Università Ca' Foscari, Venezia
Dipartimento di Scienze del Linguaggio

Dottorato di ricerca in Scienze del Linguaggio, 22° ciclo
(A.A. 2006/2007 - A.A. 2008/2009)

Semi-automatic techniques for extending the FrameNet lexical database to new languages

Settore scientifico-disciplinare di appartenenza: L-LIN/01

Tesi di dottorato di SARA TONELLI
matricola: 955294

Coordinatore del dottorato:
Prof. Guglielmo Cinque

Tutore del dottorando:
Prof. Rodolfo Delmonte
Dott. Emanuele Pianta

Acknowledgements

Many people have contributed to my research work and, even if it is almost impossible to acknowledge all of them, I will try to mention those who supported and encouraged me most during the last three years. My gratitude goes first of all to prof. Rodolfo Delmonte, who gave me the opportunity to join his group at the Laboratory of Computational Linguistics at the University of Venice seven years ago. He introduced me to research in NLP and always involved me with enthusiasm in new activities and challenges. His curiosity and his real love for research reminds me every day that it is possible to have a young researcher's attitude even after a long and glorious career.

Special thanks are devoted also to Emanuele Pianta, who supervised my work in the Human Language Technology group at Fondazione Bruno Kessler, Trento. He has patiently helped and supported me in every research step, from the basics of writing a good article to the revision of the present work. His rigorous way of working together with his friendly manners gave me daily encouragement and were a precious example to follow.

I would like to thank also Prof. Giuseppe Riccardi, who gave me the opportunity to work with his group at the Faculty of Engineering at the University of Trento, and in particular to investigate how to apply frame semantics to dialogs. I had the privilege to be introduced to the research field of spoken dialog systems and co-author papers with him and his valuable collaborators Alessandro Moschitti, Silvia Quarteroni, Marco Dinarelli and Sucheta Ghosh. A special thanks goes to Bonaventura Coppola, who gave strong motivation to my work developing his system for automatic frame annotation and was a friendly interlocutor in every discussion.

My sincere gratitude goes also to Claudio Giuliano. Not only he gave me precious insights into Wikipedia and semantic kernels, but he has also been a friend who encouraged me every day to improve my research and be determined in my work even if results are hard to achieve. His technical skills and his creativity will always represent an excellent model to me.

I am particularly grateful also to Daniele Pighin, a precious officemate and a great friend. Starting the P.h.D. adventure and going through a number of experiences together represented a valuable encouragement and helped me to face moments of difficulty and lack of motivation. Daniele's wit, variegated interests and humor have been very precious companions in my daily work during the last three years.

Many other people have played a crucial role in my research journey, especially the colleagues who have become friends during the years. Many people from the HLT group have influenced my work giving good advice, support and encouragement, and I express my gratitude especially to Elena Cabrio, Luisa Perenthaler, Luisa Bentivogli, Milen Kouylekov and Christian Girardi. I thank also the "neighbors" of the Intelligent Interfaces and Interaction group for the nice time spent together on several occasions inside and outside FBK. A special thank goes to Alberto Battocchi, who has been a great housemate and a trustworthy friend. I would also like to mention Antonella Bristot, who was my patient confidant and valuable colleague at the University of Venice.

I am never grateful enough to my family and I tend to take it for granted, but this time I am going to make an exception. They supported me every moment and always offered their precious help in "small" but relevant matters such as picking me up, cooking for me or just being there. I thank them because they always encouraged me to find my way and be independent. The last but not the least thank you goes to Marco for making things much funnier and easier also in difficult moments, and for always reminding me that there is life outside my office.

to Ilenia

Abstract

The topic of this work is the semi-automatic development of FrameNet-like resources for new languages, with a focus on Italian. Our approach is aimed at exploiting as much as possible the theoretical backbone of the English FrameNet, and to find ways to automatically populate the language-dependent part of the database, namely the lexical unit sets and the example sentences. Indeed, the fundamental assumption of this research work is that frames as defined in the English FrameNet can be re-used for the semantic analysis and representation of other languages such as Italian. Such claim proved to be true in most Italian examples considered. The few frame definitions we added mostly depend on the fact that the FrameNet database is still under construction, and FE adjustments in Italian were not frequent. A specific section is eventually devoted to the analysis of Italian spontaneous speech following the FrameNet paradigm.

This thesis is concerned with the development of a framework for the automatic extraction of frame information for new languages starting from existing resources (parallel corpora, WordNet, Wikipedia, etc.). We focus on Italian because an official project aimed at systematically creating FrameNet for this language is still missing. Besides, the development of FrameNet for Romance languages such as Spanish and French mainly relies on manual annotation, and we believe that the framework that we propose could speed up annotation by providing data with near-manual annotation quality and allow to carry out explorative studies about new FrameNets with little manual effort. Furthermore, our experiments were carried out using publicly available multilingual resources such as the Europarl corpus (available in 11 languages), MultiWordNet (5 languages) and Wikipedia (264 languages). This makes it possible to adapt our methodology to new languages, especially as regards the use of Wikipedia for the creation of FrameNets for less-resourced languages.

With this work, we want on the one hand to contribute to the theoretical debate on the extensibility of the English FrameNet model to new languages. We discuss, *among others, issues related to cross-lingual semantic parallelism and evaluation of annotation projection. On the other hand, we propose algorithms and applications for the extraction of annotated data.*

Keywords Frame semantics, automatic development of linguistic resources, multilinguality

Contents

List of Tables

List of Figures

Chapter 1

Introduction

More than 10 years have passed since Frederick Jelinek made the well-known statement "*Whenever I fire a linguist, our system performance improves*" at the Workshop on Evaluation of NLP Systems in 1998. And things must have changed if a more recent article by the same author is entitled "Some of my Best Friends are Linguists" (Jelinek, 2005). It may not be a case that exactly in 1998 the first official FrameNet project was started at Berkeley, with the aim of giving a structured lexicographic form to Charles Fillmore's theory of frame semantics. Since then, a long list of people has contributed to the lexicographic and annotation work of the FrameNet database, showing that Fillmore's case theory gets substantial support from corpus-based investigation. With this respect, the FrameNet resource is particularly interesting because on the one hand it captures linguistic phenomena in lexicographic form, and on the other hand it offers a framework for Natural Language Processing. Such elements are a sufficient motivation for developing FrameNet in new languages. There is a double advantage: the possibility to study the characteristics of a language in the light of frame semantic theory, and the integration of such information to model knowledge in NLP tasks.

The goal of this thesis is twofold: first we investigate the applicability of the FrameNet model to new languages, and then we propose semi-automatic approaches for the development of FrameNet-like resources, with a focus on Italian. This work includes also an analysis of language-dependent issues emerging for the creation of annotated data, especially because we rely on English-Italian parallel gold standards that allow for comparative corpus-based studies. However a linguistic investigation of Italian from the frame semantic perspective is not our main concern, because most part of the work is devoted to the description of various research directions proposed for the semi-automatic development of Italian FrameNet. The title of this thesis contains the word *semi-automatic* and not *fully automatic* in order to stress that we prefer to focus on the automatic annotation of small datasets with good quality, requiring some preliminary corpus preparation or eventually a quick manual correction, than to propose methodologies for large-scale acquisition of annotated data with lower accuracy. In other words, we would rather devise approaches achieving better precision than high recall, even if they require a combination of automatic and manual effort. Our proposal is aimed at avoiding resource annotation exclusively based on manual work, which has high costs from every point of view and would require long-term projects involving a large number of researchers and annotators.

We propose three approaches for acquiring Italian annotated data. The first is the projection of frame information from English to Italian based on parallel corpora. We devise two algorithms, both rule-based, which lead to the automatic annotation of Italian sentences with full frame information (both frame and frame element labels).

The second research direction is aimed at the automatic population of Italian frames with lexical units (LUs) via MultiWordNet. We first find the WordNet synset(s) that best express the meaning of an English LU in a frame, and then acquire as good LU candidates the lemmas in the Italian version of the mapped synset(s) using MultiWordNet. In this case, we propose a supervised learning framework based on kernel methods that exploits a set of semantically rich features.

The third approach is aimed at the automatic extraction of good example sentences from Wikipedia to enrich Italian frames. This task is modeled as a word sense disambiguation problem (WSD) and makes use of a WSD state-of-the-art system to find the Wikipedia article that best corresponds to the sense of an English LU. Then, we exploit the linking and redirecting strategy of the multilingual versions of Wikipedia to populate the Italian FrameNet database thanks to the mapping between frames and Wikipedia articles.

In the light of the proposed methodologies, we believe that the creation of FrameNet for Italian can benefit from being *green* in the sense proposed by Borin et al. (2009), i.e. by re-using existing resources and exploiting the structure of different multilingual resources as a bridge to pass from English to Italian. Our goal is to demonstrate the applicability of such approaches, which we limit to Italian in our experiments but could be potentially extended to other less-resourced languages.

1.1 Innovative aspects

Since current studies on Italian FrameNet are very recent, the semi-automatic creation of such resource represents a challenging, almost unexplored field. Also a task that was already performed on other languages is likely to bring to light new characteristics if applied to Italian. In the following, we briefly describe the main contribution of this thesis, differentiating between existing approaches, newly applied to Italian, or completely novel aspects.

(a) We demonstrate the applicability of the frame-semantic paradigm to Italian, after assessing a satisfying degree of English-Italian frame parallelism in aligned annotated corpora.

(b) Even if the idea of transferring frame information between parallel corpora is not new, we apply it for the first time to Italian, devising an original algorithm and proposing a variant of an existing one (Padó and Lapata, 2009) never applied to the English-Italian pair.

(c) We carry out an original study about the impact of corpus characteristics on the transfer task.

(d) We propose a novel framework for evaluating the transfer task, after analyzing the impact of different evaluation perspectives on such task.

(e) The idea of acquiring new LUs through the mapping between FrameNet and WordNet was successfully employed in the past for several languages including Swedish and Italian (Burchardt et al., 2005, Johansson and Nugues, 2007b, DeCao et al., 2008). Anyhow, we propose a kernel-based approach relying on new features such as the information about WordNet domains (Magnini and Cavaglià, 2000) and stem overlap between LU and synset definitions. In this way, we achieve state-of-the-art results in terms of precision and coverage.

(f) We propose and motivate for the first time the idea to exploit Wikipedia as a repository for FrameNet example sentences and we build a framework that carries out the task using an off-the-shelf word sense disambiguation system (Giuliano et al., 2009). We demonstrate that the extraction process can be applied to every language available in Wikipedia and we show through the evaluation of a set of Italian sentences that accuracy of the retrieved sentences reaches almost 70%.

(g) We investigate the applicability of the FrameNet paradigm to spontaneous dialogs and achieve interesting results in comparing different levels of semantic annotation. In particular, we identify a relationship between particular dialog acts and frame labels, leading the way to the use of frame information in the development of dialog systems.

1.2 Structure of the thesis

This thesis addresses the problem of resource scarcity for Italian at frame semantic level, and in particular it investigates different approaches to acquire in semi-automatic ways annotated data for Italian. The work is basically divided into two parts: the first one is more general and describes the frame semantic paradigm, ongoing projects about FrameNet development for new languages and the main NLP applications using frame information. The second part is more experimental and presents the algorithms devised for data acquisition together with the corresponding evaluations.

The thesis is structured as follows:

Chapter 2 provides an overview about frame semantic theory and the FrameNet project for English, including statistics about the latest database release. Besides, current projects about the development of FrameNet-like resources for new languages are described, including approaches based both on manual and on semi-automatic annotation.

Chapter 3 gives a general motivation of the work presented in the thesis trying to answer the basic question "Why should we develop FrameNet for Italian?". We present relevant NLP tasks in which the frame paradigm can be applied, such as question answering, textual entailment and machine translation.

Chapter 4 introduces the first of our research directions, i.e. the transfer of frame information from English to Italian exploiting parallel corpora. First we present the state of the art w.r.t. the projection of frame information, then two transfer algorithms are described and compared, highlighting the pros and cons of every approach. The two gold standards developed for the task evaluation are detailed and compared as well. Then, we illustrate the workflow applied for gold standard annotation. Finally, three evaluation frameworks are discussed, highlighting the different system performance obtained depending on different evaluation metrics. Results from this chapter have been published in Tonelli and Pianta (2008) and Tonelli and Pianta (2009b).

Chapter 5 presents the second approach for the semi-automatic creation of FrameNet for Italian. In particular, the mapping between WordNet and FrameNet is seen as a way to populate Italian frames with lexical units imported from WordNet synsets. After giving a brief description of WordNet and MultiWordNet and explaining how WordNet and FrameNet could be interconnected, we present the state-of-the-art systems performing the mapping task. Then we detail our experiments describing the dataset, the features devised, the experimental setup and the final evaluation. In the following section, we further discuss how the FrameNet - Wordnet mapping can be applied to automatically extend FrameNet for Italian, and how it can be used to add an annotation layer to the MultiSemCor corpus. Results from this chapter have been published in Tonelli and Pianta (2009a) and Tonelli and Pighin (2009).

Chapter 6 focuses on the automatic extraction of example sentences from Wikipedia to enrich FrameNet frames in English and in Italian. After describing the algorithm devised to map FrameNet frames and Wikipedia articles, we detail the mapping experiment including the experimental setup and statistics about the acquired data. Then, we discuss how such mapping can be employed to extract example sentences for English and Italian frames. For both languages, the extraction process is described and evaluated. Partial results from this chapter have been published in Tonelli and Giuliano (2009).

Chapter 7 summarizes the main results of this thesis and outlines new research directions that are still open for future work.

Appendix A presents an explorative study about the applicability of frame semantic paradigm to dialogs with the aim of developing spoken dialog systems. After an overview of the LUNA project, which supported part of the study, the annotation process is described, giving some statistics about the annotated data and reporting the newly introduced frames. Then the relationship between frames and dialog acts is discussed. A partial description of the annotation work is described in Bisazza et al. (2008) and Dinarelli et al. (2009).

Appendix B lists the frames and the lexical units annotated in the two gold standards created for evaluating the frame transfer task.

