to 71%. We selected the TQs among the total sentence pairs and we annotated them before lunching the task. We also set parameters in order to have workers with Italian language skills.[94]

[91] `https://github.com/jacopofar/italian-nlp-library`

[92] `http://www.crowdflower.com`

[93] 20% of gold data per task is the recommended amount by the Clowdflower Platform.

[94] The Clowdflower Platform enables the task manager to set a number of parameters to filter workers prior to the task and control its quality: see `http://www.crowdflower.com`.

The task was introduced by a welcome message and a list of simple instructions: these are reported in Appendix B.

3.4.3 Results for *Opposition Identification, Sense Disambiguation,* and *Type of Opposition Identification*

In this section we provide the results for each of the tasks described above.

3.4.3.1 Task A: Opposition Identification

A total of 712 pairs of sentences have been annotated with 3 judgments in almost a month, for a total of 2136 judgments (plus judgments for TQs). This represents about the 10% of the total number of sentence pairs that we retrieved with the methodology described in Section 3.4.2.

For Task A, the overall inter-annotator agreement (IAA) calculated using *Fleiss's coefficient* (Artstein and Poesio, 2008) is 0.44, with an Observed Agreement (Ao) of 72.47%. Overall, answer A1:"S2 makes sense and holds an opposition relation with S1" was chosen 30.9% of the times, answer A2:"S2 makes sense but it does not hold an opposition relation with S1" 5.1%, and answer A3: "S2 does not make sense" 64%. Data are reported in 3.13.

Consider the following sentences:
S1: The appeal must be **rejected**. (*source verb = respingere*)
S2: The appeal must be **approved**. (*target verb = approvare*)

Task A: "Would you say that S2 makes sense? If yes, would you say that there is an opposition relation between the sentences?

A1: S2 makes sense and holds an opposition relation with S1	30.9%
A2: S2 makes sense but it does not hold an opposition relation with S1	5.1%
A3: S2 does not make sense	64 %

judgments:
712 pairs x 3 judgments = 2136 judgments (plus judgments for TQs)
Inter-Annotator Agreement:
Ao: 72.47%
Fleiss's coefficient: 0.44

Figure 3.13: Results for Task A:*opposition identification*. English version.

Cases in which two sentences were judged as holding an opposition relation include

the following examples:

(3.8) S1: Lei **mormorava**, più che parlare.
(Eng. She was whispering, rather than speaking)
S2: Lei **urlava**, più che parlare.
(Eng. She was screaming, rather than speaking)

(3.9) S1: Online Shop: **acquista** online in modo sicuro, scarica e gioca subito!
(Eng. Online Shop: buy safely online, download and play now!)
S2: Online Shop: **vende** online in modo sicuro, scarica e gioca subito!
(Eng. Online Shop: sell safely online, download and play now!)

Notice in Example 3.9 many events are listed in S1 (i.e.,'to buy, to download, to play') and an opposition was however recognized between the two verbs under examination (i.e., *to buy: to sell*) . This is because we gave indication to the workers to evaluate just the minimal context in which the two verbs are used, independently of the others.[95]"

Interesting cases are also those few in which S2 are judged as "makes sense but does not hold an opposition relation with S1", such in 3.10 and 3.11.

(3.10) S1: **Verrà** creato il Livello 2.
(Eng. The Level 2 is going to be created.)
S2: **Andrà** creato il Livello 2.
(Eng. The Level 2 is going to be created.)

(3.11) S1: Filippo: Io non ti **trascurerei**, mai.
(Eng: Filippo: I will never neglect you.)
S2: Filippo: Io non ti **interesserei**, mai.
(Eng: Filippo: I will never interest you.)

In both these examples, the sense of the *target verb* is not opposite to the sense of the *source verb*. In particular, in Example 3.10 the two verbs play the same role of modal verbs expressing a future event and the two verbs are actually synonyms (the English translation is in fact the same). In Example 3.11 the *target verb* is potentially an opposite

[95]From the instruction of the task, in Appendix B: "..per decidere se due frasi sono opposte, dovete considerare il contexto minimo su cui i verbi incidono.
Frase 1: La luna **sorse** e lo abbagliò.
Frase 2: La luna **tramontò** e lo abbagliò.
Nel caso sopra, il contexto minimo di *sorgere* è "La luna **sorse**" che si oppone al contexto minimo di *tramontare* "La luna **tramontò**". In questo esempio, è quindi presente una opposizione nonostante i due eventi causino poi lo stesso effetto ("..e lo abbagliò."). Vi invitiamo a segnare "FRASE 2 ha senso e c'è opposizione con FRASE 1".

of the *source verb* if used in its reflexive form *interessarsi* (Eng. to be interested in something), thus the hypothetical opposite sentence would be:

S2: Filippo: Io non mi interesserei mai a te.
(Eng: Filippo: I will never be interest in you.)

However, due to the fact that the substitution is done automatically, it is not possible to operate this necessary modification.

Among the cases of "S2 makes no sense", we observe many cases in which a mismatch between the verb (in any of its meanings) and the new context in which the verb is inserted invalidates the sense of the sentence in its entirety. For instance, in Example 3.12, where *ridare* is the *source verb* and *trattenere* the *target verb*, the relation between the *target verb* and the direct object argument produces a "no sense" sentence (similar to what happens in the example in Figure 3.9). The pair has been judged as "Frase 2 non ha senso" by the three workers, since you can "ridare un esame" ("take an exam again") but not "trattenere un esame" (*"to hold, to keep an exam").

(3.12) S1: Posso **ridare** un esame già sostenuto come opzionale?
(Eng: Can I take an exam that I already taken as an optional one again?)
S2: Posso **trattenere** un esame già sostenuto come opzionale?
(Eng: Can I keep an exam that I already taken as an optional one again?)

Other cases in which the three workers chose "Frase 2 non ha senso" depend on the relation between the verb and other elements of the sentence which do not pertain to the verbs minimal structure. Example 2 shows a case with a coordinative structure between two events: in S1 somebody has been "imprisoned *and* deported", in S2 somebody has been "released *and* deported". We believe that workers judged the two events in S2 as incompatible and marked "Frase 2 non ha senso".

(3.13) S1: Era stato **incarcerato** e deportato. *(Eng: He was imprisoned and deported.)*
S2: Era stato **liberato** e deportato. *(Eng: He was released and deported.)*

In an initial version of the task, we also considered the idea of asking the workers the reasons why they were judging S2 as "does not make sense". For the final configuration we decided not to: first, because the reasons are various and possibly non exclusive, e.g. a syntactic mismatch (e.g. a transitive *source verbs* is substituted with an intransitive *target verb*), a semantic mismatch (e.g. Example 3.9), both a syntactic and a semantic mismatch,

no linguistic mismatch but low likelihood for the resulting event to happen, etc.; second, because explaining simply and briefly what is e.g. a "syntactic mismatch" would have been hard.

3.4.3.2 Task B: Verb Sense Disambiguation

In Task B workers were asked to disambiguate the sense of the *target verb*; the *source verb* is instead already disambiguated being associated to a specific *t-pas* in the T-PAS resource. Task B was proposed to workers only if an opposition had been identified in Task A (i.e., answer A1: "S2 makes sense and holds an opposition relation with S1"). Results (reported in Figure 3.14) were calculated for the pairs which collected a minimum of two (out of three) answers A1, for a total of 211 pairs, which correspond to 521 judgments.

We calculated the inter-annotator agreement (IAA) for each pair of sentences, considering 1 point of agreement when workers agree on selecting a sentence (i.e., a sense) and when they agree in not selecting a sentence. The overall agreement Ao, normalized by the number of workers, is 71.77%. In addiction, we calculated a *Macro Average-Fleiss'coefficient* (Mihalcea et al., 2004), where also the Expected Agreement (Ae) and the Fleiss'coefficient were determined for each pair, and then combined in an overall average. We calculated Ae *a posteriori*, considering the distribution of judgments of workers, resulting in a *Macro Average-Fleiss'coefficient* of 0.32. These values are similar to the rates reported in other Word Sense Disambiguation tasks using definitions of senses and not examples; e.g. IAA in Senseval-2 Verb Lexical Sample by expert workers is 71% (Palmer et al., 2006) and in Senseval-3 by Contributors over the Web the overall agreement is 67.3% with a Macro Average-K where of 0.35 (Mihalcea et al., 2004).

3.4.3.3 Task C: Type-of-Opposition Identification

As for Task B (*sense disambiguation*), results for the *type-of-opposition identification* step were calculated only for the pairs which collected a minimum of two (out of the three required judgments per pairs) answers "S2 makes sense and holds an opposition relation with S1" in the first *opposition identification* step. Later, judgments of two workers were removed as they answers for this task were considered unreliable: over 70% of their answers include the selection of statement 1 that was introduced as a control statement.

In order to evaluate if there is a correspondence between the four types of opposi-

Consider the following sentences:
S1: The appeal must be **rejected**. (*source verb* = *respingere*)
S2: The appeal must be **approved**. (*target verb* = *approvare*)

Task B: "Read the following sentences. In which of them *to approve* has the same meaning as in S2?"
B1: The Commission **approves** the law 2.15 of the speaker.
B2: The astronauts have *approved* the use of the TVIS in this configuration.
B3: In any case, candidates that registered absences for more than one third of the total number of the total amount of hours of class will not be *approved*.
B4: None of the previous.

judgments:
211 pairs x at least 2 judgments = 521 judgments (plus judgments for TQs)
Average Agreement:
Normalized Ao: 71.77%
M. A. Fleiss's coefficient: 0.32

Figure 3.14: Results for Task B:*sense disambiguation* task. English version.

tion (*complementarity*, *antonymy*, *converseness* and *reversiveness*, see Section 2.1 and in Section 3.4.1), and the judgments by the crowd, we clustered the judgments in four groups and used the resulting clusters as evidence of how humans recognize the four types. Moreover, we observe workers' agreement on single statements.

Clustering the data. To perform the clustering we used K-means, a well known unsupervised algorithm (MacQueen, 1967). This algorithm, given a certain number of clusters (K) and a way to measure the distance between the objects to be clustered, identifies K centroids, and splits the data in K clusters where the mean points are the centroids.

In our case, every annotation judgment is represented by a 8-tuple (an array with eight values) where the i-th coordinate is equal to 1 if the i-th answer is selected, for instance if the worker selects the second and the fourth statements the array would be (0,1,0,1,0,0,0,0). To compute the distance between arrays we used their Euclidean distance. We set K = 4 because our hypothesis is that there are 4 different types of opposition. Technically, first we used K-means++ (Arthur and Vassilvitskii, 2007) to initialize our centroids and then we defined an optimization parameter, i.e., the average of the distances between every array in the cluster and the centroid. We run the algorithm 1000 times (an arbitrary number) minimizing the optimization parameter.

The output of this clustering process is shown in Table 3.4.

Relation	Statement	cl.1	cl.2	cl. 3	cl. 4
No Opposition	**Statement 1**: A situation in which both events take place simultaneously and with the same participants is possible	0	2.16	2.05	0
Complementarity	**Statement 2**: Given a situation, if one of the two events does not take place, the other occurs	**100**	0	21.23	**100**
Antonymy	**Statement 3**: Given a situation, a "neutral" alternative in which none of the events is possible	0	**86.58**	**74.66**	**100**
	Statement 4: It is possible to use gradable modifiers as "a bit", "moderately", "a lot", etc..	5.61	17.31	26.03	27.78
Reversiveness	**Statement 5**: It is possible that the two events occur repeatedly several times one after the other	4.67	16.01	**91.78**	0
	Statement 6: One event is possible only if the other has occurred	6.54	9.09	**68.49**	11.11
	Statement 7: If one event occurs, and then the other takes place, the situation returns to the initial state	7.48	10.82	**87.67**	5.56
Converseness	**Statement 8**: the two verbs describe the same situation from two different points of view	0.09	1.73	1.37	0
	Total Judgments	107	231	146	18

Table 3.4: Clustering output. Percentage of times a statement is selected in the judgments of a cluster.

Statement 2, related to the complementarity relation, is clearly predominant in one out of the four clusters (Cluster 1, composed by 107 judgments); on the other hand, one of the statements related to the antonymy relation (statement 3) is the center of a second cluster (i.e., Cluster 2 with 231 judgments). A third cluster (Cluster 3 composed by 146 judgments) presents the statements that characterize antonyms (statement 3) and reversives (statement 5, 6, 7) as predominant. Finally, a fourth cluster is created. Analyzing the different iterations we performed, we observe that while the first three clusters remain stable, Cluster 4 tends to include judgments that do not clearly fit in the previous groups and it is always the less populated. In the iteration that we took as our final result, cluster 4 is constituted by 18 judgments (in which both statement 2 and 3 were marked together). Statement 8, related to converseness, is not prevailing in any cluster and is selected in 7 cases.

By comparing the first two clusters, results seem to suggest that workers recognized a distinction between the complementarity relation and the antonymy relation.[96] In fact,

[96]This seems in contrast with what is stated in Benotto (2014): "[..] more particular relations like contrariety and complementarity are not well distinguished by language users." (Benotto, 2014, p.30).

in Cluster 1, statement 2 (that refers to the "mutual exclusivity of the *complementaries*") is selected in all the judgments, while statement 3 (that refers to the "existence of a third alternative") is never selected, and statement 4 (that refers to "being gradual") is selected in 5.6% of the cases. On the contrary, in Cluster 2 where statement 3 is prevalent and statement 4 is selected in 17.3% of the cases (the "existence of a third alternative" and "being gradual" are both typical of *antonyms*), statement 2 characterizing *complementarity* is never chosen. According to our definition in Section 2, these two categories are actually very different with respect to the "scalar dimension" (while *antonyms* are gradable, *complementaries* are not) and with respect to their "mutual exclusivity" (while *complementaries* are mutually exclusives, *antonyms* are not) (See Section 3.4.1.1). A closer analysis of the results in this direction shows that statements 2 and 3 are selected in most of the cases (455/502) but they are selected together only in 28 cases. If we include the other statement related to antonymy (statement 4), cases of complementary and antonymy statements overlapping are 39 over 502 judgments. This seems to demonstrate that for the verb sense pairs annotated by the workers, the two categories are frequently chosen but in general not confused.

The third cluster (146 judgments) presents the features that typically characterize *antonyms* (statement 3, the "existence of a third alternative") and *reversives* (statements 5, 6, 7 referring to their "interaction with the entailment relation" and their "being an expression of a change in opposite direction") as predominant. This seems to suggest that *reversiveness* is not an exclusive relation, but in some cases co-exists with other opposition types, particularly *antonymy*. This hypothesis has already been suggested in the literature. In Section 2.1, we have seen that according to Lyons (1977), directional opposites (of which *reversives* are part) "cannot always be distinguished from the (other) three [types]" (Lyons, 1977, p. 281). Examples in which a pair of verbs is assigned to these two types can be found also in the proposed literature: the pair *lengthen / shorten* is considered by Cruse (1986) as an example of *reversives* enhancing the fact that they express a change in opposite directions[97], the same pair is provided by Ježek (2016) as *antonyms* considered the gradability of the verbs in the pair.

[97] As mentioned in Section 2.1, according to Cruse (1986), *antonyms* are mainly adjectives, and *reversives* are all verbs. The same author suggests that some *reversives* verbs can derived from *complementary* or *antonyms* adjectival opposites and, for them, "properties of the derived reversives can frequently be correlated with properties of the underlying adjectives." (Cruse, 1986, p.230).

In general, we can say that data resulting from the clustering procedure corroborate the hypothesis that some types of opposition exclude each other, while others interact, being recognized as compatible by the workers.

Analysis of worker's judgments. The majority of the workers marked different statements for different verb pairs and their judgments are in general distributed over different clusters.

A further analysis of the collected data concerns the observation of workers' agreement. Given that in our task opposition relations are associated to statements (e.g. for reversiveness, statements 5, 6, and 7), we calculated this agreement on single statements in each sense pair (e.g. joining together all judgments for statement 2 for the pair *caricare*#7-*scaricare*)[98]. More in detail, we first calculated the interannotator observed agreement on a single statement for each pair, considering a match when workers agree both on selecting and not selecting the statement. Then, for each statement separately, we summed the 135 observed agreement and we calculated the average. Values (reported in Figure 3.15) are in average over 95% for statements 1 and 8; workers mainly agreed in not selecting these statements, being both marked only in 7 pairs (see external columns of the table). For the other statements, agreement is not inferior to 56% (statement 3).

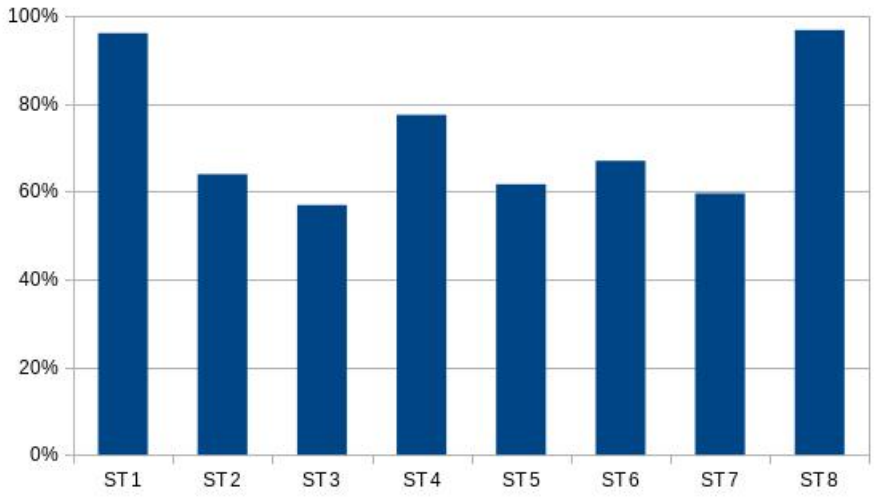

Figure 3.15: Average interannotator agreement for each statement (ST).

[98]The 502 collected judgments refer to 138 sense pairs in T-PAS.

3.4.4 Discussion on Crowd-sourcing Settings and Methodology

In this section we discuss some of the choices we made for the definition of the crowd-sourcing tasks presented in the previous sections of this chapter, and we discuss the crowd-sourcing methodology itself in the light of our experience.

Discussion on data collection setting. We showed the feasibility of collecting opposition relation among Italian verb senses through crowd-sourcing. However, more than half of the collected judgments in Task A (i.e., *opposition identification* step) are "S2 does not make sense". Even if this result was expected and it is *per se* very interesting, the amount of answers of this type can be reduced in favour of the amount of data that can be annotated in the following tasks. A possible method for improving this aspect is to filter some of these "no sense" sentences by performing a syntactic analysis before launching the task, e.g. discarding intransitive verbs that happen to be substituted in sentences where the *source verb* is a transitive one: in this case in fact the intransitive *target verb* will have a "non required" direct object. Yet, as we saw, syntactic mismatches are not the only cause that lead to create a "no sense" sentences; further work would include a deep examination of these causes.

Moreover, an alternative setting of Task A (i.e., the *opposition identification* step) would have been to ask, first, if S2 makes sense (yes/no question) and then, in a different task, to ask if an opposition is identified. This choice was however discarded so that the task would not take too long for the workers.

Concerning the setting of the Task C (i.e., the *type-of-opposition identification* step) two observations can be useful for future work or related initiatives. As we said, the statements were created to catch characteristics of the types of opposition by following the definitions in the literature. To operate in a crowd-sourcing setting in which non-expert workers cannot be trained (except for an instruction page - see Appendix B) imposes some simplifications, however. The proposed eight statements thus represent an effort to combine the complexity of the definitions in the literature with the need of a simple message for the workers. In some cases, this operation was not easy (for example, translating the "non mutual exclusivity" of *antonyms* with statement 3) and in other cases it was a limitation; for instance, we could not find a good statement to represent and explain to the workers the "again-test" that characterizes *reversives*. The methodology also imposes

brevity on the task and thus we preferred to ask a restricted number of questions perhaps limiting the chance to investigate more deeply each type of opposition (for example, by asking multiple questions for each type).

For the Task C we decided to show all 8 statements as answers of a unique question to enhance the selection of the best choice among some possible answers and reduce the cases of "second thought" of previous answers (i.e., seeing a more adequate option, the worker wants to change his/her judgments on a previous answer). Also, we intended to avoid the workers being tempted to answer repeatedly "yes" or repeatedly "no". An alternative design for this task would be to present each statement separately and ask workers whether the statement is valid for the pair of verbs they are examining (i.e., as a yes/no question). This would reduce the amount of information asked in a single question (as Mohammad et al. (2013) do), but in case of "second thought" the worker would be required to go back in the task to find the judgment s/he wants to change, change it, and finally return to the page in which s/he should continue the task; with this setting, the task would be more complex.

Although the use of examples in place of sense definitions simplified the annotation, many workers considered the tasks rather difficult. Another option would certainly have been possible. For example, in our configuration, one worker had to perform three tasks for the same pair (that is, given a pair of sentence, provide the answers for the three tasks - *all tasks per pair*); a possible optimization would have been to ask workers to first solve Task A for all the data, then to complete the entire Task B for the pairs in which "S2 makes sense and holds an opposition relation with S1" and, finally, to complete Task C . In this way the workers focus on only one task at a time (*all pairs per task*). However we have no clue if this latter setting would have simplified the operation and worked better. The *all tasks per pair* indeed can appear more complex (i.e., it requires: judging the sense of S2, evaluating the opposition, disambiguating the target verb and identify the type of opposition) but the worker is always focused on the same pair of verbs in the same context.

Discussion on the use of crowd-sourcing. One of the questions that we posed in the introduction concerns the adequacy of the crowd-sourcing methodology for collecting the data we need to populate a lexical resource with opposition relations.

In general, we believe that our experiment (schematized in Figure 3.16) proves an adequate methodology. A consideration that is worth discussing concerns the number of workers. We did not have as many workers as expected (we expected to collect a greater number of judgments in a shorter period of time). Most of them were also discarded for low accuracy in the initial page, which had only Test Questions.[99] To try to attract more workers, we increased the reward but this did not help increase their number. We decided not to decrease the threshold of accuracy in the TQs filter in order to not affect the quality of the data. A possible explanation for this low participation rate is that the total number of Italian workers is *per se* smaller than the number of contributors for other languages, e.g. for English. Also, it is possible that the number of workers with Italian skills the platform could reach was too low. Moreover, we believe that we chose a bad moment for launching the task: we began in the month of July, a period in which many people take a summer break.

The task also required continuous monitoring that we did not expect was needed: in fact, workers could interact with the task manager and could be "forgiven" for having missed a TQ if they provided a valid argument.

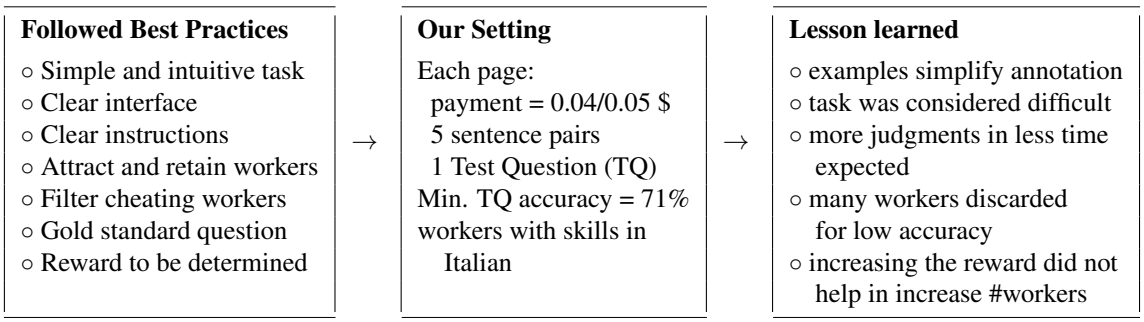

Figure 3.16: A representation of our crowd-sourcing task.

3.4.5 Contributions, Shortcomings and Future Work

In this section, we presented a crowd-sourcing task for the acquisition of the opposition relation among senses of verbs and for the acquisition of types of opposition to enrich the Italian resource T-PAS (Ježek et al., 2014). [100]

[99] As explained, a Test Question is a question for which we knew the answer. If a worker missed many TQs s/he was not permitted to continue the annotation and his/her judgments were rejected.

[100] Some of the results presented in this section have already been published in (Feltracco et al., 2015b) (Task A and B) and in (Feltracco et al., 2016c) (Task C). For this project, the author of this thesis collaborated with Elisabetta Jezek (University of Pavia), Bernardo Magnini (Fondazione Bruno Kessler, Trento) and Simone Magnolini (Fondazione

Our aim was to collect data-driven evidence about the opposition of pairs of Italian verbs in context. More specifically, the crowd-sourcing job was based on the identification of opposition by observing pairs of opposites in sentences (Task A). One verb (*source verb*) was already disambiguated (i.e., we knew which sense it corresponds to in T-PAS), while the other (*target verb*) was disambiguated by workers by comparing the verb with the same verb in other contexts, where each context involved a different sense of the verbs, as defined in T-PAS (Task B). We collected a portion of the data necessary for the annotation of opposition in T-PAS. In particular, we collected judgments for 712 pairs of sentences. The collection was organized such that for each pair of verb senses, we collected judgments on three pairs of sentences and for each pair of sentences, we collected judgments by three workers. To increase the reliability of the data, we considered the same verb pair twice: first when a verb plays the *source verb* and then when it plays the *target verb* (e.g. pair 1: source verb *caricare* - target verb *scaricare*, and pair 2: source verb *scaricare* - target verb *caricare*). We believe that this can ensure that the data are more reliable and, thus, provide us with the confidence needed for the annotation of opposition in a lexical resource. The experiment showed the feasibility of using this methodology with the proposed settings, given our desiderata.

Through the same crowd-sourcing exercises, we indirectly collected judgments on four opposition relations -*complementarity, antonymy, conversesness*, and *reversiveness*- by asking workers in the crowd to consider the pair of verbs in context and mark, among a list of eight statements, which ones are valid for that pair (Task C). These statements express individual characteristics of the four opposition relations we assumed from the literature. After clustering workers judgments, three main groups can be distinguished: one cluster includes judgments where the statement for *complementarity* is predominant, another includes judgments in which one of the *antonymy* statement is prevailing, and a third includes judgments in which both statements for *antonymy* and *reversiveness* have been marked by workers. The collected data allow us to draw interesting conclusions about the categories of oppositions and their relatedness. In fact, the results seem to confirm a main distinction between *complementarity* and *antonymy*, and suggest that the relation of *reversiveness* is not an exclusive relation, but tends to add to other types, particularly

Bruno Kessler, Trento and University of Brescia). Specifically, the author significantly contributed to designing the task, managing the experiment in the crowd-sourcing platform, monitoring the workers' jobs, and she participated in the analysis of the data.

antonymy, as these types have been indirectly recognized as compatible by the workers.

The collected data represents about 10% of all verb pairs we retrieved adopting the criteria presented in Section 3.4.2 We hypothesize that it would take about 10 times longer to collect the data for all the 210 symmetrical pairs we initially retrieved and complete the annotation of the opposition relation through out T-PAS following our methodology and settings. In Table 3.5 we schematize this hypothesis.

	Collected data	**Total effort: estimation**
verb pairs		210 symmetrical pairs (x 2)
total senses		2166
total sentence pairs	712	6498 (= 2166 pairs x 3 sentences[101])
total judgments	2136	19494 (= 6498 sentence pairs x 3 judgments)
time	almost a 1 month	*9.1 month*
money	less than 50 euros	*about 460 euros*

Table 3.5: An attempt to estimate the effort for annotating the opposition relation in T-PAS.

We also tried to quantify how many opposition relations would be annotated in T-PAS after the enrichment is completed (Table 3.6). We base this estimation on the annotation of opposition relations we performed in order to test our annotation schema presented in Section 3.2. In that case we annotated a total of 44 relations of different types for 25 pairs of verbs. We estimate that for 210 pairs, about 370 relations of opposition would be added to T-PAS.

	Data used for testing the annotation schema	**Total relation: estimation**
verb pairs (symmetrical)	25	210
total relation	44	*369*

Table 3.6: An attempt to estimate how many opposition relations will be annotated in T-PAS.

Once the judgments are collected, the analysis of judgments for each sense pair (e.g. for all three pairs S1-S2 of *t-pas* of the source verb *caricare*#7 - target verb *scaricare*) will be needed in order to tag a relation to a verb pair at the sense level. As far as we know, the extension of the annotation for all the opposite Italian verb pairs will make T-PAS the first resource for Italian systematically enriched with opposition relations using crowd-sourcing data. At the same time, it will be possible to gain important insights into the distribution of the different types of oppositions.

Concerning the use of crowd-sourcing, we have shown that the experiment demon-

[101] In our case we have few sentences less due to the unavailability of three sentences in the corpus for some senses.

strates the feasibility of using a methodology based on showing workers the verbs in context. However, less effort on the acquisition of the data was initially expected (i.e., continuous monitoring was required and more data were expected). Also, the simplicity requirement mentioned for the task must be carefully considered.

In synthesis, our contribution is manifold: i) we defined and tested a schema for the annotation of types of opposition; ii) we proposed a methodology for the acquisition of the relation of opposition, and discussed it in terms of the required effort needed to complete the acquisition of data for the annotation of the T-PAS resource; iii) we provided the analysis of a crowd-sourcing experience that can help other researchers in using this methodology; iv) we conducted an empirical study on types of opposition employing a clustering procedure: the resulting data corroborate that some types of opposition exclude others (there is a main distinction between *complementarity* and *antonymy*), and suggest that the relation of *reversiveness* is not an exclusive relation.

Chapter 4

Discourse Contrast: Background

In the previous chapters we have presented our work on opposition: a relation between two lexical units that contrast with each other. In the following two chapters we change our perspective and move from a lexical to a discourse level: we will present our work on discourse contrast relations. In particular, in this chapter we define the concept of discourse contrast relation by providing an overview of the definitions of contrast in Rhetorical Structure Theory (Mann and Thompson, 1988), in Segmented Discourse Representation Theory (Asher and Lascarides, 2003), and in the Penn Discourse Treebank (PTDB) (Prasad et al., 2007). We also propose a survey of large corpora annotated with discourse relations inspired by the PDTB.

Discourse relations are relations between parts of a discourse. By discourse we mean a coherent sequence of sentences, propositions or speech through which an idea, an argument, a message, or a story is conveyed. Hence, a discourse can be a simple sentence or even a large document. Let's consider the following examples:

(4.1) [When you press the button$_{a]}$, [the door opens by itself.$_{b]}$

(4.2) [Mary passed the exam.$_{a]}$ [John failed it.$_{b]}$

In Example 4.1, two parts of the sentence can be distinguished (i.e., the one marked by [a] and the one marked by [b]). Between the two parts a relation of *condition* holds; the situation expressed in [a] is the condition for the situation in [b] to happen (i.e., the door opens by itself if the button is pressed). Similarly, in Example 4.2 a discourse relation

holds between the two sentences but in this case there is a relation of *contrast* between Mary having passed an exam and John having failed it.

Relations such as the ones between [a] and [b] in the two examples above are called *discourse relations*. A discourse relation holds between two parts of a discourse that are called here *arguments*[102] (for instance, in Example 4.1 'When you press the button' and 'the door opens by itself' are the two arguments). Discourse relations can express different semantic relations: for example, as explained, it can be said that in 4.1 there is a *condition* relation, and in 4.2 a *contrast* relation. If the relation is lexicalized by some cue elements, these are called *discourse connectives* (Prasad et al., 2007)[103]: for instance, in 4.1, 'When' is a discourse connective for the *condition* relation. Note that in 4.2 there is no such explicit connective: in fact, discourse relations might not be overtly marked by a lexical element.

In this chapter, we aim at giving some insights on the definition of *discourse contrast relation* (e.g., Example 4.2). The first section is dedicated to *discourse relations* in general; we propose an overview of the two most influential theories on discourse relations (i.e., Rhetorical Structure Theory and Segmented Discourse Representation Theory). Moreover, we present the major corpora annotated with discourse relations and the Penn Discourse Treebank (PDTB) (Prasad et al., 2007). In the second section, we focus on the contrast relation by comparing different definitions provided in the literature, and we describe how contrast relations are defined in the PDTB, the major work of reference for discourse relations in the field of Computational Linguistics, and the major work of reference for the creation of Contrast-Ita Bank, a corpus of news which we annotated with the relation of *contrast*.[104]

The aim of this overview is to compare the definitions of the *contrast relation* in the literature and present the notion of *contrast* adopted in this thesis. This constitutes a background for the data-driven investigation of *contrast* we will present in the following chapter.

[102]The terminology we adopt is mainly the one used in the Penn Discourse Treebank (Prasad et al., 2007).

[103]We can define *connectives* as lexical elements that connect portions of text together into a coherent sequence. The notion of discourse connectives we are adopting here is further discussed in Section 5.2.1.

[104]We present Contrast-Ita Bank in Chapter 5.

4.1 Discourse Relations

In this section we propose an overview of the Rhetorical Structure Theory (RST, Mann and Thompson, 1988) and Segmented Discourse Representation Theory (SDRT, Asher and Lascarides, 2003): two influential theories on discourse relations (Section 4.1.1). We also present the major annotated corpora following these theories and the Penn Discourse Treebank (PDTB) (Prasad et al., 2007) (Section 4.1.2).

4.1.1 Discourse Relations in RST and SDRT

One of the most influential theory on discourse relations[105] is the Rhetorical Structure Theory (RST) by Mann and Thompson (1988)[106]. In RST, discourse is considered as a structured linguistic unit and discourse relations are studied as relations between parts of this structure that lead to textual coherence. In RST discourse structure is described as a tree structure in which leafs correspond to text fragments that represent the minimal units of the discourse, and the nodes of the tree correspond to contiguous text span.

Figure 4.1 shows an example of this tree structure, that is the discourse structure proposed by Mann (1999) for the beginning of a Scientific American article (title and abstract). In the figure, the text has been broken into numbered units for analysis.

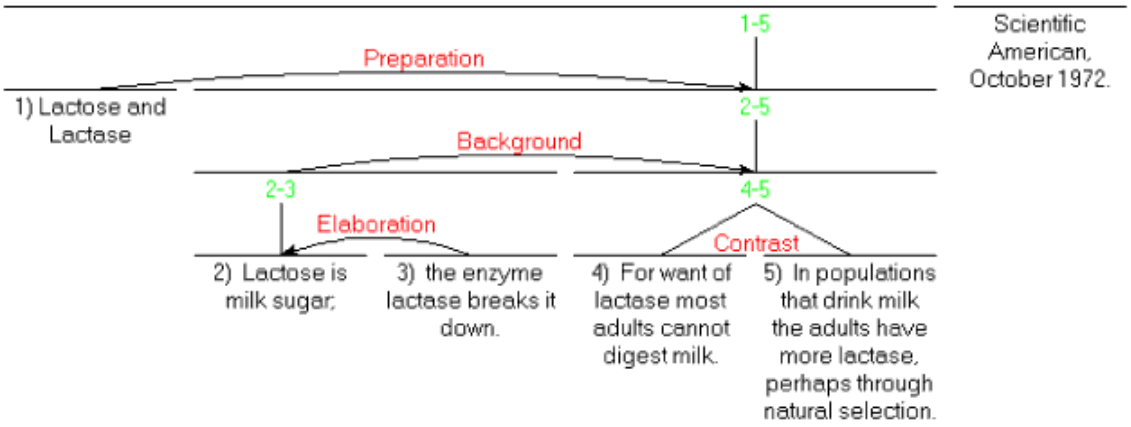

Figure 4.1: A RST representation of the *Lactose Example* in Mann (1999).

[105] Discourse relations have been discussed also under the name of *rhetorical relations* (Mann and Thompson, 1988), *coherence relations* (Hobbs, 1985).

[106] RST was originally developed as part of studies of computer-based text generation to compensate a lack of a discourse structure theory with enough detail for computational aims. See RST website `www.sfu.ca/rst/index.html`.

As can be seen, the extent of each unit is defined and relations (in red) hold between two adjacent units or spans of texts. Notice also that the entire text is a sequence of units, there is no portion of text excluded from the analysis. These units are essentially adjacent clauses or group of clauses that share no text. For example, there exists a relation named CONTRAST[107] between unit 4 and unit 5, and another relation named BACKGROUND between the span including units 2-3 and the span including units 4-5. In RST, relations are described whether or not they are lexically signalled (as an example, there is no lexical element indicating CONTRAST between 4 and 5).

The spans involved in a relation can be of two types: *nucleus* or *satellite*. The *nucleus* "is the span that is consistently more central to the writer's goals and less subject to deletion or substitution of other material" and the *satellite* is the less central and "tends to enhance the function" of the nucleus (Thompson and Mann, 1987). A discourse relation can hold between two *nuclei* (i.e., *multiple nuclei relation*) if the two parts are equally central, or between a *satellite* and a *nucleus* whether one is subsidiary to the other[108] (no matter if the *satellite* precedes or follows the *nucleus* in the text).

Examples of *multiple nuclei relation* and *nucleus-satellite relation* are respectively the above mentioned CONTRAST and BACKGROUND. CONTRAST holds between the two *nuclei*; that is, two equally central spans of text representing the two arguments of a contrast: in Figure 4.1 the relation holds between unit 4 (adults lack lactase) and 5 (adults have lactase)[109]. The relation of BACKGROUND holds between a *nucleus* that represents the text whose understanding is facilitated (by the *satellite*) and a *satellite* that represents the text for facilitating understanding (of the *nucleus*). In the example, 4-5 is the *nucleus* (discussion on the presence of lactase in adults) and 2-3 is the *satellite* (explanation of what lactose and lactase are, which simplifies the understating of the discussion about the presence of lactase in adults).

According to RST, while a *nucleus* is allowed to be involved in more than one relation

[107] We use uppercase for indicating the relation type.

[108] According to Stede (2012), Mann and Thompson suggest "a deletion test, which predicts that when removing all satellite units from a text, the main message is still recognizable [..], whereas removal of nuclei from a text leads to incoherence." (Stede, 2012, p. 84). The concepts of "more central" parts and "less central" parts have been further exploited, for example by Marcu (1998a,b) for a NLP task called *text summarization*. This task consists in, given a text, automatically producing a second text which is a summary of the original one. In his work, the author uses the discourse representation structure of a text (i.e the nucleus-satellite relations structure) to select the most important units in a text.

[109] The definition of CONTRAST relation in RST will be further discussed in Section 4.2.1.

Presentational	Subject Matter
Antithesis & Concession	Circumstance
Antithesis	Solutionhood
Concession	Elaboration
Background	Condition & Otherwise
Enablement & Motivation	Condition
Enablement	Otherwise
Motivation	Interpretation & Evaluation
Evidence & Justify	Interpretation
Evidence	Evaluation
Justify	Relations of Cause
Restatement & Summary	Volitional Cause
Restatement	Non-volitional Cause
Summary	Volitional Result
	Non-volitional Result
	Purpose
	Other relations
	Contrast
	Sequence

Table 4.1: The list of 23 relations identified by Mann and Thompson (1988).

(e.g., A is at the same time a *nucleus* for B and C), this is not the case for a *satellite*[110].

These relations are different one from the other depending on the effect that the writer intends to achieve, and this effect is translated as "constraints that operate on the *nucleus*, on the *satellite*, and on the combination of *nucleus* and *satellite*" (Mann and Thompson, 1988). To give an example, according to Mann and Thompson (1988), the BACKGROUND relation presents the following characteristics:

- *constraints on the nucleus*: the reader won't comprehend the nucleus sufficiently before reading the text in the satellite;

- *constraints on the combination of nucleus and satellite*: the satellite increases the ability of the reader to comprehend an element in the nucleus;

- *effect*: the reader's ability to comprehend the nucleus increases

A main distinction, reported in Table 4.1, can be drawn between *presentational* relations and *subject matter* relations (Mann and Thompson, 1988)[111]: *presentational* relations are those whose intended effect is to increase some inclinations in the reader;

[110] In RST, no schema of relations accounts for the case in which a *satellite* involved in more than one relation (Mann and Thompson, 1988, p.247).

[111] In the original classification by Mann and Thompson (1988), the group of Restatement & Summary is classified as *subject matter* group of relations. However in the update website of reference for RST (`http://www.sfu.ca/rst/01intro/definitions.html` the group is considered as *presentational*).

subject matter relations are those whose intended effect is that s/he recognizes the relation. For instance, BACKGROUND is a *presentational* relation: the intended effect is to increase the comprehension of the *nucleus*. CONDITION is a *subject matter* relation that holds between a conditioning situation, *satellite*, and the result from the occurrence of the conditioning situation, *nucleus* (for instance, in Example 4.1, the text signalled by [a] will be the *satellite* and [b] will be the *nucleus*). In this case, the intended effect is the understanding that the realization of *nucleus* depends on the realization of *satellite*.[112]

In Table 4.1 relations are also grouped as proposed in Mann and Thompson (1988). Relations in each group "share a number of characteristics and differ in one or two particular attribute" (Mann and Thompson, 1988). For example, ENABLEMENT and MOTIVATION belong to the same group as both involve the increasing of the probability that the reader would perform the action expressed in the *nucleus*; however, in ENABLEMENT the potential ability of the reader to perform the action is increased (i.e., the reader is enabled to do it), while in MOTIVATION, the attribute that is increased is the desire of the reader to perform the action (i.e., the reader is motivated to do it).

Mann and Thompson (1988) clarify that the list is potentially open and, in fact, this set of relations has been further enlarged in other works. For example, in a reference manual for annotating discourse relations in texts within the RST framework, Carlson and Marcu propose a larger data driven set of 78 relations grouped in 16 classes[113] (Carlson and Marcu, 2001).

Following the purpose of this chapter, in the Section 4.2.1 we will further discuss the relations of CONTRAST, ANTITHESIS and CONCESSION as identified in RST.

Another important theory that recognizes discourse structure and discourse relations is the Segmented Discourse Representation Theory (SDRT, Asher and Lascarides, 2003). SDRT extends the logical relations of discourse introduced in the Discourse Representation Theory (DRT, Kamp and Reyle, 2013) to account for discourse relations. In par-

[112]Notice that Stede (2012, p.85) reports this distinction in terms of "semantic" vs "pragatic" relations.

[113]With respect to the set proposed by Mann and Thompson (1988), this set includes new relations (such as a class for TEMPORAL relations) or further specification of the already proposed ones (e.g., the ELABORATION class includes the following relations: elaboration-additional, elaboration-general-specific, elaboration-part-whole, elaboration-process-step, elaboration-object-attribute, elaboration-set-member, example, definition). Authors say that despite the inventory is highly detailed, annotators strongly preferred keeping a higher level of granularity in their selections during the tagging process (Carlson et al., 2001).

ticular, SDRT "uses rhetorical relations to model the semantics/pragmatics interface of a discourse" (Lascarides and Asher, 2008, p.1).

In SDRT, relations are semantically defined and are distinguished in *subordinating* and *coordinating* both holding between two parts of text. In the *subordinating* relation, one part of text plays a subordinate role relative to the other; in a *coordinating* relation both parts play equally a central role.[114] For example, consider two relations identified in SDRT: ELABORATION and NARRATION. The semantic of ELABORATION "entails the events described in one part describe in more detail those described in the other part" (Lascarides and Asher, 2008) (i.e., there is change in the granularity of description) and in this sense it is *subordinating*. On the other hand, NARRATION "reflects temporal progression between the events" (Lascarides and Asher, 2008) (there is no change in the granularity of description), and it is a *coordinating* relation. In Example 4.3, provided by Asher and Lascarides (2003) (and similarly proposed also in (Lascarides and Asher, 2008)), ELABORATION is recognized between 4.3a and 4.3b: to have a great meal is a more detailed description of having a great evening; NARRATION is recognized between 4.3c and 4.3d: eating salmon and devouring cheeses are two equally central event described in progression.

(4.3) a. π_1 Fred experienced a lovely evening last night.

b. π_2 He had a fantastic meal.

c. π_3 He ate salmon.

d. π_4 He devour lots of cheese.

e. π_5 He won a dancing competition.

Example from: Asher and Lascarides 2003

Recalling DRT (Kamp and Reyle, 2013), SDRT uses also box representations. Just to give an idea, Figure 4.2 illustrates a SDRT box representation for the discourse structure in 4.3. The notation $K_{\pi i}$ stands for the DRS representing the *ith* sentence. The figure is taken from Asher and Lascarides (2003)[115].

In SDRT the same textual part of a discourse can hold different relations with different parts. In the following example by Lascarides and Asher (2008, p.18), 4.4a holds both a

[114]For a study over *subordinating* and *coordinating* relations, see (Asher and Vieu, 2005).

[115]For further discussion see also (Danlos, 2005).

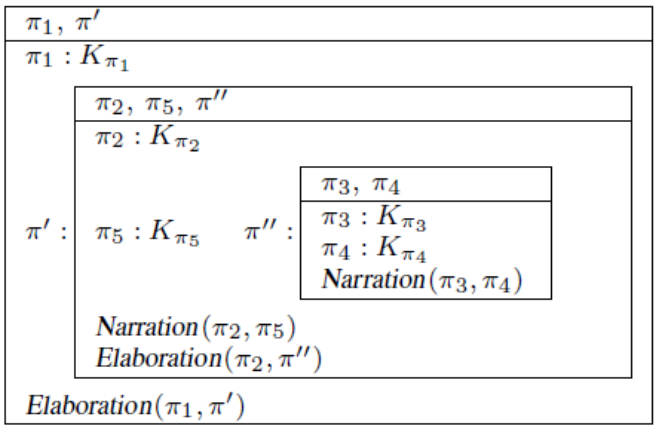

Figure 4.2: An example of SDRT box representation from Asher and Lascarides (2003).

CONTRAST relation with 4.4b and a ELABORATION relation with 4.4c.

(4.4) a. A: Max owns several classic cars.

b. B: No, he doesn't.

c. A: He owns two 1967 Alfa spiders.

Example from: Lascarides and Asher 2008, p.18

The same example proves also how, according to SDRT, a relation can hold between non continuous parts of the text (i.e., non adjacent parts) as -as just mentioned- ELABORATION is identified 4.4a and 4.4c. Furthermore, it can bee seen that (as for RST), relations do not need to be explicitly signalled by some expressions in the text (Asher, 1993), in fact there is no lexical element connecting "Max owns several classic cars." and "He owns two 1967 Alfa spiders".

In the works by Asher and Lascarides (Asher, 1993; Asher and Lascarides, 2003) the set of relations includes the just cited ELABORATION, NARRATION and CONTRAST[116], but also EXPLANATION (which involves temporal inclusion) , BACKGROUND (which imposes temporal overlap between the arguments of a relation), PARALLEL (which required arguments to be semantically and structurally similar).[117] Here some

[116]In the next section, we will further discuss the relation of CONTRAST.

[117]For completeness we report the whole list of relations in the Glossary of Discourse Relation of (Asher and Lascarides, 2003, 459): ALTERNATION, BACKGROUND, CONSEQUENCE, CONTINUATION, ELABORATION, EXPLANATION, NARRATION, RESULT, CORRECTION, COUNTEREVIDENCE, QUESTION-ANSWERING-PAIR, CONTRAST, PARALLEL. They also include cognitive-level relations: ACKNOWLEDGEMENT, INDIRECT-QUESTION-ANSWERING-PAIR, PARTIAL-QUESTION-ANSWERING-PAIR, NOT-ENOUGH-INFORMATION, PLAN-CORRECTION, QUESTION-ELABORATION, PLAN-ELABORATION.

examples for these relations by Asher and Lascarides (2003) :

(4.5) a. Max fell.
b. John pushed him.
Ex. for EXPLANATION

(4.6) a. Max entered the room.
b. It was pitch dark.
Ex. for BACKGROUND

(4.7) a. John said that Mary cried.
b. Sam did too.
Ex. for PARALLEL

Both RST and SDRT recognize the structure of the discourse. Both also recognize that the discourse relations hold between two textual elements (here *arguments*) and that these arguments can be equally central in the relation (*multiple nuclei relation; coordinating relation*) or one can be more central than the other (*nucleus-satellite relation; subordinating relation*).

However, there are also crucial differences. SDRT allows recognising more than one discourse relations between two arguments, as in the following example by Lascarides and Asher (2008, p.17) where 4.8a and 4.8b are related by both CONTRAST (a contrast is identified and signalled by *but*) and NARRATION (the temporal progression between the events is reflected). This is not the case in RST, for which only one relation can hold between two arguments (Danlos, 2005, p.8).

(4.8) a. John bought an apartment
b. but it rented it.

Furthermore, SDRT accounts for non adjacent parts of the text as seen in Example 4.4; while in RST (as proposed by Mann and Thompson (1988)) relations are identified between adjacent portions of text. Finally, according to Danlos (2005), in SDRT "preposed subordinate clauses are ignored", so the most central part always precedes the other; while RST accounts also for case in which (in SDRT terms) the *satellite* precedes the *nucleus*.

Other approaches have been proposed for the investigation of the discourse relations. As an example, an important contribution within the field of Artificial Intelligence is by Hobbs (1985). This framework, frequently discussed along the just mentioned RST and SDRT, is related to discourse interpretation. This is reflected in the proposed classification of discourse relations that, in fact, depends from the type of inferences that the reader/listener draws. Specifically Hobbs (1985) defines 4 classes of relations considering whether: i) the speaker wants to convey message, ii) the message is in service of some goal, iii) the

speaker must link what he says to what the listener already knows, iv) the speaker should ease the listener's difficulties in comprehension [118] (Hobbs, 1985).

Theories of discourse relations have lead to the creation of accordingly annotated corpora. In the next section, we introduce resources created in the RST and SDRT frameworks and the largest corpus annotated with discourse relations.

4.1.2 Corpora Annotated with Discourse Relations: the Penn Discourse Treebank

Annotated corpora are useful both for the investigation of discourse relations based on empirical data, and for NLP applications for which a discourse level analysis is required. In fact, manually annotated resources have been used, for instance, for developing methods and tools for the automatic identification and disambiguation of explicitly marked or implicitly conveyed discourse relations[119] (Pitler and Nenkova, 2009; Pitler et al., 2009; Zhou et al., 2010; Rutherford and Xue, 2015), for the identification of the spans of text that are linked by relations (discourse segmentation), for the automatic creation of a summary of a written text (text summarization) (Marcu, 2000), and for machine translation (Meyer and Webber, 2013).

Contributions for the creation of annotated corpora include works that follow the two theories presented in Section 4.1.1: both RST and SDRT have been applied for developing schemas for the annotation of resources. As mentioned in the previous section, Carlson and other researchers propose the annotation of discourse relations within the RST framework. Specifically, they create the RST Discourse Treebank (Carlson et al., 2002), a collection of 385 Wall Street Journal articles (extracted from the Penn Treebank (Marcus et al., 1993)) annotated with a list of 78 discourse relations[120]. In the framework of SDRT, a subcorpus of Wall Street Journal articles was also annotated, in the Discor project (Reese et al., 2007). Authors report the annotation of the texts with 14 different discourse relations[121]. Benamara and Taboada (2015) report the existence of annotated

[118] Class i) includes OCCASION, class ii) includes EVALUATION, class iii) includes BACKGROUND and EXPLANATION, class iv) includes PARALLEL, ELABORATION, EXEMPLIFICATION, CONTRAST, and VIOLATED EXPECTATION (Hobbs, 1985).

[119] The task of identifying discourse relations in the form of a discourse connective (explicit or implicit) taking two arguments has also been called *shallow discourse parsing* and constituted a shared task of the CONLL conference in 2015 and 2016 (Xue et al., 2015) `http://www.cs.brandeis.edu/~clp/conll16st/`.

[120] See footnote 113 of this chapter.

[121] They select: CONTINUATION, CONSEQUENCE, BACKGROUND, NARRATION, ALTERNATION, ELABORATION RESULT, EXPLANATION, CONTRAST, COMMENTARY, PARALLEL, SOURCE, PRECONDITION,

corpora following RST also in Basque, Dutch, German, English, Portuguese and Spanish and following SDRT in Arabic, French and English.

However, the most influential project for discourse relations used in the NLP field is the Penn Discourse Treebank (Prasad et al., 2008). The Penn Discourse Treebank (PDTB, Miltsakaki et al., 2004; Webber et al., 2006; Prasad et al., 2007, 2008) is a large scale corpus of English texts annotated with a layer of discourse annotation. The corpus consists of a million words and composed by the same Wall Street Journal articles on which the Penn Treebank (PTB) II corpus (Marcus et al., 1993) was built.

According to Miltsakaki et al. (2004), the PDTB is -with respect to the RST Discourse Treebank, created in the RST framework- a "more basic discourse-level annotation project [..] that aims to produce a large-scale corpus in which discourse connectives are annotated, along with their arguments" (Miltsakaki et al., 2004).

Indeed, in RST discourse is described in terms of a unique structure (a tree) and resource annotated following this theory aim at the annotation of this entire discourse structure; conversely, the PDTB is not tied to any particular discourse structure theory (it is 'pre-theoretical' in Stede's words (Stede, 2012, p.101)). In the PDTB, the identification of discourse relations is tackled using a *bottom-up* approach to discourse structure (Miltsakaki et al., 2004) that focuses primarily on the identification of local connectives[122] and their arguments, so that each relation is analysed independently of other relations (Rehbein et al., 2016). According to Scheffler and Stede (2016a), this decision "aims at taking 'one step beyond sentence syntax' but not the leap toward a discourse representation whose construction would be more difficult to annotate and involve more subjective interpretation". Moreover, while in RST the discourse relations hold primarily between adjacent spans of text (as we saw in the previous section), in the PDTB approach relations triggered by a connective can also connect non-adjacent elements (i.e., the arguments of the connective are not adjacent). The approach accounts also for *implicit relation* that are inferred among adjacent spans of text unrelated by any connective. The idea of starting the annotation by looking at specific lexical items and their arguments mainly comes from the works of Webber and other authors (Webber et al., 2003): they propose an approach in which the discourse level structure is more related to the sentence level structure.

ATTRIBUTION.

[122]The definition of connectives in the PDTB will be further clarified in this section and in the dedicated Section 5.2.1.

The *lexically-grounded* approach (Benamara and Taboada, 2015) adopted in the PDTB is also reflected in the coverage of the annotation: the RST aims at the complete annotation of a text (the entire discourse structure), while in the PDTB there is no guarantee that the entire text is annotated (Benamara and Taboada, 2015).

In the PDTB, connectives are considered as predicates of a discourse relation that takes two arguments and are annotated when they are connecting "two *abstract objects* such as events, states, and propositions " (Prasad et al., 2007, citing Asher, 1993), that can be realized mostly as clauses, nominalisations, anaphoric expressions. In 4.9, an example from the PDTB annotation manual (Prasad et al., 2007) is provided[123]:

(4.9) Although $_{[}$Georgia Gulf hasn't been eager to negotiate with Mr. Simmons and NL, a specialty chemicals concern $_{a]}$, $_{[}$the group apparently believes the company's management is interested in some kind of transaction$_{b]}$.

In this example *Although* is the connective that takes two arguments signalled respectively by [a] and by [b]. Notice that the connective is underlined: this is the notation adopted in the PDTB manual (Prasad et al., 2007) that we also adopt in this thesis.

In the PDTB, connectives belong to three syntactic classes (in PDTB 2.0, Prasad et al., 2007):

- subordinating conjunctions (e.g., *when, because, although, as soon as, now that*, etc.), also with a modifier (e.g., *just because, even though, mainly when*);
- coordinating conjunctions (e.g., *and, or, nor, but*).
- discourse adverbials, including both adverbs (e.g., *however, otherwise, instead*), and prepositional phrases (e.g., *on the other hand, as a result*). '

In the PDTB, the two arguments of a connective are called *Arg1* and *Arg2*. Differently from RST and SDRT theories, in the PDTB framework, there is no difference in terms of importance or centrality among the arguments (see Section 4.1.1). Arguments are defined as follows: "*Arg2* is the argument that appears in the clause that is syntactically bound to the connective, and *Arg1* is the other" (Prasad et al., 2007, p. 10). So for example, in 4.9, the span of text signalled with [a] is the Arg2 because it is *syntactically bound* to Although, and the one signalled with [b] is, consequently, Arg1. The example is once

[123] All the examples in English in this section are taken from the PDTB Annotation Manual (Prasad et al., 2007).

again repeated in 4.10 using the notation of the PDTB: Arg1 appears in italics, while Arg2 appears in bold.

(4.10) Although **Georgia Gulf hasn't been eager to negotiate with Mr. Simmons and NL, a specialty chemicals concern**, *the group apparently believes the company's management is interested in some kind of transaction*.

The same example shows that Arg2 precedes Arg1, however, arguments of a connectives can appear in any order and can also be located in different sentences. For example, in 4.11, Arg1 precedes Arg2 and it appears in the preceding sentence.

(4.11) *The Texas oilman has acquired a 26.2% stake valued at more than $1.2 billion in an automotive-lighting company, Koito Manufacturing Co.* But **he has failed to gain any influence at the company**.

Furthermore, Arg1 can also appear within Arg2, which is then annotated as two discontinuous spans, as in the following example:

(4.12) As an indicator of the tight grain supply situation in the U.S., market analysts said **that late Tuesday the Chinese government**, *which often buys U.S. grains in quantity*, **turned** instead **to Britain to buy 500,000 metric tons of wheat**.

There is not a limit for the extension of arguments. In the PDTB, a "minimality principle" is followed, according to which all the text that is "necessary to interpret the relation" should be included in the argument. This means that an argument can also include multiple sentences, as shown in the following example:

(4.13) *Here in this new center for Japanese assembly plants just across the border from San Diego, turnover is dizzying, infrastructure shoddy, bureaucracy intense. Even after-hours drag; "karaoke" bars, where Japanese revelers sing over recorded music, are prohibited by Mexico's powerful musicians union*. Still, **20 Japanese companies, including giants such as Sanyo Industries Corp., Matsushita Electronics Components Corp. and Sony Corp. have set up shop in the state of Northern Baja California**.

Moreover, as introduced before, arguments of an *explicitly conveyed relation* can be non-adjacent. In this case, the parts of text that are not necessary to the interpretation of a relation (i.e., they are not part of the arguments) can be found between the connective

and the arguments. Furthermore, the same text span can be an argument for two different connectives, involving two different relations [124].

The PDTB also accounts for the cases in which the discourse relation is not expressed by connectives, when it is found between adjacent sentence pairs or between complete clauses delimited by semi-colon (;) or colon (:) in the same sentence. In these cases, an *implicit relation* is recognized and a proposal of a connective that best expresses the inferred relation is provided. In the following example by Prasad et al. (2007) an implicit relation of *consequence* is recognized and the connective *so* is proposed:

(4.14) *The projects already under construction will increase Las Vegas's supply of hotel rooms by 11,795, or nearly 20%, to 75,500.* Implicit = so **By a rule of thumb of 1.5 new jobs for each new hotel room, Clark County will have nearly 18,000 new jobs**.

The PDTB also accounts for cases in which a relation is *alternatively lexicalized* by an expression other than a connective belonging to the previously mentioned categories: these cases are tagged as *AltLex* in the corpus. By definition, these are the cases where a discourse relation is inferred between adjacent sentences but where providing a suggestion of connective leads to redundancy in the expression of the relation (Prasad et al., 2007)[125]. For example, in 4.15, a temporal relation between the two arguments is inferred and conveyed through "After that" that is tagged as *Altlex*, since it does not belong to the list of connectives according to the PDTB 2.0 categories.

(4.15) And she further stunned her listeners by revealing her secret garden design method: *Commissioning a friend to spend "five or six thousand dollars . . . on books that I ultimately cut up."* AltLex [**After that**], **the layout had been easy**.

The PDTB schema proposes different *senses* for the connectives: they provide a semantic description of the relation between their arguments. The senses are annotated as *sense tags* and are organized in a hierarchical classification composed by three levels. Table 4.2 reports the hierarchy of *sense* in the PDTB 3.0 version (Webber et al., 2016).

In the first level of the hierarchy, i.e., the *class level*, sense tags are grouped in four major classes (shown in the first column of Table 4.2):

[124] This is similar to what is proposed in SDRT framework, as explained in the paragraph 4.1.1 and illustrated in Example 4.4.

[125] For further details on *AltLex* relations see (Prasad et al., 2010).

I level	II level	III level
CLASS	TYPES	SUBTYPES
TEMPORAL	Synchronous	–
	Asynchronous	Precedence
		Succession
CONTINGENCY	Cause	Reason
		Result
	Condition	Arg1-as-cond
		Arg2-as-cond
	Negative Condition	Arg1-as-negcond
		Arg2-as-negcond
	Purpose	Arg1-as-goal
		Arg2-as-goal
COMPARISON	Contrast	–
	Similarity	–
	Concession	Arg1-as-denier
		Arg2-as-denier
EXPANSION	Conjunction	–
	Disjunction	–
	Equivalence	–
	Instanciation	–
	Level-of-detail	Arg1-as-detail
		Arg2-as-detail
	Substitution	Arg1-as-subst
		Arg2-as-subst
	Exception	Arg1-as-except
		Arg2-as-except
	Manner	Arg1-as-manner
		Arg2-as-manner

Table 4.2: The PDTB 3.0 hierarchy of senses (Webber et al., 2016).

- TEMPORAL: "used when the connective indicates that the situations described in the arguments are related temporally" (Prasad et al., 2007);
- CONTINGENCY: "used when the connective indicates that one of the situations described in Arg1 and Arg2 causally influences the other." (Prasad et al., 2007). For example, the cause-effect relation and the condition relation are included in this class;
- COMPARISON: "used when the connective indicates that a discourse relation is established between Arg1 and Arg2 in order to highlight prominent differences between the two situations."(Prasad et al., 2007):
- EXPANSION: "used for those relations which expand the discourse and move its narrative or exposition forward" (Prasad et al., 2007).

The second column of Table 4.2 reports the second level of the hierarchy that specifies further information about the semantics of the class level: the *type level*. For example, in the PDTB the tag TEMPORAL.Synchronous (first type in the table) indicates the type *Synchronous* of the class TEMPORAL and is used to indicate that the events in the two arguments are simultaneous (e.g., "When she arrived, he was leaving"); differently, the TEMPORAL.*Asynchronous* tag (just below in the table) is used to indicate a before-after relation (e.g., "She arrived before he left").

The third level (*subtype level*) varies according to the role of the two arguments involved in the relation. In the PDTB 3.0 version of the schema (Webber et al., 2016), shown in Table 4.2, the third level reflects the *direction of the relations* which is associated with its asymmetry: if ARG1 and ARG2 have the same role in the relation, a sense relation is symmetric, if not it is asymmetric. For example, the type CONTINGENCY.*Cause* represents an asymmetric relation between two arguments, being one the cause, and the other the result. The tag CONTINGENCY.*Cause*.Reason (i.e subtype Reason of type *Cause* of Class CONTINGENCY) is used if the argument introduced by the connective (Arg2) is the reason for the situation in the other argument (Arg1) (e.g., "I stayed at home because it was raining"), while CONTINGENCY.*Cause*.Result is used if it represents the result/effect (e.g., "It was raining, therefore I stayed at home"). Notice that not every *type* has a further *subtype*: for example, the arguments involved in a temporal relation of type *Synchronous* do not play different roles and no subtype has been proposed.

According to the PDTB, a connective can have more than one sense, depending on the context in which it is used. For instance, in Example 4.16 by Webber et al. (2006), *since* has both a temporal and a causal sense:

(4.16) *...and domestic car sales have plunged 19%* since **the Big Three ended many of their programs Sept. 30**.

This example can be annotated with the TEMPORAL.*Asynchronous*.Succession tag (the situation described in Arg1 follows the situation described in Arg2) and also with the CONTINGENCY.*Cause*.Reason tag (the situation described in Arg1 is the effect of the situation described in Arg2 and thus Arg2 is the reason).

In the PDTB the relation of *attribution*, i.e., "the relation of 'ownership' between abstract objects and individuals or agents" (Prasad et al., 2007, p. 40) is annotated as well. The main goal is to ascribe beliefs and assertions to the agent(s) holding them (Prasad

et al., 2007). An as example, in 4.17 the span of text corresponding to the attribution phrase is shown within a box (as proposed in the PDTB Annotation Manual).

(4.17) *Advocates said* *the 90-cent-an-hour rise, to $4.25 an hour by April 1991, is too small for the working poor*, while **opponents argued** **that the increase will still hurt small business and cost many thousands of jobs**.

More detailed information on the annotation in the PDTB can be found in Prasad et al. (2007) and will be provided in Section 5.1.2, in which we describe how we adopt the PDTB schema for the annotation of the contrast relation in the Contrast-Ita Bank corpus.

4.1.2.1 Other Resources Using the Penn Discourse Treebank Annotation Schema

A number of annotated corpora similar to the PDTB have been proposed since its creation. Prasad et al. (2014) propose an overview of many related projects which include non-English corpora, e.g., resources for Czech (the Prague Discourse TreeBank, most recent version: 3.0, by Bejček et al. 2013)[126], for Chinese (the Chinese Discourse TreeBank, by Zhou and Xue 2015)[127], and for Arabic (the Leeds Arabic Discourse TreeBank, or LADTB by Al-Saif and Markert 2010).[128]

These resources can differ from the PDTB in terms of the annotated information (for instance, implicit relations are not always annotated and the annotation of multiple sense tags for a single connective, as shown in Example 4.16, is not allowed in every resource), or in terms of workflow of annotation (for instance, in the Chinese Discourse TreeBank, annotators must consider implicit relations not only between adjacent sentences or between clauses delimited by semi-colon or colon, but also between comma-delimited clauses in the same sentence), or even in terms of sense tags set (for instance, additional senses have been introduced, while other have been eliminated or modified) (Prasad et al., 2014). However, according to Prasad et al. (2014), all these resources adhere to the key ideas of the PDTB annotation in being neutral to any discourse structure, in grounding discourse relations in lexical expressions, and in considering discourse relations to hold between two and only two arguments each of which is coupled with a text span that can potentially be discontinuous (as shown in Example 4.12).

[126] `https://ufal.mff.cuni.cz/pdt3.0`

[127] `https://catalog.ldc.upenn.edu/LDC2014T21`

[128] Authors also mention resources for French, Turkish and Hindi.

For Italian, Tonelli and colleagues (Tonelli et al., 2010) are the first to use the PDTB annotation scheme for the annotation of a corpus: the LUNA conversational spoken dialog corpus, which consists of 60 real dialogs in the domain of software/hardware troubleshooting. The annotation follows the PDTB annotation schema, but with some adaptations (e.g., in the sense tags hierarchy) mainly because, as the authors said, the annotation is meant for spontaneous conversations in a different domain than the texts in the PDTB, which consist of Wall Street Journal articles (news domain).

Another project for Italian inspired by the PDTB is proposed by Pareti and Prodanof (2010). The authors specifically focus on the relation of *attribution* (which we explained at the end of the previous section) annotating the relation in 50 articles retrieved from the ISST corpus (Montemagni et al., 2003), which consists of 484 articles drawn from Italian newspapers and periodicals. For the annotation, some adaptations with respect to the PDTB schema have been proposed; for instance, the *attribution* relation is annotated independently from its connection to a discourse relation (Pareti and Prodanof, 2010).

In Chapter 5, we will present a corpus of Italian articles annotated with the contrast relation; for the annotation we adopt and adapt the PDTB schema.

4.2 Discourse Contrast

In the previous sections, we presented the two most influential theories on discourse relations (i.e., RST and SDRT), and the PDTB, the largest corpus annotated with discourse relations (for English) used in different NLP tasks, such as for retrieving information (information extraction) or for the automatic creation of summaries of a written text (text summarization). In the following sections, we focus on the *discourse contrast relation*: our goal is to study the relation of *contrast* by observing the phenomena in a corpus. Specifically, in this section we first compare different definitions of *contrast* in the literature, and then discuss the definition of *contrast* we are adopting in order to define a schema for the annotation of this relation in texts. The annotation process and its results will be presented in the next chapter.

4.2.1 Discourse Contrast in the Literature

As shown in the previous sections, different sets of discourse relations have been proposed in different theoretical frameworks and annotation schema. The presence of a relation

involving contrast seems to be, however, not controversial: in fact, both RST and SDRT list of relations and the PDTB hierarchy of senses include the relation of *contrast*.

4.2.1.1 Contrast in RST

In the RST list of relations from Mann and Thompson (1988), CONTRAST is defined as a relation involving two spans of texts equally central in the discourse structure (i.e., a *multinucleus* relation, see Section 4.1.1). Moreover, according to the same definition, the situations presented in the two spans i) have some aspects in common, ii) are different under some respects, and iii) they are compared with respect to one or more of these differences; in other words, they are compared with respect to their contrasting properties. An example by the authors is proposed in 4.18.

(4.18) Animals heal$_{[a]}$, but tree compartmentalise$_{[b]}$.[129]
CONTRAST between [a] and [b], in Mann and Thompson (1988, p. 278)

In the example, the connective *but* links two arguments that we have signalled as [a] and [b]. The two situations expressed in the two arguments are both descriptions of "strategies for wounds recovering" (i.e., arguments have something in common); however, the two presented strategies are different and are contrasted in the text, i.e., on the one hand, animals treat the wound, on the other hand, plants isolate the wound.

In the same list of relations (Mann and Thompson, 1988), the authors identify a subtype relation of CONTRAST: ANTITHESIS. The ANTITHESIS relation inherits the characteristics of the CONTRAST relation but it is considered a *nucleus-satellite* relation: i.e., a relation in which one argument is subsidiary to the other (see Section 4.1.1). It adds to the definition of CONTRAST the concept of *positive regards*. The notion of positive regard is explained as follows: "Writers pursue different sorts of goals with different texts and text spans. Some are intended to persuade, i.e., to create belief. Others are intended to create an attitude of approval or interest. Others are intended to create desire (specifically, an intention to act.)" (Thompson and Mann, 1987, p. 91). According to Mann and Thompson (1988) in ANTITHESIS there is a desired effect to cause a positive regard for the *nucleus*. An example of ANTITHESIS is proposed in 4.19.

[129] Tree compartmentalization is a strategy that trees adopt in case of wounds: trees prevent the spread of a disease by forming "walls" around the wounded area.