Chapter 2

FrameNet and Frame Semantics

2.1 Theoretical background to FrameNet

2.1.1 Frame semantics

Frame semantics is an approach to the study of lexical meaning first formalized by Charles Fillmore and his collaborators and further investigated over the past forty years. A milestone of the *frame semantics* theory is Fillmore's article "The Case for Case" (Fillmore, 1968, see) presented at the symposium "Universals in Linguistic Theory" held at the University at Texas and Austin in 1967. In this work, Fillmore described a universal set of caselike relations that play a crucial role in determining syntactic and semantic relations in all languages. In particular, he described the basic structure of a sentence as being composed by a *proposition constituent P*, that is a tenseless set of relationships involving verbs and nouns, and the *modality* constituent, which includes negation, tense, mood and aspect. The P constituent can be expanded as a verb and one or more *case categories*. The latter were described as a set of universal, presumably innate concepts used by human beings to judge the events around them. In other words, a verb had to be described first in terms of the set of semantic roles contributing to create its meaning, and then with the rules needed to convert them into grammatically realized constituents. The preliminary set of cases described included just six labels: *Agentive*, *Instrumental*, *Dative*, *Factitive*, *Locative* and *Objective*.

Fillmore's later studies on lexical semantics led to the idea that a small fixed set of 'deep' roles could not describe all complementation phenomena of lexical items. For this reason, the concept of semantic roles as described in the Case Grammar was reformulated as specific 'situational' roles that codify the conceptual structure associated with lexical items (Fillmore et al., 2004b). In this new formulation, roles do not belong to a predefined list, which may fail to capture all relevant distinctions in participants between specific situations.

Fillmore further introduced the concept of *frame* from a cognitive point of view, defining the set of frames as the internal model of the world that a language-user has created by interpreting his environment (Fillmore, 1976). Frames are used as synonyms for schemata, semantic memory or scenarios, and represent the perceptual base of our knowledge that is necessary to understand the meaning of words. For example, in order to understand the predicate 'divorce', it is necessary to be familiar with a scenario where two partners get involved in some kind of social relationship, and to know the meaning of 'marry'. Another well-known example reported in Fillmore (1977) involves the difference in meaning between 'land' and 'ground', as exemplified in the following examples:

(2.1) (a) I spent three hours on *land* this afternoon

(b) I spent three hours on the *ground* this afternoon

The background situation for the first sentence is a cruise, while the second example refers to an air travel.

This characterization of frames paved the way for a frame-based organization of the lexicon. Indeed, while at conceptual level a frame characterizes the background knowledge necessary to describe a specific situation, at linguistic level it can be seen as a semantic class containing all predicates evoking such situation. This idea was assessed through a large-scale lexicographic study of the English word 'risk' by Fillmore and Atkins (1992), where the authors describe the Risk frame and its subframes Chance and Harm and analyze the lexico-syntactic patterns where 'risk' occurs as a verb and as a noun. This study proved the applicability of frame semantics to lexicographic work and represented the first example of a frame-based lexicon, which was then systematically developed in the FrameNet project.

2.1.2 Construction Grammar

The theoretical framework to frame semantics was further developed in the early eighties by Charles Fillmore, Paul Kay and other students and researchers at the University of Berkeley, who proposed a theoretical model known as *Construction Grammar* (hence CxG) (Fillmore, 1985, 1988, Kay and Fillmore, 1999). According to CxG, the basic units of language are so-called *constructions*, a repertoire of more or less complex patterns that integrate *form* and *meaning* without derivations and semantic de-compositional models. *Constructions* are the only entities being part of linguistic knowledge and no distinctions between *deep* and *surface* structures in the sense proposed by Chomsky (1957) is made. Indeed, CxG started out as a counter-movement against the successors of transformational-generative grammars, rejecting the idea that linguistic analysis should start from the simplest fragments of language and proceed gradually to more complex structures.

In this sense, CxG is a mono-stratal, non derivational grammar, implying that there are no rewrite rules from deep to shallow structure and all information resides at the same level of representation. Since no distinction is assumed between 'rules' and 'lexical items', the list of constructions include both words and clauses and all lexical, syntactic, semantic and pragmatic information is represented within a single feature structure. Another important assumption of CxG is that no forms such as active, positive, or transitive clauses, are more central than others because all linguistic units have equal value in describing the overall grammar of a language.

Goldberg (1992) explains how a construction-based approach to arguments works in the light of some examples with the verb 'kick'. She reports eight distinct argument structures for 'kick', including the following:

(2.2) (a) Pat kicked the wall.

(b) Pat kicked Bob the football.

While 2.2.a shows the common transitive use of 'kick' with two arguments, example 2.2.b reports a ternary relation expressed through a ditransitive construction of the verb. According to the constructional approach, neither the recipient role is directly provided by the meaning of 'kick' nor the verb has different meanings. Instead, the meaning of example 2.2.b is expressed by the basic sense of 'kick' AND the skeletal ditransitive construction which conveys the ternary relation. The example is the outcome of the interaction between verb meaning and construction meaning, where the former is *integrated* into the ditransitive construction which has a meaning of its own. In this way, it is possible to account for unconventional verbal constructions analyzed by Goldberg (1992) such as *Fred sneezed the napkin off the table*. In such cases, the verb describes only the agent's action, while the movement of the napkin off the table is conveyed by the so-called *caused-motion* construction, where an agent directly causes a theme to move to a new location.

The previous example can be used also to explain the relationship between construction grammar and frame semantics. According to Goldberg (1992), frames provide the encyclopedic knowledge necessary to characterize the meaning both of verbal semantics (e.g. 'sneeze' as forceful expulsion of air) and of constructional semantics (e.g. the caused-motion scenario). As explained by Östman and Fried (2004):

"Frame Semantics has become a semantic complement to Construction Grammar, as an elaboration on the relationship between form and meaning, addressed from the perspective of lexical semantic issues". (p. 5).

It is not a case that the most common objection raised to CxG is the poor predictability, which is seen also as the main drawback of frame semantic model. As a matter of fact, in order to capture all possible semantically and pragmatically relevant combinations of sense units, one should assume an inventory of some million different constructions, which is computationally and cognitively hard to achieve (see Delmonte, 2008). The same applies also to the frame semantic paradigm, which is a usage-based model where linguistic knowledge is acquired bottom-up from real examples. This can lead to a lack of generalization, because the large number of roles introduced reduce the predictive power of the model. This drawback is relevant also from a computational point of view, because the large number of role labels to be assigned has proved to be one of the main issues in developing frame-based SLR systems. This regards in particular machine-learning approaches, which would require a huge amount of training data in order to cover all possible role labels.

2.2 The FrameNet project

2.2.1 Project versions

The first large-scale project for the creation of a general frame-based lexicon for English was funded in 1997, as described in Baker et al. (1998). In this first phase, some annotation tools were developed and the first version of the FrameNet database was built. A second phase started in 2000 and led to a broader word coverage and to the annotation of larger sets of example sentences. Since then, there have been five data releases. The last one is version 1.3, dates back to 2006 and has been downloaded more than 500 times (Lönneker-Rodman and Baker, 2009). It comprises more than 10,000 lexical units, 6,000 of which are fully annotated. The resource includes also nearly 800 semantic frames with hierarchical relations, which are instantiated in more than 135,000 example sentences. The annotation work has been going on and, although no new release has been done in the last three years, the data are being continuously updated on the project website `http://framenet.icsi.berkeley.edu/index.php`. At the moment, more than 950 frames have been defined, and new lexical units are regularly added to the database[1].

[1]To see the latest project status, visit `http://framenet.icsi.berkeley.edu/index.php?option=com_content&task=view&id=17881&Itemid=66`

2.2.2 FrameNet Structure

The FrameNet conceptual model is aimed at recording all valence properties of words in different contexts to capture their different senses. For this reason, the database is organized around three main elements, which have been developed following Fillmore's theory about frame semantics:

Semantic frame: the conceptual structure that describes a particular type of situation, object or event and the participants involved in it, for example APPLY_HEAT, COLOR, JUDGMENT. It represents a common background of knowledge against which the meanings of words are interpreted. In the FrameNet database, frames come with a definition, for example ACHIEVING_FIRST is described as "A *Cognizer* introduces a *New_idea* into society", where *Cognizer* and *New_idea* are the main participants identified in the situation. The frame definitions can include different information and don't follow strict format restrictions. Indeed, they can be very short or very informative. In some cases, they include information about the syntactic behavior of lexical units, for example for the EMPTYING frame it is specified that "The area or container can appear as the direct object with all these verbs". Some definitions also add information about the realization of frame elements. We report an example from the APPEARANCE frame: "In this class of perception words, a Phenomenon, typically expressed as External Argument, and its perceptual characteristics are given some description".

Lexical unit (LU) or target: a word, a multiword or an idiomatic expression that evokes a specific frame. Differently from WordNet synsets (Fellbaum, 1998), lexical units in the same frame can belong to various grammatical categories. In the ACHIEVING_FIRST frame, for example, the LUs include verbs, nouns and adjectives: *coin.v*, *coinage.n*, *discover.v*, *discovery.n*, *invent.v*, *invention.n*, *inventor.n*, *originate.v*, *originator.n*, *pioneer.n*, *pioneer.v* and *pioneering.a*. All such LUs are supposed to evoke the same situation, described in the ACHIEVING_FIRST definition (see above).

Frame elements (FEs): frame-specific semantic roles. With verbal LUs, they are usually realized by the syntactic dependents of the verb. FEs are divided into *core*, *peripheral* and *extra-tematic*, according to how central they are to a frame. A *core* FE is conceptually necessary to the situation described in a frame because it contributes to characterize it uniquely. For example, in ACHIEVING_FIRST, *Cognizer* and *New_idea* are core frame elements because they have to be necessarily present in the event captured by ACHIEVING_FIRST, as illustrated in Example 2.3[2].

(2.3) As Son, [Jesus $_{\text{Cognizer}}$] coined [a new word for God $_{\text{New\_idea}}$].

A *peripheral* FE does not characterize uniquely a frame. Indeed, it is usually recurring in different frames and marks such notions as *Time*, *Place*, *Means*, etc. For example, in (2.4) the *Time* and the *Basis* FEs are not necessary to characterize the ACHIEVING_FIRST frame, but they add explicit information about parameters such as the time and the origin of the invention.

(2.4) The word 'pornography' was [first $_{\text{Time}}$] coined [in 1864 $_{\text{Time}}$] [from the Greek root 'porne' $_{\text{Basis}}$].

The last FE type is called *extra-thematic* and, differently from the *peripheral* type, introduces an additional state or event. For this reason, these FEs don't conceptually belong to the frame they appear in and have a somewhat independent status. They can even evoke a larger frame embedding the reported situation. In (2.5), for example, *Reason* describes an event, the 'rising of wire costs', which caused the state of affairs expressed by the target.

(2.5) Wildcom pioneered the use of short wire segments [because of rising wire costs $_{\text{Reason}}$].

For every FE identified in the example sentences, also the corresponding phrase type (PT) and grammatical function (GF) are listed. Both are automatically assigned and eventually manually corrected. The main LU categories (verb, adjective, noun and preposition) are characterized by a particular set of possible GFs. We report in Table 2.1 the GFs employed for every category (from Ruppenhofer et al. (2006), p. 91).

Every LU is characterized by patterns called **valence patterns**, which are represented as sets of triples ⟨FE label, PT, GF⟩. For each LU evoking a frame, its valence patterns are specified and the number of occurrences in the reference corpus are reported. For example, *discover.v* in ACHIEVING_FIRST occurs with the core FEs *Cognizer* and *New_idea* in 11 sentences out of 15. These 11 occurrences belong to three patterns reported in Table 2.2.

[2]Unless explicitly specified, all example sentences in English are taken from the online version of the FrameNet database. Note that we will use different fonts for the names of frames and frame elements. Specifically, frames will always be written in capitals, while frame elements will be in italics starting with a capital letter.

LU category	GFs
Verb	Ext, Obj, Dep
Adjective	Ext, Head, Dep
Noun	Ext, Gen, Dep, Appos
Preposition	Ext, Gen

Table 2.1: Grammatical functions FrameNet 1.3

Occurrences	Cognizer	New_idea
9	NP	NP
	Ext	Obj
1	NP	Srel
	Ext	Dep
1	PP[by]	NP
	Dep	Ext

Table 2.2: Valence patterns for *discover.v*

The first pattern, with 9 occurrences, corresponds to sentences like (2.6), where *Cognizer* and *New_idea* are both expressed by NPs which are respectively the subject (Ext) and the object (Obj) of the LU.

(2.6) It is 500 years since [Columbus $_{\text{Cognizer}}$] <u>discovered</u> [America $_{\text{New\_idea}}$].

In (2.7), which instantiates the second valence pattern in Table 2.2, the *Cognizer* FE is still an NP bearing the subject role, but *New_idea* is a finite relative clause (Srel) and a sentential complement (Dep) of the main verb.

(2.7) So [a team of Japanese scientists $_{\text{Cognizer}}$] appears to have <u>discovered</u> [how to weigh smells $_{\text{New\_idea}}$].

The third pattern is shown in (2.8), with *New_idea* being an NP subject (Ext) and *Cognizer* being a PP-dependent:

(2.8) Under natural conditions, [all of the six elements $_{\text{New\_idea}}$] <u>discovered</u> [by the Darmstadt center $_{\text{Cognizer}}$] are unstable.

The FrameNet framework also accounts for **polysemous and homophonous** words, which are classified as different LUs, each belonging to a frame and thus representing a separate word sense. For example, *discover.v* corresponds to two different LUs, one evoking the ACHIEVING_FIRST frame and the other the BECOMING_AWARE frame.

In order to disambiguate a LU sense, it is necessary to read the frame definition, look at the example sentences and consider other LUs sharing the same frame. An additional help comes from the lexicographic definition given for every LU in the FrameNet database. For example, we report in (2.9) the definitions for the two senses of *discover.v*:

(2.9) *discover.v* in ACHEVING_FIRST: be the first to find or observe (a place, substance, or scientific phenomenon).
discover.v in BECOMING_AWARE: become aware of (a fact or situation).

In some cases, frames, LUs and FEs are enriched with a **semantic type** label, which records some general semantic constraints. A semantic type assigned to a FE defines the type of semantic head of the constituent expressing the FE label, which has to be constant for that FE across frames. For example, the *Cognizer* FE is labeled as Sentient. This means that, in all frames where this FE is involved, such as BECOMING_AWARE, CATEGORIZATION, COGITATION, etc. the head of the constituent expressing the *Cognizer* FE has to meet the requirement of being a Sentient. Not every FE in the FrameNet database is annotated with a semantic type: the 28 labels defined for FEs are attached to 3,678 FEs out of 7,124 (Lönneker-Rodman, 2007, p. 11).

As for LUs, semantic types can be specified to highlight semantic variations within the same frame. For example, in the FRUGALITY frame, *austerity.n* and *squander.v* are labeled with the Negative_judgment semantic type, while *economical.a* and *thrift.n* are classified as Positive_judgment. These differences reflect the attitude of the *Judge*, a FE involved in the frame to assign a judgment about how a *Resource_controller* spends money for a particular purpose. In this case, the semantic type is very useful because it allows to distinguish between antonymous LUs belonging to the same frame

The semantic types for frames can be either general semantic types or framal types. The former are assigned to a frame whose LUs are supposed to bear this or a more specific semantic type. For example, the LOCALE frame bears the "Landform" semantic type, so all LUs in this frame designate the shape of a land, such as *area.n*, *grouds.n*, *place.n*, *region.n*, etc. Framal types, instead, can be assigned to a frame itself and for the moment include two labels: *non-lexical* and *non-perspectivalized* frame. The former characterizes frames that don't have any lexical unit and are inserted just to connect two or more frames semantically (Ruppenhofer et al., 2006,

p. 113). For example, the RECIPROCALITY frame was created only as a background for a number of lexical frames, including CHATTING, SIMILARITY and EXCHANGE. The non-perspectivalized label, instead, is used for frames whose LUs share a kind of scene as a background but are very different from each other. These frames don't have a consistent set of FEs and a consistent point-of-view shared by all LUs, and might be split up into smaller frames in the future. An example of this type of frame is CHANGE_OF_LEADERSHIP, which contains such diverse LUs as *coronate.v*, *coup.n*, *insurrection* and *elect*. In FrameNet 1.3, only 10 frames are non-perspectivalized.

Another important information encoded in the FrameNet database are **frame-to-frame** relations (Fillmore et al., 2004a). They are asymmetric relations involving two frames, where one is the so-called Super_frame because it is more general and less dependent, while the other, the Sub_frame, is more precise and more dependent. We report in Table 2.3 the 8 relation types with the specific names for Super and Sub_frames (from Ruppenhofer et al. (2006), p.104).

Relation	**Sub**	**Super**
Inheritance	Child	Parent
Perspective_on	Perspectivized	Neutral
Subframe	Component	Complex
Precedes	Later	Earlier
Inchoative_of	Inchoative	State
Causative_of	Causative	Inchoative/State
Using	Child	Parent
See_also	Referring Entry	Main Entry

Table 2.3: Frame-to-frame relations with Super and Sub_frames

Inheritance is the strongest relation because the Child inherits the semantics of the Parent w.r.t. FE membership (except for extra-thematic ones), FE relationships and semantic types, and frame relations. The **Perspective_on** relation indicates the presence of two different points-of-view that can be taken on a event. For example, GET_A_JOB and HIRING are perspectives of BEING_EMPLOYED. The **SubFrame** relation connects sequences of states to a more general event, expressed by a Superframe. For example, the complex frame CRIMINAL_PROCESS is linked to several other frames that correspond to different steps of the process such as ARREST, ARRAGINEMT, TRIAL and so on. Two Component frames of a single Complex frame can be related by the **Precedes** relation, like for example ARREST and ARRAIGNMENT. As for **Causative_of** and **Inchoative_of** relations, they are used to connect stative frames with their inchoative and causative version. For example, CAUSE_CHANGE_OF_CONSISTENCY has a causative relation with CHANGE_OF_CONSISTENCY, as illustrated in the following sentences:

(2.10) The ground was <u>hardened</u> by a sharp frost.
(CAUSE_CHANGE_OF_CONSISTENCY)

(2.11) The vinyl may <u>harden</u> and crack after 10 to 15 years.
(CHANGE_OF_CONSISTENCY)

The Inchoative_of relation, instead, is exemplified by the couple CAUSE_TEMPERATURE_CHANGE and INCHOATIVE_CHANGE_OF_TEMPERATURE, as shown in the following examples:

(2.12) The house had <u>cooled off</u> by a few more degrees by midnight.
(INCHOATIVE_ CHANGE_OF_TEMPERATURE)

(2.13) The chill air <u>cooled</u> her face. (CAUSE_TEMPERATURE_CHANGE)

The **Using** relation refers to cases in which part of the scene described by the Child refers to the Parent frame. For example, the COMPLIANCE frame uses OBLIGATION_SCENARIO because the former describes a particular event that can occur in the framework of the situation described by COMPLIANCE. Finally, the **See_also** relation denotes frames that are similar and need to be compared for a better comprehension of their meaning. The relation is highlighted also in the frame definitions of the Referring and the Main Entry in order to make their distinction clear. For example, the OPIONION frame has a See_also relation to AWARENESS, whose online definition specifies that "this frame is undergoing some degree of reconsideration. Many of the targets will be moved to the Opinion frame".

Frame-to-frame relations can be visually displayed with Frame Grapher (`http://framenet.icsi.berkeley.edu/FrameGrapher`), an online tool available on the FrameNet website. We report in Fig.2.1 the graph obtained for the LUCK frame, with thick arrows corresponding to inheritance relations and dotted arrows indicating Using relations.

2.2.3 Annotation workflow

The database creation in the framework of English FrameNet consists of several steps. In order to develop semantic frames, the lexicographers produce an initial informal description of the situation described by the frame and the identification of the possible frame elements. Then, some lists of words that could belong

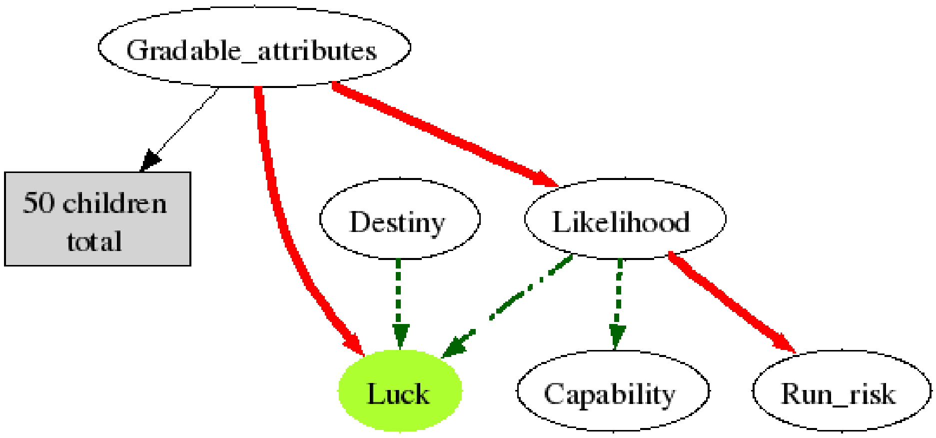

Figure 2.1: Frame-to-frame relations of LUCK

to that frame are compiled. Finally, some corpus evidence is extracted and analyzed in order to verify the consistency of the list and the valence patterns for the lexical units. This workflow was adopted at the beginning of the FrameNet project, principally during the preliminary study carried out in the DELIS project (`http://www.ims.uni-stuttgart.de/projekte/delis/`) focused on multilingual frame analyses of communication and perception verbs. With the time, the Berkeley group has developed automatic processes to extract subcorpora for annotation, to add grammatical function and phrase type labels and produce corpus-based descriptions of valence patterns. Human annotators, instead, are in charge of choosing representative instances of each LU and to annotate them with all frame information required. The sentences are taken from the 100-million-word *British National Corpus* and from U.S. news texts provided by the *Linguistic Data Consortium*. All LUs come with definitions from the *Concise Oxford Dictionary* or, if not available, with a definition written by a FN staff member.

In addition to lexicographic work, the FrameNet project has started the annotation of continuous text, the so-called **full-text annotation** (Baker, 2008). Lexicographers select one by one all content words in a text, identify a frame for each of them and then annotate the relevant constituents. On the one hand, the work was carried out on five texts taken from the PropBank corpus (Palmer et al., 2005), which allowed also for analyzing the relation between PropBank and FrameNet annotation. On the other hand, some texts about weapons programs produced for the Nuclear Threat Initiative were also annotated and made available at `http:`

//framenet.icsi.berkeley.edu/index.php?option=com_wrapper&Itemid=84.

In order to support annotation work and allow quality check, several tools have been developed by the Berkeley group. The primary annotation work is carried out using **FN Desktop** (Baker and Sato, 2003), that displays parallel aligned layers of annotation such as grammatical function and phrase type, beside the LU and the FE layer. It allows lexicographers to mark up these layers with appropriate label sets, to add new layers and to record information on them. Besides, other tools have been made available to search the database content and visualize the data. **FrameSQL** (Sato, 2003), for instance, was developed as a web-based application that carries out SQL searches from a web-browser. The tool allows the search for frames, LUs and FEs according to different criteria and connects FN data to Spanish, Japanese and German FrameNet entries. Another important visualization tool is FrameGrapher, which was briefly introduced in Section 2.2.2.

2.2.4 FrameNet Statistics

We report in Table 2.4 some statistics on FrameNet 1.3. Note that the current online version, which is the most up-to-date, contains more than 960 frames, 11,600 lexical units and more than 150,000 annotated sentences.

Total n. of frames	795
Non-lexical frames	75
N. of lexical units	10,195
Average LU polysemy	1.22
N. of annotated sentences	139,439
LU Category:	
Verbs	4,095
Nouns	4,172
Adjectives	1,770
Adverbs	63
Interjections	1
Preposition	1
N. of frame elements	7,124
N. of core frame elements	2,452
N. of peripheral frame elements	3,280
N. of extra-thematic frame elements	1,392

Table 2.4: Statistics on FrameNet 1.3

Even if the FrameNet paradigm was first developed focusing on the predicate-argument structure of verbs, nominal LUs are the most frequent category in the database, and the ongoing work is aimed at adding more adverbs and prepositions.

The polysemy value expresses the average number of frames where a LU can occur. In WordNet (Fellbaum, 1998), the average polysemy for every lemma in version 3.0 is 1.51.[3], and 83% of all words appear only in one lemma, which means that the majority of lemmas is monosemous, but that the other 17% is highly polysemous. In FrameNet, instead, 69% of LUs belong only to one frame, but the average polysemy of the rest is 2.32. This difference can help us highlight some FrameNet features: first, the low polysemy value indicates that the database is still under construction and the polysemy may be higher in the next release, with richer frame sets. Second, WordNet is more fine-grained and several synsets for a given lemma could correspond to the same lexical unit in a frame. All the attempts to map the two resources will have to take into account this feature (for further details, see Chapter 5).

We report in Fig.2.2 the distribution of example sentences per LU in the FrameNet database. The chart shows that more than 34% of all LUs have no example sen-

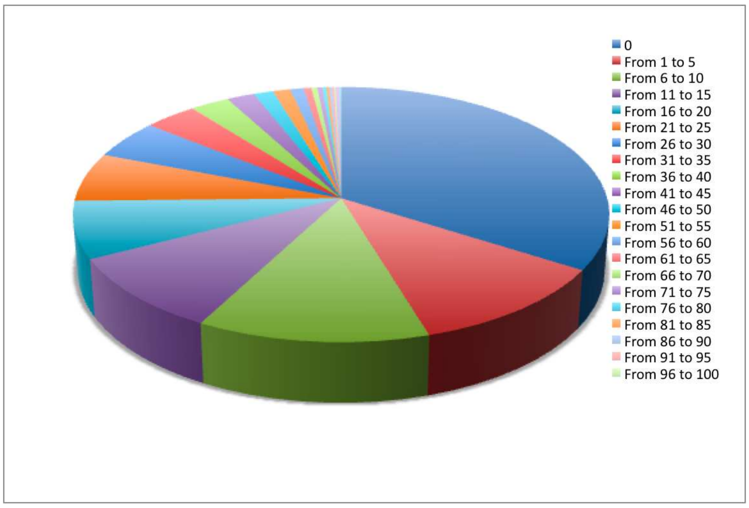

Figure 2.2: Distribution of example sentences per LU

tences, that about 11% have from 1 to 5 corpus instances, that more than 12% of all LUs have between 6 and 10 example sentences, and so on. This unbalanced distribution proves the incomplete status of the resource and shows that it is worth attempting some kind of automatic acquisition of missing information (we will propose in the following part of this work an extension via Wikipedia, see Chapter

[3]Polysemy is expressed by the average number of synsets for every lemma. WordNet 3.0 measures about polysemy are reported at `http://wordnet.princeton.edu/wordnet/man/wnstats.7WN.html`

6).

As for frame elements, we have drawn some statistics about the frequency of core, peripheral and extra-thematic roles. We report our results in Fig. 2.3, which shows how many frame elements divided into the three categories are typically part of a frame. The majority of frames tends to be characterized by 2 or 3 core FEs, while the most frequent number of peripheral FEs is 5, shared by 100 frames. As for extra-thematic FEs, they are mostly between 1 and 3. The chart shows also that the highest number of core FEs in a frame is 11 (EDUCATION_TEACHING), peripheral FEs can amount to 17 and extra-thematic ones can be a maximum of 15.

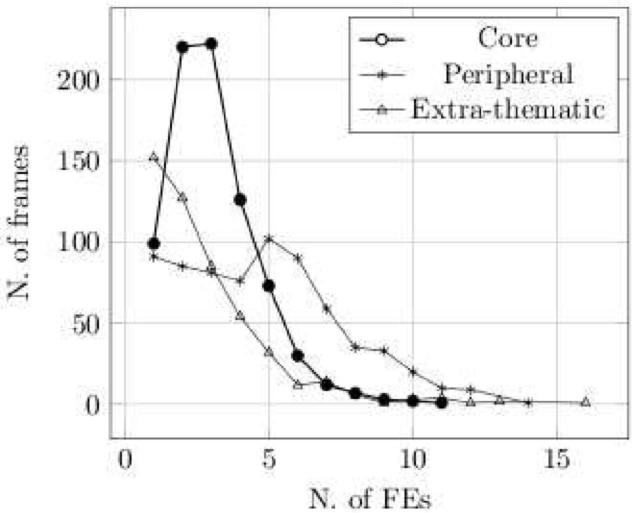

Figure 2.3: Frequency of core, peripheral and extra-thematic FEs

We computed also the most frequent core, peripheral and extra-thematic FEs. The 10 top FEs in the frequency list are reported in Table 2.5. Note that these statistics do not refer to FE occurrences in the annotated examples but to their distribution across frames.