(4.19) I recently purchased a text which purported to be a guide to Pascal for engineers.[a]
It totally ignored the subtleties of the language and made no bones about it.[b]
ANTITHESIS between [a] and [b], (Thompson and Mann, 1987, p. 91)

In this case, the two arguments, signalled as [a] and [b], both refer to a textbook that has been bought; however, one argument conveys the idea that the volume is (thought to be) a guide to Pascal for engineers, the other conveys the idea that it is not. These two situations are contrasted in the text. According to Thompson and Mann (1987), the writer has a positive regard for [b] and intends the reader to have positive regard for it as well, by recognizing the incompatibility between the two arguments.

Compare Example 4.19 to the following one:

(4.20) According to Paul the text is a guide to Pascal for engineers;[a] according to Mary the text is a cookbook.[b]

In Example 4.20, the descriptions of the volume are still incompatible (i.e., in one case is a guide, in the other is a cook book), but the desired effect (causing the reader to have a positive regard for one of the two arguments of the relation) is not conveyed, as it happens in Example 4.18.

In this thesis we also discuss the relation of CONCESSION which is considered a type of *contrast* according to a number of proposals that will be presented. In line with this position, we believe that CONCESSION captures a contrast between the arguments it involves: this will be discussed in Section 4.2.2.

In RST, a CONCESSION relation is a relation between a *nucleus* and a *satellite*, in which the writer "acknowledges a potential or apparent incompatibility between the nucleus and the satellites, but regards them as compatible" (Mann and Thompson, 1988), and desires the reader to have positive regards toward the *nucleus* (as in the ANTITHESIS relation). We can discuss this definition by analysing an example proposed by the authors reported in 4.21.

(4.21) Concerns that this material is harmful to health or the environment may be misplaced.[a]
Although it is toxic to certain animals,[b]
evidence is lacking that it has any serious long-term effect on humans beings.[c]
CONCESSION between [b] and [c], (Mann and Thompson, 1988, p. 255)

Mann and Thompson (1988) recognized a CONCESSION relation between [b] and [c]: they argue that in this example the writer acknowledges that what is expressed in [b] is potentially incompatible with what is expressed in [c], since toxicity in animals often implies toxicity in humans. However, the two situations are regarded as compatible: the fact that something is toxic to animals is claimed to be compatible with the lack of evidence that it is also toxic to humans. The desired effect is for the reader to have positive regards for the *nucleus* (i.e., situation in [c]) by recognizing this compatibility between the two arguments.

Thus, both ANTITHESIS and CONCESSION involve the concept of positive regards: the difference lies in how it is conveyed. ANTITHESIS relies on a semantic contrast that causes incompatibility between the situations expressed in the arguments (e.g., in Example 4.19, the incompatibility is between the book being a guide, and being not a guide), while for CONCESSION the situations are presented as compatible, even if the writer recognises a certain incompatibility (e.g., in Example 4.21, toxicity to animals is compatible with no evidence of toxicity to humans, although the potential incompatibility between them is acknowledged).

As we mentioned in Section 4.1.1, Carlson and Marcu (2001) use the set of relations defined by Mann and Thompson (1988) to annotate a corpus with discourse relations. In their guidelines, the difference between CONTRAST, ANTITHESIS, and CONCESSION is further specified. CONTRAST is defined as a relation in which "two or more nuclei come in contrast with each other along some dimensions. The contrast may happen in only one or few respects, while everything else can remain the same in other respects." (Carlson and Marcu, 2001, 53). They also specify that the ANTITHESIS relation differs from CONCESSION because in CONCESSION "the situation indicated in the nucleus is contrary to expectation in the light of the information presented in the satellite. In other words, a CONCESSION relation is always characterized by a violated expectation"(Carlson and Marcu, 2001). In the Annotation Manual, they propose the following example for CONCESSION in which [a] is the nucleus and [b] is the satellite.

(4.22) Still, today's highest-yielding money funds may beat CDs over the next year$_{[a]}$ even if rates fall$_{[b]}$.
CONCESSION between [a] and [b],
(Carlson and Marcu, 2001, p. 50)

We believe that the expectation that raises from the satellite [b] is that the falling of rates will have negative consequences, and this expectation is *violated* by the nucleus [a] in which a positive situation is described.

This definition of CONCESSION is in line with the definition provided in other proposals, including the one provided in the PDTB project, which will be presented in 4.2.1.3. Interestingly, Carlson and Marcu (2001) group CONTRAST, ANTITHESIS and CONCESSION in the same class of relation called *Contrast* (Carlson and Marcu, 2001).[130]

Table 4.3 summarises the characteristics of the three relations described according to Mann and Thompson (1988); Thompson and Mann (1987), and Carlson and Marcu (2001).

Relation	Mann and Thompson (1988) Thompson and Mann (1987)	Carlson and Marcu (2001)
CONTRAST	• multi-nuclear relation • situations are: i) the same in many respects ii) differing in a few respects iii) compared with respect to differences	• multi-nuclear relation • two or more nuclei come in contrast with each other along some dimension. The contrast may happen in only one or few respects, while everything else can remain the same in other respects.
CONTRAST ANTITHESIS	• nucleus-satellite relation • is a subtype of CONTRAST • situations are: i) the same in many respects ii) differing in a few respects iii) compared with respect to differences • positive regard for nucleus due to: nucleus-satellites incompatibility deriving from CONTRAST	• nucleus-satellite relation • the situation in the nucleus comes in contrast with the situation in the satellite. The contrast may happen in only one or few respects, while everything else can remain the same in other respects.
CONCESSION	• nucleus-satellite relation • positive regard for nucleus due to: nucleus-satellites potentially incompatible but regarded as compatible	• nucleus-satellite relation • it is characterized by a violated expectation: the situation in the nucleus is contrary to expectation in light of the information in the satellite.

Table 4.3: CONTRAST, ANTITHESIS and CONCESSION in RST.

4.2.1.2 Contrast in SDRT

In the SDRT framework, Asher and Lascarides (Asher, 1993; Asher and Lascarides, 2003) define *contrast* as a coordinating relation that involves "structurally similar but semanti-

[130]Relations in classes share some types of "rhetorical meaning"(Carlson and Marcu, 2001, 32).

cally dissimilar objects" (Asher and Lascarides, 2003) [131]. According to them, CONTRAST involves a "partially isomorphic mapping" (Asher and Lascarides, 2003, p. 168) between the constituents structure (i.e., they are structurally similar) but also a "contrasting theme" (Asher and Lascarides, 2003, p. 465) [132] (i.e., they are semantically dissimilar).

The following example for CONTRAST is proposed:

(4.23) A girl that I met dislikes Sam. But I like him.
(Asher, 1993, p.285)

In Example 4.23, there is parallelism between two entities (*girl, me*) liking/disliking the same entity (*Sam*) and "the themes of liking Sam and disliking Sam are plausibly complementary"(Asher, 1993, p.285), in other words, they are contrasting. This is not the case in the following example by Asher (1993), where there is not a connection between "envies Sam" and "find him attractive".

(4.24) A girl that I met envies Sam. But I find him attractive.
(Asher, 1993, p.286)

The author also points out that contrast is a scalar relation: that is, some contrast relations are stronger than others. Lascarides and Asher (2008) define this "strength" as "dependent on the extent to which the semantics of the connected propositions are dissimilar [..] *John loves to collect classic cars. But his favourite car is a 1999 Ford Mondeo* is a 'better' contrast than *John loves to collect classic cars. hates football*)"(Lascarides and Asher, 2008, p.13) . The strongest case is when one constituent negates a default consequence of the other (Asher and Lascarides, 2003, p.465).

In SDRT, CONTRAST is of two types: *formal contrast* and *violation of expectation*. An example of formal contrast is the one in 4.23; an example of *violation of expectation*, in which there is denial of expectation, is the following:

(4.25) John loves sports. But he hates football.
(Asher and Lascarides, 2003, p.168)

[131] The relation is described together with Parallelism, which involves two semantically and structurally similar objects (Asher, 1993, p.285).

[132] Theme is defined as sentence topic in (Reese et al., 2007).

Since football is a kind of sports, "John loves sports" *entails* "John loves football" (Asher and Lascarides, 2003, p.168), but this is denied in the second argument.

According to the authors, while *formal contrast* can also be inferred without an explicit connective, explicit cues are, instead, necessary when *violation of expectation* is conveyed (Asher and Lascarides, 2003, p.167). In their manual for the annotation of discourse relation in the framework of SDRT, Reese et al. (2007) suggest that strong cues for the contrast relation are *but, however, on the other hand, nevertheless*.

Table 4.4 summarizes the characteristics of CONTRAST in SDRT.

Relation	Characteristics (Asher, 1993; Asher and Lascarides, 2003; Lascarides and Asher, 2008)
CONTRAST	• constituents are structurally similar but have *contrasting themes* • is a scalar relation. Strongest case: one constituent negates a default consequence of the other • Types: *- formal contrast: does not require explicit cues to hold* *- violation of expectation: requires explicit cues to hold* • Most frequent connectives: *but, however, on the other hand, nevertheless*

Table 4.4: CONTRAST in SDRT.

4.2.1.3 Contrast in the Penn Discourse Treebank

In the PDTB2.0 (Prasad et al., 2008; Prasad et al., 2007) hierarchy of relations, *Contrast* is a type (II level of the hierarchy) of the class COMPARISON (I level of the hierarchy). Table 4.5 shows the class COMPARISON in the PDTB 2.0 that includes both *Contrast* and *Concession*.

Contrast applies to cases in which the two arguments of the relation "share a predicate or a property and the difference between the two situations described in the arguments is highlighted with respect to the values assigned to this property" (Prasad et al., 2007, p. 32), as in the following example:

(4.26) *John paid $5* <u>but</u> **Mary paid $10.**

Contrast in PDTB (Prasad et al., 2007, p. 32)

In the example, the two arguments[133] share the predicate *paid* that takes two different values (*$5*, *$10*) and also applies to two entities (*John*, *Mary*).

[133] We remind the reader that we are following the convention proposed in the PDTB 2.0 (Prasad et al., 2007) according to which the connective is underlined, Arg1 is in italics, Arg2 is in bold.

PDTB 2.0 (Prasad et al., 2007)		
Class	Type	Subtypes
COMPARISON		
→	*Contrast*	
		→ Juxtaposition
		→ Opposition
→	*Pragmatic Contrast*	
→	*Concession*	
		→ Expectation
		→ Contra−expectation
→	*Pragmatic Concession*	

Table 4.5: *Contrast* and *Concession* in PDTB 2.0 (Prasad et al., 2007).

Authors specify that *Contrast* is a relation in which "neither argument describes a situation that is asserted on the basis of the other one." In this sense, the relation is symmetric and there is no directionality among its arguments (unlike, e.g., the *Cause* sense that links a cause and an effect - see Section 4.1.2). In the example, the two arguments "John paid $5" and "Mary paid $10" do not have different roles in the relation.

In the PDTB 2.0, two subtypes of *Contrast* are defined[134]: *juxtaposition* and *opposition* depending if the values assigned to some shared property are taken to be alternatives (i.e., *juxtaposition*, Example 4.26) or are the "extremes of a gradable scale, that is they are antonyms"(Prasad et al., 2007) (i.e., *opposition*, Example 4.27).

(4.27) *Most bond prices fell* [..]. **Junk bond prices moved higher**, however.
Contrast.opposition in PDTB2.0 (Prasad et al., 2007, p. 32)

Authors do not provide explicit definition for what should be intended as alternatives or extremes of a gradable scale[135], but do advise that this difference "strongly depends on the context where the sentence is uttered" (Prasad et al., 2007, p.33).

The PDTB 2.0 aims at registering also cases in which "the connective indicates a contrast between one of the arguments and an inference that can be drawn from the other, in many cases at the speech act level."(Prasad et al., 2007, p.33). These cases are tagged as

[134]Notice that these two subtypes are not reported in Table 4.2 as the table makes reference to the more recent PDTB 3.0 schema. Further in this section we will see that in this most recent version the distinction has been eliminated.

[135]This distinction remind us the distinction about *complementary opposites* and *antonyms opposites*, however we believe that the first group of alternative values (that lead to *juxtaposition*) can included also incompatible terms, not necessarily opposite (see Section 2.1.1).

Pragmatic Contrast. Authors propose the following example explaining that the contrast is between Arg1, in italics, (i.e., quantity is explained) and the inference from Arg2, in bold, (i.e., quantity is not the only thing that needs to be explained: quality needs to be explained as well)(Prasad et al., 2007).

(4.28) "It's just sort of a one-upsmanship thing with some people," added Larry Shapiro. "They like to talk about having the new Red Rock Terrace one of Diamond Creek's Cabernets or the Dunn 1985 Cabernet, or the Petrus. Producers have seen this market opening up and they're now creating wines that appeal to these people." *That explains why the number of these wines is expanding so rapidly*. But **consumers who buy at this level are also more knowledgeable than they were a few years ago**.
Pragmatic Contrast in PDTB (Prasad et al., 2007, p. 32)

The annotation manual of the PDTB 2.0 accounts also for a *Concession* relation, that is -as *Contrast*- a type (II level of the hierarchy) of the class COMPARISON (I level of the hierarchy), as shown in Table 4.5.[136] *Concession* applies to cases in which one of the arguments describes a situation that "triggers a set of potential consequences, while the other denies one or more of them"(Prasad et al., 2007). For instance, in Example 4.29, Arg2 (in bold) triggers the expectation that the economical situation would lead to a recession, but Arg1 (in italics) denies it.

(4.29) Although **the purchasing managers' index continues to indicate a slowing economy**, *it isn't signaling an imminent recession* [..].
Concession.expectation in PDTB (Prasad et al., 2007, p. 34)

Differently from *Contrast*, *Concession* is an asymmetric relation in which the two arguments play a different role: one arguments triggers a consequence, the other denies it. Two subtypes of *Concession* are defined depending on which argument (Arg1 or Arg2) creates the expectation / triggers the consequence, and which denies it. Specifically, *expectation* is when Arg2 (the argument bound to the connectives) creates an expectation that Arg1 denies, as in Example 4.29, and *contra-expectation* is when Arg1 creates an expectation that Arg2 denies, as in Example 4.30 in which Arg1 triggers the consequence

[136] We already clarified before in this section that we are discussing *Concession* as we believe that it represents a kind of contrast.

that having made an important purchase, the oilman would have gained influence, but Arg2 denies it.

(4.30) *The Texas oilman has acquired a 26.2% stake valued at more than $1.2 billion in an automotive-lighting company, Koito Manufacturing Co.* But **he has failed to gain any influence at the company**.
Concession.contra-expectation in PDTB (Prasad et al., 2007, p. 34)

As for *Contrast*, the PDTB 2.0 accounts for the *Pragmatic Concession*, when the relation is between one of the two arguments and an inferred situation drawn from the other argument.

The set and the hierarchy of relations used in the PDTB project have varied from version 2.0 (Prasad et al., 2007) to 3.0 (Webber et al., 2016). In fact, in the most recent PDTB 3.0 schema of relations presented in Table 4.2, *Contrast* and *Concession* are proposed with some variations with respect to the PDTB 2.0 schema[137]. Table 4.6 reports the section of the hierarchy that refers to *Contrast* and *Concession* in PDTB 2.0 (Prasad et al., 2007) and in PDTB 3.0 (Webber et al., 2016).

PDTB 2.0 (Prasad et al., 2007)			PDTB 3.0 (Webber et al., 2016)		
Class	Type	Subtypes	Class	Type	Subtypes
COMPARISON			COMPARISON		
	→ *Contrast*			→ *Contrast*	
		→ Juxtaposition			
		→ Opposition			
	→ *Pragmatic Contrast*				
	→ *Concession*			→ *Concession*	
		→ Expectation			→ Arg1.as.denier
		→ Contra-expectation			→ Arg2.as.denier
	→ *Pragmatic Concession*				

Table 4.6: *Contrast* and *Concession* in PDTB 2.0 (Prasad et al., 2007) and in PDTB 3.0 (Webber et al., 2016).

In particular, as can be seen from Table 5.1, the third level of the hierarchy for *Contrast* has been eliminated (i.e., there is no *Contrast* subtypes and thus no difference between *juxtaposition* and *opposition*). The second level of pragmatic relations has also been eliminated: instead, features signalling that the relation is between an argument and what it

[137]These variations are mainly meant to simplify the annotation (Webber et al., 2016).

is inferred by the other argument have been adopted. The two subtypes for *Concession* are simply renamed *Concession*.Arg1.as.denier (*expectation* in PDTB 2.0) and *Concession*.Arg2.as.denier (*contra-expectation* in PDTB 2.0) to better reflect the directionality of the relation.

4.2.1.4 Other definitions of Contrast

Other theoretical frameworks have proposed definitions for the contrast relation. For example, Hobbs (1985) (see Section 4.1.1) defines *contrast* as the relation in which "contrasting predications are made about similar entities [..or..] the same predication is made about contrasting entities" (Hobbs, 1985). He proposes the following example to explain how contrasting predication is made.

(4.31) You are not likely to hit the bull's eye, but you are likely to hit the bull's eyes than any other equal area.
CONTRAST in Hobbs (1985, p. 21)

According to the author, in the example there is a contrast relation between the first clause ("the probability of hitting the bull's eye is less than whatever probability counts as alike") and the second clause ("the probability is greater -and thus not less than- the typical probability of hitting any other equal area")(Hobbs, 1985). He also defines the *violated expectation* relation, in which "what is inferred from an argument (assertion) is *contradicted* by the other." (Hobbs, 1985, p. 22). In Example 4.32, reading the first clause one would expect that John is a dishonest (since he is a layer), but the second clause violates this expectation (Hobbs, 1985).

(4.32) John is a layer, but he's honest.
VIOLATED EXPECTATION in Hobbs (1985, p. 22)

Also Lakoff (1971), in her analysis of *but* (a very frequent connective of contrast), distinguishes between two uses of the connective both involving the concept of combining *similarity* (i.e., elements conjoined by *but* have something in common), and *difference*. The first one is the *semantic opposition* use, in which the elements are "directly *opposed* to each other in a particular property" (Lakoff, 1971, p. 133). Example 4.33 by the author shows a case of *semantic opposition*: the arguments are contrasted (i.e., *opposed* in Lakoff's terms) for a particular property (that is, being tall and being short).

(4.33) John is tall but Bill is short.

Semantic opposition in Lakoff (1971, p. 133)

The second one is the *denial of expectation* use, in which the expectation (i.e., *presupposition* in Lakoff's terms) rises from an element that is denied by the other one (Lakoff, 1971). In Example 4.34 the clause "he's no good at basketball" denies the expectation rising from the first clause "John is tall".

(4.34) John is tall but he's no good at basketball.

Denial of expectation in Lakoff (1971, p. 133)

4.2.1.5 Effort to combine definitions

There have been efforts to combine the broad panorama of different classifications, and this frequently involves trying to merge different definitions of contrast. For example, Hovy and Maier (1997) propose a hierarchy of approximately 70 semantic relations that fuses and taxonomizes more than 400 relations from different sets of relations. In their hierarchy, CONTRAST and CONCESSION are in two different groups: CONTRAST is in the Ideational group (the group of relations that express some experience of the world), while CONCESSION is the Intentional group (the group of relations by which the author attempts to affect addressee's beliefs, attitudes, desires, etc.). It is also worth mentioning that in the 30 classifications they collect, 16 of them account for a CONTRAST relation. Among others, they provide a clear example of CONCESSION defined as follows: "One of the text segments raises expectations which are contradicted / violated by the other" (Hovy and Maier, 1997).

Also Benamara and Taboada (2015) try to merge different classifications. They propose a unified hierarchy, both intentionally and semantically driven, in order to map different annotations made in the frameworks of RST and SDRT theories. As a result, they create a class (called *ARGUMENTATIVE - Opposition*) which groups relations where "the segments have similar semantic structures, but contrasting themes, i.e., sentence topics, or when one constituent negates a default consequence of the other", thus grouping together the concepts of *Contrast* and *Concession*. [138]

[138] In this panorama, it is also worth mentioning the work by Scheffler and Stede (2016a) that focus on mapping the annotation of connectives in a corpus for German made by following the PDTB3.0 schema, with their corresponding labels in RST framework; even if they do not directly map the list of relations, according to the authors the method allows for a data-driven mapping of the relations from one taxonomy to another.

To summarise what we have seen so far, it can be said that with reference to the contrast relation a major distinction is largely shared in the literature: on the one hand, the relation that has been called *contrast* (Mann and Thompson, 1988; Prasad et al., 2007), *formal contrast* (Asher, 1993), *semantic opposition* (Lakoff, 1971) and it is characterised by *a contrast along dimensions* (Carlson and Marcu, 2001); on the other hand, the relation that has been called *concession* (Carlson and Marcu, 2001; Prasad et al., 2007) or *violation of expectation* (Carlson and Marcu, 2001; Asher and Lascarides, 2003; Lakoff, 1971).

We have seen that in some cases these two are conceived as two different relations (Mann and Thompson, 1988; Carlson and Marcu, 2001; Prasad et al., 2007; Webber et al., 2006), although belonging to the same higher level semantic class (Carlson and Marcu, 2001; Prasad et al., 2007; Webber et al., 2006); in other cases they are considered as two types of a unique relation (Asher, 1993; Asher and Lascarides, 2003). In the next section, we will discuss the notion of *contrast* we are adopting in this thesis.

4.2.2 The Notion of Contrast in this Thesis

As introduced at the beginning of this chapter, one of the contributions of our research is the creation of a corpus for Italian annotated with contrast (see Section 5.1). The overview proposed in this chapter was meant to clarify the concept of *discourse contrast relation* in order to define the schema we will adopt in the annotation process.

We decided to take as our main work of reference the PDTB project (Prasad et al., 2007; Webber et al., 2016) that produced the largest and most used annotated corpora with discourse relations (see Section 4.1.2), which has been proven to be compatible with other frameworks. In an effort to maximise the compatibility and similarity with the PDTB project, in the annotation process we adopted the PDTB approach and schema of annotation. This has some major implications. First, differently from RST, we do not consider whether an argument in a relation is a *nucleus* or a *satellite* (i.e., if one is more central or important) and we are not defining the contrast relation based on these characteristics [139]. Thus, there is no distinction of the kind *multiple nuclei* / *nucleus-satellite* relation. We also exclude the distinction between the *coordinating* / *subordinating* relation proposed in SDRT. Following the PDTB, we identify the two arguments as *Arg1* and *Arg2*, and there is no difference between the two in terms of importance or centrality with

[139] This is similar to the approach adopted by Benamara and Taboada (2015).

respect to a discourse structure. Second, we follow the definitions provided for the *senses* (i.e., a semantic description of the relation between the arguments of the connectives) in the PDTB 2.0 manual. Thus, we do not consider, for instance, the concept of *positive regards* that characterizes ANTITHESIS and CONCESSION according to the RST proposal (see Section 4.2.1). Additional details on the practical aspects that the choice of adopting the PDTB schema implies will be provided in Section 5.1.2.

Consistent with this approach, and in order to annotate the *contrast* relation, the most straightforward choice is to annotate what is called COMPARISON.*Contrast* in the PDTB schema. As we have seen in Section 4.2, in the PDTB 2.0 Annotation Manual (Prasad et al., 2007), *Contrast* indicates that Arg1 and Arg2 "share a predicate or a property and the difference between the two situations described in the arguments is highlighted with respect to the values assigned to this property" (Prasad et al., 2007). Even though we do not adopt the RST or the SDRT approach, it is important to highlight that the definition of *Contrast* in the PDTB has some overlap with the ones proposed in RST and in SDRT frameworks (Section 4.2.1), although they stem from different approaches. In fact, in these frameworks the relation of CONTRAST captures cases in which the arguments in the relation *have some aspects in common* (Mann and Thompson, 1988; Carlson and Marcu, 2001), *share a property* (Prasad et al., 2007), or *have a similar structure* (Asher, 1993), but they *differ in some respect* (i.e., *some values of the property* (Prasad et al., 2007), or *contrasting themes* (Asher, 1993)) and *are compared with respect to these differences*.[140] In the PDTB it is also specified that neither argument involved in a CONTRAST relation describes a situation that is asserted on the basis of the other one, so the two arguments are independent and neither of the two is necessary for the interpretation of the other. [141]

The notion of contrast in this thesis is, however, not limited to what is intended as COMPARISON.*Contrast* in the PDTB: we consider also COMPARISON.*Concession*, which applies to cases in which one of the arguments describes a situation that "triggers a set of potential consequences, while the other denies one or more of them"(Prasad et al., 2007). The definition agrees with the definition of CONCESSION proposed by

[140]Notice that the property of the arguments of a contrast relation of "having some aspects in common but being different in some respect" is also a property of opposites at the lexical level, as presented in Section 2.1.1. We will further discuss this analogy in Chapter 6.

[141]This seems to overlap with the concept of "equally central" importance of the arguments recognized for the contrast relation in RST, i.e., contrast is a *multi-nuclei relation*.

Carlson and Marcu (2001) and with the type *violation of expectation* proposed by Asher and Lascarides (2003) (Section 4.2). All these definitions indicate that *an expectation* (or *potential consequences*) *rises from an argument and the other argument denies it* (or *contributes to violate it*). As a consequence, we believe that there is a *contrast* between the expectation raised and the other argument. For example, in 4.25 ("John loves sports. But he hates football.") there is a (non explicit) contrast between the expectation "John loves football" (as he loves sports) and "he hates football".

Given this analysis, we will consider both the senses *Contrast* and *Concession* of the PDTB in the annotation of Contrast-Ita Bank (see Chapter 5) as types of a broader concept of *contrast relation*.

Chapter 5

A Data-driven Analysis of the Contrast Relation

This chapter is dedicated to the presentation of a data-driven analysis of the contrast relation in Italian in terms of frequency and explicitness/implicitness of the contrast relation as a result of the annotation of the relation in a corpus of news (Contrast-Ita Bank). We also present an investigation on the connectives of contrast in Italian, starting from a lexicon we have created (LICO, Lexicon for Italian COnnectives).

In this chapter we present our investigation on *contrast* focusing on three areas. The first area of investigation (Section 5.1) concerns the analysis of how the relation is realized in Italian. We gather data on how frequently contrast is implicitly conveyed and provide some insights over the distribution of the two different kinds of contrast in a corpus: CONTRAST and CONCESSION (which, following the PDTB, will be called here *senses of relations*). We believe this could also be an interesting result for applications in the field of text understanding: in fact, depending on whether the relation under study is more frequently explicit or whether it is mostly implicitly conveyed, different techniques for the automatic identification of the contrast relation can be adopted. To address this analysis, and since we noted a lack of corpora annotated with discourse relations in Italian, we created Contrast-Ita Bank, a corpus of news in Italian annotated with the contrast relation. This will allow us to collect quantitative data from corpora and carry out our investigation.

The second area of investigation (presented in Section 5.2) regards the connectives of contrast, which are those lexical elements that convey the relation of contrast in a discourse. In particular, we collect an exhaustive list of these connectives and gather quan-

titative data about their frequency, their polysemy and the kind of contrast they convey (CONTRAST or CONCESSION). This information can benefit the automatic identification of contrast in texts. For this analysis, we take advantage of two resources: LICO, the Lexicon of Italian Connectives (Section 5.2.3) and the above mentioned Contrast-Ita Bank. LICO is a lexicon in which connectives for different discourse relations are listed together with orthographic, syntactic, semantic information and examples of their usage. Contrast-Ita Bank is the corpus that we annotated with contrast, and in which we also annotated connectives.

5.1 Contrast-Ita Bank: a Corpus Annotated with the Contrast Relation

This section is dedicated to the investigation of the *contrast* relation. We perform this analysis through the observation of *contrast* in Contrast-Ita Bank[142]: a pre-existing corpus of news for Italian which we annotated with the discourse relation of *contrast*. We present the resource and discuss the data we collected in terms of frequency and explicitness/implicitness of the relation.

The section is structured as follows. In Section 5.1.1 we motivate and clarify the choice of adopting the PDTB schema for the annotation of contrast in Contrast-Ita Bank, and in 5.1.2 we describe in detail the guidelines we followed. In Section 5.1.3 we present the corpus of documents that we annotated, and the annotation tool that we used. Data on *contrast* emerging from Contrast-Ita Bank are discussed in Section 5.1.4, followed by an inter-annotator agreement evaluation in Section 5.1.5.

5.1.1 Adopting the Penn Discourse Treebank Schema

The creation of Contrast-Ita Bank (a corpus annotated with the relation of *contrast*) requires a scheme for annotation of the phenomena and the definition of the guidelines annotators have to follow. As mentioned in the chapter introduction, we aim at creating a resource that is parallel to the Penn Discourse Tree Bank (PDTB) (Prasad et al., 2007), the largest and most used corpus annotated with discourse relations. As described in Section 4.1.2, relations in the PDTB are identified in terms of *connectives* that take two *argu-*

[142]Some of the results presented in this section are presented in (Feltracco et al., 2017). Contrast-Ita Bank is freely distributed resource under a CC-BY-NC licence at: `https://hlt-nlp.fbk.eu/technologies/contrast-ita-bank`.

ments, and are tagged with *senses* that provide their semantic description. Each relation is considered in isolation and independently from the others. This approach is compatible with our goal (i.e., to identify only *contrast* in documents), since it allows us to focus just on the *contrast relation* with no need for identifying the entire discourse structure, that is, for identifying all the discourse relations in a document.[143]

Following the PDTB, the annotation of *contrast* in a corpus involves: i) the identification of *connectives* that convey the discourse relation of contrast (in case the relation is actually made explicit by a connective), ii) the identification of the arguments of the relation (*Arg1* and *Arg2*[144]), iii) the tagging of the *sense* of the connective.

Example 5.1 is extracted from the PDTB Annotation Manual (Prasad et al., 2007) and shows how discourse relations are conventionally reported in the PDTB Manual[145]: the *connective* (if present) is underlined, *Arg1* appears in italics, *Arg2* appears in bold. In this thesis, we follow the same convention.

(5.1) *Most bond prices fell on concerns about this week's new supply and disappointment that stock prices didn't stage a sharp decline*. **Junk bond prices moved higher**, however.
(sense tag: COMPARISON:*Contrast*)

5.1.2 The Penn Discourse Treebank Guidelines and our Adaptations

We defined the guidelines for the annotation of the relation of *contrast* mainly following the Penn Discourse Treebank 2.0 Annotation Manual (Prasad et al., 2007) and the most recent proposal by Webber et al. (2016), adapting these guidelines to our goal. These adaptations are mainly due to the fact that we are focusing on the *contrast* relation and not on all the phenomena that the PDTB accounts for (e.g. other relation types, other relation senses). More specifically, the guidelines for the annotation of Contrast-Ita Bank are defined in terms of *types of relations*, *connectives, arguments*, and *sense tags*.

Type of relations.

Following the guidelines in the PDTB2.0, we annotated *explicit relations* by identifying the discourse connectives that trigger the relations, and the respective arguments, as

[143] This is different from other approaches that aim at annotating the entire structure of a discourse. See Section 4.1.2.

[144] As explained in Section 4.1.2, *Arg2* is the argument that is syntactically bound to the connective, and *Arg1* is the other one (Prasad et al., 2007).

[145] Many other examples have been presented in the previous chapter.

in Example 5.1 and in Example 5.2 from Contrast-Ita Bank[146].

(5.2) *Il ministro del Lavoro e delle Pensioni britannico, Andrew Smith, ha rassegnato ieri le dimissioni* nonostante **i tentativi del premier Tony Blair di convincerlo a rimanere**. [doc:5397]
Eng.: *British Secretary of State for Work and Pensions Andrew Smith resigned yesterday* despite **Prime Minister Tony Blair's attempts to persuade him to stay.**[147]

Similarly to the PDTB, we annotated cases of *implicit relation* as well. These are cases in which the relation lacks an explicit discourse connective and is found between adjacent sentences. We also applied a similar annotation procedure, which consists in identifying the relation, tagging the first token of Arg2 to signal the relation, and identifying the connective(s) that can make the inferred discourse relation(s) explicit, as we reported in Example 5.3, in which the first token of Arg2 appears underlined.

(5.3) [..] *La Cdu ha migliorato il risultato di cinque anni fa passando dal 45,5% al 47,5% dei voti*. **Il partito socialdemocratico è crollato al 30,8% dal 44,4%**. [doc:5396]
- implicit: invece
Eng.: [..] *The Cdu has improved its five-year-before result passing from 45.5% to 47.5% of the votes*. **The social democratic party has dropped to 30.8% from 44.4%**.

Differently from the PDTB, we extended the annotation to arguments which appear in the same sentence and are separated by a comma[148], as shown Example 5.4.

(5.4) *le Generali hanno chiuso con un rialzo dello 0,59%*, **le Ras con un calo dello 0,31%**. [doc:5704]
implicit: invece
Eng.: *Generali closed with an increase of 0.59%*, **(the) Ras with a loss of 0.31%.**

We also accounted for cases of *alternatively lexicalization* (*AltLex*), in which a discourse relation is inferred between adjacent sentences and it is signalled by an expression (which is not in the set of explicit connectives), so that providing a suggestion of connective leads to redundancy in the expression of the relation (Prasad et al., 2007). An

[146]The examples are presented following the PDTB conventions: *connectives* are underlined, *Arg1* in italics, *Arg2* in bold. The number in square brackets refers to the Contrast-Ita Bank document in which the example can be found.

[147]The English version of the examples are produced by the author of this thesis and is intended only for readability purposes.

[148]Also Tonelli et al. (2010) adopt a similar strategy in the annotation of the LUNA corpus - see Section 4.1.2.1. The fact that implicit discourse relations have been not annotated within a sentence (except between clauses connected by a semicolon) is discussed by Prasad et al. (2014) as a gap in the PDTB annotation that authors plan to address.

example of AltLex from the Contrast-Ita Bank is shown in Example 5.5; in this case, the insertion of a connective, such as *ma* or *tuttavia* (Eng. 'but' and 'nevertheless'), sounds redundant. The span of the expression that constitutes the *alternatively lexicalization* (i.e., the expression that makes us inferring the relation) has been annotated.

(5.5) *Sei-righe-sei per dire che questa volta va bene, che si può andare avanti.* [..] **Peccato che questo succinto comunicato segua di poche ore la pubblicazione di una intervista ad una Margherita Cogo mai così agguerrita**. [doc:405610]
altlex: Peccato che
Eng.: *Six-lines-six to say that this time it is ok, that it is possible to go on.* [..] **It is a pity that this brief public notice follow a few-hours-before release of an interview to Margherita Cogo, never so hardened**.

In Contrast-Ita Bank, we did not limit the annotation of *AltLex* to adjacent sentences, rather we included non-adjacent ones.

The annotation of the PDTB also accounts, on the one hand, for cases in which no discourse relation can be inferred between adjacent sentences, i.e., the second sentence only serves to provide some further description of an entity in the first sentence (*EntRel*), and, on the other hand, for cases in which neither a discourse relation nor entity-based coherence can be inferred (*NoRel*).[149] However, since our aim was to exclusively annotate cases in which there is a (explicit, implicit or alternatively lexicalized) discourse relation of *contrast*, we did not annotate these cases and the labels *EntRel* and *NoRel* have not been adopted. We also did not annotate the relation of *attribution* or material that is relevant but not necessary for the interpretation of the relation, i.e., *supplements* (see previous chapter).