As expected, the most frequent FEs are also the most general and peripheral ones such as *Time*, *Manner* and *Place*, which occur in more than 50% of the frames. Some FEs are very frequent both as peripheral and as extra-thematic, for example *Degree* and *Duration*. The difference between peripheral and extra-thematic role of *Degree* is shown in the following examples:

(2.14) The sailor was <u>treated</u> for venereal disease [for three months $_{\text{Duration}}$]. (CURE)

(2.15) The balloon <u>floated</u> [for hours $_{\text{Duration}}$]. (MOTION)

Core	Freq.	**Peripheral**	Freq.	**Extra-thematic**	Freq.
Agent	124	Time	461	Depictive	171
Entity	63	Manner	430	Circumstances	129
Theme	56	Place	396	Result	106
Cause	45	Means	286	Reason	97
Topic	44	Degree	266	Explanation	76
Goal	43	Purpose	186	Purpose	74
Cognizer	42	Duration	125	Frequency	52
Source	40	Instrument	77	Subregion	32
Speaker	35	Speed	44	Degree	26
Medium	30	Path	39	Duration	25

Table 2.5: 10 most frequent core, peripheral and extra-thematic FEs

In the CURE frame (2.14) and in many other state or activity frames, *Duration* is typically extra-thematic and shows a somewhat independent status, while in MOTION (2.15) it introduces some explicit information about parameters that are inherent in the evoked scene (for further examples, see Ruppenhofer et al. (2006) p. 135).

2.3 FrameNet projects for new languages

A number of research groups have expressed interest in building FrameNet for other languages and are working at the creation of such databases. The approaches are basically two: for some languages, the corpus annotation and the development of the lexical database are carried out by hand as is the case with German, Spanish, Japanese and Hebrew FrameNet. Other groups, instead, use semi-automatic and automatic approaches to create parallel lexicon fragments, for example in the French and the Chinese FrameNet. Details about the two research directions and the languages involved are reported in the following subsections.

2.3.1 Manual annotation

For German two projects are ongoing. At the University of Texas, Austin, the **German FrameNet** project (Boas, 2006)[4] aims at building a lexical resource for German starting from the Berkeley FrameNet database and cutting out all English specific information, while conceptual information about frames, FEs and their re-

[4] `http://www.laits.utexas.edu/gframenet/index.html`

lations should be left in place. The lexical database is repopulated with German-specific information using FN-Desktop and the Berkeley procedure. The example sentences are extracted from the *LDC German newspaper corpus* and the *Datenbank gesprochenes Deutsch*.

The other project for German is the **SALSA**(*Saarbrücken Lexical Semantics Annotation*) project[5] (Burchardt et al., 2009a), which is aimed at extending the existing *Tiger* treebank (Brants et al., 2002), a syntactically annotated corpus from German newspapers (about 1,5 mio words), with a frame information layer, focusing on verbal lexical units. The SALSA methodology comprises both the manual annotation of all verbs, deverbal nouns and multiword expressions in the corpus with frame semantic information and the development of techniques for the wide-coverage statistic-based semantic annotation of texts. The project started in 2002, and the first version of the annotated corpus, which was released in 2007, contains about 500 German predicates corresponding to more than 1,300 lexical units, and about 20,000 annotated instances. SALSA follows a corpus-based approach, aiming at covering all possible instances of a particular lemma in the *Tiger* corpus. Before annotation, a small sample of sentences is extracted from the corpus and analyzed in order to check FrameNet coverage and spot missing frames. In case the predicate instance is not covered by FrameNet, a lemma-specific frame is defined, the so-called proto-frame, which is not intended as a final description of the given sense but allows to be easily integrated in the annotation process, even if it doesn't show the same generalization level provided by FrameNet frames. As reported by Burchardt et al. (2009a), the average number of frames per lemma is 2.33, composed of 1.6 FrameNet frames and 0.73 SALSA proto-frames. This means that the English FrameNet could not cover about one third of the lemma senses in SALSA.

German FrameNet so far includes 1,105 frames, i.e. about 300 more than the English version, divided as follows: 609 of them are identical to the original frames (Berkeley database, version 1.2), 12 frames are slightly different because they account for richer realization possibilities in German, 453 frames are lemma-specific proto-frames and 31 are taken from FrameNet 1.3 because they are missing in the previous version.

The Assistance frame, for example, was already present in the Berkeley FrameNet but has been modified for the German version. The definition remains the same for both languages, but FEs were changed: while in the English version there were four core FEs, namely *Benefitted_party*, *Focal_entity*, *Goal* and *Helper*, the German ones include also the *Instrument* FE and replace *Focal_entity* with the more general *Ac-*

[5] `http://www.coli.uni-saarland.de/projects/salsa/page.php?id=index`

tivity role. In several cases, a new peripheral FE called *Beneficient* was introduced in the German FrameNet to account for a very frequent use of German dative in sentences like "Sie zapfte *ihm* ein Bier" (She spilled a beer *for him*).[6]

A peculiarity of the SALSA approach regards the annotation of special phenomena such as limited compositionality and underspecification. Metaphors, for example, are a special case of limited compositionality, in which frame or argument choice diverge from a straightforward mapping between syntactic and semantic structure and a figurative reading is recoverable from the literal meaning of the metaphorical expression. While for the moment the Berkeley FrameNet does not account for this kind of phenomena in a systematic way, the SALTO group has chosen to annotate metaphors with two frames, a source one corresponding to the literal meaning of the expression and the target one representing the figurative meaning. For example, the metaphorical expression "unter die Lupe nehmen" (lit. to take under a magnifying glass, i.e. to focus on) would have a double annotation: as for the literal meaning, the verb "nehmen" would be categorized as a LU of the PLACING frame, while the whole expression "unter die Lupe nehmen" would be annotated also as a LU of SCRUTINY (see Burchardt et al. (2009a), pp. 216–218). Another kind of double-annotation allowed in the SALSA corpus involves cases of vagueness in semantic annotation, where the assignment of one single label for frames or FEs would not be appropriate. In such cases, annotators can assign more than one label and link them by an underspecification link.

The **Spanish FrameNet** (SFN) project[7] (Subirats-Rüggeberg and Petruck, 2003, Subirats, 2009) is developed at the Universidad Autónoma de Barcelona in cooperation with the Berkeley FrameNet project, starting from a 370 mio. word Spanish corpus. The lexicographic and annotation work is similar to the lemma-by-lemma annotation of the English FrameNet: for every LU, a preliminary study of the occurrences of the lemma in the corpus is carried out and the grammatical constructions of such lemma are identified and extracted using regular expressions. Then, 30 sentences are randomly selected from the subcorpus containing all occurrences of the lemma. This allows to induce the set of all syntactic constructions involving the lemma and then to pre-process the 30 sentences for annotation through PoS-tagging and lemmatization. The proper annotation is then performed using the same software (FN Desktop) and database structure as in the Berkeley project.

The SFN project started in 2003 and at the moment of writing comprises about

[6] A similar case can be found also in Italian in sentences like "*Mi* apriresti la finestra?" (Would you please open the window *for me*?). For this reason, it will be useful to employ the same FE label also in Italian FrameNet.

[7] `http://gemini.uab.es:9080/SFNsite`

1,100 lexical units divided into 325 frames. Almost 10,500 sentences have been annotated, which means that every LU has about 10 attestations in the corpus. Also for Spanish, the annotation of metaphors has been particularly challenging because they cannot be interpreted literally. Following the original Berkeley FrameNet project, SFN assigns to metaphorical expressions only the frame label referring to the figurative meaning and marks such cases with a specific sentence-level tag that indicates the presence of a metaphorical interpretation. This assignment procedure differs from the one adopted in SALSA, which encodes information for both literal and figurative meaning of metaphors.

The manual creation of FrameNets for new languages does not involve only Indo-European languages; it has been extended also to languages that are not typologically related to English, for example Japanese and Chinese, which makes the study of cross-lingual parallelism much more challenging. The **Japanese FrameNet** (JFN) project[8] (Ohara et al., 2003) at Keio University started in 2000 with a preliminary investigation of the applicability of the frame-semantic approach to the analysis of the Japanese lexicon. In a pilot experiment, the frames and the FEs for the verb *osou* (*assault, attack, hit, pound, strike*) were semi-automatically identified using a bilingual corpus. The large-scale JFN project was launched in 2002 and is aimed at manually creating a FrameNet-style database of Japanese lexical units with valence descriptions and corpus attestations. The annotation so far, which has been carried out on a subcorpus from the *Kyoto University Annotated Text Corpus* (1.6 mio. words), has demonstrated that it is necessary to insert new frames and new subcategorizations with respect to the BFN frame ontology. This is due among others to differences in the English and the Japanese verb construction. Indeed, verbs in Japanese and English that are translational equivalents of each other can evoke different frames. For instance, the stative verb *lay* in the sentence "He lay on the floor" is translated in Japanese as *fall* + a resultative auxiliary, suggesting movement (Ohara, 2008).

Another project involving a language that is genetically distinct from English is **Hebrew FrameNet** (HFN) (Petruck, 2009), which is still in a preliminary phase and aims at creating an online lexical resource for contemporary Hebrew through the manual full-text annotation of sentences in a newspaper corpus. Similarly to SFN and JFN, annotation work will be carried out using FrameNet Desktop after some minor software adaptations.

In general, the 'traditional' approaches described above for different languages all consist of a lot of lexicographic work with manual annotation. The current

[8] `http://jfn.st.hc.keio.ac.jp/`

projects basically try to import or adapt English frames and FEs, while developing their own methods for creating LUs and for extracting sentences. In all projects, except for SALSA, annotation is carried out using FrameNet Desktop, the Berkeley software, which guarantees a complete compatibility and comparability between the databases in different languages. This tool encodes three annotation layers, namely grammatical function, phrase type and FE label, but only for those chunks that bear relevant frame information. The Saarbrücken group, instead, developed its own tool, SALTO (Burchardt et al., 2006), where frame information is added on top of parse trees, pointing to tree nodes. In this case, no grammatical function is encoded, while a preliminary parsing is required. Anyhow, the SALSA corpus was released in the Berkeley format in order to facilitate the linking between resources and the creation of a multilingual database with frame information. To this purpose, FrameSQL (Sato, 2003), the tool for searching the FrameNet database, was adapted in order to integrate the Spanish, Japanese and German data (Sato, 2008) and to carry out cross-lingual search.

2.3.2 Semi-automatic annotation

Beside the traditional approach, some experiments have been carried out to see if the FrameNet framework can be (partly) automatized in order to speed up data collection, example selection, pre-processing and annotation. Some approaches focus on the reuse and the merging of existing monolingual resources while others investigate the automatic mapping of frame information from English texts into new languages by exploiting existing bilingual resources. The results so far seem quite encouraging.

Johansson and Nugues (2006) have developed a frame-based semantic role labeler for **Swedish** starting from the English-Swedish parallel bitext in *Europarl* (Koehn, 2005). First, a FrameNet parser for English was created by training a kernel-based classifier on the FrameNet database and was used to automatically annotate English *Europarl* with frames. Then, the bracketing, the frames and the FEs of the English sentences were projected onto the Swedish side of the bitext using Giza++ word aligner (Och and Ney, 2003), and the resulting annotation of 100,000 sentences was exploited to train a semantic role labeller for Swedish. Besides, a research group at the University of Gothenburg is currently working at the merging of existing lexical resources for Swedish (Borin et al., 2009) into a FrameNet-like database comprising different lexica, historical dictionaries, the Swedish WordNet, the Swedish Wiktionary and the Lund University frame list.

For **German**, the SALSA group has proposed a statistic-based approach developed in parallel to the manually annotated resource and has released **Shalmaneser**

(Erk and Padó, 2006), a software for shallow semantic parsing and frames assignment with pre-trained classifiers for English and German. The system consists of a toolchain of independent modules communicating through a common XML format, whose output can be inspected with the SALTO tool (Burchardt et al., 2006).

Besides, Padó and Lapata (2009) have proposed a methodology to transfer frame annotation from an English to a German text starting from a parallel subcorpus of about 1,000 sentences extracted from *Europarl*[9]. They focus in particular on frame element transfer and model this task as a constituent alignment problem based on bipartite graph optimization. Different word alignment strategies are taken into account, as well as noise reduction techniques for constituent alignment.

A similar projection was carried out by Padó and Pitel (2007) on an English-**French** bitext. The sentences were first annotated by hand and then used to evaluate the automatic projection of frames following the projection-based framework introduced by Padó and Lapata (2005). The experiment shows that this approach has considerable potential in reducing the manual effort required to create annotated resources. Anyhow, the transfer performance obtained with the English-German pair is higher because this language pair is more semantically and syntactically similar than the English-French pair.

Another approach for the automatic transfer of frame information from English to French was proposed by Pitel (2009) and was based on a bilingual vector space built with the Latent Semantic Analysis (LSA) generalization method. This method relies on lexical similarity and can be easily applied to languages for which no preprocessing tools are available. Besides, it achieves better results than by applying the projection model by Padó and Pitel (2007) to French (F1 65.0 vs. 63.1).

A complete different strategy was presented by Chen and Fung (2004) in the **BiFrameNet** project[10]. They proposed to map LUs in the Berkeley FrameNet with entries listed in *HowNet*, a Chinese ontology with semantic relations for each word sense, using a bilingual English-Chinese lexicon. Second, they searched for monolingual Chinese sentences that contain predicates instantiating these concepts and whose POS-tag sequences are similar to those in the Berkeley FrameNet corpus. In the last step, they transferred the FEs from the English corpus to the Chinese one. This approach is based on the assumption that sentences having the same predicate and a similar syntactical structure share similar semantic roles and, unlike the other projection systems, does not rely on a parallel bilingual corpus. Despite the high accuracy value, this approach cannot be easily applied to other languages,

[9] This approach is thoroughly discussed in Chapter 4

[10] `http://www.cse.ust.hk/~hltc/BiFrameNet/`

because it requires the existence of an ontology whose design is compatible with frame distinctions made in FrameNet, and we presume that such resources are quite rare.

2.4 Summary

In this introductory chapter, we have presented an overview of frame semantics and the FrameNet project. In Section 2.1, we have outlined the genesis and the main features of frame semantics as described in Charles Fillmore's early works and we have introduced the fundamentals of construction grammar and its interaction with the frame semantic model. In Section 2.2 we have described the FrameNet project in detail: after a short introduction about the project history (2.2.1), we have illustrated the database structure (2.2.2). Then, we have described the semantic types included in the annotation and the eight frame-to-frame relations. In Section 2.2.3 we have presented the annotation workflow of the FrameNet project, describing the tools used for annotation and for data extraction. Then, in Section 2.2.4 we have presented a quantitative analysis of the data in the latest FrameNet release (version 1.3), discussing in particular the distribution of example sentences and of FE types in the corpus. The last Section deals with the main ongoing FrameNet projects for new languages and is divided into two Subsections: in 2.3.1 we detail current efforts based on manual annotation, which include the German, Spanish, Japanese and Hebrew FrameNet projects. In 2.3.2, we describe the semi-automatic approaches, dealing with Swedish, French, German and Chinese FrameNets.

This overview shows that the framework applied to the creation of English FrameNet is being largely extended to other languages. Besides, we notice that Italian is the only language among the main European languages for which this kind of project is missing. This proves that it is worth starting this new effort. In this respect, the ongoing projects on other languages show that two ways are possible, namely manual and semi-automatic annotation/database development. In this work, we focus particularly on the latter approach and we investigate ways to speed-up annotation. On the other hand, we are aware that there cannot be a release of the Italian FrameNet database without a manual validation and an accurate quality-check of the annotated data. For this reason, we believe that the best and most effective model for the creation of Italian FrameNet should be similar to the method applied for German, where a systematic, quality-checked manual annotation is carried out in parallel with some automatic processing.

Chapter 3

Is FrameNet useful?

3.1 Introduction

In recent years, many research groups have focused their work on the development of FrameNet-like resources, or on the enrichment of the existing FrameNet for English. In 2004, the 3rd Senseval evaluation campaign (`http://www.senseval.org/senseval3`) introduced for the first time the task called "Automatic Labelling of Semantic Roles" (`http://www.clres.com/SensSemRoles.html`) using frame elements as role repository (Litowski, 2004). The basic task consisted in the identification and labelling of FEs within a sentence, given the target word and its frame, and was based on the FrameNet database version 1.1. Eight teams participated in the task, achieving in some cases a precision and recall above 0.90.

In 2007, the following SemEval campaign for semantic evaluation (`http://nlp.cs.swarthmore.edu/semeval/index.php`) included the task called "Frame Semantic Structure Extraction" (`http://framenet.icsi.berkeley.edu/semeval/FSSE.html`) based on FrameNet 1.3. The annotation to be provided was more complex than in Senseval-3, including target identification and frame recognition on running text, assignment of FE labels and also the identification of the closest known frame in case a frame occurred in the test but not in the training data (Baker et al., 2007). This time, three groups participated in the frame recognition task, with best precision of 0.86 and best recall of 0.66, and two groups took part to the combined frame and FE identification task, with best precision 0.67 and best recall 0.46.

The latest SemEval campaign (`http://semeval2.fbk.eu/semeval2`), which has been launched this year and will close in April 2010, includes again a FrameNet-related task called "Linking Events and their Participants in Discourse" (`http://www.coli.uni-saarland.de/projects/semeval2010_FG/`). The main task is more similar to the Senseval 2004 proposal, including role recognition and labelling given the frame. Besides, a new subtask has been introduced, requiring the identification of links between null instantiation of roles and the wider discourse context (Ruppenhofer et al., 2009). For example, sentence (3.1), taken from Ruppenhofer et al. (2009), p. 107, shows how the identification of null instantiations should be carried out. In particular, the LU '*won*' belonging to the FINISH_COMPETITION frame is expected to have a dependent corresponding to the *Competition* FE, which in this case is uninstantiated but is expressed in the previous sentence. The task can be resolved by assigning to '*Last night's debate*' the *Competition* label.

(3.1) *Last night's debate*$_{\text{Competition}}$ was eagerly anticipated. Two national flash polls suggest that [Obama]$_{\text{Competitor}}$ <u>won</u> $_{\text{FINISH\_COMPETITION}}$ $0_{\text{Competition}}$.

This kind of task can help highlighting the interplay between local argument structure and the surrounding discourse and to extend semantic evaluation beyond sentence boundaries. In the task, both FrameNet and PropBank (Palmer et al., 2005) paradigms can be applied, in order to investigate the relationship between a more general and coarse-grained approach to semantic role labelling (PropBank) and a more specific and situation-dependent one (FrameNet).

The above mentioned evaluation campaigns demonstrate that the frame paradigm has raised more and more interest in the NLP community and show which new research directions are being investigated, for example detection of null instantiations. In this chapter we will try to point out why frame semantics can be useful in computational linguistics and why it has deserved so much attention. We will focus on some relevant questions about the utility and the applicability of the FrameNet paradigm, such as: is FrameNet useful? If yes, in which fields? What is the major contribution of this framework? What are the main problems related to it? What is the impact of poor coverage?

We will try to answer the questions above giving an overview of the studies that propose to integrate frame information in a NLP systems to solve various tasks. We will focus on three main tasks, such as question answering (QA), textual entailment and machine translation. In Section 3.2 we will discuss the ideas that originally motivated the theorization of frame semantics and we will briefly introduce some applications using frame information. Then, in 3.3 we will describe how the FrameNet paradigm has been applied to question answering and which results have been obtained. In 3.4, instead, we will give a more detailed description of the impact of frame information on textual entailment. We will introduce a recent analysis by Burchardt et al. (2009b) aimed at assessing the influence of frame semantics on the entailment task, and then present a new study carried out with VENSES (Delmonte et al., 2005), the entailment system developed at the Laboratory of Computational

Linguistics of the University of Venice. In Section 3.5 we will discuss the potential impact of frame semantics on machine translation and finally we will draw some conclusions in 3.6.

3.2 FrameNet as a general framework for Semantic Analysis

From its beginning, the frame paradigm was not just a theoretical proposal about how to interpret word senses and usages in their contexts, but had also cognitive motivations. For example, Fillmore (1976) suggested that frames could contribute to the investigation of *abstraction* capabilities in language speakers and, more generally, to the analysis of thinking strategies enhanced by the use of language. The author also proposed to study the steps of language acquisition in a child using the frame system as a parameter to see which frames are conceptual prerequisites to others. Besides, frames can help to understand cultural and historical differences between languages and more generally between societies. The author reports as an example the fact that, in a society that practices free love and where the institution of marriage is missing, there can be no word for "cuckold" and no frame describing this event.

Since then, frame information has been exploited for different purposes. Atkins (2008, 2009) highlighted the importance of FrameNet-like resources in the lexicographic work, since they can be used as a repository of the possible senses of an entry and provide a summary of the main construction patterns. Ellsworth and Janin (2007), instead, developed a system to generate **paraphrases** using syntactic and semantic information extracted from FrameNet. Another study by McConville and Dzikovska (2005) shows how to harvest a wide-coverage lexicon of English verbs from the FrameNet database to carry out deep parsing and enrich a **natural language understanding** system. Also Bos and Nissim (2008) proposed to combine FrameNet roles and the formalism of Discourse Representation Theory to carry out deep **semantic parsing**. Shi and Mihalcea (2004), instead, developed a rule-based semantic parser by integrating FrameNet and WordNet in order to identify semantic relations between words in open text. Also Giuglea and Moschitti (2006) proposed to integrate different resources, namely FrameNet, VerbNet and PropBank, to create a broad knowledge base for robust semantic parsing. All these studies confirm the utility of extensive FrameNet-like resources and demonstrate that they can effectively contribute to knowledge acquisition and representation.

3.3 FrameNet and Question Answering

One of the main fields where the FrameNet paradigm has proved to give an effective contribution is **question answering** (QA). This aspect is mentioned also in one of the first formulation of frames, since Fillmore (1976) proposed that "Comprehension can be thought of as an active process during which the comprehender - to the degree that it interests him - seeks to fill in the details of the frames that have been introduced [...] by asking his interlocutor to say more".

The first study in this direction was carried out by Narayanan and Harabagiu (2004), who highlighted the importance of semantic roles in answering complex questions and proposed to merge information from PropBank and FrameNet to deal with the task.

Kaisser and Webber (2007) went a step further by comparing the impact of different lexical resources such as FrameNet, PropBank and VerbNet on question answering. They devised two different methods for using semantic roles in QA. One exploits the data in such resources to generate answer templates and send them to a search engine. The second method, instead, performs a keyword-based search and then matches the potential answers to the dependency structure of examples found in PropBank and FrameNet. The overall accuracy achieved suggests that semantic roles contribute to obtaining better performance in QA tasks. However, FrameNet proved to have the richest encoded information but the poorest coverage with regard to PropBank and VerbNet.

Also Shen and Lapata (2007) showed that frame information can improve open-domain factoid question answering. They introduced a graph-based methodology for answer extraction and compared the contribution of frame information in the QA task to a baseline obtained exploiting solely syntactic information. The evaluation carried out on the TREC datasets showed that FrameNet-based semantic role analysis applies to about 35% of the data. This means that the extraction module of the QA system could not rely only on FrameNet to find the correct answers, although the frame paradigm provided useful information to cope with the task. On the other hand, they proved that about 19% of the questions include frames or predicates that are missing in the FrameNet database v. 1.3, and that 40% of the questions on average evoke a frame that is different from the frame evoked by the answer. In other words, the coverage problem of FrameNet is relevant but is not the major one in the QA task. In general, the authors suggested to combine the proposed model with a syntax-based system to obtain an effective improvement in performance.

Lately, Coppola et al. (2009) have proposed to apply FrameNet-like information to model **dialog managers**, claiming that such information may be useful to generalize beyond dialog applications and to learn from annotated corpora. In this respect, they developed a system for automatic frame annotation of conversational speech. The system achieves good results in the annotation task, but the development of a frame-based dialog system still remains an open issue.

3.4 FrameNet and Textual Entailment

The textual entailment recognition (RTE) task (Dagan et al., 2006) has been proposed to the NLP community in the last five years in order to develop systems able to "recognize, given two text fragments T and H, whether the meaning of one text is entailed (can be inferred) from the other text" (Dagan and Glickman, 2004). More specifically, T is defined as the entailing text, and H the entailing hypothesis. T entails H if the meaning of H can be inferred from the meaning of T according to common human understanding. We report in (3.2) an example of entailing pair from Dagan et al. (2006), with H being clearly implied by the meaning of T:

(3.2) **T**: Cavern Club sessions paid the Beatles £15 evenings and £5 lunchtime.
H: The Beatles performed at Cavern Club at lunchtime.

A few systems that participated in the RTE challenges (Burchardt and Frank, 2006, Burchardt et al., 2007) used frame semantic information about T and H to check if there was an entailing relation between the two sentences. In the light of such experiences, a study has been presented by Burchardt et al. (2009b) to assess the impact of frame semantics on textual entailment. Despite the expectations, they have shown that FrameNet coverage is not the main problem limiting the applicability of frame semantics to the entailment task. Instead, they proved that, given a frame-based semantic analysis, the main hurdle is that the current entailment systems do not deliver a knowledge model that is effective enough to cope with the entailment problem. Indeed, a model that just considers frames or FEs overlaps in T and H like in Burchardt et al. (2007) does not deliver strikingly good results, since this overlap measure does not significantly outperform other measures applied at lexical and syntactic level. Another important setback is the quality of frame semantic analysis delivered by state-of-the-art systems for automatic frame annotation like *Shalmaneser* (Erk and Padó, 2006): the complex textual material used to build RTE datasets limits the discriminative power of automatic frame annotation over the textual entailment task. Burchardt et al. (2009b) reported also some evaluations carried out on the FATE corpus of entailment pairs (Burchardt and Pennacchiotti, 2008), built over the RTE-2 challenge test, with manually annotated frame information (available at `http://www.coli.uni-saarland.de/projects/salsa/fate/`). The corpus comprises 800 *(T,H)* entailment pairs with frame and FE labels. Differently from the FrameNet corpus, annotation was carried out only on words that are intuitively *relevant* to the described situation(s), possibly skipping elements that are not central to the given context. Since only parts of the texts of entailing pairs usually contribute to the inferential process that allows to derive *H* from *T*, annotators were asked to mark these parts, called *spans*. We report in (3.3) an example of *T* and *H* extracted from the FATE corpus where the words in italics express the entailing span. Frame-evoking lexical units are underlined and followed by the label of the evoked frame.

(3.3) **T**: The holy Shiite city south of Baghdad was ravaged by *<u>fighting</u>* HOSTILE_ENCOUNTER *Thursday and Friday between American <u>forces</u>* MILITARY *and radical <u>cleric</u>* LEADERSHIP *Muqtada al-Sadr's Mahdi Army <u>militia</u>* MILITARY that left scores dead.
H: *Al-Sadr's Mehdi Army <u>militia</u>* MILITARY *and U.S. <u>forces</u>* MILITARY *have been <u>fighting</u>* HOSTILE_ ENCOUNTER.

Span information is important because it can be used to concentrate the analysis of entailment phenomena on the parts of *T* and *H* relevant for entailment. Example (3.3), for instance, shows a typical entailment case, where *H* can be completely subsumed into *T*. Note that all frame labels in H occur in the entailing span of *T*.

In order to understand to what extent the problem of knowledge modeling can impact on the performance of entailment systems, we carried out a preliminary investigation about the usefulness of frame information using VENSES (Delmonte et al., 2005), an acronym of *Venice Semantic Evaluation System*, which participated in most RTE challenges. To this purpose, we processed the RTE2 dataset with the semantic analysis module of VENSES; then we enriched the output of VENSES with frame information from the FATE corpus, and checked if this could improve the system performance on the textual entailment task.