Connectives.

We followed the PDTB also concerning the syntactic classes of the connectives. According to Prasad et al. (2007), a connective should belong to the following classes:

- subordinating conjunctions (e.g. *when, because, although, etc.*, and for Italian *quando, nonostante, sebbene, anche se, affinché*)
  For example, *although* in: "Although he came, she left." ;

[149] For detailed annotation guidelines refer to Prasad et al. (2007).

- coordinating conjunctions (e.g. *and, or, but*, and for Italian *e, ma, però, eppure*);
  For example, *but* in: "He came, but she left."

- adverbs (e.g. *however, otherwise, instead*, and for Italian *così, allora, invece, ciononostante, però, tuttavia*)
  For example: "He came. However, she left."

- prepositional phrases (e.g. *on the other hand, as a result*, and for Italian *a causa di, invece di, a dispetto di*).
  For example: "He came. As a result, she left.'

It is worth mentioning that we did not approach the annotation of the corpus with a list of connectives of contrast that can potentially convey contrast (e.g. from LICO, see Section 5.2.3); instead, we annotated the contrast relation by discovering each time which are the connectives that convey it.[150]

Arguments.

We mainly followed the Annotation Manual of the PDTB2.0 also concerning the annotation of the arguments. In PDTB, relations are annotated when they connects "two *abstract objects* such as events, states, and propositions" (Asher, 1993), that can be realised mostly as clauses, normalisations, anaphoric expressions.

Arguments in the PDTB are generally composed by at least a full clause (as in the examples proposed so far). There are no restrictions on how many or what types of clauses can be included in an argument; however, "only as many clauses and/or sentences should be included in an argument selection as are minimally required and sufficient for the interpretation of the relation": this is called Minimality Principle (Prasad et al., 2007). For instance, in the following example from Contrast-Ita Bank, Arg1 includes two sentences.

(5.6) *Al CdA di via Segantini, ai comitati tecnici, ai vertici dei consorzi di secondo grado e delle società partecipate o collegate verrà chiesto di indicare quali sono le aspettative sulla Federazione, i suoi punti critici, gli obiettivi prioritari per migliorarne il ruolo. Ai responsabili dei reparti e servizi interni, ai loro vice nonché ai principali collaboratori sarà invece domandato di indicare quali sono le attese e le criticità nei rapporti interni alla Federazione e tra quest'ultima e*

[150] As we will see in Section 5.2.4.3, a side result of the annotation of Contrast-Ita Bank is a corpus-based list of connectives of contrast that can be integrated to LICO.

le cooperative socie e allo stesso tempo clienti. Ma **la centrale del movimento cooperativo si aprirà anche alla società trentina**. [doc:5705]

Eng.: *The BoD of Segantini street, the technical committees, the top managers of the second-degree consortium and of the affiliated companies will be asked to indicate which are their expectation about the Federation, the shortcomings, the main goals to improve its position. The heads of departments and of the internal services, their vices and the main collaborators will be asked, on the other hand, to indicate their expectations and criticalities in the internal relation of the Federation and between the Federation and the member (and also costumer) cooperatives.* However **the main office of the cooperative movement will open to the trentinian society as well.**

As detailed in the Annotation Manual of the PDTB2.0 (Prasad et al., 2007), nominalizations can also be annotated as arguments[151], as in the following example from Contrast-Ita Bank:

(5.7) *Su tutte, la Ferrari F.lli Lunelli, che ha consolidato le proprie posizioni,* nonostante **la non facile congiuntura nazionale e internazionale**, *con un utile netto pari a 4,2 milioni di euro* .. [doc:5437]

Eng. *Among all of them, Ferrari F.lli Lunelli, that has reinforced its position,* although **the not-easy national and international situation**, *with a net income equals to 4.2 milions of euro* ..

In the annotation, we included conjoined VPs as well, as proposed by Webber et al. (2016). This includes avoiding the annotation of the span of text that can be referred to both arguments in case of inter-sentencial VP conjoined arguments. In Example 5.8 from Webber et al. (2016) "Give the EPA more flexibility to" is not annotated since it is referred both to Arg1 and to Arg2. Example 5.9 reports a similar example from Contrast Ita-Bank.

(5.8) Give the EPA more flexibility to *declare a pesticide an immediate hazard* and **pull it from the marketplace**.

(5.9) Un gruppetto di persone *condanna quanto compiuto dai terroristi a Beslan* ma **sottolinea che non si può non evidenziare le crudeltà compiute dagli uomini di Mosca ai danni del piccolo paese caucasico**. [doc:5385]

Eng.: A group of people *condemns what the terrorist did in Beslan* but **highlights that it is not possible not to emphasize the cruelties made by the Moscow men to the little Caucasian country.**

[151] According to the PDTB guidelines, nominalizations can be annotated as arguments of connectives when they allow for an existential interpretation, and when they involve a clearly observable case of a derived nominalization. For further details, see the PDTB Annotation Manual (Prasad et al., 2007).

According to Webber et al. (2016), in these cases the spans of both arguments should be parallel (e.g. both bare infinitives, to-infinitive, or tensed clauses): in Example 5.8, the two arguments are constructions with the infinitive form of the verb, in Example 5.9 both are tensed clauses.

In the PDTB, verbs of attribution along with their subject ('He said', 'they announce..') are excluded, as in Example 5.10 below by Webber et al. (2016), except when they contribute to the interpretation of the relation, as in Example 4.17, here reported in 5.11 (from Prasad et al., 2007, p.40) [152]:

(5.10) The company, based in San Francisco, said it *had to put down a crude-oil pipeline in the Bay area to check for leaks* but added that **its refinery in nearby Richmond, Calif., was undamaged**.

(5.11) *Advocates said* *the 90-cent-an-hour rise, to $4.25 an hour by April 1991, is too small for the working poor*, while **opponents argued** **that the increase will still hurt small business and cost many thousands of jobs**.

Final conventions on punctuation and parenthesis were also specified for the annotators. In particular, punctuation at the boundaries is excluded from the extent of the arguments; text in parenthesis is included in case of appositions, nouns, adverbials, but excluded in case of propositions that are not required for the interpretation of the relation.

As specified in the previous chapter, in the PDTB arguments of a discourse relation can be found in any order and this is the same for Contrast-Ita Bank. For instance, notice that in example 5.8, 5.9 , 5.10, 5.11 the text of Arg2 is following Arg1, while in 5.7 Arg2 is inside the span of Arg1.

Moreover, in the case of *explicit* or *alternatively lexicalized* relations, there is no restriction on how distant arguments can be[153], as in the following example from Contrast-Ita Bank:

(5.12) *La capomissione, Simona Torretta, di Baghdad si innamorò nel '94, al suo primo viaggio. E da cinque anni praticamente vive lì.* È una donna «tosta » e concreta. Esperta, anche. Non ha risparmiato, mai, critiche a chiunque sia andato in Iraq con un'arma in mano, siano essi americani, inglesi, italiani. «La notte e il giorno di Baghdad non riservano più sorprese: ore e ore di bombardamenti che

[152] As specified in Section 4.1.2, the span of text corresponding to the attribution phrase is shown within a box.

[153] As explained before, this is an adaptation with respect to the PDTB guidelines, according to which this is only possible for explicit relations.

ci costringono a rifugiarci dove capita» raccontava in una intervista all'Ansa nei primi giorni di guerra, con la rabbia di non poter continuare a lavorare. Con i suoi volontari iracheni, inoltre, ha portato migliaia di litri d'acqua a Falluja e Najaf, entrando nelle città assediate e sotto le bombe. «Sparano a chiunque si muova - disse dopo la missione a Najaf- hanno occupato anche l'ospedale pediatrico. È una situazione drammatica». **Simona Pari**, invece, **in Iraq ci sta da un anno, per occuparsi soprattutto di progetti con i bambini**. [doc:5658]

Eng.: *The expedition leader, Simona Torretta, fell in love with Baghdad in '94, during her first journey. And, in practice, since five years she live there*. She is a «tough» and practical woman. Also an expert. She never avoided criticize anyone who went to Iraq with a weapon, no matter if America, English or Italian. «Nights and days in Baghdad do not hold surprises anymore: hour and hour of bombing which impose us to seek refuge where possible» she said in an interview to Ansa in the first days of the war, angry because it was not possible to continue to work. Moreover, with her Iraqis volunteers, she brought thousand of liters of water to Falluja and Najaf, entering the besieged cities, under the bombs. «They shoot to anythings that moves she said after the mission in Najaf- they occupy even the pediatric hospital. This is a tragic situation ». **Simona Pari**, on the other hand, **lives in Iraq since just one year, mainly handling project for children**.

Implicit relations are, instead, limited to adjacent sentences or to intra-sentencial adjacent parts of text separated by a semi-column or a comma[154] as in Example 5.3 and Example 5.4.

Senses of relations.

In Contrast-Ita Bank we annotated the two types COMPARISON.*Contrast* and COMPARISON.*Concession* with its subtypes (Arg1.as.denier and Arg2.as.denier), as defined in the PDTB 3.0 schema of relations (Webber et al., 2016): in our opinion these are all cases of contrast, as explained in the previous chapter (Section 4.2.2). Table 5.1 synthesizes the definitions for these sense tags presented in detail in Section 4.2.

We thus annotated the following three sense tags: CONTRAST, CONCESSION.Arg1-as-denier and CONCESSION.Arg2-as-denier. Table 5.2 shows the mapping between the sense tags in the PDTB and the simplified version we used in Contrast-Ita Bank.

The examples from Contrast-Ita Bank that we presented so far are annotated with

[154] As explained before, to annotate comma-separated clauses in a sentence is an adaptation with respect to the PDTB guidelines.

Relation	Definition
Contrast	→ the two arguments "share a predicate or a property and the difference between the two situations described in the arguments is highlighted with respect to the values assigned to this property."
Concession.Arg1.as.denier	→ expectation(s) raised by Arg2 are then denied by Arg1
Concession.Arg2.as.denier	→ expectation(s) raised by Arg1 are then denied by Arg2

Table 5.1: Definition of CONTRAST and CONCESSION in the PDTB 2.0 and 3.0.

	PDTB tags		→	Contrast-Ita Bank tags
CLASS:	TYPES:	SUBTYPES:		
COMPARISON	*Contrast*		→	CONTRAST
	Concession	Arg1-as-denier	→	CONCESSION.Arg1-as-denier
		Arg2-as-denier	→	CONCESSION.Arg2-as-denier

Table 5.2: CONTRAST and CONCESSION in the PDTB and in Contrast-Ita Bank.

these three sense tags in the resource. For instance, Example 5.3, that we report here as Example 5.13, is annotated in Contrast-Ita Bank as CONTRAST.

(5.13) [..] *La Cdu ha migliorato il risultato di cinque anni fa passando dal 45,5% al 47,5% dei voti*. **Il partito socialdemocratico è crollato al 30,8% dal 44,4%**. [doc:5396]
- implicit: invece
CONTRAST
Eng.:[..] *Cdu has improved its five-year-before result passing from 45.5% to 47,5% of the votes*. **The social democratic party has dropped to 30.8% from 44.4%**.

Example 5.13 has been annotated with the tag CONTRAST because the elements in the two arguments share a property but a difference in the value of the property is highlighted. In both arguments, the variance of the total electoral votes of the two political parties (with respect to a precedent situation) is described: the two parties thus share the fact that they register a new result of their votes. The difference lies in the fact that votes increased for one party (Arg1), and decreased for the other (Arg2), and this difference is highlighted in the text.

Example 5.14 is annotated with the tag CONCESSION.Arg1.as.denier.

(5.14) *Il ministro del Lavoro e delle Pensioni britannico, Andrew Smith, ha rassegnato ieri le dimissioni* nonostante **i tentativi del premier Tony Blair di convincerlo a rimanere**. [doc:405397]

CONCESSION.Arg1.as.denier

Eng.: *British Secretary of State for Work and Pensions Andrew Smith resigned yesterday* despite **Prime Minister Tony Blair's attempts to persuade him to stay.**

The situation in Arg1 (the fact that the British Secretary of State for Work and Pensions resigned) happened although what it is expressed from Arg2 (the premier tried to convince him to stay): in other words, Arg1 is denying the expectation that rises from Arg2.

On the other hand, in Example 5.15, it is Arg2 that denies the expectation from Arg1: the text in Arg1 makes a detailed reference to the involvement of several administrations in a project; in Arg2 is however stated that these are not the only subjects involved (as one could expect) and other local institutions will be affected. The example is annotated with as CONCESSION.Arg2.as.denier.

(5.15) *Al CdA di via Segantini, ai comitati tecnici, ai vertici dei consorzi di secondo grado e delle società partecipate o collegate verrà chiesto di indicare quali sono le aspettative sulla Federazione, i suoi punti critici, gli obiettivi prioritari per migliorarne il ruolo. Ai responsabili dei reparti e servizi interni, ai loro vice nonché ai principali collaboratori sarà invece domandato di indicare quali sono le attese e le criticità nei rapporti interni alla Federazione e tra quest' ultima e le cooperative socie e allo stesso tempo clienti*. Ma **la centrale del movimento cooperativo si aprirà anche alla società trentina**. [doc:5705]

CONCESSION.Arg2.as.denier

Eng.: *The BoD of Segantini street, the technical committees, the top managers of the second-degree consortium and of the affiliated companies will be asked to indicate which are their expectation about the Federation, the shortcomings, the main goals to improve its position. The heads of departments and of the internal services, their vices and the main collaborators will be asked, on the other hand, to indicate their expectations and criticalities in the internal relation of the Federation and between the Federation and the member (and also costumer) cooperatives.* However **the head office of the cooperative movement will open to the trentinian society as well.**

The annotation guidelines of PDTB 2.0 (Prasad et al., 2007) account also for cases in which a discourse connective can have more than one sense (i.e., it can indicate more than one semantic relation): the annotation of multiple sense for the same connective in the same context is thus allowed. In line with this, in Contrast-Ita Bank, we allowed annotators to mark, when necessary, two senses for the same connective (e.g. both CONTRAST

and CONCESSION_Arg1.as.denier). An example of "double" relation annotation from Contrast-Ita Bank is reported in 5.16.

(5.16) *Per un anno il nostro lavoro in Iraq è andato avanti bene*. Però **da marzo di quest' anno, con la rivolta di Al Sadr a Falluja la situazione è degenerata**. [doc:5664]
CONTRAST, CONCESSION.Arg2.as.denier
Eng.: *For one year, our work in Iraq was going on well.* However **since March of this year, with the riot of Al Sadr in Falluja the situation degenerated**.

In this example, the author describe how a work situation of a group of people has changed in time. In Arg1 the previous positive work situation is described, in Arg2 the actual dramatic one: the two situations are in contrast, and the contrasting property is highlighted. At the same time, Arg2 represents a denial for the expectation rising in Arg1: indeed, one would expect that since the situation was positive (Arg1), it still is; but this is not the case since the situation changed (Arg2).

Finally, according to the schema proposed in Webber et al. (2016), it is possible to assign "features to indicate an inference of *implicit belief* or a *speech act* associated with arguments". These are cases in which what is involved in a relation is the speaker's/writer's belief or the speech-act associated to the argument. In Contrast-Ita Bank we did not assign these features.[155]

5.1.3 The Corpus and the Annotation Tool

Following the PDTB 2.0 Annotation Manual Prasad et al. (2007), the proposal by Webber et al. (2016) and the few adaptations that we described in the previous section, we created Contrast-Ita Bank by annotating the *contrast relation* in a corpus of 169 news stories, for a total of 65,053 tokens (average length: about 385 tokens per document). The documents correspond to articles published in a newspaper in two different days and include reports, news about politics, news about economics, sport results. They contain narrations and quotes of oral interviews. The documents are the same contained in the Fact-Ita Bank corpus, a corpus annotated with factuality information (Minard et al., 2014) and, partially, with negation (Fact-Ita Bank-Negation, Altuna et al., 2017)[156]. Originally, the documents were selected from the larger corpus Ita-TimeBank (Caselli et al., 2011): a

[155]Further details on the guidelines of the PDTB can be found in Prasad et al. (2007) and Webber et al. (2016).

[156]Find both resources at `https://hlt-nlp.fbk.eu/technologies/fact-ita-bank`

language resource manually annotated with temporal and event information. The same documents are also part of the I-CAB corpus (Magnini et al., 2006): a corpus of Italian news annotated with temporal expressions and different types of entities (i.e., persons, organizations, locations, and geo-political entities).

Despite the already annotated information in these documents, the annotation of contrast was carried out on raw text.

The corpus is annotated using CAT (Content Annotation Tool)[157], a general-purpose text annotation tool developed by Bartalesi Lenzi et al. (2012). Figure 5.1 reports the screen-shot of the annotation of Example 5.2, here reported as 5.17, with the tool. In the figure, the *connective* (reported as underlined in the text) is in purple, *Arg1* (in italics in the text) is in yellow, *Arg2* (in bold in the text) is in blue.

(5.17) *Il ministro del Lavoro e delle Pensioni britannico, Andrew Smith, ha rassegnato ieri le dimissioni* nonostante **i tentativi del premier Tony Blair di convincerlo a rimanere**. [doc:5397]
Eng.: *British Secretary of State for Work and Pensions Andrew Smith resigned yesterday* despite **Prime Minister Tony Blair's attempts to persuade him to stay**.

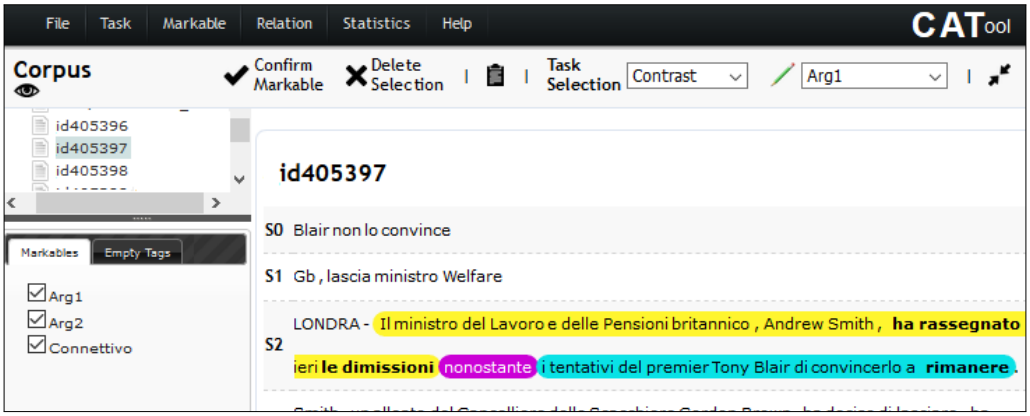

Figure 5.1: A screen-shot of the CAT tool used for the annotation of contrast.

5.1.4 Analyzing Discourse Contrast in Contrast-Ita Bank

One expert annotator[158] concluded the annotation of *Explicit*, *Implicit*, and *Altex* relations of contrast in the 169 documents of the corpus in almost three weeks, for a total 372 relations, the average being thus 2.16 relations per document. In this section we discuss the

[157] `https://dh.fbk.eu/resources/cat-content-annotation-tool`

[158] The annotator is the author of this thesis.

results of the annotation (reported in Table 5.3) in terms of type of the relation (*explicit, implicit, altlex*) and in term of sense of the relation (CONTRAST and CONCESSION). An analysis of the connectives used to convey explicit relations will be proposed in Section 5.2.4.1.

Types of relations / Senses of relations	Explicit	Implicit	AltLex	Total
CONTRAST	87	12	3	102
CONCESSION.Arg1-denier	21	0	1	22
CONCESSION.Arg2-denier	201	8	3	212
CONTRAST - CONCESSION.Arg1-denier	3	0	0	3
CONTRAST - CONCESSION.Arg2-denier	28	4	0	32
CONCESSION.Arg1-denier - CONCESSION.Arg2-denier	1	0	0	1
Total	341	24	7	372
Density (relations / documents)	0.0052	0.0003	0.0001	0.0056

Table 5.3: Relations of contrast in Contrast-Ita Bank per type and sense.

From the table, it is possible to notice that the explicit relations (second column) are the most common and correspond to more than 91% of all the relations, with an average of 2 relations per document. We register a maximum number of 15 explicit relations in one document.

Implicit relations (third column) are less frequent and occur 15 times inter-sentencially and 9 times infra-sentencially for a total of 24 annotations. This is different from the PDTB2.0, as in the English version the ratio between explicit and implicit concerning CONTRAST and COMPARISON, and their subtypes, is 0.45 (without counting when the relations are tagged with others), while in Contrast-Ita Bank is ten time less. This might be due to the fact that in Contrast-Ita Bank the annotator mark just the contrast relation, and it is possible that they simply failed to capture implicit relations, while in the PDTB2.0 annotators were asked to mark also cases where no relation is found, thus analysing in detail if a certain relation appears between every pair of sentences[159].

AltLex relations (fourth column) seem rarer: they have been annotated in 7 cases. This is also the rarest case for the sense tags in the PDTB 2.0. Relations in this case have been *alternatively lexicalized* inter-sentencially and infra-sentencially by: *anche al netto*

[159] According to the PDTB 2.0 guidelines, the tag *NoRel* has to be used where neither a discourse relation can be inferred between the adjacent sentences (Prasad et al., 2007).

di, certo, il punto è che, non, peccato che, quella sì, macché (Eng. 'even after deducting', 'for sure', 'the point is that', 'not', 'now, that', 'as if'). By definition none of these expressions are connectives in Italian and to provide a suggestion of connective leads to redundancy in the expression of the relation, as in Example 5.5 and in the following one:

(5.18) *Alla fine, commenti positivi , dal diessino Bondi* [..] *al margheritino Casagranda («Spirito costruttivo»), passando per il Verde Marco Boato*. Certo, **non tutto è risolto** [doc:5608]
CONCESSION.Arg2.as.denier - altlex: Certo
Eng.:*At the end, positive comments came from Bondi, of DS,* [..] *from Casagranda, of Margherita («Productive Mood»), from Marco Boato, of Verdi,*. Of course, **not everything is solved**.

We believe that these are interesting results not only because they provide some insights of the contrast relation in Italian, but also from a machine learning point of view. For example, a system that relies only on the presence of a connective of contrast for the identification of contrast relation will not be able to detect the cases of implicit or altLex relation, but will definitely cover a high amount of cases. In the next section we will discuss further this observation, when discussing which are the connectives that convey contrast.

Table 5.3 also shows that the per token density of *contrast* in the corpus is 0.0056, similar to the PDTB (i.e., 0.0072).

The same table reports the data concerning the three sense tags used in Contrast-Ita Bank: CONTRAST, CONCESSION.Arg1-as-denier, CONCESSION.Arg2-as-denier; and multiple sense tags were allowed.

The most frequent type of contrast (here *sense*) is CONCESSION.Arg2-as-denier (fourth row), that is when Arg2 is denying an expectation that rises from Arg1. This cover about 0.57% of the cases. CONTRAST (second row) covers almost a quarter of the cases and has been annotated together with CONCESSION.Arg2-as-denier 32 times (28 for explicit type and 4 for implicit type, see row six). CONCESSION.Arg1-as-denier is far less frequent both as only type and as with other relations, and has been annotated less than 10% of the cases. This sense is associated to a limited set of connectives: despite the list of connectives of contrast marked in total Contrast-Ita Bank is of 19, 7 of them (e.g. *nonostante*, Eng. 'although') signal CONCESSION.Arg1-as-denier all the times. We will

discuss more about the list of connectives in Section 5.2.4.1.

The case in which CONCESSION.Arg1-as-denier and CONCESSION.Arg2-as-denier have both been annotated is reported here.

(5.19) "Gli organizzatori del Palio stanno lavorando così bene che *non posso tirarmi indietro* anche se, potete immaginarlo, **sono brutti momenti..."**[doc:5848]
CONCESSION.Arg1-denier, CONCESSION.Arg2.as.denier
Eng.:"The Palio organizers are working so hard that *I cannot pull back* even if, as you can imagine, **circumstances are bad.."**

In Arg1 of Example 5.19, the speaker says that he would not pull back his participation and in Arg2 he says that circumstances are bad. A double interpretation of the arguments is possible. One interpretation recognises Arg2 as the denier; he is participating but do not expect that this means that everything is fine: circumstances are bad. Another interpretation conceives Arg1 as the denier; circumstances are bad but do not expect this is stopping him: he will not pull back.

Another interesting result that can be seen in Table 5.3 concerns the sense for *implicit relations*. Many cases are tagged with the sense CONTRAST in which the two arguments share some properties and a difference between the two is enlightened. This characteristics can be express through a very clear structural parallelism of the two arguments that does not require an explicit connective. As we saw in Section 4.2.1, this structural similarity has already been discussed by (Asher and Lascarides, 2003) as a feature of the *contrast* relation. For instance, in Example 5.3, reported here above, there is a parallelism in the two arguments (i.e., party's name - results increasing/decreasing - value of the variance).

(5.20) *La Cdu ha migliorato il risultato di cinque anni fa passando dal 45,5% al 47,5% dei voti*. **Il partito socialdemocratico è crollato al 30,8% dal 44,4%**. [doc:5396]
- implicit: invece
CONTRAST
Eng.: [..] *Cdu has improved its five-year-before result passing from 45.5% to 47.5% of the votes*. **The social democratic party has dropped to 30.8% from 44.4%**.

Notice that in the example the contrasting feature (i.e., the highlighted difference) is expressed through the use of antonymical expression *ha migliorato il risultato* (Eng.'it has improved its results) and *è crollato* (Eng. 'it has dropped'). We will discuss in Chapter 6 the relationship between discourse contrast and lexical opposition in terms of co-

occurrence.

The presence of implicit relations tagged with CONCESSION, on the other hand, seems contrasting with the definition proposed by Asher and Lascarides (2003) for the subtype of contrast *violated expectation*: the authors say that this type of contrast needs always to be conveyed by an explicit cue (Asher and Lascarides, 2003, p.167). However, we found this happening in 12 cases. Specifically, in 4 cases the relation is detected together with CONTRAST as in Example 5.21, where in addition to the parallelism of CONTRAST (i.e., in the example two contrasting versions about the reason for a redeployment are described), Arg2 has been interpreted as denying the expectation that rises from Arg1 (that the official version was the actual and accepted one).

(5.21) [..] *è stato trasferito alle Volanti su decisione del Questore. Ufficialmente per «esigenze di servizio e per miglior riassetto interno dei vari uffici»*; **per il sindacato autonomo di polizia (Sap) si è trattato «di un trasferimento di carattere esclusivamente "punitivo" e che nulla ha a che vedere quindi con il miglior riassetto interno** [..] [doc:5900]
- implicit: invece, nonostante ciò
CONTRAST, CONCESSION.Arg2.as.denier
Eng.: [..] *he has been redeploy to the Volanti following the commissionaire decision. Officially due to service requirements and to improve the internal reorganization of the various offices»*; **for the police trade union (Sap) it was a « merely "punitive" redeployment that, thus, has nothing to do with the internal reorganization** [..]

In the other cases (such as in Example 5.22) only the CONCESSION.Arg2.as.denier tag has been assigned to an implicit relation.

(5.22) *La risorsa maggiore del Trentino è l' ambiente*. **Da solo non basta**. [doc:5920]
- implicit: tuttavia
CONCESSION.Arg2.as.denier
Eng.: [..] *The major resource for the Trentino area is the enviroment*. **Alone it is not enough**.

In the example, in Arg1 the fact that (natural) environment is the most important resource for a place (i.e., for tourism sector to be prosperous) triggers a positive expectation that it is a sufficient resource, but Arg2 denies this expectation.

5.1.5 Evaluating Contrast Annotation with Inter Annotator Agreement

In order to estimate the complexity of annotating the corpus with the contrast relation, we compute the agreement (IAA) between two annotators on 18 documents, 10.6% of the whole corpus. This means that we asked a second expert annotator[160] to annotate 18 documents, following the guidelines reported in Section 5.1.2. The second annotator was trained on some examples which are not included in the evaluation we report here.

In this sub-corpus we created for the evaluation, annotators marked a total of 51 and 57 relations respectively. We calculate the IAA for five configurations: *relation identification*, *arguments*, *sense type* (i.e., CONTRAST or CONCESSION), *CONCESSION subtypes* (i.e., Arg1.as.denier or Arg2.as.denier), and finally *connectives identification*. Figure 5.2 reports the procedure that we adopt for calculating the five configurations. In the figure, the arrows connect two sequential steps: for instance, we calculate *sense type* after calculating (and for the data in which there is) the agreement on *relation identification*.

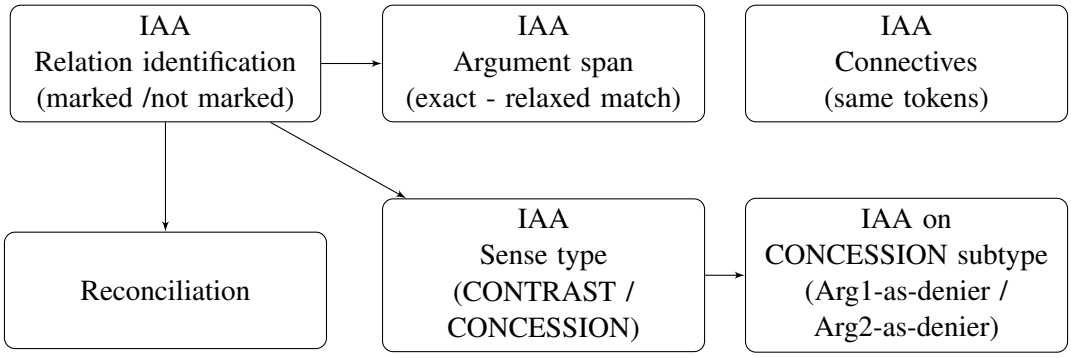

Figure 5.2: Procedure for calculating Inter Annotator Agreement.

Relation identification. In the first configuration, *relation identification*, we measure the agreement on recognising *explicit*, *implicit* or *altlex* contrast relations. This is possible because during the annotation process annotators marked, for each relation, the token(s) that represents i) the connective of the relation in case of *explicit* relation, ii) the first token on Arg2 in case of *implicit* relation, iii) the token(s) corresponding to the alternative lexicalization in case of *altlex* relation.

To calculate the agreement for *relation identification*, we consider as cases of agree-

[160] We acknowledge Manuela Speranza for her contribution.

ment when the span of text marked by one annotator to signal a relation corresponds to or is part of the span marked by the other annotator (for instance, if both marked *ma*, or if one marked *se* and the other *anche se* to signal the presence of a contrast relation). We calculate the final score for this configuration calculating the Dice's coefficient (Rijsbergen, 1979)[161] for the sets of relations identified by the annotators (i.e., the spans of text marked for signalling the presence of *contrast* by the two annotators). We divide the total number of cases in which annotators agreed in identifying a relation (i.e., they marked the same span or part of the same span of text), by the number of relations identified by each annotator. The result for this configuration *relation identification* is that annotators agree on recognizing the contrast relation in 37 cases, for a Dice value of 0.68. We consider this result reasonable given the difficulty of the task which has not to be underestimated. To identify *contrast* in a document means, first of all, to distinguish cases in which a lexical element is playing the role of connective from cases in which it is not (for example, when it is used as a cue element for signalling a change int the topic or to introduce an exclamation). Once a discourse relation is identified, the other important step is to recognize if the relation is a relation of contrast or if it is not. Moreover, the task does not just aim at the identification of *explicit* relations (that in turns means to judge if a lexical items in the text is conveying a relation, and of contrast type), it also accounts for *implicit relations* that by definition are not marked in the text. The task in this case is to identify the relation of contrast when there is no linguistic element that facilitates the operation.

Due to the difficulty of the task, and in order to perform a quantitative analysis of the annotated data and exclude cases of human error tagging or human misunderstanding of the guidelines, annotators were asked to meet and discuss their annotation choices with the possibility to revise their selections in light of the discussion. Cases of disagreement consist in cases in which one annotator marked a relation that the other did not mark. Several of these cases were "reconciled": this means that the annotators agreed in revise their choices and found an agreement. These cases included, for instance, cases of implicit or altlex relations that one annotator marked and the other simply failed to capture, or cases in which one annotator found it difficult to interpret the meaning of the text and

[161] As we saw in Section 3.2.2.1, Dice's coefficient measures how similar two sets are by dividing the number of shared elements of the two sets by the total number of elements they are composed by (i.e., Given two sets of annotation A and B, Dice's value = [2 x (A ∩ B) / (A + B)]). This produces a value from 1, if both sets share all elements, to 0, if they have no element in common.

decided not to mark the presence of a contrast: after the discussion (in which the other annotator exposed her interpretation) the presence of a contrast became clearer.

In other cases disagreement remained. These includes mainly cases in which both annotators agreed in recognizing a discourse relation (also conveyed by a connectives) but one interpreted the relation to be of the kind contrast, while the other did not. We notice that in many of those cases, the connective that conveys the relation is the coordinating conjunction *e* (Eng. 'and'), as in the following example, in which we report the annotation made by one annotator. The other did not mark the presence of a contrast relation since she interpreted the two arguments to be two alternatives, instead of being two arguments of a contrast.

(5.23) *Lascia perdere il Vietnam* e **parla di economia**. [doc:5401]
CONTRAST vs NON-MARKED
Eng.:*Leave Vietnam apart* and **speak about economics**.

Similarly, in Example 5.24 one annotator recognized a contrast between the arguments (we report her annotation), and the other interpreted the two arguments to be two non-contrasting parts of a description.

(5.24) [..] *sono portatori sani di Talassemia Mayor* e **il loro bambino, Luca, cinque anni, è talassemico**. [doc:5402]
CONTRAST vs NON-MARKED
Eng.:[..] *they are carrier of Talassemia Mayor* and **their son, Luca, five years old, is thalassaemic**.

Other cases of disagreement include the ones in which an annotator did not mark a relation of contrast due to a lack of contextual information, or due to the fact that too many inferences were required to identify a contrast, and she did not feel enough confident to annotate the relation, as for Example 5.25:

(5.25) *Due mesi fa avevo chiesto la verifica* ma **mi avevano dato del Follini** [doc:5608]
CONCESSION.Arg2.as.denier vs NON-MARKED
Eng.:*Two months ago I asked for a verification* but **they called me Follini**

In this case, both annotators recognised that they do not understand fully the meaning of Arg2 (i.e., they do not who Follini is, and what does it means to be compare to him). However, one annotator interpreted and marked the relation as being of contrast, being

helped by the presence of *ma* (Eng. 'but') and guessing that the request for a verification (Arg1) had not been accepted (Arg2); conversely, the other annotator decided to not annotate contrast, having not enough information.

Argument span. For the cases in which there is agreement on the *relation identification* (37 cases) we evaluate the *arguments* agreement: we calculate if the same portion of text has been recognised to be an argument of a *contrast* relation. We calculate the overall agreement for Arg1 and for Arg2 in two ways: *exact match* and *relaxed match*. In the *exact match* mode, we have agreement if, for the same relation (for which we already know that there is agreement), the two annotators considered the same exact span of text as Arg1. The same is calculated for Arg2. In the *relaxed match* mode, we have agreement if the text span identified by the annotators matches at least for its 50%.