The enrichment process was quite straightforward. In fact, frame annotation in FATE is carried out on top of parse trees, with frame and frame elements labels pointing to tokens or tree nodes corresponding to constituents. Besides, for each frame element the semantic head is explicitly indicated. This representation is compatible with the Augmented Head Dependent Indexed Structure produced by VENSES, where predicate-argument structures are represented as a dependency relation between the predicate and the head of the argument(s). For example the sentence "*Kerry says that Bush lied about Yucca Mt.*" is represented in VENSES by the following set of structures:

(3.4) subj-actor(say-cl6, 'Kerry'-sn5).
ccomp-prop(that, lie, say-cl6).
iobj-goal(about, lie-cl5, 'Mt'-sn2).
subj-actor(lie-cl5, 'Bush'-sn1).

The same sentence in FATE reports a STATEMENT frame pointing to *say*, with *Kerry* holding a *Speaker* label, and the PREVARICATION frame evoked by *lied* with *Bush* as *Speaker* and *about Yucca Mt.* as *Topic*. Thus, we could induce the following enriched annotation. Note that frame information is reported between brackets, with the frame label followed by FE label:

(3.5) subj-actor(say-cl6, 'Kerry'-sn5). [STATEMENT - *Speaker*]
ccomp-prop(that, lie-cl5, say-cl6). [STATEMENT - *Message*]
iobj-goal(about, lie-cl5, 'Mt'-sn2). [PREVARICATION - *Topic*]
subj-actor(lie-cl5, 'Bush'-sn1). [PREVARICATION - *Speaker*]

In the enrichment step, we were able to assign 69% of all FE labels in FATE with a verbal lexical unit to constituents identified by VENSES. This depends partly on errors by the system, partly on the nature of frame elements annotated in FATE. In fact, annotation was carried out by hand, thus some frame elements are not directly linked to their lexical units and/or their relationship is implicit. This kind of dependency can hardly be captured by an automatic system.

In a second step, we investigated whether the additional information could be used to improve the system output. To this purpose, we extracted from VENSES output those sentence pairs that were wrongly classified and divided them into false negatives (46,5%) and false positives (53,5%), trying to devise a different recovery strategy for every group.

For **false negatives**, that is sentence pairs that were wrongly classified as non entailing, we considered sentence pairs where H and T evoke the same frame or frames linked by a direct relationship (relationships between frames are defined in the FrameNet database). We made the hypothesis that, if the frames are matching/directly related and all frame elements of H are present in T, entailment should hold. This simple strategy for error recovery allowed us to correct 4% of false negative assignments and could cope with variants of the same expression in T and H. For example, the *Activity* frame element of the ACTIVITY_START frame has been assigned to the following constituents respectively in T and H of the same sentence pair:

(3.6) **T**: obj-theme_aff(open-cl1, relation-sn2).
H: obj-actor(start-cl2, relation-sn4).

This allowed us to assess a match between the two FEs, even if the predicate is different ('open' vs. 'start').

Although this simple matching strategy has delivered an improvement over false negatives, it would need further lexical constraints to achieve better matching precision. We explain this issue in the light of the following non-entailing example:

(3.7) **T**: Greek coastguard officials say they have found a body on a boat
H: Coastguard officials have found a dead man

Since *H* adds new information (the male body of a dead man was found, which is more specific than just a body), the entailment does not hold. However, our strategy would consider the constituents 'obj-theme_bound(find-cl2, body-sn5)' in *T* and 'obj-theme_bound(find-cl7, man-sn2)' in *H* as matching frame elements because they have the same *Sought_entity* label, and thus the pair would be wrongly classified as entailing.

As for **false positives**, that is sentence pairs that were wrongly classified as entailing, we tested a more restrictive version of the previous strategy. We considered the sentence pairs with *H* and *T* evoking the same frame, and took into account only cases where all FEs of *H* are contained in *T*. Then we checked the compatibility between *T* and *H* using constraints based on predicate unification and argument identity. For example, we analyzed with VENSES the following sentence pair:

(3.8) **T**: X-rays had been discovered [...] some years before that Hertz had discovered radio waves.
H: Hertz discovered X-rays.

The system produced an Augmented Head Dependent Indexed Structure, and we enriched it frame information from the FATE corpus. Note that we report frame information between parenthesis, with the frame label followed by FE information:

(3.9) **T**: subj-theme_unaff(discover-cl1, 'X-rays'-sn1). [BECOMING_AWARE1 - *Phenomenon*]
obj-theme(discover-cl2, wave-sn10). [BECOMING_AWARE2 - *Phenomenon*]
subj-theme_unaff(discover-cl2, 'Hertz'-sn9). [BECOMING_AWARE2 - *Cognizer*]

(3.10) **H**: subj-actor(discover-cl3, 'Hertz'-sn1). [BECOMING_AWARE - *Cognizer*]
obj-theme_unaff(discover-cl3, 'X-rays'-sn2). [BECOMING_AWARE - *Phenomenon*]

We could map both FEs in H with two FEs in T, specifically those referring to the same predicate ('discover', with index *cl-2*). Without any lexical constraint, the frame information available would lead us to assess an entailing relation between H and T because there is a case of frame identity and all FE labels in H are present in T. Instead, we should look for an identity relation between the mapped arguments, in particular between their heads. Only a constraint identifying the incompatibility between 'X-rays' in H and 'radio waves' in T would help us exclude an entailing relation.

In conclusion, our study on VENSES output confirms the assessment by Burchardt et al. (2009b): frame information can convey a semantic layer useful for the interpretation of T and H. The main problem is that no strategy for modeling this kind of knowledge has proved to be strikingly successful so far. At the moment, most entailment systems integrate lexical information from WordNet and similar resources because they can be straightforwardly implemented in such systems exploiting synonymy and is-a relations. We believe that a strategy for exploiting frame information in the entailment task would be successful only if it could effectively integrate the rich set of information encoded in the database such as LU valence patterns, frame-to-frame relations and semantic types.

3.5 FrameNet and Machine Translation

One of the most challenging but less explored applications of frame semantics is machine translation. Even if many researchers working on FrameNet claim that machine translation is a very promising field in which frame information can be successfully employed, to our knowledge no such MT system has been developed yet.

Most work in this direction has dealt with the *lexicographic* side of MT frameworks, namely the development of FrameNet-based dictionaries for machine translation (Boas, 2002). The idea was to create dictionary entries for the English lexical units and link them with their counterparts in another language, in order to create an electronic resource of translation equivalents. The syntactic and semantic information of this contrastive database could then be applied to automatically determine the best translation for a given lexical unit. The idea was further developed by Boas (2005b), who proposed a corpus-based procedure to create parallel lexicon fragments for Spanish, German and Japanese, provide them with the same kind of information of the English entries and link the multilingual records to each other via semantic frames. In this approach, frames are seen as a kind of interlingual representation, particularly those describing very general and language-independent scenarios and events (e.g. JUDGMENT, STATEMENT, APPEARANCE).

However, the issue of integrating FrameNet lexical resource in an MT system and of improving translation quality is left open. The *generalization* degree of FrameNet represents a main issue which we detail in the remainder in the light of some examples.

In some cases, FrameNet frames offer some level of generalization over individual lexical items, since frames do not contain only synonyms belonging to the same grammatical category, like WordNet synsets. Indeed, they include LUs that can have related meanings but different category, for example *call.v* and *call.n* in CONTACTING. This aspect can be exploited in MT systems because it could be used for cross part-of-speech synonymic translations. For instance, an MT system could produce as equivalent translations "*to call someone*" and "*to make a call to someone*", or to recognize such expressions as synonymic in English source sentences. The generalization degree offered by frames allows also for taking into account translations that may imply a departure from the source meaning. In other words, it could license free translations, for example the Italian *chiamare.v* (*to call*) could be translated as *reach (by telephone)*, considering that the CONTACTING frame includes *to call*, *to reach*, *to contact* and many other LUs.

The degree of generalization offered by some frames can also turn into a disadvantage, because in some cases the abstraction level is to high. Consider for example the CLOTHING frame, in which several garments are listed. It is clear that such LUs are not interchangeable, and that it is not correct to use *bikini.n* or *blazer.n* as synonyms for a translation, even if they appear in the same frame. The same applies to frames containing antonymous lexical units, such as *activate.v* and *deactivate.v* or *turn_on.v* and *turn_off.v* in the CHANGE_OPERATIONAL_STATE frame.

In some other cases, the generalization degree offered by frames is not sufficient to capture translation equivalence. An example is reported below. The parallel sentences are extracted from the English-Italian bitext of *Europarl*. The frame label has been manually added and is reported between parenthesis.

(3.11) Let me **say** it again quite clearly, we have not brought up the question of privatization. [STATEMENT]
Desidero ancora una volta **sottolineare** che non abbiamo affrontato la questione della privatizzazione. [CONVEY_IMPORTANCE]
I want to underline once again that we have not brought up the question of privatization.

Under a frame-semantic perspective, the English sentence is classified as belonging to the STATEMENT frame, with *say* being the corresponding lexical unit. The Italian translation, instead, is an example of CONVEY_IMPORTANCE frame evoked by the lexical unit *sottolineare* (*underline*). Example 3.11 clearly corresponds to a free translation which is however correct, because the meaning of the Italian sentence and of the English one are equivalent. In this case, the FrameNet paradigm cannot capture such semantic equivalence, because two different frames are assigned. The main problem is that the English lexical unit is *say* and not the phrase "*say it quite clearly*". For the moment, such phrases are not taken into account in FrameNet as frame-evoking elements, and the only multiword expressions considered as lexical units are idioms and phrasal verbs.

Our analysis shows that lexicographic work aimed at connecting lexical units in different languages through frames has been investigated and has proved to be useful to develop electronic dictionaries. If we consider the problems connected to different generalization degrees in FrameNet frames, we can conclude that such dictionaries can be successfully implemented in MT systems only if they are developed on a per-LU basis, connecting LUs in different languages that are translation equivalents. Single frames are either too general or too fine-grained, which makes it impossible to exploit information about frames in a consistent way to generate translations.

3.6 Summary

In this chapter, we discussed the motivation of our research work, in particular we explained why it is worth developing FrameNet for new languages and extending the coverage of existing FrameNet-like resources. In the light of some examples, we showed that FrameNet is considered a relevant resource by the NLP community, as confirmed by the last Senseval/SemEval campaigns. Besides, we described some applications that could benefit from the integration of frame information, such as question answering, textual entailment and machine translation. We described the state of the art regarding frame information in the three fields and we presented some original contributions about the integration of frame semantics in the VENSES entailment system. As for machine translation, we pointed out why the integration of frame information in MT systems is for the moment still 'potential' and we discussed some generalization issues related to frames.

Chapter 4

Frame information transfer from English to Italian

4.1 Introduction

In this Chapter, we present the first of the three approaches proposed for the semi-automatic creation of FrameNet for Italian, namely the *annotation transfer* from English texts. The basic idea is to use a parallel corpus where the English side has been annotated with frame information and to transfer the annotation onto the other side, i.e. the corpus in a new language. The transfer methodology is one of the most common unsupervised approaches applied to create new FrameNets by exploiting the existing English database and has been employed for Swedish, German and French (see Section 2.3.2). Besides, it has been applied in the past also to the projection of a wide range of linguistic information, from parts of speech and chunks (Yarowsky and Ngai, 2001) to coreference chains (Postolache et al., 2006) and WordNet synsets (Bentivogli and Pianta, 2005). The annotation projection across languages, especially involving frame information, is based on the assumption that translation preserves semantic information. This is in most cases true, in particular for the languages that are typologically similar, but there are some exceptions that will be discussed in the next sections.

For Italian, we have developed and evaluated two different transfer algorithms. Besides, we have created two English-Italian parallel corpora with different characteristics in order to assess the impact of the corpus on transfer quality. Finally, we have discussed some issues about evaluation strategies adopted in the past and we have proposed our own evaluation framework.

The Chapter is organized as follows: in Section 4.2 an overview of the main studies available about frame annotation projection is given, while Section 4.3 introduces the general transfer task applied to Italian. Then, in Section 4.4 we describe the two algorithms developed for annotation projection. In Section 4.5 we detail and compare the characteristics of the gold standards created for evaluation. Then, in Section 4.6 three evaluation methodologies are presented: the first and the second one are existing approaches commonly applied to transfer experiments, while the third one is our proposal for a new evaluation under real-world conditions. Finally, we discuss our contribution to the transfer task and draw some conclusions in Section 4.7.

4.2 Related work

As described in Section 2.3.2, several approaches developed for the creation of FrameNet-like resources are based on annotation projection. In particular, Padó and Lapata (2009) use an English-German parallel corpus for creating semantic alignments between constituents and transferring *manual annotation* from source to target language. The search for the best alignment is seen as an optimization problem of graph matching. The authors test different filtering techniques for constituent alignment and several word alignment models, achieving very promising results (Precision 86.6, Recall 75.2, F1 80.5). However, evaluation focuses only on parallel sentences with the same frame label, and the sentences in the gold standard have been selected from the EUROPARL corpus in order to maximize their semantic similarity. Besides, evaluation concerns only FE transfer and does not take into account frame-evoking lexical units. The same projection model was applied also to English-French parallel texts (Padó and Pitel, 2007), achieving with the best model Precision 66.2, Recall 60.3 and F1 63.1.

A related approach was used for the English-Swedish pair by Johansson and Nugues (2006), who however did not start from manual annotation. Indeed, they automatically annotated the English side of the whole EUROPARL corpus using a kernel-based classifier trained on the FrameNet database. The annotation transfer was based on word alignment carried out by the Giza++ tool (Och and Ney, 2003). Alignment was then combined with heuristics to identify the correct constituent bracketing. Then, the transfer output, which was basically the Swedish side of EUROPARL enriched with frame information, was used to train a semantic role labeller for Swedish. Since the creation of the frame labeller for Swedish was the main goal of the project, no manual gold standard was developed for the evaluation of the transfer task, so the results achieved by the system can be hardly compared to those by Padó and Lapata (2009) and Padó and Pitel (2007). Besides, the close relationship between English and Swedish probably made the task easier.

As for Italian, a few projects are currently aimed at exploring new approaches to speed up manual annotation or convey fully automatic annotation. Basili et al. (2009) have proposed a methodology to automatically transfer frame information on an English-Italian parallel corpus based on a statistical machine translation step augmented with a rule-based post-processing. Results will be thoroughly commented in the Evaluation Section. Coppola et al. (2009) have trained and tested a system for automatic frame element detection using a corpus of Italian dialogs manually annotated with frame information.

4.3 General transfer framework

The task of transferring frame annotation between two languages using a parallel corpus is generally composed of several steps which include some pre-processing and annotation of the source corpus, the word-level alignment of the parallel texts, then the induction of constituent alignment, and finally the transfer of annotation from source to target language. At some points, this framework can present variants, for example the source corpus annotation can be manual, as in our case, or automatic (Johansson and Nugues, 2006). Other possible options involve the strategy for the induction of constituent alignment, for instance it can be graph-based (Padó and Lapata, 2009) or rely on language-specific heuristics (Johansson and Nugues, 2006).

In Figure 4.1 we report the general transfer workflow adopted for our experiments on the English-Italian pair:

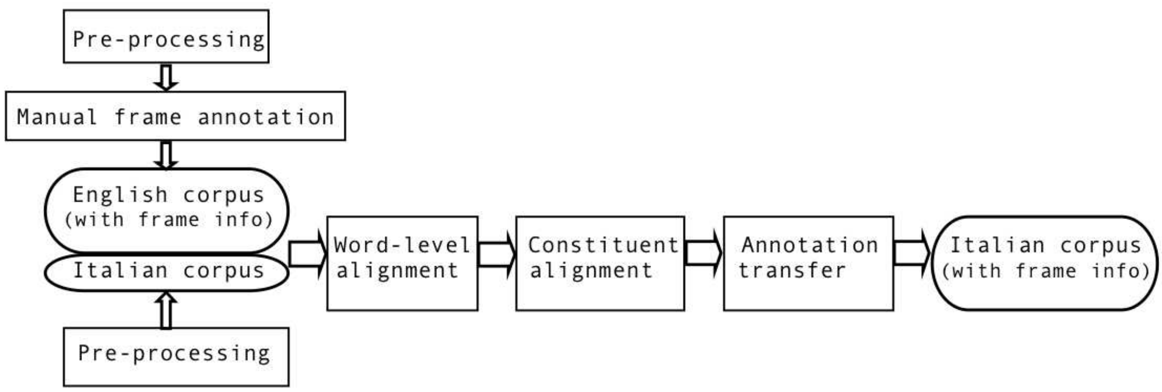

Figure 4.1: Annotation transfer workflow

Since some syntactic information is generally needed for the transfer task, the pre-processing usually comprises part-of-speech tagging and parsing or chunking of both corpora. Then, the English side of the parallel corpus is manually annotated with frame information. Although source corpus annotation is usually carried out by hand because no automatic system for frame annotation has proved to deliver near-manual annotation quality so far, automatic annotation would greatly increase the amount of data suitable for the transfer, and for this reason this strategy has been investigated in several works (Padó, 2007, Padó and Pitel, 2007, Johansson and Nugues, 2006). However, since our main concern is the creation of an annotated corpus for Italian with high-quality frame information, we focused only on experiments based on manual annotation of the source text.

After pre-processing, the parallel sentences are automatically aligned at word level, and this alignment output is used to develop some strategy for constituent alignment. The latter can exploit word alignment only or, when available, also syntactic structure information. In our framework, we chose to introduce and exploit syntactic information because it proved to yield substantial improvements over relying on word alignment alone (Padó and Lapata, 2005). To this purpose, we developed and tested two different constituent alignment strategies: one exploits syntactic parsing and semantic head alignment, while the other takes into account some syntactic information but is mostly based on word alignment. Details are given in Section 4.4.

After all words and constituents bearing some frame information in English have been coupled with the corresponding words and constituents in Italian, the semantic labels can be transferred, so that in the end we obtain an Italian corpus with frame information.

4.4 The Transfer Algorithms

In the following sections the two algorithms for frame information transfer are described and compared. Note that they differ in the constituent alignment step, whereas the alignment of frame-evoking lexical units is the same in both algorithms and exploits simple word-alignment.

4.4.1 Algorithm 1

The first transfer experiments were carried out applying a new algorithm that uses the semantic head of the English frame element as a bridge for constituent alignment. The starting point is the availability of a parallel English-Italian corpus, with the English side being fully parsed and annotated with frame information. The algorithm is based on four steps: first, the Italian text is parsed, then the parallel sentence pairs in English and Italian are aligned at word level, then the semantic head of every annotated constituent in the English corpus side is automatically extracted and finally annotations are projected from English to Italian constituents using aligned semantic heads as a bridge. The four transfer steps are described in the following subsections.

Italian corpus preparation

Italian texts are first parsed with Bikel's phrase-based statistical parser trained on the Turin University Treebank (Bosco, 2007) in constituency format (Corazza et al., 2007)[1]. The parser delivers a shallow syntactic analysis based on a small set of constituency labels.

Word alignment

The English-Italian corpus is aligned at word level with KNOWA (KNowledge-intensive Word Aligner) (Pianta and Bentivogli, 2004), a word aligner relying mostly on information contained in the Collins' bilingual dictionary, but also on a morphological analyzer and a multiword-recognizer. We chose KNOWA because with this language pair it outperforms GIZA++, in particular w.r.t. alignment of content words (85.5 precision vs. 53.2 of GIZA++ in the EuroCor task, which was carried out on a subset of English and Italian texts from Europarl as reported in Pianta and Bentivogli (2004)). This is important because our algorithm relies on information projection between semantic heads, which are mostly content words.

Semantic head extraction

The strategy devised for constituent alignment is based on the *semantic head*, which is the word that bears the main semantic features of the constituent. We took into account semantic heads because the definition of frame elements in the FrameNet database is more focused on semantics than on their syntax. Another reason for choosing semantic heads as transfer bridge is that KNOWA aligner achieves a particularly good performance on content words, the typical category of semantic heads, while it is poorer on function words. This implies that our preference for semantic heads should have a positive influence also on the task performance. Furthermore, FEs are often assigned to semantic types. So, if we were able to couple Italian semantic heads with FE semantic types, this could be used as a filter for the alignment

[1]The parser developed by Corazza et al. obtained the best score in the 1st EVALITA evaluation campaign for Italian NLP tools with 67.97 f-measure.

of source and target semantic heads.

The basic idea of annotation transfer via semantic heads is the following: annotations on the English side refer to syntactic constituents such as NP, VP, PP etc., which are maximal projections of a given lexical category; any such constituent has only one semantic head, and we expect that the Italian translation of such head be the semantic head of the Italian phrase corresponding to the English annotated constituent.

Since the English corpus is PoS-tagged and parsed with Collins' parser, we implemented his algorithm for head extraction (see Collins, 1999, pp. 238-240) which, for every non-terminal node, indicates a list of child-nodes that can be the semantic head of the parent, sorted according to a priority ranking. A direction is also specified to indicate if the search should start from the left or the right end of the nodes dominated by the given non-terminal. The head-extraction rules used in our algorithm are a combination of Collins' rules and those implemented in the CHAOS parser for English (Basili and Zanzotto, 2002)[2] Besides, we added rules for subjectless sentences (SG) and basal NP nodes (NPB and NX), which were missing in the original head extraction rules. For details, see Table 4.1:

Parent Non-terminal	Direction	Priority List
ADJP	left	JJ, NNS, QP, NN, $, ADVP, VBN, VBG, ADJP, JJR, NP, JJS, DT, FW, RBR, RBS, SBAR, RB
ADVP	right	RB, RBR, RBS, FW, ADVP, TO, CD, JJR, JJ, IN, NP, JJS, NN, NNP
CONJP	right	CC, RB, IN
LST	right	LS, ":"
NAC	left	NN, NNS, NNP, NNPS, NP, NAC, PRP, EX, $, CD, QP, PRP, VBG, JJ, JJS, JJR, ADJP, FW
PP	right	NP, NPB, VP, S, VBG, VBN, RP, FW, SBAR, SG, TRACE, PP, IN, TO
PRT	right	RP
QP	left	$, IN, NNS, NN, JJ, RB, DT, CD, NCD, QP, JJR, JJS
RRC	right	VP, NP, ADVP, ADJP, PP

[2]Thanks to Fabio Massimo Zanzotto for providing us with the extraction rules for semantic heads developed for CHAOS.

Parent Non-terminal	Direction	Priority List
S	left	VP, S, SBAR, ADJP, UCP, NP, NPB, TO, IN
SG	left	VP, S, SBAR, ADJP, UCP, NP, TO, IN
SBAR	left	S, SQ, SINV, SBAR, Ss, FRAG, WHNP, WHPP, WHADVP, WHADJP, SG, IN, DT
SBARQ	left	SQ, S, SINV, SBARQ, FRAG
Ss	left	VP, S, SBAR, ADJP, UCP, NP, NPB, TO, IN
SINV	left	VBZ, VBD, VBP, VB, MD, VP, S, SINV, ADJP, NP
SQ	left	VP, VBZ, VBD, VBP, VB, MD, SQ
VP	left	VBD, VB, VP, VBP, VBZ, VBG, VBN, MD, ADJP, NN, NNS, NP, NPB, AUX, TO
FRAG	left	WHADVP
WHADJP	left	WRB, JJ, ADJP, CC
WHADVP	right	WRB, CC
WHNP	left	WDT, WP, WP$, WHADJP, WHPP, WHNP
WHPP	left	WDT, WP, WHNP, FW, IN, TO
NP	right	NN, NNP, NNPS, NNS, NPB, NX, JJR, NP, PRP, ADJP, PRN, CD, JJ, JJS, QP, RB, $, POS
PRN	right	NN, NNP, NNPS, NNS, NPB, NX, JJR, NP, PRP, ADJP, PRN, CD, JJ, JJS, QP, RB, $, POS
NPB	right	NN, NNP, NNPS, NNS, NPB, NX, JJR, NP, PRP, ADJP, PRN, CD, JJ, JJS, QP, RB, DT, WDT, EX, $, POS, IN
NX	right	NN, NPB, NNP, NNPS, NNS, NX, POS, JJR

Table 4.1: Rules for semantic head extraction

Cross-lingual transfer

Frame information is conveyed by two different components: lexical units are usually single words or multiword expressions, whereas frame elements are usually expressed by more complex constituents. For this reason, the cross-lingual transfer is two-fold. The transfer of frame targets involves only a lexical unit, usually a verb, on both sides of the corpus, so it relies basically on word alignment. This implies that the target transfer performance is strongly influenced by the word aligner performance. Instead, a different transfer strategy is required for frame elements.

In case of algorithm 1, we developed the following original procedure: after extracting the semantic head of the English constituent annotated with frame element information, we get the Italian aligned semantic head, when available; then, we find the highest syntactic projection of the Italian head compatible with the annotated English constituent. Finally we transfer annotation from the English maximal projection to the Italian constituent. We define a table of compatibility between English and Italian constituents, assessing for example that the same FE can be realized by two different constituent types in the two languages. For instance, English NPs can correspond to either NPs or PPs in Italian.

The transfer procedure with algorithm 1 is exemplified in Figure 4.2.

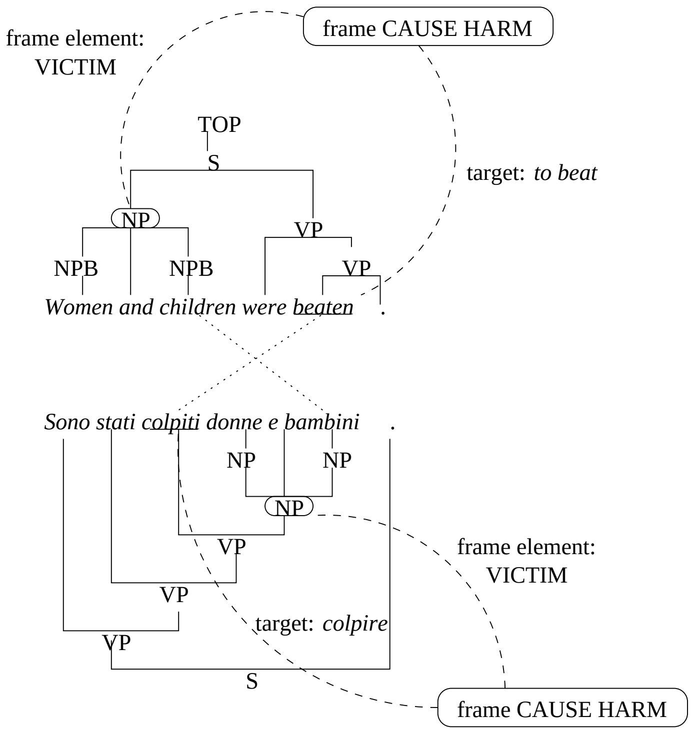

Figure 4.2: Example of cross-lingual transfer with Algorithm 1

We start from the English sentence syntactically parsed and manually annotated and the corresponding Italian translation. In this case, "*beaten*" is the lexical unit evoking the CAUSE_HARM frame and goes with a FE labeled as *Victim*, i.e. "*Women and children*". The Italian translation is faithful, even if its syntactic structure is different from the English version because the subject is postponed.

First, the Italian sentence is syntactically parsed, so that we obtain the constituent tree displayed below the sentence. Then, the parallel sentences are automatically word-aligned. In the example, dotted lines connect aligned words, so that "*beaten*" is aligned to "*colpiti*" and "*children*" to "*bambini*".

As for the lexical unit, this step is enough to infer that "*colpiti*" in the Italian sentence is the target word evoking the CAUSE_HARM frame, and more generally to infer that the Italian verb "*colpire*" can evoke this frame.

As for the frame element, we first identify "*children*" as the semantic head of "*women and children*" by applying the rules in Table 4.1. In particular, we consider the rule defined for NP maximal projection and we start from the right-end of the children sequence to look for a semantic head according to the priority list. We discard NN, NNP, NNPS and NNS because they are not among the children nodes, but we stop our search at "*bambini*" because NPB is a possible head as listed in the ranking list. Then, since "*children*" is aligned with "*bambini*", we find the highest syntactic projection of the head compatible with the annotated English constituent, i.e. the uppermost NP. We stop the search at the NP node dominating "*donne e bambini*" because the upper VP node would include the target word and would not have been compatible with the English NP, as defined in our algorithm. So, "*donne e bambini*" is automatically annotated with the *Victim* label.

With this strategy, only the head of the constituent is required to be correctly aligned in order to carry out the whole FE transfer.

4.4.2 Formalization of Algorithm 1

The formalization of algorithm 1 is reported in Figure 4.3.