Agreement in the *exact match* mode for Arg1 is 0.51 and for Arg2 is 0.70; the agreement in the *relaxed match* mode for Arg1 is 0.89 and for Arg2 is 0.91. We expected the *exact match* agreement difficult to reach. In fact, as described in 5.1.2, following the guidelines of the PDTB 2.0 (Prasad et al., 2007), we adopted the Minimality Principle, according to which "only as many clauses and/or sentences should be included in an argument selection as are minimally required and sufficient for the interpretation of the relation". This means, for example, there is no constrain on the length of an argument or that more than a sentence can be annotated as an argument (i.e., punctuation is generally not a limiting constrain). The selection of the arguments span thus significantly relies on the interpretation of the annotators and cases in which there is no exact match can be frequent. For instance, in the following example which the annotators discussed in an initial phase of training, they disagree on marking the same span for Arg1.

(5.26) *Più inquietante è ciò che è accaduto alla collega Anna Politkovskaya di Novaya Gazeta, anch'essa in viaggio per Beslan.* [*Qualcuno avrebbe avvelenato il caffè servitole a bordo dell' aereo, provocandone l' immediato ricovero all'atterraggio*]. Mentre **il corrispondente di Al Jazira, Amr Abdel Hamid, egiziano di nazionalità russa è stato arrestato** - [doc:5383]

Eng.: *What happen to the colleague Anna Politkovskaya of Novaya Gazeta, in a journey to Beslan as well, is more disturbing.* [*Somebody would have poisoned the coffee they offered in her flight, causing her immediate recovery when they lended.*]. While **the journalist of Al Jazira, Amr Abdel Hamid, Egyptian of Russian nationality, has been**

arrested.

One annotator suggested that Arg1 extent corresponds to the first two sentences (i.e., the span of text in italics) interpreting that they are both needed to emphasis the similarity between the situation in the arguments: the two persons were colleagues. The other annotator suggested that Arg1 is limited to the second sentence (i.e., the span of text in italics and in square brackets) and that information in the first sentence are not necessary for the interpretation of the relation.

Sense type: CONTRAST - CONCESSION. For the cases in which there is agreement on *relation identification*, we calculate the agreement on *sense type*: we consider the agreement in identifying CONTRAST, CONCESSION (not distinguishing between the two subtypes Arg1.as.denier and Arg2.as.denier) or both. Data are calculated counting 1 point if annotators agreed to assign or not assign the same relation(s) (i.e., they made the exact same choice), 0.5 if one annotated a relation and the other both (i.e., partial agreement), 0 point if one annotated CONTRAST and the other CONCESSION (i.e., total disagreement). Agreement is obtained summing the points for each annotation and dividing by the total of 37 relations that both annotators identified.

Agreement for *sense type* is 0.73, showing that recognizing the sense type of contrast can be a controversial decision among annotators. However, we believe that this result is fair, considering that the annotation regards two senses of the same class, and that they are not mutually exclusive. We report here two examples of disagreement. In 5.27, one annotator selected CONTRAST, while the other selected both CONTRAST and CONCESSION.Arg2-as-denier emphasizing that the first argument creates an expectation (i.e., precise information are not communicated) that the second denies.

(5.27) [..] *non precisa il Paese dei 10 presunti mujaheddin «arabi» .* Mentre **conferma che solo uno sarebbe sopravvissuto** [doc:5381]
CONTRAST vs CONTRAST-CONCESSION.Arg2.as.denier
Eng.:[..] *it does not specify the Country of the 10 supposed «Arabs» mujaheddin.* While **it confirms that just one survived.**

In Example 5.28, one annotator selected CONTRAST capturing the parallelism between what is expressed in Arg1 (i.e., everybody being with soul and heart in Beslan) and Arg2 (i.e., he-Putin- not being in Beslan). Conversely, the other selected the sense

tag CONCESSION.Arg2-as-denier believing that Arg1 triggers an expectation that Arg2 denies (i.e., since he underlines that the grief for what have happened in Beslan is shared by everybody, one would also expect him to attend the funerals; but he was not there).

(5.28) *«Con l'anima e col cuore siamo tutti a Beslan», ha detto Putin* , anche se **lui non si è recato ai funerali**. [doc:5381]
CONTRAST vs CONCESSION.Arg2.as.denier
Eng.:*«We are all in Beslan, heart and soul», said Putin* , even if **he did not attend the funerals**.

CONCESSION subtypes: Arg1.as.denier - Arg2.as.denier. For the 10 cases in which there is agreement on selecting the type CONCESSION (i.e., both annotator has marked at least one of the subtypes of CONCESSION, Arg1.as.denier or Arg2.as.denier), we calculate the *CONCESSION subtypes* agreement (see Figure 5.2). We apply the same calculation adopted for *sense type* (i.e., 1 point for agreement, 0.5 for partial agreement, 0 for complete disagreement). Agreement for *sense type* is of 0.9. This means that annotators agreed in 9 cases to mark the CONCESSION relation but in one cases they disagreed over the direction of the relation. The case is reported in 5.29.

(5.29) La famiglia aveva detto *che Clinton non aveva subito un attacco cardiaco*, ma **che alcune arterie risultavano ostruite**. [doc:5401]
CONCESSION.Arg1.as.denier vs CONCESSION.Arg2.as.denier
Eng.: The Clinton family said *that Clinton did not suffer a heart attack*, but **that some arteries were obstructed**.

For this example, both annotators agreed in recognising a contrast relation of type CONCESSION. One annotator considered it a case of Arg2.as.denier: the text signalled in bold denies the situation in Arg1, signalled in italics: the fact that Clinton did not go under surgeon is creating the expectation that Clinton is fine, but in Arg2 it is specified that some arteries are obstructed. The second annotator suggested exactly the opposite, interpreting the fact that not going under surgeon is denying the expectation rising from the fact that some arteries are obstructed.

Connectives identification. In the *connectives identification*, we have agreement if both annotators agreed on recognising the same explicit relation and the same exact span of

text to be the connective that conveys the relation. This is very similar to *relation identification* configuration but it is actually a more fine grained version, since we exclude from this calculation cases of altlex, for which the portion of text performing the "alternative lexicalization" has been annotated, and implicit, for which the first token of the Arg2 has been annotated. Moreover, we verify if both annotators annotated the same lexical elements, which can actually be composed of more than one token and be discontinuous. As a consequence, we consider as cases of agreement only cases in which annotators agreed in marking the same exact tokens as connectives; for instance, if one marked *se* and the other *anche se* is a not agreement. In these terms, the cases of agreement for *connectives identification* are a subset of the cases of agreement already captured by the *relation identification* agreement.

Also for this configuration we adopt the Dice's coefficient: the resulting value is of 0.68. Some of the consideration made discussing the agreement for *relation identification* are also valid for this configuration: the difficulty of the task lies in distinguishing cases in which a lexical element is actually playing the role of connective, and if the conveyed relation is the one of contrast. A deeper analysis shows that the difference with the *relation identification* (that however considers also altlex and implicit relations) is of just one case in which two connectives of contrast were adjacent in the text and was easily solved in the *reconciliation step*.

(5.30) [..] se da una parte *si mobilita con iniziative di preghiera per le vittime innocenti di Beslan* , dall' altra **non ha dubbi: « Le religioni non possono mai utilizzare il nome di Dio per insegnare l' odio.** [..] [doc:5396]
Eng.: [..] if on the one hand*he organizes prayer meetings for the Beslan innocent victims*, on the other hand **he has no doubt: «Religions can never use the name of God to teach hate.** [..]

In this case one annotator selected *se* as a coordinating conjunction that conveys contrast, the other selected *se da una parte.. dall' altra* identifying an explicit relation conveyed by a discountinuos connective: the two annotators thus marked the presence of a relation of contrast but did not agree on the span of the connective. Notice, however, that the second annotator forced the definition of connective provided in the guidelines, according to which connectives are subordinating or coordinating conjunction, adverbs or preposition and cannot include two of them. *A posteriori*, this case of disagreement was

solved by marking two relations: one conveyed by *se*, the other conveyed by *da una parte .. dall' altra*.

Table 5.4 reports a synthesis of the results we have detailed in this section. [162]

Number of relations: Annotator1 = 57; Annotator2 = 51; A ∩ B = 37	
IAA on:	
relation identification	
Dice's value	0.68
relation identification - post reconciliation	
Dice's value	0.84
arguments span - exact match	
(exact matches / total cases)	Arg1= 0.51; Arg2 = 0.70
arguments span - relaxed match	
(matches at 50% / total cases)	Arg1= 0.89 ; Arg2 = 0.91
sense type: CONTRAST - CONCESSION	
(1 agreement, 0.5 partial agreement, 0 disagreement / total cases)	0.73
CONCESSION subtypes:Arg1.as.denier - Arg2.as.denier	
(1 agreement, 0.5 partial agreement, 0 disagreement / total cases)	0.9
connectives identification	
Dice's value	0.68

Table 5.4: Inter-annotator Agreement for Contrast-Ita Bank.

5.1.6 Contributions, Shortcomings and Further Work

In the previous sections we presented our analysis of contrast in Italian, through the Contrast-Ita Bank corpus[163]: a corpus of 169 news annotated with the discourse relation of contrast. We provided quantitative data on the contrast relation in terms of frequency and explicitness/implicitness of the two types of contrast we distinguished (CONTRAST and CONCESSION). In the corpus, we found that CONCESSION.Arg2-as-denier (i.e., when Arg2 is denying an expectation that rises from Arg1) is the most

[162]Prasad et al. (2008) report higher values for the annotation in the PDTB 2.0. They report an overall agreement for argument identification of 90.2% for explicit relation and 85.1% for implicit (we do not calculate the value considering this granularity); when relaxing the match to partial overlap, the two values increase to 94.5% and to 85.1%. concerning sense tags, authors report an agreement of 94% for sense class, of 84% for sense type, and of 80% for the subtype level. See also (Miltsakaki et al., 2004).

[163]The work presented in this section has also been presented in Feltracco et al. (2017). In this project, the author of this thesis collaborated with Elisabetta Jezek (University of Pavia) and Bernardo Magnini (Fondazione Bruno Kessler, Trento). Specifically, the author defined the annotation guidelines, annotated the corpus and participated in the IAA evaluation. We also acknowledge Anne-Lyse Minard for her contribution in extracting the values for IAA directly from the CAT tool output.

frequent kind of contrast. CONTRAST covers almost a quarter of the cases, while the kind CONCESSION.Arg1-as-denier is far less frequent. The relation is most frequently explicit, that is, is conveyed by a connective (91% of the cases). We saw as many as 15 explicit relations in one document (average length of the examined documents is 387 tokens) and an average of 2 relations per document.

Overall, the inter annotator agreement exercise highlights that the main difficulties of annotating *contrast* concern: i) the *contrast identification*, especially for the relations conveyed by the coordinating conjunction *e* (Eng. 'and') and for the *implicit* and *altlex* ones; ii) the extent of the *arguments*: in many cases, the two annotators did not mark exactly the same tokens but it is very likely that their annotations match at least 50%; iii) *sense type*, since it seems that one annotator tended to annotate more frequently the CONCESSION_Arg2.as.denier relation and, for the cases in which she annotated the sense together with CONTRAST, the other annotator did not capture the CONCESSION_Arg2.as.denier.

We believe that the Contrast-Ita Bank corpus is a useful resource for the linguistic study of *contrast* in Italian, a frequent relation in discourse. It can be used in Natural Language Processing, where automatic identification of semantic relations between portions of text (such as contrast) is a main task. Contrast-Ita Bank also represents a first step towards a corpus of discourse relations for Italian, compatible with the Penn Discourse Treebank (PDTB) project, the largest and most used corpus annotated with discourse relations in the field of NLP.[164] Indeed, in the future, the resources can be further enriched by annotating all the sense tags listed in the PDTB3.0 schema (Webber et al., 2016).

5.2 Connectives of Contrast in Italian

In the previous section we have seen that relations of contrast in Contrast-Ita Bank are mostly conveyed by a connective. We thus believe that it is important to analyze which are the connectives that mark contrast, also considering the importance that these lexical cues can have for the automatic identification of contrast in texts. In this section, we aim at discussing which are those connectives. First, we will provide a definition for the concept of *connective* (Section 5.2.1) and an overview of related works in terms of existing lexicons for connectives (Section 5.2.2) including LICO, Lexicon for Italian COnnectives (Section 5.2.3). Then, we discuss the list of connectives that is retrieved

[164] In Section 4.1.2.1, we already presented similar attempts by Tonelli et al. (2010) and Pareti and Prodanof (2010).

from the Contrast-Ita Bank annotation (Section 5.2.4.1) and the cases of *false positives* that we found (Section 5.2.4.2). Finally, a comparison between the list of connectives in LICO and in Contrast-Ita Bank is provided.

5.2.1 A Definition for Discourse Connectives

We have been referring to discourse connectives as lexical markers that are used to express relations between the parts of the discourse. Similarly, to clarify the concept of *connective*, whose definition is controversial in the literature, Ferrari (Ferrari and Zampese, 2000; Ferrari, 2010) describes a *connective* as "each of the invariable forms [...], which introduce relations that structure "logically" the meanings of the sentence and of the text"[165]. She clarifies that the relations marked by connectives hold between events or assertions, and includes as arguments for the relation also nominalisations (e.g. "after the pressing invitation ...'). On the other hand, she excludes those grammatical elements that introduce relative clauses or pronouns (as *who* in "I don't know who you are.") to be connectives. This is in line with the definition provided for the *arguments* of a *connective* in the PDTB2.0 project, and adopted in Contrast-Ita Bank, for which connectives connect two events, states, and propositions, that can be realised mostly as clauses, nominalisations, anaphoric expressions (Prasad et al., 2007) (See Section 4.1.2).

In her definition, Ferrari (2010) does not include uses of the connectives in which they loose their "logic meaning" (*significato logico* in Ferrari's terminology), that is, they loose their function of introducing a discourse relation. These cases are also known as *discourse markers*. Bazzanella (1995) defines *discourse markers* (*segnali discorsivi*) as those elements that do not entirely lose their meaning, but are used, for instance, to take the turn in a conversation (interactive function) or as indicators of reformulation (metatextual function). *Discourse markers* thus include words/phrases that have not the function of connectives, but are used, for example, to change the topic in a discourse or to initialise it, such as *ma* in Example 5.31 by Ferrari (2010), or to introduce the question that follows as in Example 5.32 from Contrast-Ita Bank.

(5.31) "ma che freddo che fa!" example by Ferrari (2010)
Eng.: Lit.: "(But) how cold it is!"

[165] Original text: "Il termine *connettivo* indica in linguistica ciascuna delle forme invariabili [...], che indicano relazioni che strutturano 'logicamente' i significati della frase e del testo" (Ferrari, 2010).

(5.32) [..] passaporto italiano, quindi straniero, quindi nemico. Ma chi lo spiega ai bambini di Jameela? [doc:5658]

Eng.: [..] Italian passport, therefore foreign, therefore enemy. But who is going to explain this to the children in Jameela?

Notice that this is compatible with the indications in the PDTB 2.0 (Prasad et al., 2007) (presented in Section 4.1.2), for which "*cue phrases* or *discourse markers* [..] have not been annotated since they serve to signal the organisational or focus structure of the discourse, rather than relate Abstract Objects." (Prasad et al., 2007).

Finally, the definition provided by Ferrari (2010) is restrictive, as it does not include variable forms, i.e., those forms which are subject to morphological modifications, such as "it follows that" in Example 5.33 since the verb "to follow" is subject to morphological modifications.

(5.33) From his description of the facts it follows that John was not there when the fire started.

According to Ferrari (2010), connectives belong to different syntactic classes, that are the same defined in the PDTB schema:

- subordinating conjunctions or subordinating expressions;
- coordinating conjunctions or coordinating expressions;
- adverbs or adverbial expressions;
- prepositions or prepositional expressions.

Ferrari (2010) also proposes a non hierarchical classification for connectives depending on the "type of logical relation they convey" that can be e.g. *temporal* and *causal*. [166]In the set of relations she identifies, there are also the relations of *opposition* (or *adversative*, or *of contrast*) and of *concession*[167]. Even if a definition for these types is not provided, we believe that these relations correspond respectively to CONTRAST and the

[166]In the original Italian version, relations are the following: relazioni temporali, relazioni causali, relazioni di consecuzione, di concessione (o limitazione), di condizione, di rielaborazione linguistica o semantica, di opposizione (avversativa, di contrasto, di sostituzione, ecc.), di aggiunta, di dispositio (Ferrari, 2010).

[167]In the original Italian version: relazioni di opposizione (avversativa, di contrasto, di sostituzione, ecc.), relazioni di concessione (o limitazione) (Ferrari, 2010). In a previous proposal, Ferrari and Zampese (2000) consider both the relation of contrast (*di contrasto*) and the relation of concession (*di concessione*) -together with substitution (*di sostituzione*)- subtypes of the relation of opposition (*di opposizione*).

CONCESSION in the PDTB 3.0 (Webber, 2013). In fact, the author proposes for these relations examples 5.34 and 5.35 [168] that we think can be positively tagged as CONTRAST and CONCESSION respectively.

(5.34) mentre i dialetti mediani e (alto) meridionali sono parlati nelle zone anticamente occupate dalle popolazioni italiche, quelli meridionali estremi caratterizzano aree di influenza greca
relazione di opposizione in (Ferrari, 2010) -> CONTRAST
Eng.: while Middle and (upper) south dialect are spoken in the areas which were historically occupied by Italic population, the extreme south ones characterize areas of Greek influence.

(5.35) il loro [dei foni] numero varia da lingua a lingua, ma è comunque sempre ristretto rispetto al numero dei suoni producibili
relazione di concessione in (Ferrari, 2010) -> CONCESSION
Eng.: [phones] number varies from language to language, however it is always limited with respect to the number of sounds one can produce.

She also notices that there is no relation between the syntactic category of the connective and the semantic of the relation it conveys; indeed, she observes that, for instance, the *concession relation* can be conveyed by a subordinating or a coordinating conjunction, or by adverbs.

In line with this definition, Stede (2012) distinguishes discourse connectives as never inflected, closed-class lexical items, which belong to the above mentioned syntactic categories. He also specifies that these lexical elements can only be interpreted successfully when they appear in a relation between two discourse segments.

5.2.2 Lexicons for Connectives

Information on connectives can be found in grammars (e.g. Ferrari and Zampese, 2000), encyclopedias (e.g. Ferrari, 2010), lexical resources (e.g. Sabatini-Coletti, 2005), and other linguistic resources. The PDTB 2.0 (Prasad et al., 2007), for example, certainly represents a useful resource for the study of connectives. In fact, by annotating discourse relations in texts, authors collect a list of connectives for English along with the semantics relations that they convey in the documents. However, the PDTB was not originally conceived as a lexicon for connectives.

[168] Ferrari retrieved the examples from: P. D'Achille, *Breve grammatica storica dell'italiano*, Roma, Carocci, 2001

As a matter of fact, few projects have so far aimed at collecting connectives exhaustively, despite the fact that a resource in which connectives are listed together with the relations they can convey, can be useful for NLP tasks, such as the already mentioned *shallow parsing* [169].

An example of lexicon of connectives, created for being used also in the field of NLP, is DIMLEX for German (Scheffler and Stede, 2016b; Stede and Umbach, 1998; Stede, 2002)[170]. DIMLEX is an XML-encoded resource that provides information on orthographic variants, syntactic category, semantic relations in terms of PDTB3.0 (Webber et al., 2016) sense tags, and usage examples for 274 connectives in German. It is used for automatic discourse parsing, and also for semiautomatic text annotation using the ConAno tool (Stede and Heintze, 2004).

A similar repositories for French is LEXCONN (Roze et al., 2012)[171], which contains more than 300 connectives with their syntactic categories and discourse relations from Segmented Discourse Representation Theory (Asher and Lascarides, 2003). The lexicon has been constructed manually, using a corpus as empirical support.

For the Italian language, we created LICO, Lexicon for Italian COnnectives (Feltracco et al., 2016b) that, according to our knowledge, is the highest coverage resource of discourse connectives available for Italian. We present LICO in the following section.

5.2.3 LICO: A Lexicon for Italian Connectives

To address the lack of a resource of discourse connectives for Italian that can be used for NLP purposes, the author of the thesis and other colleagues created LICO, Lexicon for Italian COnnectives (Feltracco et al., 2016b)[172] which is freely distributed under a CC-BY licence[173].

Inspired by the DimLex project for German (Stede, 2002) and aiming at being compat-

[169] See footnote 119 in the previous Chapter.

[170] `https://github.com/discourse-lab/dimlex/` and `http://connective-lex.info/`

[171] `http://www.linguist.univ-paris-diderot.fr/~croze/D/Lexconn.xml/` and `http://connective-lex.info/`

[172] For this project, the author of this thesis collaborated with (in alphabetic order) Elisabetta Jezek (University of Pavia), Bernardo Magnini (Fondazione Bruno Kessler, Trento) and Manfred Stede (University of Potsdam). Specifically, the author collected the list of connectives, and the information related to them (except the examples and the pointers to entries in lexicon for other languages). We acknowledge Denise Pangrazzi, for her contribution in identifying the Italian equivalents of the German connectives, and proposing an Italian version of the German examples; Felix Dombek (University of Potsdam), for his contribution in aligning LICO entries with DimLex entries (Stede, 2012); Manfred Stede (University of Potsdam), for having stimulated our work on LICO.

[173] `https://hlt-nlp.fbk.eu/technologies/lico` and `http://connective-lex.info/`

ible with it, in LICO connectives are listed together with orthographic, syntactic, semantic information and also possible alignments with lexicon of connectives in other languages. LICO is organized in entries, each one corresponding to a connective and its orthographic or lexical variants. Currently, it counts 173 entries.

5.2.3.1 Retrieving connectives for LICO

In LICO, connectives are listed together with information concerning lexical or orthographic variants. In fact, we partly dropped the invariability criteria proposed by (Ferrari, 2010) and discussed in Section 5.2.1: we did not include forms which exhibit morphological inflection or conjugation, but we did include connectives which show a certain degree of lexical variability that is, multi-word expressions which are not totally rigid from a lexical point of view, e.g. *ad esempio/per esempio* (Eng.'for example').

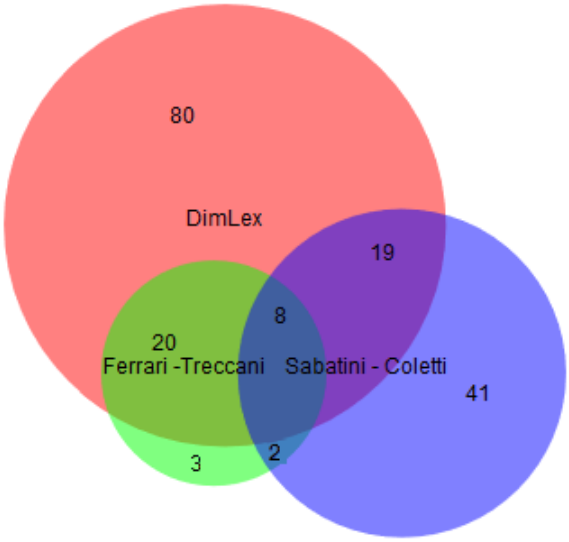

Figure 5.3: Overlap between the resources used to composed the list of connectives in LICO.

In order to compile this list we used a number of grammatical and lexical resources for Italian and for other languages. First, we retrieved the list of connectives mentioned by Ferrari (2010) for the entry *connettivi*[174] for a total of 33 connectives. Then, we retrieved the list of connectives tagged as *congiunzione testuale* in Sabatini Coletti 2006 (Sabatini-Coletti, 2005) discarding the ones of literary use, for a total of 70 entries. Finally, we benefited from the already mentioned DimLex resource for German, as we enriched our

[174]http://www.treccani.it/enciclopedia/connettivi_(Enciclopedia-dell'Italiano)/

list by identifying the equivalent Italian terms of the German connectives[175]. This process was facilitated by the presence of examples in the German resource in which the connective is displayed in context: only the Italian candidates that maintain the sense of the German connectives were added to LICO. We kept trace of this "German-Italian" links (e.g. *aber* → *ma*) and we used this information to enrich the characteristics of the entry in LICO. A total of 127 entries were collected from this third source. Figure 5.3 shows the overlap between the three resources and Table 5.5 shows a sample of the connectives in LICO and the respective sources.

LICO Entries	**Resources**		
	Ferrari Treccani	**Sabatini Coletti**	**DimLex (equivalent)**
dopo	dopo	dopo	dopo
dopo di che dopodiché		dopodiché	dopo di che
dopotutto		dopotutto	
dunque	dunque	dunque	dunque
e	e		e
ebbene		ebbene	
eccetto			eccetto
eppure		eppure	eppure

Table 5.5: Sample of connectives in different resources used to composed the list of connectives in LICO.

5.2.3.2 LICO Structure

For each entry LICO specifies:

- possible orthographic variants: e.g. *ciò nonostante* and *ciononostante* (Eng. 'nevertheless');
- possible lexical variants: e.g *dopo di ché* (Eng. 'thereafter') and *dopo di ciò* (Eng. 'after that'). Notice that in some cases this lexical variants determine a different syntactic environment, such as *in modo da* (Eng. 'in order to') and *in modo che* (Eng. 'so that'), the first being followed by infinitive form, the following by a subjunctive form;

[175] https://github.com/discourse-lab/dimlex

- whether the connective (or its variants) is composed by a single token ("part = single") or by more than one token ("part = phrasal"), e.g. "part = single": *perché* (Eng. 'because'); "part = phrasal": *di conseguenza* (Eng. 'as a consequence');

- whether the connective is composed by correlating part ("orth = discont") or not ("orth = cont") and the specification of the two correlating parts, e.g. "orth = discont": *da una parte*, *dall'altra* (Eng. 'on the one hand', 'on the other hand'); "orth = cont": *perché* (Eng. 'because');

- the syntactic category (as identified by Ferrari (2010) and by Prasad et al. (2007) among others): adverbs, preposition, subordinating or coordinating conjunctions;

- the semantic relation(s) that the connective indicates, according to the PDTB 3.0 schema of relations (Webber et al., 2016; Rehbein et al., 2016) (see Section 4.1.2), as proposed in the DimLex resource (Stede, 2002), which is our main reference resource. In this process we took advantage from the information in the resources we used for building the list. In fact, the DimLex resource provides this information for the German connectives, and both the Italian resources previously mentioned provide useful information about the semantic relation triggered by the connective.[176] A total of 23 different PTDB relations have been used to describe LICO entries;

- examples of the connectives for the semantic relation. The examples in the first version of the resource are translation of the German examples already present in the DimLex resource Scheffler and Stede (2016b); Stede (2002); Stede and Umbach (1998);

- possible alignments with lexicon of connectives in other languages. In the current version entries are aligned with the DimLex (Stede, 2002) entries.

Table 5.6 shows the entry for *quando* (Eng. 'when'), which presents more than one semantic relation, and Table 5.7 shows the entry for *ciononostante, ciò nonostante, nonostante ciò* (Eng. 'nevertheless'), as example of a connective with orthographic variants in LICO.

[176]In particular, in the online version of Sabatini Coletti `http://dizionari.corriere.it/dizionario$_$italiano/D/dizionario.shtml` the semantic relations the connectives can trigger are described in the definition of the connective itself, e.g. "*quindi*, cong. testuale: Con valore deduttivo-conclusivo, perciò, di conseguenza, per questo motivo, dunque". On the other hand, Ferrari (2010) in the Enciclopedia Treccani proposes a non hierarchical classification which we presented in Section 5.2.1.

▷ entry-id	146
▷ orth	cont
▷ part	single
	quando
▷ POS	subordinating
▷ sem relation	TEMPORAL: Synchronous
	ex.: Quando lasciò l'appartamento, arrivò la chiamata
▷ Ref.	DimLex.xml/id="k202"/coh-relation=simultaneous
▷ sem relation	CONTINGENCY:Condition
	ex.: Quando si preme sul bottone, la porta si apre da sola.
	ex.: Quando me lo chiedi, lo lascerò stare.
▷ Ref.	DimLex.xml/id="k295"/coh-relation=condition

Table 5.6: The connective *quando* in LICO.

▷ entry-id	30
▷ orth	cont
▷ part	single
▷ variant	orthographic
	ciononostante
▷ orth	cont
▷ part	phrasal
▷ variant	orthographic
	ciò nonostante
▷ orth	cont
▷ part	phrasal
▷ variant	orthographic
	nonostante ciò
▷ POS	coordinating
▷ sem relation	COMPARISON:Concession:Arg2-as-denier
	ex.: La procura ha ordinato la restituzione dell'esemplare confiscato. Ciononostante l'istruttoria prosegue.
▷ Ref.	DimLex.xml/id="k62", "k78", "k107", "k143"

Table 5.7: The connectives *ciononostante, ciò nonostante, nonostante ciò* in LICO.

5.2.3.3 Inter-annotator Agreement on Semantic Relations

In order to validate the tagging of semantic relations, we conducted a research by observing examples of the use of the connectives in corpora. We intended to verify whether the relation that a connective introduces in a portion of text is one of the relations already tagged for that same connective in the list. In particular, we searched for 20 connectives

in the ItWac corpus (Baroni et al., 2009) and we retrieved occurrences with 400 characters on both sides of the connective. We limited our observation to 5 retrieved segments of text in which the connective is actually playing such a role. We finally tagged each connective in each portion of text with the semantic relation it indicates.

This tagging task was performed by two annotators (one being an expert annotator, the other not). We then calculated the inter-annotator agreement between the two annotators adopting the Dice's coefficient (Rijsbergen, 1979)[177] for three configurations, one for each level of the PDTB relation schema (presented in Section 4.1.2 and reported in 5.8): *class agreement*, *type agreement*, *subtype agreement*.

We considered that there was agreement if both annotators identified exactly the same *class*, *type*, *subtype* respectively. The Dice values result in 0.78 for *class agreement* and 0.71 for both *type agreement* and *subtype agreement*.[178] Results are shown in Table 5.9.

Interestingly, cases of disagreement include the COMPARISON:Contrast relation (on one hand) and the EXPANSION:Substitution relation (on the other hand). For instance, the connective *anziché* 'rather than' in Example 5.36 has been annotated as COMPARISON:Contrast by one annotator and as EXPANSION:Substitution:Arg1-as-subst by another annotator: the first highlighted the contrast between "emissione attraverso il Tesoro" and "usare il tradizionale sistema", the second emphasised that Arg2 represents the alternative to the Arg1.[179]

(5.36) [..] chiedeva l' emissione di dollari in banconote statunitensi attraverso il Tesoro *anziché* usando il tradizionale sistema della Federal Reserve.
Eng.: [..] asked for U.S.- dollar banknotes emission throughout the Treasury *instead of* using the Federal Reserve traditional system.

Another case of disagreement concerns the disagreement between the relations belonging to different classes: TEMPORAL:Asynchronous:precedence (in which Arg2 follows Arg1) and CONTINGENCY:Cause:Result (in which Arg2 is the result of Arg1), being the two strictly connected (i.e., in a cause-effect relation, the effect follows the

[177] As mentioned in Section 5.4, Dice's coefficient measures how similar two sets are by dividing the number of shared elements of the two sets by the total number of elements they are composed by. This produces a value from 1, if both sets share all elements, to 0, if they have no element in common.

[178] Notice the *type agreement* value is very similar to the Contrast-Ita Bank *Senses of relations* agreement of 0.73. In that case, given that the two annotators had recognised a relation of contrast, we measure the agreement in marking it as CONTRAST or CONCESSION; see Section 5.1.5.

[179] Notice that we have similar cases of disagreement in Contrast-Ita Bank, in which one annotator recognised CONCESSION or CONTRAST and the other non marked the relation as such; see Section Example 5.23 discussed in 5.1.5.

I level	II level	III level
CLASS	TYPES	SUBTYPES
TEMPORAL	Synchronous	–
	Asynchronous	Precedence
		Succession
CONTINGENCY	Cause	Reason
		Result
	Condition	Arg1-as-cond
		Arg2-as-cond
	Negative Condition	Arg1-as-negcond
		Arg2-as-negcond
	Purpose	Arg1-as-goal
		Arg2-as-goal
COMPARISON	Contrast	–
	Similarity	–
	Concession	Arg1-as-denier
		Arg2-as-denier
EXPANSION	Conjunction	–
	Disjunction	–
	Equivalence	–
	Instanciation	–
	Level-of-detail	Arg1-as-detail
		Arg2-as-detail
	Substitution	Arg1-as-subst
		Arg2-as-subst
	Exception	Arg1-as-except
		Arg2-as-except
	Manner	Arg1-as-manner
		Arg2-as-manner

Table 5.8: The PDTB 3.0 hierarchy of senses (Webber et al., 2016).

cause). As an example, in 5.37 one annotator marked the connective as indicator of the temporal sequence of Arg1 and Arg2, while the other preferred to mark it as an indicator of the cause-effect relation.

(5.37) Il bello è che i tipi hanno pure accennato a prendersela con me, *al che* io gli ho abbaiato contro una sequela di insulti
Eng.:The best part is that the guys start to get angry with me, *so that* I have bark a series of insult to them.

In general, the relations that were initially assigned to these connectives were confirmed by the corpus-based exercise (i.e., at least one annotator assigned the tag in at

Data: *20 verbs* *5 segments for verb*	
IAA on:	Dice's value
class agreement	0.78
type agreement	0.71
subtype agreement	0.71

Table 5.9: Inter-Annotator Agreement for LICO.

least one portions of text); viceversa, in some cases one of the two annotators assigned a relation that was not initially identified.[180]

5.2.3.4 Contributions, Shortcomings and Further Work

LICO fills a gap with respect to similar resources existing for other languages and it is a freely distributed resource under a CC-BY licence[181]. However it is still under construction under several aspects, which include the completion of the lexical entries with corpus derived examples and the observation of the connectives in Italian corpora. This will allow us to acquire more information about the semantic relations that each connective can indicate and thus extend the annotation of the semantic relations in LICO. These shortcomings are partially addressed for what concerns the connectives of contrast: in fact, as discussed in Section 5.1, the annotation of the contrast relations (i.e., CONTRAST and CONCESSION) in Contrast-Ita Bank included the annotation of the connectives that convey these types of the relation. As we will see in Section 5.2.4, the list of connectives in LICO can thus be integrated with information from Contrast-Ita Bank.

This list of connectives of contrast from Contrast-Ita Bank is the object under discussion in the following section.

5.2.4 Analyzing Connectives of Contrast in Contrast-Ita Bank and LICO

In this section we focus on the connectives that introduce contrast. As announced, one of our aim is to collect an exhaustive list of these connectives by taking advantage of both Contrast-Ita Bank (Section 5.1) and LICO (Section 5.2.3).

[180] These "new" relations are not included in the first available version of LICO.

[181] `https://hlt-nlp.fbk.eu/technologies/lico`

5.2.4.1 Connectives of Contrast in Contrast-Ita Bank

In Section 5.1, we presented Contrast-Ita Bank: a corpus of news for Italian annotated with the contrast relation. The large majority of the relations in Contrast-Ita Bank are explicit relations that are conveyed by a connective (over 90% of the relations, for a total of 341).

In Table 5.10 the list of the connectives ordered by frequency is shown (first column) along with the total number of relations they conveyed in Contrast-Ita Bank (second column) and the relative percentage over the 341 explicit relations annotated in the corpus (third column). The table also indicates the percentage each connective have been tagged with every *sense of relation* tag, for instance *ma* (Eng. 'but) has been tagged with CONTRAST the 4.3% of times.

connective	**#**	**%**	**CONTRAST**	**CONC.Arg1-denier**	**CONC.Arg2-denier**	**CONTRAST-CONC.Arg1-denier**	**CONTRAST-CONC.Arg2-denier**	**CONC.Arg1-denier - CONC.Arg2-denier**
ma	164	48.09	4.3%		87.2%		8.5%	
invece	41	12.02	78%		9.75%	2.5%	9.75%	
mentre	36	10.56	88.9%		2.8%		8.3%	
però	35	10.26	2.9%		85.7%		11.4%	
nonostante	11	3.23		100%				
anche se	10	2.93			90%			10%
e	8	2.35	75%				25%	
se	8	2.35	75%			25%		
eppure	7	2.05		100%				
comunque	4	1.17		100%				
pur	4	1.17		100%				
tuttavia	4	1.17			100%			
a dispetto di	2	0.59		100%				
seppure	2	0.59		100%				
al contrario	1	0.29	100%					
al contrario di	1	0.29	100%					
da una parte.. dall'altra	1	0.29	100%					
in verità	1	0.29					100%	
in realtà	1	0.29		100%				

Table 5.10: List of connectives of contrast in Contrast-Ita Bank along with: total number, percentage over the total cases, percentage of cases per sense tags.

Not surprisingly, *ma* accounts for almost the half of the cases (the equivalent 'but' is also the most used for these senses in the PDTB 2.0), and *invece* (Eng. 'instead'), *mentre* (Eng. 'while'), *però* (Eng. 'but') for about a 10% each one, followed by *nonostante* (Eng. 'nevertheless') and *anche se* (Eng. 'even if') with 11 and 10 cases respectively. Other connectives, are more rare and appear less than 2.5%.

Table 5.10 shows that, as it happens for content words, the most frequent connectives are the most polysemous ones even if with a general preference. For example, *invece* and *mentre* are mostly used to convey CONTRAST, *ma* and *anche se* are mostly used for CONCESSION.Arg2-as-denier. On the other hand, less frequent connectives are less ambiguous: e.g. *eppure*, *comunque*, *pur* (Eng. 'yet', 'however', 'even if') appear respectively 7, 4 and 4 times and are all marked as CONCESSION.Arg1-as-denier, while *al contrario*, *al contrario di*, *da una parte [..] dall'altra* (Eng. 'conversely', 'contrary to', 'on the one hand [..] on the other hand') seem to convey essentially CONTRAST.

As we mentioned in Section 5.1.4, when discussing data about the sense types (CONTRAST or CONCESSION), CONCESSION.Arg1-as-denier is not very frequent; however, compared to the other sense, this sub-type seems to be marked by a numerous but defined set of connectives, e.g. *nonostante*, *eppure*, *comunque*, *seppure*.

Another interesting observation regards the syntactic nature of the connectives. Data seem to confirm that in general there is no relation between the syntactic category of a connective and the relation it conveys (see definition by Ferrari (2010) in Section 5.2.1): in fact, different syntactic categories are found for both CONTRAST and CONCESSION, e.g the coordinating conjunction *ma* and the adverbial expression *mentre*. However, we can observe that no subordinate conjunction is used for conveying CONTRAST while they are used for conveying CONCESSION (cf. *nonostante*, *pur*, *seppure*, *anche se*).

By analysing the data, we can also make observations concerning individual connectives. For instance the conjunction *e* (Eng. 'and') is not very frequent as a connective of contrast but it constitutes one of the main cause for the disagreement presented in Section 5.1.5. *E* is in fact notoriously highly ambiguous: it can express temporal relation ("Ho mangiato *e* dormito"), for presenting alternatives ("Mangia *e* non dormire"), for parallelism ("Ho mangiato *e* anche Luisa"), cause-effect relation ("Stava piovendo *e* ho aperto l'ombrello"). Another interesting case concerns the conjunction *se* (Eng. 'if') that is always registered in its coordinating use and not as subordinating element; it is often

found at the beginning of sentences, selecting as arguments clauses which are in the same sentence (instead, no specific position has been registered for the other connectives), and it is often associated with other elements that help the interpretation of the CONTRAST parallelism such as: Se *prima [..], poi [..]*, *Se da una parte [..], dall'altra [..]* (Eng. 'If before [..], then [..]', 'If on the one hand [..], on the other [..]').

5.2.4.2 Connectives of Contrast: False Positives

We said before that, given the high number of explicit contrast relations, the performances of a system that detects contrast by relying of this information are expected to have a high *recall*[182]; in other words, we expect the system to find many cases. However, it might not register a high *precision*, tagging as connectives of contrast, lexical elements that are not such. Indeed, the list of connectives which we presented in the previous section corresponds to a list of textual markers that can potentially signal the presence of a contrast relation, but actually they do not serve as connectives of contrast all the times. For instance, the word *ma* (Eng. 'but') can also be used as a cue element for introducing an exclamation (as we have seen for Example 5.31 in Section 5.2.1) and in Example 5.38, *mentre* (Eng. 'while') does not convey a relation of contrast:

(5.38) L'elicottero dell' Aiut Alpin è intervenuto a Passo Fedaia per il soccorso ad un motociclista germanico caduto in moto mentre dal Passo stava dirigendosi a Canazei. [doc:5901]
Eng.: The Aiut Alpin helicopter intervened at Fedaia Pass to rescue a Germanic motorcyclist that has fallen from his moto while he was driving from Pass towards Canazei.

In order to collect more information about the connectives in Contrast Ita-Bank (in Table 5.10), we annotated in the same corpus also cases in which the most common connectives do not serve as such. More practically, we annotated all the instances of the five most frequent connectives of contrast in the corpus distinguishing cases in which the lexical unit is an explicit connective of contrast (see Sections 5.2.4.1 and 5.1.4) and cases in which it is not. We are calling these latter cases "False Positives". Specifically, they are cases of in which *ma* (Eng.'but'), *invece* (Eng.'instead'), *mentre* (Eng.'while'), *però* (Eng.'but'), *nonostante* (Eng.'nevertheless'):

[182] As we mentioned in Footnote 31, *precision* is the number of correct positive results divided by the number of all positive results, and *recall* is the number of correct positive results divided by the number of positive results that should have been returned.

- do not serve as discourse connectives, since they are used, for instance, as a cue elements for signalling a change in the topic or to introduce an exclamation (*False positive - type 1*);
- do not serve as discourse connectives, since they connect two PP or NP (*False positive - type 2*);
- convey a discourse relation other that contrast (*False positive - type 3*)

The results of this analysis are reported in Table 5.11 and are described below.

connective	**#**	**%**	Type 1 Other function	Type 2 Not discourse relation	Type 3 Non contrast	Type 2 - 3
ma	164	48.09	25	35	-	14
invece	41	12.02	8	1	-	-
mentre	36	10.56	-	-	12	-
però	35	10.26	3	-	-	-
nonostante	11	3.23	-	-	-	-

Table 5.11: Data on False positives.

False positive - type 1. This group includes lexical elements that are called *discourse markers* (Bazzanella, 1995; Ferrari, 2010), or *cue phrases* or *pragmatic markers* in the PDTB2.0 and "serve to signal the organizational or focus structure of the discourse, rather than relate AOs."(Prasad et al., 2007, p. 8). (See Section 5.2.1). We annotated 25 cases of *ma* of this kind, 8 for *però*, 3 for *mentre*. An example of this type is provided in Example 5.32, here repeated in 5.39.

(5.39) [..] passaporto italiano, quindi straniero, quindi nemico. Ma chi lo spiega ai bambini di Jameela? [doc:5658]
Eng.: [..] Italian passport, therefore foreign, therefore enemy. But who is going to explain this to the children in Jameela?

In the example, *ma* does not convey a discourse relation of contrast, but it is used to introduce the question that follows in which the topic slightly changes.

False positive - type 2. This group includes cases in which the lexical element we are analysing connects, for instance, two PP or NP and not two clauses (as required in the

PDTB2.0)[183]. We annotated 35 cases for *ma* and one for *mentre*. An example in which *ma* coordinates two adjectival phrases is reported in 5.40:

(5.40) Sarà un consiglio di amministrazione denso di argomenti ma con poche decisioni definitive quello della Trentino Spa che si annuncia per questo pomeriggio [doc:5712]
Eng.: This afternoon Board of Directors of Trentino Spa will be dense of topics but with few definitive decisions. [doc:5712]

False positive - type 3. We planned the annotation of False positive of type 3 (i.e., connective that convey relations other than contrast) in order to study the polysemy of connectives of contrast. We register 12 cases conveyed by *mentre*, in which the conjunction indicates a temporal relation (of type TEMPORAL.Synchronous according to the PDTB). An example has been presented in 5.38 and is reported here.

(5.41) L'elicottero dell' Aiut Alpin è intervenuto a Passo Fedaia per il soccorso ad un motociclista germanico caduto in moto mentre dal Passo stava dirigendosi a Canazei. [doc:5901]
Eng.: The Aiut Alpin helicopter intervened at Fedaia Pass to rescue a Germanic motorcyclist that has fallen from his moto while he was driving from Pass towards Canazei.

False positive - type 2 and 3 Other 14 cases concerns the conjunction *ma* in which, according to our guidelines, the conjunction does not convey contrast and coordinates two elements that are not arguments of a discourse relation. An example is reported in 5.42. In this case *ma* has been interpreted as connective that conjoins nouns which are not in contrast. In all these cases, *ma* it is followed by *anche* (Eng.'also').

(5.42) [..] indica per ora solo l' etnia di alcuni (ceceni, ingusci, ma anche kazaki e tatari) [doc:5381]
Eng.: [..] for the moment it indicates just the ethnicity of some (Chechens, Ingusits, but also Kazakhs and Tatars)

Observing Table 5.10 and Table5.11, it is possible to notice that *ma* is the most controversial one. Other than being the most frequent connectives in the corpus (48.09% of the cases), it also the most frequent False Positive. The most polisemous is *mentre* that

[183]We remind the reader that we mainly followed the PDTB 2.0 Annotation Manual (Prandi, 2017) for the annotation of Contrast-Ita Bank.

in 36 cases is a connectives of contrast and for 12 is of temporal[184]. On the other hand, *nonostante* always serves as a connective of contrast.

5.2.4.3 Adding Data on Connectives from Contrast-Ita Bank into LICO

Contrast-Ita Bank represents a compatible resource to the lexicon LICO (Feltracco et al., 2016b) that we presented in Section 5.2.3, and provides it two benefits. The first concerns the enrichment of the list of connectives of contrast in LICO. Table 5.12 reports the list of the 19 connectives marked as CONTRAST, CONCESSION.Arg1-as-denier and CONCESSION.Arg2-as-denier in Contrast-Ita Bank (second column) and the list of the 38 connectives marked with the same sense tags in LICO (third column).

connective	Contrast-Ita Bank	LICO	**connective**	Contrast-Ita Bank	LICO
a dire il vero		x	invece	x	x
a dispetto di	x	x	ma	x	x
a onor del vero		x	malgrado		x
ad ogni modo		x	malgrado ciò		x
al contrario	x	x	mentre	x	x
al contrario di	x		nondimeno		x
anche		x	nonostante	x	x
anche se	x	x	per contro		x
benché		x	per quanto		x
bensì		x	però	x	x
ciononostante		x	persino		x
cionondimeno		x	pur / pure	x	x
comunque	x	x	quantunque		x
con tutto questo		x	se	x	
contrariamente a		x	sebbene		x
da un canto .. dall'altro		x	sennonché		x
da un lato .. dall'altro lato		x	seppure	x	
da una parte .. dall'altra parte	x	x	solamente		x
e	x		solo		x
eppure	x	x	tuttavia	x	x
in verità (a dire la verità)	x	x	viceversa		x
in realtà	x		Total	19	38

Table 5.12: Connectives of contrast in Contrast-Ita Bank and in LICO.

[184]This is different from the behaviour of *while* in the PDTB2.0 (Prasad et al., 2007) which is mostly found as a connective of the subtype *TEMPORAL:synchrony*. However according to Asr and Demberg (2012), it is also a reliable cue for *CONTRAST:opposition*, since it is seen in many instances of this relation relative to its total number of occurrences.

As can be seen from the table, 14 out of the 19 connectives of contrast in Contrast-Ita Bank are also tagged in LICO. The other 5 connectives, which are not in LICO, are: *e* (Eng.'and'), which in LICO is tagged with EXPANSION:*Conjunction* and TEMPORAL:*Asynchronous*:precedence senses; *in realtà* (Eng.'actually'), tagged as EXPANSION:*Specification*:Arg2-as-detail; *se* (Eng.'if'), tagged as a connective of the relation CONTINGENCY:*Condition*; and *seppure* (Eng.'altough') and *al contrario di* (Eng. 'contrary to') which are not present in the resource. The information we collected from Contrast-Ita Bank can be added to LICO in a new version of the resource. Moreover, we believe that the result of this merging represent also an exhaustive list of connectives of contrast for Italian.

The second important benefit of considering Contrast-Ita Bank for the enrichment of LICO is the access to a list of corpus examples[185]; in fact, as we mentioned in Section 5.2.3, the examples in the first version of the resource are not corpus derived.[186] Through these corpus examples we can observe, for instance, whether a connective is used at the beginning of a sentence, or if it requires the following verb to be in a conjunctive form. Another advantage concerns the information of how the connective is used with reference to the relation it conveys: in Contrast-Ita Bank, each connective is annotated with its arguments. This information can be inserted in LICO, enriching the structure of the lexicon to account for this (i.e., inserting the specification of <arg1> or <arg2> in the example item).

It should be notice, however, that the integration of Contrast-Ita Bank can provide reliable data for just some of the connectives listed in LICO (i.e., LICO registers a total of 38 connectives of contrast, 24 of which were not found in Constrast-Ita Bank).

5.2.5 Contributions, Shortcomings and Further Work

From Contrast-Ita we collected 19 connectives of contrast: some are very frequent (*ma* accounts for almost the half of the cases), while others are more rare and appear just once in the corpus. The most frequent connectives are the most polysemous ones even if

[185]This is a work in progress. In particular, we take advantage of the data annotated in Contrast-Ita (the connectives and their arguments) and pick additional examples from a larger version of the corpus from which Contrast-Ita Bank is derived. The work has been submitted for presentation and publication to The Eleventh International Conference on Language Resources and Evaluation (LREC 2018). See (Feltracco et al., 2018).

[186]The examples are translations of German examples of the DimLex resource (Scheffler and Stede, 2016b; Stede, 2002; Stede and Umbach, 1998).

they appear more frequently with a relation with respect to another, while less frequent connectives are less ambiguous. We also show that, in general, there is no relationship between the syntactic category of a connective and the relation it conveys. A detailed analysis of the five most frequent connectives shows that *ma* is also frequently used as a non-connective and that *mentre* is used in a temporal sense for the 25% of times it appears. This type of analysis can be extended to all the connectives.

In this same section we presented LICO, the Lexicon of Italian COnnectives, a resource for connectives in which the reported information is derived from other resources. In LICO, 38 connectives of contrast (i.e., tagged with CONTRAST or CONCESSION sense tags) are listed. While LICO is very important because it provided a complete list of connectives for Italian, we believe that the merging of the information from Contrast-Ita Bank would provide LICO important benefits for what concerns the enrichment and the validation of the list of 38 connectives found in LICO (of the 19 connectives tagged in Contrast-Ita Bank, 5 are not in LICO) and for adding a number of corpus examples to the resource. This highlights the importance of a corpus investigation in enriching a lexical resource.

Chapter 6

Lexical Opposition and Discourse Contrast: Co-occurrence and Representation

In this chapter we focus on the connection between lexical opposition and discourse contrast. In particular, we investigate the co-occurrence of the two linguistic phenomena in a corpus. We also propose an exploratory research about how the co-occurrence of opposition and contrast can be formally represented.

In this chapter we present our investigation on the co-occurrence of opposition and contrast. For instance, in example 6.1, a pair of opposites (double underlined) appears in the context of a contrast relation (whose arguments are enclosed in square brackets).

(6.1) [The price of this book increased], while [the price of that one decreased.']

Through the example we can see how opposition participates in contrast; in fact the opposites *to increase / to decrease* convey the difference for which the two mentioned entities (i.e., the books) are compared, leading to a contrast relation.

Our analysis starts with the observation that both linguistic phenomena involve two elements that are similar in many aspects, but that differ in others. We will refer to studies that highlight this co-occurrence both in the theoretic literature and in the computational field, in which opposition is used as a feature for identifying contrast, and viceversa (Section 6.1). In Section 6.2, we investigate the opposition-contrast intersection adopting a corpus-based approach; in particular, we observe how frequently opposites are found in the arguments of a contrast relation in Contrast-Ita Bank, a corpus introduced in Section

5.1. In addition, we study the cases in which the two phenomena co-occur, in order to understand the contribution of opposition to discourse contrast; we run an inter-annotator agreement and an analysis of its results. Finally, in Section 6.3, we present our proposal for a formal representation of the opposition-contrast co-occurrence by comparing a format based on the PDTB schema of annotation to a format based on the Abstract Meaning Representation (a semantic representation language). In particular, we attempt to represent Italian sentences, in which opposites appear in the arguments of a contrast relation, with the two formats, discussing also the limits we observe.

6.1 Lexical Opposition in Discourse Contrast

As we have seen in Chapter 2, according to Cruse opposition indicates the relation in which *two terms differ along only one dimension of meaning: in respect to all other features, they are identical* (Cruse, 1986, p.197). This definition is consistent with those proposed in RST and in SDRT frameworks for the *contrast* relation and in the PDTB2.0 (Prasad et al., 2007) for the sense tag CONTRAST (Section 4.2). In fact, in these theories, the relation of contrast captures cases in which the arguments in the relation *have some aspects in common* (Mann and Thompson, 1988; Carlson and Marcu, 2001), *share a property* (Prasad et al., 2007), or *have a similar structure* (Asher, 1993), but they *differ in some respect* (i.e., *some values of the property* (Prasad et al., 2007), or *contrasting themes* (Asher, 1993)) and *are compared with respect to these differences*. Both relations thus involve comparing two elements that are similar in many aspects, but that differ in others: lexical opposition captures contrast at the lexical level, while discourse contrast captures it at the discourse level.

In the previous chapters, we presented the two phenomena as distinct, but they can co-occur; in particular, opposition represents a communicative strategy for the creation of a discourse contrast. The use of opposition in order to convey contrast is, for instance, mentioned by Lakoff (1971) (see Section 4.2.1), when describing the *semantic opposition* use of 'but'. According to Lakoff, in *semantic opposition* the elements conjoined are "*opposed* to each other in a particular property" (Lakoff, 1971, p. 133): she provides the example reported in 6.2 in which the difference (or *opposition*) is "apparently present overtly, in two of the lexical items in each sentence, which form a pair of antonyms" (Lakoff, 1971, p. 133) (i.e., *tall* and *short*):

(6.2) John is tall but Bill is short. *Semantic opposition* in Lakoff (1971, p. 133)

In other words, Lakoff (1971) notices that opposition (*antonymy* in her terms) is a source of contrast.

Also the Penn Discourse Treebank (Prasad et al., 2007) acknowledges lexical opposition participates in CONTRAST. As reported in Section 4.2.1, the PDTB2.0 schema defines two subtypes of CONTRAST[187], depending on whether the *contrast* is created by alternative values (i.e., *juxtaposition*) or by antonym values (i.e., *opposition*), thus recognizing the importance of (a type of) lexical opposition as a feature for contrast. For the *opposition* type, the author propose the example reported in 6.3, in which *fell* and *moved higher* are antonyms:

(6.3) *Most bond prices fell* [..]. **Junk bond prices moved higher**, however.
Contrast.opposition in PDTB (Prasad et al., 2007, p.32)

In the area of NLP, the co-occurrence of the two phenomena has been considered, for instance, by Roth and Im Walde (2014), which use what they call *discourse markers* for distinguishing paradigmatic relations, including opposition (see Section 2.2).

Other contributions in the same area use lexical opposition as feature for detecting contrast. As an example, Harabagiu et al. (2006) base the identification of contrast on the opposition relation (antonyms in their terms) given that in some examples "[..] the presence of opposing information contributes more to the assessment of a CONTRAST than the presence of a cue phrase" (Harabagiu et al., 2006).[188]

The lexical-discourse intersection is exploited also by Marcu and Echihabi (2002). They create a system to identify implicit relations (e.g., an implicit relation of contrast) under the hypothesis that some lexical item pairs can "provide clues about the discourse relations that hold between the text span in which the lexical items occur".[189]

In a cross-lingual evaluation for English and Swedish, Murphy et al. (2009) show that opposites (*antonyms* in their terminology) are used for different functions: the most common is the one of "creat[ing] or highlight[ing] a secondary contrast within the sentence/discourse" (they call this the *ancillary function*).

[187]The distinction is however not confirmed in the PDTB 3.0 schema (Webber et al., 2016).

[188]They describe their effort for recognizing contradictions between texts by deriving linguistic information that identifies negations, contrasts, or oppositions.

[189]They however do not use lexical pairs judged as opposites *a priori* but collect them by automatically searching for specific lexical units, e.g. 'but', in a corpora; thus the words in a pair may in the end not correspond to the definition of opposites.

On the contrary, Spenader and Stulp (2007) give evidence that opposition is not a strong feature for contrast. In particular, they evaluate whether adjectival opposition is actually a *source of contrast* (i.e., how frequently opposition contributes to contrast) or a *predictor of contrast* (i.e., how frequently opposition is a cue for contrast). They calculate the co-occurrence of opposite adjectives in the contrast relations marked or non-marked by *but* in a corpus. They collect the list of opposites by retrieving the list of direct and indirect antonyms from WordNet 2.0[190]. They then show that opposition is not common in cases of explicit contrast conveyed by *but* (1% of co-occurrence among the sentences observed, of which contrast is licensed in 81% of the cases) (*opposition as a source of contrast*), and it is also not that frequent in cases of non-*but* marked contrast (opposition conveys contrast in 15% of the cases) (*opposition as a predictor of contrast*). In a similar way, we intend to evaluate whether opposition is a key feature for contrast.

6.2 Co-occurrence of Opposition and Contrast in Contrast-Ita Bank

We aim at providing some insights of the role of lexical opposition in creating contrast. Specifically, we intend to observe the co-occurrence of the two linguistic phenomena analyzing cases in which pairs of opposites (such as *to increase* / *to decrease*) are used in the context of a contrast relation, both in the presence and in the absence of a connective. We perform this investigation in the Contrast-Ita Bank corpus.

6.2.1 Annotating Opposition in Contrast-Ita Bank

In order to verify the co-occurrence of opposition and contrast, we calculate how frequently opposites are found in the arguments of a contrast relation in Contrast-Ita Bank. Technically, we manually annotated the opposites[191] by marking *opposite1* and *opposite2* when the former is part of *Arg1* and the latter is part of *Arg2*.

In this exercises, we do not limit our annotation to "good", prototypical opposites[192] or to pairs of mono-token (nuclear) words, but we manually mark also broader expressions, including cases similar to Example 6.4, in which opposites are doubly underlined.

[190] See Section 2.3 for a definition of *direct* and *indirect* antonyms in WordNet.

[191] The annotation was carried on by the author of this thesis.

[192] See Section 2.1.1 and Cruse (1986, p.262).

(6.4) *Il ministro del Lavoro e delle Pensioni britannico, Andrew Smith, ha rassegnato ieri le dimissioni* nonostante **i tentativi del premier Tony Blair di convincerlo a rimanere**. [doc:5397]
CONCESSION.Arg1.as.denier
Eng.: *British Secretary of State for Work and Pensions Andrew Smith resigned yesterday,* despite **Prime Minister Tony Blair's attempts to persuade him to stay.**

In the example, the light-verb construction *rassegnare le dimissioni* (Eng.'to resign') is considered as the opposite of *rimanere* (Eng.'to remain') and the two are found respectively in *Arg1*, conventionally reported in italics, and *Arg2*, conventionally reported in bold.

Figure 6.1 shows the annotation of the example in the CAT tool (Bartalesi Lenzi et al., 2012). In the figure, opposites are annotated in green, the connective in purple, *Arg1* in yellow, and *Arg2* in blue. Portions of text labeled as opposites (i.e., *ha rassegnato le dimissioni* and *rimanere*) appear in bold to signal that they are annotated with two layers: the contrast layer (*Arg1* or *Arg2* label) and the opposition layer (*opposite1* or *opposite2* label).

Figure 6.1: A screen-shot of the CAT tool used for annotating opposition in the contest of contrast.

We do not rely on any automatic procedure for the extraction of the opposites and we include in the annotation also 'opposites in context', as it happens in Example 6.5. In the example *perde* and *sale* are semantically opposite in the specific context they are used (i.e., Eng. 'to rise of x', 'to fall by x'), even if they are not typically marked as opposites in lexical resources [193].

[193]For example, they are not registered as opposite in the Dizionario dei Sinonimi e dei Contrari - Rizzoli Editore, http://dizionari.corriere.it/dizionario_sinonimi_contrari

(6.5) *Sul Nuovo Mercato, Tiscali perde lo 0.05% a 2,23*, **E. Biscom sale dell'1,09% a 41,44**. [doc:5436]
- implicit = invece
CONTRAST
Eng.: *On the New Market, Tiscali looses 0.05% to 2.23*, **E. Biscom rises by 1.09% to 41.44**.

It might be claimed that these 'contextual opposites' lack one of the requirement described by Cruse (1986, 2011) for defining contrast: *inherently binary incompatibility* (i.e., opposites are 'logically' two incompatible elements). We believe this is not the case. As we have seen in the dedicated Section 2.1.1, Cruse counterposes *inheritance* to *accidentality*, providing the example of *tea* and *coffee* (Cruse, 2011, p.144) that can accidentally be the only drinks offered in a meeting; they are not opposite since they are not the only *logical* choices for drinks. On the contrary, *up / down* are opposites since they are logically the only two possible movements in a linear axis. Similarly, in the proposed Example 6.5, *perdere / salire* are not *accidental*: *perdere* in the sense of 'loosing (some value)' and *salire* in the sense of 'increasing (of some value)' are the only two logical possibilities of movement in a scale of 'changing values'. In this sense, they can be considered (*contextual*) opposites.[194]

6.2.2 Results of the Annotation

We identified a total of 20 cases in which opposites are in the arguments of a contrast relation in Contrast-Ita Bank.[195] We find out that lexical opposition is present in the context of different parts of speech. For instance, in Examples 6.4 and 6.5, opposition occurs between the predicate of the arguments, in Example 6.6 it occurs between the two adjectives *ricco* and *povero*, and in Example 6.7 between the two nouns *rialzo* and *calo*.

(6.6) [..] *uno dei due è ricco di cellule staminali*, **l'altro ne è povero**. [doc:5402]
- implicit = invece CONTRAST
Eng.: *[..] one is rich in stem cells*, **the other is poor of them.**

[194]This can also be considered as a case of *modulation* (see Section 2.1); the sense of the verb *perdere* ('to lose') is enhanced for the trait 'having no more or having less of (something)' and it is thus opposite of *salire* ('to increase') in its trait of 'having more of (some value)'.

[195]The annotation was not an easy operation considering that we are not relying on a pre-existing list of opposites and we are including also 'opposite in context' in the annotation. We report here the data for which we are more confident, leaving the discussion about controversial cases to a further annotation exercises involving multiple annotators.

(6.7) *..le Generali hanno chiuso con un rialzo dello 0,59%*, **le Ras con un calo dello 0,31%**. [doc:5704]
- implicit: invece CONTRAST
Eng.: *Generali closed with an increase of 0.59%*, **(the) Ras with a loss of 0.31%.**

Table 6.1 shows interesting data concerning the distribution of the 20 cases of *lexical opposition*. Opposition is present both in *explicit* and *implicit* relations, and both when CONTRAST or CONCESSION have been marked (see Examples 6.4 and 6.5). However, lexical opposition co-occurs in 18% of the cases of *implicit* relations, especially in those marked as CONTRAST. Relations of type CONTRAST are in general the ones in which there is a higher number of opposites in the arguments (4 explicit, 4 implicit): these 8 cases represent 7.8% of all the cases of CONTRAST (i.e., 8 over 102 cases).

Types of relations / Senses of relations	Explicit	Implicit
CONTRAST	4	4
CONCESSION	6	0
CONTRAST - CONCESSION	5	1
Total	15	5
over #explicit / implicit	4%	18%

Table 6.1: Lexical opposition in the context of contrast in Contrast-Ita Bank.

Overall, the limited number of opposition identified (20 cases over 327 cases of contrast, 5.3%) seems to suggest that the presence of opposites is not an impacting feature for the identification of contrast relation in the Italian language. It is, however, quite frequent for *implicit* relations, suggesting that the use of opposition can be a strategy to convey contrast when there is a lack of an explicit connective that lexicalizes the relation.

We conducted a deeper investigation in order to evaluate whether the opposites in the arguments of a contrast relation actually contribute to it. We performed an inter annotator agreement exercises among two annotators in order to understand whether to distinguish cases in which the opposites in the arguments of a contrast relation contribute in conveying the discourse relation (and cases in which they do not) is an easy operation. One annotator is the author of this thesis, the second annotator, who has some familiarity with linguistic tasks, was provided with simple oral instructions through which we ask her to

judge the contribution of the opposites when in the context of a contrast relation.[196] We calculated the IAA with the Dice's coefficient (Rijsbergen, 1979)[197] obtaining a value of 85% (disagreement in 3 cases out of 20). After a reconciliation step (in which annotators compared their annotations, and could revise their decisions) the value increased to 95%: one case was solved due to a rethinking, the other was due to a misunderstanding of the portion of text to observe.

In general, the IAA results show that to judge whether opposites contribute to contrast seems not to be controversial operation, even if some disagreement remains. Moreover, we register that out of the 20 cases we annotated, in 17 both annotators agree in considering opposites as crucial for the contrast relation, in 2 they agree in judging it as non affecting the contrast relation and in 1 we registered disagreement. In conclusion, it seems that when opposites appear in the context of a contrast relation they actually participate to the discourse contrast phenomena.

We propose an analysis of a selection of the 20 cases judged by the two annotators in light of the definition of contrast that we followed through this thesis. In some cases, the contrast relation holds because two entities that share a property are compared with respect to different values that this property takes.[198] We register that the different values can be expressed through opposites. Example 6.6 is one of these cases. Figure 6.2 below reports our attempt to schematize Example 6.6 figuring out which are the entities compared, the properties they share, and the values assumed by those properties. At the same time, we made an effort to keep in the representation the lexical opposite units. In the figure, we indicated where the contrast lies and opposites are reported double underlined.

[cordone ombelicale].. *uno dei due è ricco di cellule staminali*, **l'altro ne è povero**.
Eng.: *[umbilical cord].. one is rich in stem cells*, **the other is poor of them.**

ENTITY: UMBILICAL_CORD_1		ENTITY: UMBILICAL_CORD_2
▷ PROPERTY: STEM_CELLS_LEVEL		▷ PROPERTY: STEM_CELLS_LEVEL
▷ VALUE: RICH	⟷ CONTRAST	▷ VALUE: POOR

Figure 6.2: Analysis of opposition in the context of contrast. Example 1.

[196] We acknowledge Enrica Troiano for collaborating as second annotator.

[197] For an explanation of the measure, see Section 3.2.2.1.

[198] The format Entities-Properties-Values is common to the tradition in Knowledge Representation in the field of Artificial Intelligence.

The figure helps us in showing that the compared values that trigger the contrast are the two opposites *rich* and *poor*. Opposition is thus central for the discourse contrast in the text in the example.

Other examples includes case in which the contrast relation stem from a comparison between two values of a property assigned to the same entity, as happens for the example in Figure 6.3.

Il commercialista *doveva essere il cavaliere bianco chiamato a salvare la Chini* e, invece, **è stato quello che l' ha affossata**.
Eng.: The accountant should be the white knight designated to save the Chini and, on the contrary, **he has been the one that ruined it**.

ENTITY: ACCOUNTANT

▷ PROPERTY: ROLE	⟷	▷ PROPERTY: ROLE
▷ VALUE: TO SAVE	CONTRAST	▷ VALUE: TO RUIN

Figure 6.3: Analysis of opposition in the context of contrast. Example 2.

In the example, the contrast arises from the comparison between the opposite roles of the participant: *to save (something) / to ruin (something)* .

Annotators agree in recognizing that in this kind of examples opposition is relevant for the origination of a contrast relation. This is not the case in Example 6.4, for which annotators agree in judging opposites not as a source for the discourse contrast relation.

A dispetto degli **sforzi della pubblica amministrazione [..],** *gli investimenti privati* in termini di *istruzione sono ancora bassi.*
Eng.: Despite **public administration efforts [..],** *private investments in terms of education are still low*

ENTITY: INVESTMENT_1		ENTITY: INVESTMENT_2
▷ PROPERTY: NATURE_OF_THE_EFFORT		▷ PROPERTY: NATURE_OF_THE_EFFORT
▷ VALUE: UNDEFINED		▷ VALUE:MONEY
▷ PROPERTY: INVESTMENT_SOURCE		▷ PROPERTY: INVESTMENT_SOURCE
▷ VALUE: PUBLIC ADMINISTRATION		▷ VALUE: PRIVATE SECTOR
▷ PROPERTY: LEVEL	⟷	▷ PROPERTY: LEVEL
▷ VALUE: POSITIVE	CONTRAST	▷ VALUE:LOW, NEGATIVE

Figure 6.4: Analysis of opposition in the context of contrast. Example 3.

In the example, the opposite adjectives *pubblico / privato* (Eng. 'private / public')

are two different values for the property INVESTMENT_SOURCE of the two entities. Although one can say that the participants do have opposite characteristics, the contrast relation does not stem from this opposite property. Rather, it is based on the comparisons between the 'positive efforts' on the one hand and the 'low investments' on the other hand. Recalling the PDTB definition, we can say that "the difference between the two situations described in the arguments is highlighted with respect to the values assigned to this property" (Prasad et al., 2007). We believe that in this case the property of being 'private' or 'public' is not 'highlighted', and thus the opposites are not at the center of the discourse contrast.

One case of disagreement remained and it is reported and analyzed in Figure 6.5. In the example, one annotator (*ann.1* in the figure) considered that the contrast among the two situations in the arguments originates from the different value assigned to the property STARTING_DATE of the two timetables for different services (i.e., the date of their entering into force), and that it does not stem from the comparison of the opposites *suburban / urban* (similarly to what happens in Example 6.4). Conversely, the other annotator (*ann.2* in the figure) considered that also these values of the property SERVICE are highlighted as contrasting and recognized the opposites *suburban / urban* as sources of the resulting discourse contrast.

A decorrere da domenica 12 entra in vigore il nuovo orario invernale per il servizio extraurbano e la Trento - Malè. **Da lunedì 13 entra** invece **in vigore il nuovo orario invernale 2004 / 2005 per il servizio urbano di Trento e Rovereto**

Eng.: *Starting from Sunday 12 the new winter timetable for the suburban service and for the Trento - Malè enters into force.* **From Monday 13** instead**the new winter timetable 2004 / 2005 for the urban service of Trento and Rovereto enters into force**.

ENTITY: WINTER TIMETABLE_1		ENTITY: WINTER TIMETABLE_2
▷ PROPERTY: STARTING_DATE	⟷	▷ PROPERTY: STARTING_DATE
▷ VALUE: SUNDAY_12	CONTRAST ann.1 and ann.2	▷ VALUE: MONDAY_13
▷ PROPERTY: SERVICE		▷ PROPERTY: SERVICE
▷ VALUE: SUBURBAN; TRENTO-MALÈ	⟷ CONTRAST ann.2	▷ VALUE: URBAN_TRENTO URBAN-ROVERETO

Figure 6.5: Analysis of opposition in the context of contrast. Example 4.

The case of disagreement thus stems from the identification of what is compared and contrasted in the relation, but not in the definition of the two entities and their properties. Moreover, as we anticipated, this is the only case of disagreement we register in our annotation exercise.

6.2.3 Contributions, Shortcomings and Further Work

Through the annotation of opposites in the arguments of a contrast relation in Contrast-Ita Bank, we provided evidence that lexical opposition is not a key feature for the CONCESSION relation and that it is more frequent in cases of CONTRAST; in addition, the presence of opposites is of limited frequency in cases of *explicit* relations, and it is (proportionally) more frequent in cases of *implicit* relation. This is in line with the conclusion drawn by Spenader and Stulp (2007), who claim that opposition is not a strong *source of contrast*.

Moreover, limited to the 20 cases we observed, we performed an inter-annotator agreement exercise and conducted an analysis in terms of contrasted entities, properties and values, in order to understand the contribution of opposites when in the context of a contrast relation: we registered that if opposites appear in a contrast relation they frequently participate to it.

Finally, we have seen that opposition, depending on the context, can hold among pairs of lexical units that are not conventionally intended as opposites. The acquisition of these data was only possible through the corpus investigation we run, and it represents an important result of the data-driven methodology we adopt in this thesis.

Future work includes the annotation of other discourse relations (e.g. temporal, causal) and of opposition in the context of these relations in order to understand whether opposition co-occurs with other relations, or if, in the words of Spenader and Stulp, it is a *predictor for contrast*. The idea is that if opposition co-occurs with other relations, it can not be interpreted as a specific signal for a contrast relation.

6.3 Towards Representing Intersections between Contrast and Opposition: a preliminary proposal

We have seen that, even if not that common, opposition at lexical level can occur with contrast at discourse level. An interesting area of investigation that we intend to explore

concerns the formal representation of the contrast relation, accounting also for the opposition relation. In other words, we investigate if it is possible to represent together the discourse level of contrast and the lexical/semantic level of opposition. We present two options: the integration of the annotation of lexical opposition in the PDTB schema, and the adoption of the AMR format. As we have seen in Section 4.1.2, the PDTB schema has been defined for the annotation of many different discourse relations: we attempt to expand it including the annotation of lexical relations. On the other hand, the AMR format (as we will see in Section 6.3.2) aims at representing a whole sentence: we attempt to investigate how the contrast relation is encoded and to expand the format including the opposition relation.

6.3.1 Integrating Opposition in the PDTB Schema

One of the representation that we indirectly presented (and tested) implicates to add a layer of lexical annotation (annotation of opposite lexical units) on top of the annotation of the contrast relation as suggested by the PDTB schema.

In Section 6.2.1 we reported two examples (Examples 6.4 and 6.5) from Contrast-Ita Bank where a relation of opposition is identified: one of the member of an opposite pair is in Arg1 and the other in Arg2. For instance, in Example 6.4, the light-verb construction *rassegnare le dimissioni* (Eng. 'to resign') is the opposite of *rimanere* (Eng. 'to remain') and the two are found respectively in Arg1 and in Arg2 of the contrast relation.

The PDTB representation permits to capture important aspects of the two relations: on the one hand, it is possible to tag discourse relations and their characteristics (for instance, the span of text involved in the relations, the connectives, etc.); on the other hand, lexical units can be annotated with no limitation about the number of tokens that can be involved (so one can tag also multi-words units). Moreover, the layer of lexical annotation is simply added to the discourse annotation thus not causing the two annotations to be incompatible. This means that there is no need for a modification of the PDTB schema, that has been vastly accepted and adopted (see Section 4.1.2.1). However, the PDTB schema does not allow the annotation of information "inside" the arguments of a relation: it only accounts for the span of text involved, and for its relation with the connective (that leads to distinguish between Arg1 and Arg2)[199]. Yet, we believe that it would be interesting

[199]The annotation on the same corpus on which the Penn Treebank (PTB) II corpus Marcus et al. (1993) was built, but the PDTB schema *per se* does not include the annotation of this kind of information.

to explore more in depth the arguments involved in a discourse relation, for example to acquire information about the argument structure of sentences in the arguments, and find a way to represent it. In this way, it would be possible to observe where the opposition lies (e.g. are there opposite predicates?, do we find opposite attributes?).

6.3.2 Adopting Abstract Meaning Representation

Background. A formalism that we started investigating for the representation of both opposition and contrast is Abstract Meaning Representation (AMR) (Banarescu et al., 2013). Specifically, we investigated if AMR is an adequate format for representing contrast and cases of opposition in the context of contrast (as in Examples 6.4 and 6.5 in Section 6.2.1).[200]

AMR is a semantic representation language that aims at representing the meaning of sentences abstracting from syntactic notions, such as word category (verbs and nouns), word order, morphological variations, and function words. The main motivation that leads to AMR is the need for encoding different semantic pieces of information (such as co-references and semantic roles) in one semantic representation.

In AMR, each sentence is represented as a rooted, directed, acyclic graph (AMR graph), with a root that "represents the focus of the sentence or phrase" (Banarescu et al., 2013), labels on edges that represent relations, and leaves that represent concepts. In Figure 6.6, we report an example of AMR graph (i.e., an example of representation) for the sentence 'The cat wants to eat the little mouse.', in which 'want-01' is the root, and 'c / cat' is a leaf which refers to an instance called 'c' of the concept 'cat'. ARG0 is a relation that links 'w / want-01' with '(c / cat)'. It is important to highlight that AMR does not annotate individual words in a sentence, rather it represents concepts.

AMR concepts are mainly described by English words (e.g.'cat'), PropBank framesets (Kingsbury and Palmer, 2002; Palmer et al., 2005) (e.g. 'want-01')[201]. AMR relations

[200] Other possible representational formats that can be explored for representing the presence of an opposition relation in the arguments of a contrast relation are the ones proposed in the RST and SDRT theories (see Section 4.1.1). A study in these directions however needs to take into accounts that in RST relations there is no representation of the semantic structure of the arguments involved in the discourse relation. SDRT actually does provide an internal representation of the arguments involved in a relation, but with a more logical approach that the one we pursue here. (SDRT extends the logical relations of discourse introduced in the Discourse Representation Theory (DRT, Kamp and Reyle, 2013). Nevertheless, as we said, what we are reporting here is mainly a exploratory research, thus we do not exclude researches in these directions.

[201] Recall that in Section 2.3 we have already presented the Proposition Bank (Kingsbury and Palmer, 2002; Palmer et al., 2005). PropBank is a large corpus in which the arguments of each predicate are annotated with the semantic

Sentence:	The cat wants to eat the little mouse.

AMR graph:	`(w / want-01`
	`:ARG0 (c / cat)`
	`:ARG1 (e / eat-01`
	`:ARG0 c`
	`:ARG1 (m / mouse`
	`:mod (l / little))))`

Figure 6.6: Example of an AMR graph.

(preceded by a semicolon) can be PropBank frame arguments (in the example, the *:ARG0* relation connects the upper node 'w / want-01' with its ARG0 'c / cat'), general semantic relations (e.g. *:mod, :location, :name, :cause, :concession, :condition,* etc.) and other relations. AMR represents also named entities, modality, negation, questions, and quantities. [202] AMR gives information about semantic roles, and also includes coreference; for instance, in Example 6.6 'c' stands for 'cat' in relation with 'w/want-01' and it is repeated in relation with 'e/eat-01'.

However, it presents also some limitations, highlighted by Banarescu et al. (2013). AMR does not represent quantifier scope, or universal quantification. It drops grammatical number, tense, aspect, quotation marks, etc. (for example 'c / cat' can represent 'a cat','the cat', 'cats').

The first result of the application of AMR format is the AMR Bank: a set of English sentences paired with simple, readable semantic representations[203].

We are considering the adoption of the AMR format for representing opposition and contrast, since it permits to capture semantic information; it captures 'who is doing what to whom' in a sentence (Banarescu et al., 2014), thus facilitating the analysis of the internal structure of the arguments. We believe that this information can be useful for the identification and 'localization' of the opposites in the arguments that contribute to the contrast relation.

roles. In PropBank, it is possible to explore verbs *frameset*s, that is the set of syntactic frames, each one associated with its set of roles (e.g. ARG1, ARG2). For example, for 'eat.01' (to consume) the identified roles are ARG0 (Agent = consumer, eater) and ARG1 (Patient = meal).

[202] For further details see `https://github.com/amrisi/amr-guidelines/blob/master/amr.md`.

[203] For further information: `http://amr.isi.edu/`

The contrast relation in AMR. In order to evaluate if AMR is a valid format for representing discourse contrast and lexical opposition, first we verified how contrast is represented in AMR. [204]

Observing examples in the AMR Bank, we register that PropBank makes available the frame 'c/contrast-01'. The frame takes two arguments: one argument is linked via :ARG1 relation, the other via :ARG2 relation, being ARG1 the first item in comparison, and ARG2 the second item in comparison.[205] Figure 6.7 reports a sentence, its AMR representation from the AMR Bank, and its annotation that we made following the PDTB schema (and the PDTB convention for representation)[206]. In the AMR graph 'c/contrast-01' is the root and it is linked via :ARG1 to '(f/feel-01..)'and via :ARG2 to '(f2/feel-01..)'.

Sentence:	Some women feel the burka empowers them, but other feel trapped by it.

AMR graph:	`(c / contrast-01`
	`:ARG1 (f / feel-01`
	`:ARG0 (w / woman`
	`:quant (s / some))`
	`:ARG1 (e / empower-01`
	`:ARG0 (b / burqa)`
	`:ARG1 w))`
	`:ARG2 (f2 / feel-01`
	`:ARG0 (o / other)`
	`:ARG1 (t / trap-01`
	`:ARG0 b`
	`:ARG1 o)))`

PDTB annotation:	*Some women feel the burka empowers them*, but **other feel trapped by it**.

Figure 6.7: Example of *c/contrast-01* in a AMR graph and our annotation following the PDTB schema.

In the AMR bank we also find examples in which the frame *h / have-concession-91* is used. Figure 6.8 shows two sentences from the AMR Bank and their AMR representations: in the first (top box), *h / have-concession-91* allows two arguments (connected via relations *:ARG1* and *:ARG2*, as for contrast-01), while in the second (bottom box) the *:concession* relation links two elements, i.e., 'g / game' and the frame 'r / rain-01'.

[204] The main idea behind this proposal arose during a visiting period at the CLEAR center at University of Colorado at Boulder. We acknowledge Martha Palmer and Timothy O'Gorman for their significant help in discussing this formalism.

[205] See `https://verbs.colorado.edu/propbank/framesets-english/contrast-v.html`

[206] As we have seen in Chapter 4 and in Chapter 5, the connective is underlined, Arg1 is in italics, Arg2 is in bold.

Sentence:	The daily violence has continued despite a ceasefire declaration in areas affected by the 08 October 2005 earthquake.
AMR graph:	(h / have-concession-91 :ARG1 (c / continue-01 :ARG1 (v / violence :frequency (r / rate-entity-91 :ARG3 (t / temporal-quantity :quant 1 :unit (d / day))))) :ARG2 (d2 / declare-02 :ARG1 (c2 / cease-01 :ARG1 (f / fire-01)) :location (a / area :ARG1-of (a2 / affect-01 :ARG0 (e / earthquake :time (d3 / date-entity :year 2005 :month 10 :day 8)))))).
PDTB annotation:	*The daily violence has continued* despite **a ceasefire declaration in areas affected by the 08 October 2005 earthquake.**

Sentence:	The game continued, despite of the rain.
AMR graph:	(c / continue-01 :ARG1 (g / game) :concession (r / rain-01))
PDTB annotation:	*The game continued*, despite **of the rain**.

Figure 6.8: Example of *h / have-concession-91* and of *:concession* in AMR graphs and our annotation following the PDTB schema.

We have found no detailed guidelines for the use of *c / contrast-01*, *h / have-concession-91* and *:concession*. However, examining examples of usage, they seem to be parallel to the two types of contrast we identified (that are indeed named as *contrast* and *concession*, see Section 4.2.1.4). In fact, as can be seen from Figure 6.7 and Figure 6.8, if the proposed sentences would be annotated following the PDTB guidelines (detailed in Section 5.1.2), the elements connected to the concepts *c/contrast-01* and *h / have-concession-91* via :ARG1 and :ARG2 will be parallel to what is called *arg1* and *arg2* in the PDTB schema (or *arg2* and *arg1*, depending which one is linked to the connectives); similarly, what the relation of *:concession* does connect would be annotated as *arg1* and *arg2*.

Despite this similarity, the connection between discourse contrast (and its types) and

its representation in AMR needs not be undertaken. One of the main characteristics of AMR is that it links concepts. This means that AMR can represent the arguments of a discourse relation but it does not report their textual realization. For instance, in the example in Figure 6.8 the string of text 'a ceasefire declaration' is represented with '(d2 / declare-02 :ARG1 (c2 / cease-01 :ARG1 (f / fire-01))' in graph, rather than with the string of text itself. In other words, while in the PDTB format the arguments of the relation are portions of the original text connected by an explicit or an implicit connective, AMR graphs represents concepts.

Notice that as a consequence of this abstraction, connectives are not in the graph (we have no *but* or *despite*), but the relation they convey is represented.

AMR for Italian. As announced, our aim is to evaluate if the AMR format can be a valid representational framework for representing Italian sentences in which there is an opposition in the context of a contrast relation, and to highlight the limits of the representation. To do this, we attempted to represent a number of these sentences. Before presenting and discussing our attempt, in the shape of two examples extracted from Contrast-Ita Bank, one last specification needs to be made concerning the language. AMR is geared toward English and the vocabulary of English, and it makes extensive use of the Propbank resource. There is no such a resource for Italian. However, we believe that the T-PAS resource (Ježek, 2016) (see a detailed description in Section 3.1.1) presents some similarities with Propbank, and can be used for the purpose. In particular, the list of frames in PropBank is similar to the list of *t-pas* in T-PAS: they both provide information about the semantic structures of the predicates. The main difference lies in that PropBank lists semantic roles in the frames (e.g. ARG0, ARG1), while T-PAS uses semantic types that generalize over the lexical items found in a corpus for that argument position (e.g. [[Human]], [[Building]]).

Given this background, we tried to represent the two sentences proposed in Section 6.3.1 from Contrast-Ita Bank; they are shown in Figures 6.9 and 6.10.

In both figures, we report the examples from Contrast-Ita Bank, our graph representation and the annotation following the PDTB schema. Notice that in both cases the opposition relation is among lexical units that typically do not hold this relation, i.e., they are 'opposites in context' (See Section 6.2.1).

Sentence: Il ministro del Lavoro e delle Pensioni britannico, Andrew Smith, ha rassegnato ieri le dimissioni nonostante i tentativi del premier Tony Blair di convincerlo a rimanere.
Eng: British Secretary of State for Work and Pensions Andrew Smith resigned yesterday, despite Prime Minister Tony Blair's attempts to persuade him to stay.

AMR graph:

```
(h / have-concession-91
    :ARG1 (d / dimettere-t-pas#1 - OPP#1
        :ARG1 (p / persona
            :wiki ‘‘Andrew_Smith’’
              :name (n / name :op1 ‘‘Andrew’’ :op2 ‘‘Smith’’)
            :ARG0-of (h / have-org-role-91
                :ARG1 (p1 / paese :wiki ‘‘United_Kingdom’’
                  :name (n / name :op1 ‘‘Britain‘‘))
                :ARG2 (m / ministro)
                :topic (a / and
                    :op1 (l / lavoro)
                    :op2 (p2 / pensioni))))
        :time (i / ieri))
    :ARG2 (c1 / cercare-t-pas#2
        :ARG1(p3 / persona
            :wiki ‘‘Tony_Blair’’
              :name (n / name :op1 ‘‘Tony’’:op2‘‘ Blair’’)
            :ARG0-of (h / have-org-role-91
                :ARG2 (p3 /premier)))
        :ARG2 (c2 / convincere-t-pas#1
            :ARG1 p3
            :ARG2 p
            :ARG3 (r / rimanere-t-pas# - OPP#2))))
```

PDTB annotation: *Il ministro del Lavoro e delle Pensioni britannico, Andrew Smith, ha rassegnato ieri le dimissioni* nonostante **i tentativi del premier Tony Blair di convincerlo a rimanere**.
Sense tag in Contrast-Ita Bank: CONCESSION.Arg1.as.denier

Figure 6.9: AMR for contrast and opposition in Italian. Example 1.

The two examples highlight important issues that need to be considered. The first concerns the language to use as metalanguage (e.g. n / name or n / nome?) and the resources to use as pointers (e.g. English wikipedia or Italian wikipedia?).

The second issue concerns the use of T-PAS as a knowledge base for retrieving information about the frames (parallel to PropBank for English AMR). The example shows the feasibility of using T-PAS. However, there is no exact mapping between T-PAS and its

Sentence: Sul Nuovo Mercato , Tiscali perde lo 0,55% a 2,33 , E.Biscom sale dell' 1,09% a 41,44.
Eng: On the New Market, Tiscali looses 0.05% to 2.23, E. Biscom rises by 1.09% to 41.44.

AMR graph:

```
(c / contrast-01
    :ARG1 (p / perdere-t-pas#1 - OPP#1
        :ARG1 ( o / organization
            :name (n / name
                :op1 ‘‘Tiscali’’)
            :ARG2 (p1 / percentage-entity :value 0,55)
        :destination(q / quantity :value 2,33)))
    :ARG2 (s / salire-t-pas#10 - OPP#2
        :ARG1 (o1 / organization
          :name (n1 / name
                 :op1 ‘‘E.Biscom’’))
        :ARG2 (p2 / percentage-entity :value 1,09)
    :ARG4 (q1 / quantity :value 41,44))
      :location (m / mercato
          :mod (n / nuovo)))
```

PDTB annotation: *Sul Nuovo Mercato, Tiscali perde lo 0.05% a 2,23*, **E. Biscom sale dell'1,09% a 41,44**.
- implicit connective = invece, Sense tag in Contrast-Ita Bank: CONTRAST

Figure 6.10: AMR for contrast and opposition in Italian. Example 2.

argument positions, and PropBank and its semantic roles. In PropBank, the labels ARG0, ARG1, etc. identify semantic roles in syntactic frames and an effort was made to use ARG0 for "prototypical agent" and ARG1 for "prototypical patient" (Loper et al., 2007). T-PAS organizes the frame in argument positions describing the semantic type for each slot, but does not assign semantic roles. However, both resources permit to inspect the structure of a frame, thus being useful for our goal. In the examples we provided, we simply progressively assigned a number to the argument slots found in the *t-pas* (i.e., ARG1 is the first argument we encountered for the *t-pas*, that is usually -but not necessary- the subject of the frame). Moreover, T-PAS is a useful resource as far as the predicates are concerned. Unlike PropBank, in T-PAS there is no *contrast* frame, and this information needs to be possibly integrated differently. Also, the coverage of the resource needs to be extended. The actual version of T-PAS lists 1000 verbs and does not include, for example, *rimanere* (needed for completing the graph in 6.9). Also, it can be that a suitable *t-pas* for

a verb is not found, e.g., due to its rare usage: a decision about whether enriching T-PAS with this new *t-pas* will have to be taken.

Opposition in AMR. As far as opposition is concerned, the labels OPP#1 and OPP#2 can be assigned to the opposites. Ideally, on a larger scale, this information would be retrieved using other resources that register the opposition relation (see the work presented in Chapter 3, which goes in this direction). Moreover, we show that also 'opposites in context' can be annotated, as in the proposed examples (Figures 6.9 and 6.10). What is clear is that in the AMR framework *opposition* has to be interpreted as a relation among concepts, not among words forms. To understand this important aspect, consider the opposition tagged in Example 6.9: it is annotated between OPP#1 and OPP#2, being OPP#1 the frame 'd / dimettere-*t-pas*#1', abstraction of 'rassegnato le dimissioni', and OPP#2 the frame 'r / rimanere-t-pas#'. Another important element that does not emerge from the examples is that AMR abstracts from the word category. This can favour the identification of opposition across word categories, e.g. between *costruire* (Eng. 'to build') and *distruzione* (Eng. 'destruction'), as they will be both represented with the verbal frame, respectively 'costruire-*tpas*#' (Eng. 'to build') and 'distruggere-*tpas*#' (Eng. 'to destroy').

AMR for contrast and opposition. In conclusion, we have conducted an exploratory investigation in order to evaluate the possibility of representing opposition in the arguments of a contrast relation for Italian. We have highlighted some positive aspects of adopting the AMR representation, possibly using T-PAS; the most important one is the capability of AMR to provide a semantic representation of the arguments involved in a contrast relation. Yet further discussion is required, for instance, about multi-sentences representation, the difference between contrast and concession in AMR, the use of the T-PAS resource, the difference between *concession* as a frame or as a relation (*h / have-concession-91* and *:concession*). Certainly, one of the most important aspects that needs to be considered is that AMR is not text linked, thus opposition has to be interpreted as a relation among concepts. Moreover, the information about the connective that conveys the relation is not captured by AMR, but represented in the graph structure.

The discussion about whether AMR is a valid format for representing discourse and lexical semantic relations is open, and will be pursued in further work.

Conclusions

In this thesis we have investigated lexical opposition and discourse contrast in Italian. Opposition is the relation between two lexical units that contrast each other with respect to one key aspect of their meaning and are similar for other aspects (e.g. *to pass / to fail* an exam). Contrast is a discourse relation, i.e., a relation between two parts (*arguments*) of a coherent sequence of sentences, propositions or speeches; in a relation of contrast, the two arguments are similar in many aspects but different in one aspect for which they are compared (e.g. '$_{[}$Mary passed the exam$_{\text{arg1}]}$. $_{[}$John failed it $_{\text{arg2}]}$.') or one argument is denying an expectation that is triggered from the other argument (e.g. 'Although $_{[}$it was raining$_{\text{arg2}]}$, $_{[}$we went to the beach $_{\text{arg1}]}$.'). We considered both cases in which this relation is conveyed explicitly by mean of lexical *connectives* (e.g. *but, while, although*) and cases in which it is implicitly conveyed.

In this chapter, we review the main contributions of this thesis and discuss future work.

Opposition

Towards the annotation of opposition in a lexical resource: schema and population

Contributions. We defined a schema for annotating opposition in lexical resource and collected judgments about opposition among verbs of the Italian language at the sense level. The approach is based on the automatic substitution of a verb in a sentence and the comparison of the initial sentence with the newly created one (which thus differs from the first one just in the substituted verb). We chose this setting so that annotators could observe the verbs under discussion in context. We did not use definitions of the sense of the verbs.

Shortcomings and future work. The most important shortcoming of this acquisition concerns its degree of completeness; we collected a part of the data necessary for completing the enrichment of the T-PAS resource with the opposition relation. In Section

3.4.5 we tried to quantify the effort still needed.

A methodology for acquiring opposite verb pairs.

Contributions. As presented in Section 3.4, we experimented with a crowd-sourcing strategy in order to collect human annotated data in a short period of time. The experiment we run demonstrates the feasibility of using this methodology for collecting data, provided that the task is properly designed.

Shortcomings and future work. As highlighted in Section 3.4.4, the task requires continuous monitoring. In addition, we expected more data: few workers participated (possibly because we launched the task during the summer break). The task was also judged to be rather difficult by the workers, and we observed that increasing the rewards did not help in increasing the number of workers. Potential further work (and suggestions for future experiences) includes trying to use other settings in order to make the task even simpler.

Empirical analysis of opposition types.

Contributions. Through the crowd-sourcing experiment, we provided evidence that different types of opposition are recognized by Italian speakers. In particular, we focused on four types of opposition, i.e. *complementarity, antonymy, converseness*, and *reversiveness*, identified in the lexical semantics literature, including Cruse (1986, 2011); Lyons (1977); Prandi (2004); Murphy (2003); Ježek (2016). Results confirm that some types of opposition exclude each other (there is a main distinction between *complementarity* and *antonymy*), while the relation of *reversiveness* seems not to be an exclusive relation and tends to add to other opposition relations, particularly *antonymy*. Through another annotation exercise by expert annotators (Section 3.2), we have also shown that even if the inter-annotator agreement concerning the identification of opposite senses of two verbs is very high, it is less solid, but still fair, when we asked the annotators to identify the type of oppositions they hold. This lead us to point out that the annotation of opposition relations in lexical resources is feasible provided that the designed annotation schema considers the co-existence of more than one type of opposition relation.

Shortcomings and future work. One possible criticism of the approach we used for acquiring information about types of opposition through the crowd-sourcing experiment

is that we did not ask participants direct questions about the types of opposition, but we operated instead through the use of eight statements, which possibly do not capture all the characteristics of the different types. As explained in Section 3.4.1, the crowd-sourcing methodology requires that the task is brief and clear, thus forcing us to reformulate a direct sentence of the kind "The two verbs are *complementaries*" with "Given a situation, if one of the two events does not take place, the other occurs, e.g. to fail/to pass (an examination)". However, we believe we managed to reach a good compromise between the clearness that the methodology requires and the direct request we would have liked to ask.

As far as future work is concerned, the complete acquisition of the data for the annotation of the opposition relation through out the T-PAS resource will provide important insights on the distribution of the different types of oppositions will be collected.

An analysis of non conventional opposites in context.

Contributions. Through the annotation of opposition relations in a corpus (for verifying their co-occurrence with the contrast relations - Chapter 6), we have also provided examples from the corpus of non conventional pairs of opposites, that is, pairs of lexical units that trigger a relation of opposition when in a specific context (e.g. "salire del #%" - "perdere il #%"). These opposite pairs are observable only through a corpus investigation.

Shortcomings and future work. It would also be extremely interesting to expand the annotation to other texts in order to collect other cases of 'opposites in context'.

Contrast

An annotated corpus: Contrast-Ita Bank.

Contributions. We created the freely available resource Contrast-Ita Bank[207], a corpus of 169 news documents annotated with relations of contrast (see Section 5.1). Specifically, we used the sense tags CONTRAST and CONCESSION according to the Penn Discourse Treebank (PDTB) (Prasad et al., 2007) schema of relations (see Section 4.1.2). We annotated the same corpus with the relation of opposition when occurring in the arguments of a contrast relation. Contrast-Ita Bank also represents the first step in creating a corpus

[207] `https://hlt-nlp.fbk.eu/technologies/contrast-ita-bank`

annotated with discourse relations for Italian that is compatible with the PDTB project (Prasad et al., 2007).

Shortcomings and future work. A potential future work regards the annotation of all the discourse relations of the PDTB schema in the same corpus of document of Contrast-Ita Bank. This will lead us to draw some conclusions on contrast relation in comparison to other relations (e.g. frequency, connectives for conveying the relation), and to create a resource for Italian similar and compatible to the PDTB.

A corpus-based analysis of contrast.

Contributions. We provide quantitative data about: frequency, explicitness/implicitness and connectives for the two types of contrast we distinguished (CONTRAST and CONCESSION) as a result of the annotation of Contrast-Ita Bank. In our data, we register that the sense tag CONCESSION.Arg2-as-denier (i.e., when Arg2 is denying an expectation that rises from Arg1, as in '$_{[}$It was her birthday$_{arg1]}$, however $_{[}$she did not organize any party$_{arg2]}$') is the most frequent type of contrast. CONTRAST covers almost a quarter of the cases. CONCESSION.Arg1-as-denier is far less frequent both as a single type and with other relations; in contrast with the first, this subtype is associated with a limited set of connectives. CONTRAST and CONCESSION have also been annotated together. We also notice that contrast is mainly explicit (91%) and we register up to 15 explicit relations in one document (average length of the examined documents is 387 tokens) and an average of 2 relations per document. Concerning the connectives, *ma* (Eng.: 'but') is the most frequent connective, followed by *invece* (Eng.: 'instead', 'on the other hand'), *mentre* (Eng.: 'while'), and *però* (Eng.: 'but', 'yet').

Shortcomings and future work. Future work includes the detailed comparison of the data from Contrast-Ita Bank with quantitative data for English or other languages in a multi-lingual perspective. Preliminary results seem to suggest that there are some differences: for example, we observed that in Contrast-Ita Bank the ratio between *explicit* and *implicit* relations concerning CONTRAST and COMPARISON (and their subtypes) is about ten times less than in the PDTB 2.0 (Prasad et al., 2007).

A corpus-based analysis of the connectives of contrast.

Contributions. Another contribution deriving from the annotation of Contrast-Ita

Bank concerns the connectives of contrast (Section 5.2.4.1). We collected 19 connectives: some are very frequent (*ma*, Eng.: 'but', accounts for almost half of the cases), while others are rarer and appear just once in the corpus. The most frequent connectives are the most polysemous ones even if they show a general preference. On the other hand, less frequent connectives are less ambiguous. We also gave evidence that in general there is no relationship between the syntactic category of a connective and the relation it conveys (as also Ferrari and Zampese (2000) point out). A detailed analysis of the five most frequent connectives (Section 5.11) shows that *ma* is also frequently used not as a connective, and that 25% of the time when *mentre* is used in the corpus, it conveys a temporal relation.

Shortcomings and future work. A possible extension of this work is the annotation of the connectives of contrast in other documents. As we mentioned, some of them are very rare in Contrast-Ita Bank, and the extension of the annotation will give us a chance to study more precisely the behaviour of these connectives, e.g. frequency, position in a sentence, and type of contrast conveyed (CONTRAST or CONCESSION). Moreover, the analysis of the cases of False Positives (i.e., lexical items that can be connectives of contrast, but are not; such as *mentre* when it conveys a temporal relation) can be extended to all the connectives we have identified.

A lexicon of Italian connectives.

Contributions. In order to respond to a lack of a free resource for connectives in Italian that can also be used for NLP purposes, we created LICO[208] (See Section 5.2.3). In LICO, Lexicon of Italian Connectives, connectives are listed together with orthographic, syntactic, semantic information (according to the PDTB 3.0 schema of relations) and possible alignments with lexicon of connectives in other languages. LICO is organized as a set of entries (173), each one corresponding to a connective and its orthographic or lexical variants. The list of connectives is retrieved from other grammatical and lexical resources for Italian and for other languages.

Merging the list of connectives from Contrast-Ita Bank and the list of connectives registered with the tag CONTRAST or CONCESSION in LICO allowed us to create a more exhaustive list of the textual connectives for contrast in Italian (See Section 5.2.4.3).

[208] `https://hlt-nlp.fbk.eu/technologies/lico`

Shortcomings and future work. LICO is still under construction. We believe that its main shortcoming is that the resource does not include information from corpora. Thus, an ongoing work is the integration of the information in LICO with information retrieved from the analysis of the connectives in a corpus, such as frequency and usage examples. A first step in this direction is the integration of the data from Contrast-Ita Bank into LICO: this allows us to validate the list of connectives of contrast already proposed in LICO, and to add them with examples from the corpus.

Opposition and Contrast

A corpus-based analysis of the co-occurrence of opposition and contrast.

Contributions. We provide evidence that in Contrast-Ita Bank opposition does not co-occur very often with discourse contrast when the contrast relation is explicitly marked by a connective (4%), but it is more frequent when no connective is present (18%). Also, it seems more frequent in cases of CONTRAST than in cases of CONCESSION. Limited to the 20 cases we observed, we performed an IAA exercise and conducted a qualitative analysis in order to understand the role of opposites in the context of a contrast relation. We show that to judge whether opposites contribute in contrast seems not to be controversial operation and we registered that if opposites appear in a contrast relation they frequently participate in it.

Finally, we have seen that opposition, depending on the context, can hold among pairs of lexical units that are not conventionally intended as opposites.

Shortcomings and future work. The main future work of this contribution concerns the annotation of all the opposite pairs in the arguments of other discourse relations, in a corpus. This will make it possible to verify their general frequency, and to understand if opposition can predict the presence of a contrast relation.

A proposal for representing the co-occurrence of opposition and contrast.

Contributions. We propose our exploratory research about how the contrast relations can be formally represented, even with the specification of lexical opposites in the arguments of the relation. We discuss the PDTB schema of annotation and the AMR formalism. In particular, we tested whether they are valid formats for representing discourse contrast and lexical opposition in Italian. We evaluate positively the capability of AMR of repre-

senting internal structures of the arguments of a contrast relation, in order to annotate also opposition, but further analysis is required.

Shortcomings and future work. The discussion is open, and will be object of further work. Starting from our attempt, a broader investigation could also be conducted concerning the representation of other discourse relations (e.g., temporal, causal, etc.) and other lexical relation (e.g., synonymy, hyperonymy, etc.) with the AMR format.

Appendix

A. Example of *t-pas* for *abbattere*

```
<?xml version="1.0"?>
<pattern_set>
  <verb_stem>abbattere</verb_stem>
    <pattern num="2" id="2">
    <template adverbial_class="" adverbial_none="" object_none=""
        adverbial_opt="" auxiliary="">
      <verb_form>abbattere</verb_form>
      <subject>
        <argspec det_quant="" ocl_ing="" ocl_wh="" type="" headword=""
            ocl_to="" ocl_quote="" none="" ocl_that="">
          <BSO_type name="Human"/>
          <subspec/>
        </argspec>
        <alternation>
          <argspec det_quant="" ocl_ing="" ocl_wh="" type="" headword=""
              ocl_to="" ocl_quote="" ocl_that="">
            <BSO_type name="Event"/>
            <subspec/>
          </argspec>
        </alternation>
      </subject>
      <object>
        <argspec det_quant="" ocl_ing="" ocl_wh="" type="" headword=""
            ocl_to="" ocl_quote="" none="" ocl_that="">
          <BSO_type name="Building"/>
          <subspec/>
        </argspec>
      </object>
      <clausal/>
      <oclausals to="" quote="" optional="" ing="" that="" wh=""/>
      <clausals to="" quote="" optional="" ing="" that="" wh=""/>
      <comment type=""/>
    </template>
    <primary_implicature phrasal="" idiom="">[[Human | Event]] demolisce,
        distrugge, butta giu; eliminando [[Building]]</primary_implicature>
    <framenet/>
    <domain/>
    <register/>
    <comment type=""/>
    <exploitation/>
    <semclass/>
</pattern_set>
```

B. Original Crowd-Sourcing Tasks

The appendix reports the original text for introducing the tasks to the workers and an example of the original task are reported.

Buongiorno,
stiamo conducendo una ricerca per capire quando due verbi sono opposti e che tipo di opposizione creano. Il questionario vi chiederà di giudicare una coppia di frasi (**Frase 1** e **Frase 2**) e di dirci se esse sono opposte.
Per ogni coppia di frasi giudicate come opposte, verranno proposte delle domande sui rapporti di opposizione tra i due verbi che appaiono in esse. Ogni coppia è formata dalla Frase 1 e dalla Frase 2.
La Frase 1 è stata estratta automaticamente da un insieme di testi mentre la Frase 2 è identica alla Frase 1, tranne che per un verbo che è stato sostituito anch'esso automaticamente.

ATTENZIONE!
Essendo la Frase 1 estratta automaticamente, essa potrebbe essere incomprensibile senza il suo contesto.
Esempio:
FRASE 1: Poi usa il bibi congelato ed infine prova a **calarlo** al tramonto.
FRASE 2: Poi usa il bibi congelato ed infine prova a **salirlo** al tramonto.

Vi invitiamo a segnare che la Frase 2 non ha senso.

INOLTRE.. In altri casi, mentre la Frase 1 è perfettamente comprensibile, la Frase 2 potrebbe non avere assolutamente senso oppure apparire strana a seguito della sostituzione del verbo.
Per un esempio di Frase 2 che non ha un senso, osservate la seguente coppia:
Frase 1: La chiesa **sorge** su una collina.
Frase 2: La chiesa **tramonta** su una collina.

In questo caso, la Frase 2 non ha un senso poichè un edificio può *sorgere* in un luogo non può *tramontare* in un luogo. Per un esempio di Frase 2 semanticamente strana, considerate invece:
Frase 1: La luna **sorse** e lo abbagliò.
Frase 2: La luna **tramontò** e lo abbagliò.

In questo caso, nonostante sia poco probabile che la luna abbagli con la sua luce nel momento del tramonto, diremo che la Frase 2 ha senso poichè la luna può *sorgere* ma anche *tramontare*.

INFATTI.. ..per decidere se due frasi sono opposte, dovete considerare il contesto minimo su cui i verbi incidono. Nel caso sopra, il contesto minimo di *sorgere* è "La luna **sorse** " che si oppone al contesto minimo di *tramontare*: "La luna **tramontò**". In questo esempio, è quindi presente una opposizione nonostante i due eventi causino poi lo stesso effetto ("..e lo abbagliò.").
Vi invitiamo a segnare "FRASE 2 ha senso e c'è opposizione con FRASE 1".

Grazie per la vostra collaborazione.

Confronta le seguenti frasi. Nella Frase 2 è stato sostituito un verbo in modo automatico.

Frase 1: L' appello va, pertanto, **respinto**

(annotated example of T-PAS 1 of the Source Verb = *respingere*)

Frase 2: L' appello va, pertanto, **approvato**.)

(Target Verb = *approvare*)

Task A: "Diresti che la Frase 2 ha un senso compiuto?

Se sì, diresti che c'é una relazione di opposizione tra le Frasi?"

A1: Frase 2 ha senso e e c'è opposizione con Frase 1

A2: Frase 2 ha senso ma non e c'è opposizione con FRASE 1

A3: Frase 2 non ha senso

Task B: "Leggi le seguenti frasi. In quali frasi **approvare** ha lo stesso senso della Frase 2?"

B1: La Commissione **approva** l'emendamento 2.15 del relatore.

B2: Gli astronauti hanno **approvato** l' uso del TVIS in questa configurazione.

B3: In ogni caso , non verranno **approvati** i candidati che abbiano registrato assenze superiori a un terzo del numero complessivo di ore di lezione previste.

B4: Nessuna delle precedenti

Task C: Considera ancora **respingere** e **approvare** nelle FRASI 1 e 2.

Quali delle seguenti affermazioni sono vere:

1: è possibile una situazione in cui entrambi gli eventi possono verificarsi contemporaneamente e con gli stessi partecipanti (esempio: io respiro, io scrivo)"

2: data una situazione, se non si realizza un evento, si realizza l'altro (esempio: se uno studente non viene promosso, viene bocciato)

3: data una situazione, esiste un' alternativa "neutrale" in cui nessuno dei due eventi si verifica (esempio: la temperatura nè si alza, nè si abbassa ma resta com'è!)

4: è possibile usare modificatori, come "un po"', "moderatamente", "molto" (esempio: il suo progetto è migliorato molto, il suo progetto è peggiorato molto)

5: esiste la possibilità che le azioni si susseguano più volte ripetutamente (esempio: aprire la porta, chiudere la porta, aprire la porta,..)

6: uno dei due eventi è possibile solo se l'altro si è già verificato (esempio: distruggere un edificio implica che sia stato prima costruito)

7: se uno dei due eventi si verifica e successivamente si verifica l'altro, la situazione ritorna quella di partenza (esempio: se sbottono la giacca, e poi la abbottono)

8: gli eventi descrivono la stessa situazione da due punti di vista diversi (esempio: dare un regalo - ricevere un regalo)

Lascia un commento

Example of the original version of the task. Task B and Task C were shown just in case of selection of answer A1, otherwise a new pair of sentences was shown.

Bibliography

Heike Adel and Hinrich Schütze. Using mined coreference chains as a resource for a semantic task. In *Proceedings of the 2014 Conference on Empirical Methods on Natural Language Processing (EMNLP 2014)*, pages 1447–1452, 2014.

Eneko Agirre and Philip Edmonds. *Word sense disambiguation: Algorithms and applications*, volume 33. Springer Science & Business Media, 2007.

Eneko Agirre, Carmen Baneab, Claire Cardiec, Daniel Cerd, Mona Diabe, Aitor Gonzalez-Agirrea, Weiwei Guof, Inigo Lopez-Gazpioa, Montse Maritxalara, Rada Mihalceab, et al. Semeval-2015 task 2: Semantic Textual Similarity, English, Spanish and Pilot on Interpretability. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval '15)*, pages 252–263, 2015.

Eneko Agirre, Aitor Gonzalez-Agirre, Inigo Lopez-Gazpio, Montse Maritxalar, German Rigau, and Larraitz Uria. Semeval-2016 task 2: Interpretable semantic textual similarity. *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval '16)*, pages 512–524, 2016.

Amal Al-Saif and Katja Markert. The leeds arabic discourse treebank: Annotating discourse connectives for arabic. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC '10)*, 2010.

Begoña Altuna, Manuela Speranza, and Anne-Lyse Minard. The Scope and Focus of Negation: A Complete Annotation Framework for Italian. *SemBEaR 2017*, page 34, 2017.

David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.

Ron Artstein and Massimo Poesio. Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–596, 2008.

Nicholas Asher. *Reference to Abstract Objects in Discourse*. Kluwer Academic Publishers, Dordrecht, 1993.

Nicholas Asher and Alex Lascarides. *Logics of conversation*. Cambridge University Press, 2003.

Nicholas Asher and Laure Vieu. Subordinating and coordinating discourse relations. *Lingua*, 115(4):591–610, 2005.

Fatemeh Torabi Asr and Vera Demberg. Measuring the strength of linguistic cues for discourse relations. In *Proceedings of the Workshop on Advances in Discourse Analysis and its Computational Aspects (ADACA)*, pages 33–42, 2012.

Collin F Baker, Charles J Fillmore, and John B Lowe. The berkeley framenet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*, pages 86–90. Association for Computational Linguistics, 1998.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. Abstract meaning representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, 2013.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. Abstract meaning representation (amr) 1.2 specification. 2014. `https://www.isi.edu/~ulf/amr/help/amr-guidelines.pdf`.

Marco Baroni and Adam Kilgarriff. Large linguistically-processed web corpora for multiple languages. In *Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics: Posters & Demonstrations*, pages 87–90. Association for Computational Linguistics, 2006.

Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. The WaCky

wide web: a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation*, 43(3):209–226, 2009.

Valentina Bartalesi Lenzi, Giovanni Moretti, and Rachele Sprugnoli. Cat: the celct annotation tool. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC '12)*, pages 333–338, 2012.

Carla Bazzanella. I segnali discorsivi. *Grande grammatica italiana di consultazione*, 3: 225–257, 1995.

Eduard Bejček, Eva Hajičová, Jan Hajič, Pavlína Jínová, Václava Kettnerová, Veronika Kolářová, Marie Mikulová, Jiří Mírovský, Anna Nedoluzhko, Jarmila Panevová, Lucie Poláková, Magda Ševčíková, Jan Štěpánek, and Šárka Zikánová. Prague Dependency Treebank 3.0. `http://ufal.mff.cuni.cz/pdt3.0/`, 2013.

Farah Benamara and Maite Taboada. Mapping different rhetorical relation annotations: A proposal. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics,*SEM*, volume 2012, pages 147–152, 2015.

Giulia Benotto. *Distributional Models for Semantic Relations: A study on Hyponymy and Antonymy*. PhD thesis, University of Pisa, 2014.

Douglas Biber, Susan Conrad, and Randi Reppen. *Corpus linguistics: Investigating language structure and use*. Cambridge University Press, 1998.

Lynn Carlson and Daniel Marcu. Discourse tagging reference manual. *ISI Technical Report ISI-TR-545*, 54:56, 2001.

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue-Volume 16*, pages 1–10. Association for Computational Linguistics, 2001.

Lynn Carlson, Mary Ellen Okurowski, and Daniel Marcu. *RST discourse treebank*. Linguistic Data Consortium, University of Pennsylvania, 2002.

Tommaso Caselli, Valentina Bartalesi Lenzi, Rachele Sprugnoli, Emanuele Pianta, and Irina Prodanof. Annotating events, temporal expressions and relations in Italian: the

It-TimeML experience for the Ita-TimeBank. In *Proceedings of the 5th Linguistic Annotation Workshop*, pages 143–151. Association for Computational Linguistics, 2011.

Walter G Charles and George A Miller. Contexts of antonymous adjectives. *Applied psycholinguistics*, 10(3):357–375, 1989.

Timothy Chklovski and Patrick Pantel. Verbocean: Mining the web for fine-grained semantic verb relations. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '04)*, volume 2004, pages 33–40, Barcelona, Spain, July 2004.

Kenneth Ward Church and Patrick Hanks. Word association norms, mutual information, and lexicography. *Computational linguistics*, 16(1):22–29, 1990.

Silvie Cinková and Patrick Hanks. Validation of corpus pattern analysis–assigning pattern numbers to random verb samples. 2010. `http://ufal.mff.cuni.cz/spr/data/publications/annotation\_manual.pdf`.

D Alan Cruse. *Lexical semantics*. Cambridge University Press, 1986.

D Alan Cruse. *Meaning In Language: An Introduction To Semantics And Pragmatics*. Oxford University Press, USA, 2011.

Ido Dagan, Bill Dolan, Bernardo Magnini, and Dan Roth. Recognizing textual entailment: Rational, evaluation and approaches. *Natural Language Engineering*, 15(4): i–xvii, 2009.

Laurence Danlos. Comparing RST and SDRT discourse structures through dependency graphs. *Proceedings of Constraints in Discourse*, pages 46–53, 2005. `http://www.constraints-in-discourse.de/Papers/20/Danlos.pdf`.

Francesca Della Moretta, Anna Feltracco, Elisabetta Ježek, and Bernardo Magnini. Tagging Semantic Types for Verb Argument Positions. In *Proceedings of the Fourth Italian Conference on Computational Linguistics (CLIC-it 2017)*, Rome, Italy, 2017. To appear.

Katrin Erk and Sebastian Padó. A structured vector space model for word meaning in context. In *Proceedings of the Conference on Empirical Methods in Natural Language*

Processing (EMNLP '08), pages 897–906. Association for Computational Linguistics, 2008.

Katrin Erk and Sebastian Padó. Exemplar-based models for word meaning in context. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL '10) - short papers*, pages 92–97. Association for Computational Linguistics, 2010.

Stefan Evert. The statistics of word cooccurrences: word pairs and collocations. 2005.

Christiane Fellbaum. Co-occurrence and antonymy. *International journal of lexicography*, 8(4):281–303, 1995.

Christiane Fellbaum. *WordNet*. Wiley Online Library, 1998.

Anna Feltracco, Elisabetta Ježek, and Bernardo Magnini. Opposition relations among verb frames. In *Proceedings of the 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, Denver, Colorado, June 2015a. Association for Computational Linguistics.

Anna Feltracco, Elisabetta Ježek, Bernardo Magnini, and Simone Magnolini. Annotating opposition among verb senses: a crowdsourcing experiment. In *Proceedings of the Second Italian Conference on Computational Linguistics (CLIC-it 2015)*, Trento, Italy, December, 2015b.

Anna Feltracco, Lorenzo Gatti, Simone Magnolini, Bernardo Magnini, and Elisabetta Ježek. Using WordNet to Build Lexical Sets for Italian Verbs. In *Proceedings of the Eighth Global WordNet Conference (GWC '16)*, Bucharest, Romania, January, 2016a.

Anna Feltracco, Elisabetta Ježek, Bernardo Magnini, and Manfred Stede. LICO: A Lexicon of Italian Connectives. In *Proceedings of the Third Italian Conference on Computational Linguistic (CLiC-it 2016)*, page 141, 2016b.

Anna Feltracco, Simone Magnolini, Elisabetta Ježek, and Bernardo Magnini. Acquiring opposition relations among italian verb senses using crowdsourcing. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC '16)*, May 2016c.

Anna Feltracco, Bernardo Magnini, and Elisabetta Ježek. Contrast-Ita Bank: A corpus for Italian Annotated with Discourse Contrast Relations. In *Proceedings of the Fourth Italian Conference on Computational Linguistic (CLiC-it 2017)*, 2017. To appear.

Anna Feltracco, Elisabetta Ježek, and Bernardo Magnini. Enriching LICO with Corpus-based Data. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC '18)*, Miyazaki, Japan, 2018. Submitted.

Angela Ferrari. Connettivi. In *Enciclopedia dell'Italiano*. diretta da Raffaele Simone, con la collaborazione di Gaetano Berruto e Paolo D'Achille, Roma, Istituto della Enciclopedia Italiana, 2010.

Angela Ferrari and Luciano Zampese. *Dalla frase al testo: una grammatica per l'italiano*. Zanichelli, 2000.

Charles J Fillmore. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280(1):20–32, 1976.

J.R. Firth. A synopsis of linguistic theory 1930-1955. *Studies in Linguistic Analysis*, pages 1–32, 1957.

Derek Gross, Ute Fischer, and George A Miller. The organization of adjectival meanings. *Journal of memory and language*, 28(1):92–106, 1989.

Patrick Hanks. Corpus pattern analysis. In *Proceedings of the Eleventh EURALEX International Congress*, Lorient, France, Universite de Bretagne-Sud, 2004.

Patrick Hanks. *Lexical analysis: Norms and exploitations*. Mit Press, 2013.

Patrick Hanks and James Pustejovsky. A pattern dictionary for natural language processing. *Revue française de linguistique appliquée*, 10(2):63–82, 2005.

Sanda Harabagiu, Andrew Hickl, and Finley Lacatusu. Negation, contrast and contradiction in text processing. In *Association for the Advancement of Artificial Intelligence (AAAI)*, volume 6, pages 755–762, 2006.

Zellig S Harris. Distributional structure. *Word*, 10(2-3):146–162, 1954.

Jennifer Hay, Christopher Kennedy, and Beth Levin. Scalar structure underlies telicity in" degree achievements". In *Semantics and linguistic theory*, volume 9, pages 127–144, 1999.

Jerry R Hobbs. On the coherence and structure of discourse.Technical report CSLI-85-37, Center for the Study of Language and Information, Stanford University. 1985.

Malka Rappaport Hovav. Lexicalized meaning and the internal. *Theoretical and crosslinguistic approaches to the semantics of aspect*, 110:13, 2008.

Eduard H Hovy and Elisabeth Maier. Parsimonious or profligate: How many and which discourse structure relations. *Discourse Processes*, 1997.

Elisabetta Ježek. *The Lexicon. An Introduction.* Oxford: Oxford University Press, 2016.

Elisabetta Ježek, Bernardo Magnini, Anna Feltracco, Alessia Bianchini, and Octavian Popescu. T-PAS; A resource of Typed Predicate Argument Structures for linguistic analysis and semantic processing. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC '14)*, Reykjavik, Iceland, May 2014.

Elisabetta Ježek, Anna Feltracco, Lorenzo Gatti, Simone Magnolini, and Bernardo Magnini. Mapping Semantic Types onto WordNet Synset. In *Proceedings of the Twelfth Joint ACL - ISO Workshop on Interoperable Semantic Annotation (Isa '12)*, Portorose, Slovenia, May, 2016.

Steven Jones, Carita Paradis, M Lynne Murphy, and Caroline Willners. Googling for 'opposites': a web-based study of antonym canonicity. *Corpora*, 2(2):129–155, 2007.

John S Justeson and Slava M Katz. Co-occurrences of antonymous adjectives and their contexts. *Computational linguistics*, 17(1):1–19, 1991.

Hans Kamp and Uwe Reyle. *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and discourse representation theory*, volume 42. Springer Science & Business Media, 2013.

Paul Kingsbury and Martha Palmer. From TreeBank to PropBank. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC '02)*, pages 1989–1993, 2002.

Barbara Ann Kipfer. *Roget's 21st Century Thesaurus: in Dictionary Form.* The Philip Lief Group, Inc, 2005.

Karin Kipper-Schuler. *VerbNet: A broad-coverage, comprehensive verb lexicon*. PhD thesis, Computer and Information Science Dept., University of Pennsylvania, Philadelphia, 2005.

Robin Lakoff. If's, and's and but's about conjunction. *C. Fillmore and D. T. Langendoen (Eds.), Studies in linguistic semantics. New York: Holt, Rinehart, and Winston*, pages 115–150, 1971.

Alex Lascarides and Nicholas Asher. Segmented discourse representation theory: Dynamic semantics with discourse structure. In *Computing meaning*, pages 87–124. Springer, 2008.

Alessandro Lenci, Federica Busa, Nilda Ruimy, Elisabetta Gola, Monica Monachini, Nicoletta Calzolari, Antonio Zampolli, Emilie Guimier, Gaëlle Recourcé, Lee Humphreys, Ursula Von Rekovsky, Antoine Ogonowski, Clare McCauley, Wim Peters, Ivonne Peters, Robert Gaizauskas, and Marta Villegas. "SIMPLE Work Package 2 - Final Linguistic Specifications" deliverable D2.2, workpackage 2, LE-SIMPLE (LE4-8346). 2000. URL `http://webilc.ilc.cnr.it/AZ_bibliography/Z176.PDF`.

Alessandro Lenci, Gabriella Lapesa, and Giulia Bonansinga. LexIt: A Computational Resource on Italian Argument Structure. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC '12)*, pages 3712–3718, 2012.

Ganna Volodymyrivna Lobanova. *The Anatomy of Anatonymy: a Corpus-driven Approach*. PhD thesis, University of Groningen, 2012.

Edward Loper, Szu-Ting Yi, and Martha Palmer. Combining lexical resources: mapping between propbank and verbnet. In *Proceedings of the 7th International Workshop on Computational Linguistics, Tilburg, the Netherlands*, 2007.

John Lyons. Semantics, vol. i. *Cambridge: Cambridge*, 1977.

Ismail El Maarouf. Methodological aspects of Corpus Pattern Analysis. In *Proceedings of*

the 34th International Computer Archive of Modern and Medieval English Conference (ICAME 2013), Santiago de Compostela, Spain, 2013.

James MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA., 1967.

Bernardo Magnini, Emanuele Pianta, Christian Girardi, Matteo Negri, Lorenza Romano, Manuela Speranza, Valentina Bartalesi Lenzi, and Rachele Sprugnoli. I-CAB: The Italian content annotation bank. In *Proceedings of Proceedings of the Language Resources and Evaluation Conference (LREC '06)*, pages 963–968, 2006.

William C Mann. An introduction to Rhetorical Structure Theory (RST). 1999. `http://www.sil.org/mannb/rst/rintro99.htm`.

William C Mann and Sandra A Thompson. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, 8(3):243–281, 1988.

Daniel Marcu. Improving summarization through rhetorical parsing tuning. In *The 6th Workshop on Very Large Corpora*, pages 206–215, 1998a.

Daniel Marcu. *The rhetorical parsing, summarization, and generation of natural language texts*. PhD thesis, University of Toronto, 1998b.

Daniel Marcu. The rhetorical parsing of unrestricted texts: A surface-based approach. *Computational linguistics*, 26(3):395–448, 2000.

Daniel Marcu and Abdessamad Echihabi. An unsupervised approach to recognizing discourse relations. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 368–375. Association for Computational Linguistics, 2002.

Mitchell P Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. Building a large annotated corpus of English: The Penn Treebank. *Computational linguistics*, 19(2): 313–330, 1993.

Thomas Meyer and Bonnie Webber. Implicitation of discourse connectives in (machine) translation. In *Proceedings of the 1st DiscoMT Workshop at ACL 2013 (51st An-*

nual Meeting of the Association for Computational Linguistics), number EPFL-CONF-192528, 2013.

Rada Mihalcea, Timothy Chklovski, and Adam Kilgarriff. The Senseval-3 English lexical sample task. In *Proceedings of SENSEVAL-3, the Third International workshop on the evaluation of systems for the semantic analysis of text*, 2004.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

George A Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J Miller. Introduction to WordNet: An on-line lexical database*. *International journal of lexicography*, 3(4):235–244, 1990.

Eleni Miltsakaki, Rashmi Prasad, Aravind Joshi, and Bonnie Webber. Annotating discourse connectives and their arguments. In *Proceedings of the HLT/NAACL Workshop on Frontiers in Corpus Annotation*, pages 9–16. Boston, MA., 2004.

Anne-Lyse Minard, Alessandro Marchetti, and Manuela Speranza. Factuality in italian: Annotation of news stories from the ita-timebank. In *Proceedings of the First Italian Conference on Computational Linguistic (CLiC-it 2014)*, 2014.

Saif M Mohammad, Bonnie J Dorr, Graeme Hirst, and Peter D Turney. Computing lexical contrast. *Computational Linguistics*, 39(3):555–590, 2013.

Massimo Moneglia, Susan Brown, Francesca Frontini, Gloria Gagliardi, Fahad Khan, Monica Monachini, and Alessandro Panunzi. The IMAGACT Visual Ontology. An Extendable Multilingual Infrastructure for the Representation of Lexical Encoding of Action. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC '14)*, Reykjavik, Iceland, May 2014.

Simonetta Montemagni, Francesco Barsotti, Marco Battista, Nicoletta Calzolari, Ornella Corazzari, Alessandro Lenci, Antonio Zampolli, Francesca Fanciulli, Maria Massetani, Remo Raffaelli, et al. Building the italian syntactic-semantic treebank. *Treebanks*, pages 189–210, 2003.

Gregory Leo Murphy and Jane M Andrew. The conceptual basis of antonymy and synonymy in adjectives. *Journal of memory and language*, 32(3):301–319, 1993.

M Lynne Murphy. *Semantic relations and the lexicon: Antonymy, synonymy and other paradigms*. Cambridge University Press, 2003.

M Lynne Murphy. *Lexical meaning*. Cambridge University Press, 2010.

M Lynne Murphy, Carita Paradis, Caroline Willners, and Steven Jones. Discourse functions of antonymy: a cross-linguistic investigation of Swedish and English. *Journal of pragmatics*, 41(11):2159–2184, 2009.

Kim Anh Nguyen, Sabine Schulte im Walde, and Ngoc Thang Vu. Integrating distributional lexical contrast into word embeddings for antonym-synonym distinction. *arXiv preprint arXiv:1605.07766*, 2016.

Alessandro Oltramari, Guido Vetere, Isabella Chiari, Elisabetta Ježek, Fabio Massimo Zanzotto, Malvina Nissim, and Aldo Gangemi. Senso Comune: a collaborative knowledge resource for Italian. In *The People's Web Meets NLP*, pages 45–67. Springer, 2013.

Masataka Ono, Makoto Miwa, and Yutaka Sasaki. Word embedding-based antonym detection using thesauri and distributional information. In *Proceedings of the Human Language Technology Conference of the NAACL*, pages 984–989, 2015.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. The Proposition Bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1):71–106, 2005.

Martha Palmer, Hwee Tou Ng, and Hoa Trang Dang. Evaluation of WSD systems. *Word Sense Disambiguation: Algorithms and Applications*, pages 75–106, 2006.

Alessandro Panunzi, Irene De Felice, Lorenzo Gregori, Stefano Jacoviello, Monica Monachini, Massimo Moneglia, Valeria Quochi, and Irene Russo. Translating Action Verbs using a Dictionary of Images: the IMAGACT Ontology. In *Proceedings of the XVI EURALEX International Congress: The User in Focus*, pages 1163–1170, 2014.

Carita Paradis, Caroline Willners, and Steven Jones. Good and bad opposites: using textual and experimental techniques to measure antonym canonicity. *The Mental Lexicon*, 4(3):380–429, 2009.

Silvia Pareti and Irina Prodanof. Annotating attribution relations: Towards an italian discourse treebank. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC '10)*, 2010.

Andreana Pastena and Alessandro Lenci. Antonymy and canonicity: Experimental and distributional evidence. In *Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers (COLING '16)*, page 166, 2016.

Nghia The Pham, Angeliki Lazaridou, and Marco Baroni. A multitask objective to inject lexical contrast into distributional semantics. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL '15)*, pages 21–26, 2015.

Emanuele Pianta, Luisa Bentivogli, and Christian Girardi. MultiWordNet: developing an aligned multilingual database. In *Proceedings of the First Global WordNet Conference (GWC '02)*, volume 152, pages 55–63, 2002.

Emily Pitler and Ani Nenkova. Using syntax to disambiguate explicit discourse connectives in text. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 13–16. Association for Computational Linguistics, 2009.

Emily Pitler, Annie Louis, and Ani Nenkova. Automatic sense prediction for implicit discourse relations in text. 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*, pages 683–691. Association for Computational Linguistics, 2009.

Octavian Popescu. Building a resource of patterns using semantic types. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC '12)*, pages 2999–3006, Istanbul, Turkey, May 2012.

Octavian Popescu, Ngoc Phuoc An Vo, Anna Feltracco, Elisabetta Ježek, and Bernardo Magnini. Toward disambiguating typed predicate-argument structures for italian. *Proceedings of the First Italian Conference on Computational Linguistics (CLIC-it 2014)*, 2014.

Michele Prandi. *The building blocks of meaning: Ideas for a philosophical grammar*, volume 13. John Benjamins Publishing, 2004.

Michele Prandi. *Conceptual Conflicts in Metaphors and Figurative Language*. Routledge, 2017.

Rashmi Prasad, Eleni Miltsakaki, Nikhil Dinesh, Alan Lee, Aravind Joshi, Livio Robaldo, and Bonnie L Webber. The Penn Discourse Treebank 2.0 Annotation Manual. 2007.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind K Joshi, and Bonnie L Webber. The Penn Discourse TreeBank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC '08)*, Marrakech, Morocco, May 2008.

Rashmi Prasad, Aravind Joshi, and Bonnie Webber. Realization of discourse relations by other means: alternative lexicalizations. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 1023–1031. Association for Computational Linguistics, 2010.

Rashmi Prasad, Bonnie Webber, and Aravind Joshi. Reflections on the Penn Discourse Treebank, comparable corpora, and complementary annotation. *Computational Linguistics*, 2014.

James Pustejovsky. Events and the semantics of opposition. *Events as grammatical objects*, pages 445–482, 2000.

James Pustejovsky, José Castaño, Robert Ingria, Roser Saurí, Rob Gaizauskas, Andrea Setzer, Graham Katz, and D Radev. TimeML: A specification language for temporal and expressions. In *Proceedings of the Fifth International Workshop on Computational Semantics (IWCS-5)*, page 193, 2003.

Alessandro Raganato, Jose Camacho-Collados, and Roberto Navigli. Word sense disambiguation: A unified evaluation framework and empirical comparison. In *Proceedings of European Chapter of the Association for Computational Linguistics (EACL '17)*, pages 99–110, 2017.

Andrea Amelio Ravelli, Lorenzo Gregori, and Anna Feltracco. Evaluating a rule based strategy to map IMAGACT and T-PAS. In *Proceedings of the Fourth Italian Conference on Computational Linguistics (CLIC-it 2017)*, Rome, Italy, 2017. To appear.

Brian Reese, Julie Hunter, Nicholas Asher, Pascal Denis, and Jason Baldridge. Reference manual for the analysis and annotation of rhetorical structure. Technical report, University of Texas, Austin, 2007.

Ines Rehbein, Merel Scholman, and Vera Demberg. Annotating Discourse Relations in Spoken Language: A Comparison of the PDTB and CCR Frameworks. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC '16)*, Portorož, Slovenia, May 2016.

Joseph Reisinger and Raymond J Mooney. Multi-prototype vector-space models of word meaning. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 109–117, 2010.

Cornelis van Rijsbergen. *Information retrieval*. Butterworth, London, 1979.

Michael Roth and Sabine Schulte Im Walde. Combining word patterns and discourse markers for paradigmatic relation classification. In *Proceedings of the 52th Annual Meeting of the Association for Computational Linguistics (ACL '14)*, pages 524–530, 2014.

Adriana Roventini, Antonietta Alonge, Nicoletta Calzolari, Bernardo Magnini, and Francesca Bertagna. ItalWordNet: a Large Semantic Database for Italian. In *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC 2000)*, 2000.

Charlotte Roze, Laurence Danlos, and Philippe Muller. LEXCONN: a French lexicon of discourse connectives. *Discours. Revue de linguistique, psycholinguistique et informatique.*, (10), 2012.

Josef Ruppenhofer, Michael Ellsworth, Miriam RL Petruck, Christopher R Johnson, and Jan Scheffczyk. *FrameNet II: Extended theory and practice*. retrieved November 12, 2013, 2010.

Attapol Rutherford and Nianwen Xue. Improving the inference of implicit discourse relations via classifying explicit discourse connectives. In *Proceedings of the 2015 Confer-*

ence of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL HLT 2015)), pages 799–808, 2015.

Francesco Sabatini and Vittorio Coletti. Dizionario della lingua italiana 2008 (2007). *Milano: Rizzoli Larousse*, 2007.

Il Sabatini-Coletti. Dizionario della lingua italiana 2006, con CD-ROM. *Milano, Rizzoli Larousse*, 2005.

Marta Sabou, Kalina Bontcheva, Leon Derczynski, and Arno Scharl. Corpus annotation through crowdsourcing: Towards best practice guidelines. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC '14)*, pages 859–866, 2014.

Enrico Santus, Qin Lu, Alessandro Lenci, and Chu-Ren Huang. Taking antonymy mask off in vector space. In *Proceeding of The 28th Pacific Asia Conference on Language, Information and Computation (PACLIC '14)*, pages 135–144, 2014a.

Enrico Santus, Qin Lu, Alessandro Lenci, and Churen Huang. Unsupervised antonym-synonym discrimination in vector space. In *Proceedings of the First Italian Conference on Computational Linguistics (CLiC-it 2014) and of the Fourth International Workshop EVALITA 2014*, Pisa, Italy, December 2014b.

Tatjana Scheffler and Manfred Stede. Mapping PDTB-style connective annotation to RST-style discourse annotation. *Bochumer Linguistische Arbeitsberichte*, page 242, 2016a.

Tatjana Scheffler and Manfred Stede. Adding Semantic Relations to a Large-Coverage Connective Lexicon of German. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC '16)*, Portorož, Slovenia, May 2016b.

Sabine Schulte Im Walde and Maximilian Köper. Pattern-based distinction of paradigmatic relations for german nouns, verbs, adjectives. In *Language Processing and Knowledge in the Web*, pages 184–198. Springer, 2013.

Hinrich Schütze. Automatic word sense discrimination. *Computational linguistics*, 24 (1):97–123, 1998.

John Sinclair. *Corpus, concordance, collocation*. Oxford University Press, 1991.

Jennifer Spenader and Gert Stulp. Antonymy in contrast relations. In *Seventh International Workshop on Computational Semantics*, volume 3, page 100, 2007.

Manfred Stede. Dimlex: A lexical approach to discourse markers. In *Exploring the Lexicon - Theory and Computation*. Edizioni dell'Orso, Alessandria, 2002.

Manfred Stede. *Discourse processing*. Morgan & Claypool Publishers, 2012.

Manfred Stede and Silvan Heintze. Machine-assisted rhetorical structure annotation. In *Proceedings of the 20th International Conference on Computational Linguistics*, pages 425–431, Geneva, 2004.

Manfred Stede and Carla Umbach. Dimlex: A lexicon of discourse markers for text generation and understanding. In *Proceedings of the 17th international conference on Computational linguistics-Volume 2*, pages 1238–1242. Association for Computational Linguistics, 1998.

Yu Takabatake, Hajime Morita, Daisuke Kawahara, Sadao Kurohashi, Ryuichiro Higashinaka, and Yoshihiro Matsuo. Classification and acquisition of contradictory pairs using crowdsourcing. In *Proceedings of the 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 99–107, 2015.

Sandra A Thompson and Wlliam C Mann. Rhetorical structure theory: A frarnework for the analysis of texts. *IPM Papers in Pragmatics I*, (1), 1987.

Sara Tonelli, Giuseppe Riccardi, Rashmi Prasad, and Aravind K Joshi. Annotation of discourse relations for conversational spoken dialogs. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC '10)*, 2010.

Peter D Turney. A uniform approach to analogies, synonyms, antonyms, and associations. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 905–912. Association for Computational Linguistics, 2008.

Peter D Turney and Patrick Pantel. From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, 37:141–188, 2010.

Bonnie Webber. What excludes an alternative in coherence relations. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)*, pages 276–287, 2013.

Bonnie Webber, Matthew Stone, Aravind Joshi, and Alistair Knott. Anaphora and discourse structure. *Computational linguistics*, 29(4):545–587, 2003.

Bonnie Webber, Aravind Joshi, Eleni Miltsakaki, Rashmi Prasad, Nikhil Dinesh, Alan Lee, and Katherine Forbes. A short introduction to the penn discourse tree bank. *COPENHAGEN STUDIES IN LANGUAGE*, 32(9), 2006.

Bonnie Webber, Rashmi Prasad, Alan Lee, and Aravind Joshi. A discourse-annotated corpus of conjoined vps. *LAW X*, page 22, 2016.

Nianwen Xue, Hwee Tou Ng, Sameer Pradhan, Rashmi Prasad, Christopher Bryant, and Attapol Rutherford. The CoNLL-2015 Shared Task on Shallow Discourse Parsing. In *CoNLL Shared Task*, pages 1–16, 2015.

Yuping Zhou and Nianwen Xue. The chinese discourse treebank: A chinese corpus annotated with discourse relations. *Language Resources and Evaluation*, 49(2):397–431, 2015.

Zhi-Min Zhou, Yu Xu, Zheng-Yu Niu, Man Lan, Jian Su, and Chew Lim Tan. Predicting discourse connectives for implicit discourse relation recognition. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 1507–1514. Association for Computational Linguistics, 2010.