We take the English corpus annotated with frame information C_{en} and align it at word level to the Italian corpus C_{it}, whose sentences have been previously parsed. In the initial state, two parallel and aligned sentences are considered, s_{en} and s_{it}, with s_{en} being annotated with a frame label $info_{frame}$ pointing to some lexical unit, and with a set of frame element labels FE_{en} pointing to its constituents. The algorithm is divided into two sub-procedures, one for LU alignment and the other for FE alignment, which are independent.

The first step is quite straightforward: if the English target word $lexunit_{en}$ is aligned with an Italian word, we assume that the latter is the Italian lexical unit

Given two aligned sentences s_{en} and s_{it}, s_{en} annotated with $info_{frame}$ and FE_{en}
$take\ lexunit_{en} \in s_{en}$
$if\ exists\ alignment_{lexunit_{en}}$
  $take\ aligned\ lexunit_{it} \in s_{it}$
  $transfer\ info_{frame}\ from\ lexunit_{en}\ to\ lexunit_{it}$
  $return\ lexunit_{it+infoframe}$
$else$
  $return\ false$
$for\ each\ fe_{en} \in FE_{en}$ pointing to c_{en}
  $extract\ semantic\ head\ he_{en}$
  $if\ exists\ alignment_{he_{en}}$
    $take\ aligned\ semantic\ head\ he_{it} \in c_{it}$
    do
      $take\ upper\ constituent\ node\ c_{it}$
    $until\ c_{it}\ is\ compatible\ with\ c_{en}$
      $Cand_{best} = c_{it}$
      $transfer\ fe_{label}\ from\ fe_{en}\ to\ Cand_{best}$
      $return\ Cand_{best+fe}$
  $else$
    $return\ false$
$end\ for$

Figure 4.3: Transfer algorithm 1

$lexunit_{it}$, so we transfer the frame label $info_{frame}$ from $lexunit_{en}$ to $lexunit_{it}$.

In the FE transfer step, for every English frame element fe_{en} bearing a FE information fe_{label}, we consider the constituent c_{en} it is pointing to and we extract its semantic head he_{en}. If he_{en} is aligned with an Italian word, then we assume that the latter is the semantic head he_{it} of an Italian constituent $Cand_{best}$ to be identified. We search such constituent by visiting every upper constituent node c_{it} dominating he_{it} and checking if it is compatible with the English FE node fe_{en}. We carry out the search until the compatibility is assessed. In this case, the visited node c_{it} is the node that should become the Italian frame element $Cand_{best+fe}$ aligned with fe_{en}.

4.4.3 Algorithm 2

We present here a second transfer algorithm which is more similar to the best model presented by Padó and Lapata (2009) in that the alignment between constituents is not based on the semantic head but on the best percentage of aligned words and takes into account the syntactic dependents of the target word through argument selection. However, we propose an algorithm for the contextual transfer of target and FE information from English to Italian, while Padó and Lapata focus on FE transfer only[3].

The steps required for the complete frame information transfer are the following. Note that the first and the second ones are the same as for Algorithm 1. In this case, we start from the English side of the parallel corpus being chunked and manually annotated with frame information.

Italian corpus preparation

Italian texts are parsed with Bikel's phrase-based statistical parser trained for Italian.

Word alignment

The English-Italian corpus is aligned at word level with KNOWA.

Target transfer

In this algorithm, target and FE transfer are not carried out independently. Indeed, the former is necessary for FE alignment and transfer. Like in algorithm 1, the frame label is straightforwardly transferred from the English to the Italian target relying on word alignment.

FE alignment

This procedure is aimed at aligning English FEs and Italian constituents. Given that in the previous step the Italian target word has been identified, we first extract its syntactic dependents following different selection rules constrained by the grammatical class of the target word. Then, we compute the number of aligned words that every English FE shares with each of such dependents. The Italian constituent sharing the highest number of aligned words with a given English FE is the best candidate for alignment at constituent level and for bearing the FE label in the Italian sentence.

[3] In the work proposed by Padó and Lapata (2009), it is not clear how the argument selection step is carried out, in particular how it is dealt with if the target transfer is wrong or missing.

FE transfer

The FE label of every English constituent that has been aligned with an Italian constituent is transferred from the source to the target constituent. In the end, the Italian sentence bears both frame and FE information.

The transfer procedure with algorithm 2 is exemplified in Figure 4.4.

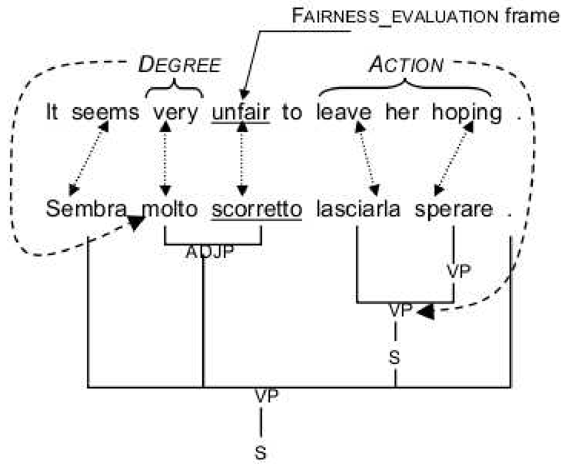

Figure 4.4: Example of cross-lingual transfer with Algorithm 2

In the initial step, the English sentence is chunked and frame information is manually added. In this case, "*seems*" is the lexical unit of FAIRNESS_EVALUATION and has two FEs corresponding resp. to the *Degree* label, i.e. "*very*", and to the *Action* label, i.e. "*leave her hoping*". The Italian translation is faithful to the original sentence. The only difference lies in the presence of the expletive pronoun "*it*" in English, which is missing in Italian.

First, the Italian sentence is syntactically parsed, so that we obtain the constituent tree displayed below the sentence. Then, the parallel sentences are automatically aligned. In the example, dotted lines connect aligned words, so that "*seems*" is aligned to "*Sembra*", "*very*" to "*molto*", and so on.

As for the lexical unit, this step is enough to assess that *"scorretto"* is the target word in the sentence that evokes the FAIRNESS_EVALUATION frame, and to define it as a lexical unit for the frame.

As for the frame element, we first identify the constituents that are syntactically linked to the Italian target word. The general rule employed for the extraction of such constituents selects for each target word t all nodes that are siblings of the parent node of t, and all nodes that are siblings of t. Then, some further selections are carried out depending on the part of speech of t. In this case, the selected

Given two aligned sentences s_{en} and s_{it}, s_{en} annotated with $info_{frame}$ and FE_{en}
$take\ lexunit_{en} \in s_{en}$
$if\ exists\ alignment_{lexunit_{en}}$
$\quad take\ aligned\ lexunit_{it} \in s_{it}$
$\quad transfer\ info_{frame}\ from\ lexunit_{en}\ to\ lexunit_{it}$
$\quad return\ lexunit_{it+infoframe}$
$\quad extract\ D_{it}\ from\ s_{it}$
// D_{it} = set of constituents syntactically linked to $lexunit_{it}$ in s_{it}
$\quad for\ each\ fe_{en} \in FE_{en}$
$\quad\quad Score_{best} = 0$
$\quad\quad Cand_{best} = empty$
$\quad\quad for\ each\ d_{it} \in D_{it}$
$\quad\quad\quad calculate\ Score_{it}$
// $Score_{it}$ = n. of aligned words between fe_{en} and d_{it}
$\quad\quad\quad if Score_{it} > Score_{best}$
$\quad\quad\quad\quad Score_{best} = Score_{it}$
$\quad\quad\quad\quad Cand_{best} = d_{it}$
$\quad\quad\quad end\ if$
$\quad\quad end\ for$
$\quad\quad return\ Score_{best}$
$\quad\quad return\ Cand_{best}$
$\quad end\ for$
$else$
$\quad return\ false$

Figure 4.5: Transfer algorithm 2

constituents are the modifier "*molto*" and the infinitive clause "*lasciarla sperare*".

For every FE in the source sentence, we compute the number of words aligned with the tokens in the Italian candidate constituents and we select as best candidate the constituent with the maximum alignment. In the example sentence, "*very*" has one overlap with "*molto*" and zero overlaps with "*lasciarla sperare*". On the other hand, "*leave her hoping*" shares no overlapping words with "*molto*" and two with "*lasciarla sperare*". This means that the *Degree* label is transferred to "*molto*" and the *Action* label to "*lasciarla sperare*".

4.4.4 Formalization of Algorithm 2

The formalization of algorithm 2 is reported in Figure 4.5.

We take the English corpus annotated with frame information C_{en} and align it at word level to the Italian corpus C_{it}, whose sentences have been previously parsed. For each sentence $s_{en} \in C_{en}$, we take the annotated lexical unit $lexunit_{en}$ and find the Italian aligned word, that we assume to be the target lexical unit $lexunit_{it}$. If no alignment is available, the transfer fails, otherwise the English frame label is assigned to the Italian $lexunit_{it}$. Then, for every English frame element fe_{en}, we take all syntactic dependents D_{it} of $lexunit_{it}$ and compute the number of aligned words between fe_{en} and $d_{it} \in D_{it}$. We consider the Italian dependent with most aligned words $Cand_{best}$ as the best candidate for annotation projection.

4.4.5 Algorithm comparison

A comparison between the two algorithms is necessary to understand the differences in their performance and to highlight the respective pros and cons. The main difference concerns the FE transfer process because, as we have seen in the previous sections, the LU transfer step is identical.

Algorithm 1 has the advantage to carry out the two steps independently, so that one can be successfully carried out even if the other one fails. The main disadvantage is that it strongly relies on semantic head alignment, which means that if the English head is not aligned, the whole FE transfer fails. Another problem is that it is based on language-specific compatibility rules between Italian and English, which means that it cannot be generalized or applied to other languages as it is. A third issue is that it requires both corpus sides being fully parsed, which can introduce noise in the pre-processing and impact on the transfer task.

As for algorithm 2, the main problem is that target and FE transfer cannot be carried out independently. This implies that, if the target word in Italian is not found because of a missing alignment, no FE can be transferred. On the other hand, it requires less pre-processing on the English side, because the alignment procedure for FEs requires only the source text to be chunked. Besides, it is more generalizable because it does not rely on language-specific syntactic information.

In order to highlight the differences between the two algorithms, two examples are reported: in Fig.4.6 we show a transfer case that is successful when applying algorithm 1 but not algorithm 2. In Fig.4.7 an opposite case is reported. In both examples, dotted arrows connect aligned words and the lexical units are underlined. In the first example, since "*morire*" is correctly aligned with the target "*dying*", it becomes the Italian lexical unit of the DEATH frame. Then, if algorithm 1 is applied, "*children*" is identified as the semantic head of the *Protagonist* frame element, then it is connected to "*bambini*", and finally the NP node dominating "*Donne e bambini*"

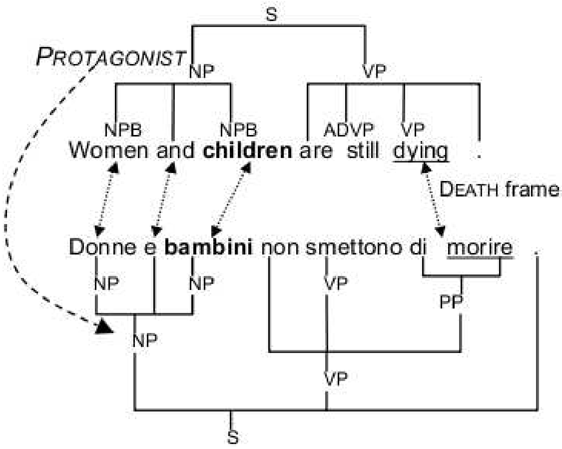

Figure 4.6: Correct transfer with Algorithm 1. Lit. translation: "*Women and children don't stop dying*"

is selected as the best Italian constituent because it represents the highest syntactic projection of the Italian head compatible with the annotated English constituent. Algorithm 2 would not deliver any FE transfer on the same couple of sentences, as it cannot identify "*Donne e bambini*" as dependent of "*morire*", due to the different syntactic structure of the Italian sentence.

In Figure 4.7 we report the output of the second transfer algorithm applied to two parallel sentences from the EUROPARL corpus (Koehn, 2005). Note that we do not exploit any syntactic information on the English side and that FE labels point to flat chunks, whereas in Figure 4.6 the sentences have been parsed on both sides. In this example, "*demonstrated*" is the target of the REASONING frame, and two frame elements are present, namely *Content* and *Arguer*. Both frame elements point to the correct constituent nodes in Italian, that are the syntactic dependents of the target "*dimostrato*". The *Content* frame element is correctly transferred even if only one word (*dialogue* - *dialogo*), which is not the semantic head of the constituent, has been aligned. This algorithm can cope with a different syntactic structure of the sentence in Italian, since the English secondary clause "*that we want dialogue*" is translated as "*la sua volontà di dialogo*" (i.e. *its wish for dialogue*). With algorithm 1 the transfer of the *Content* label would have failed because the semantic head of the constituent, "*want*", has not been aligned by KNOWA to the noun "*volontà*".

In some cases, both algorithms fail because they cannot deal with particular syntactic phenomena or some differences in the source and the target sentence.

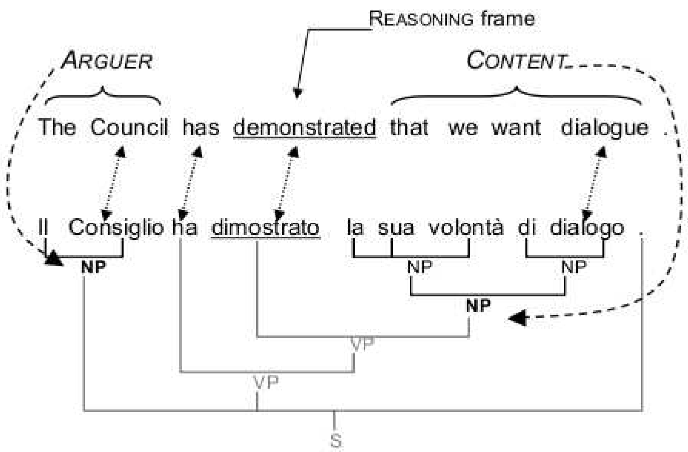

Figure 4.7: Correct transfer with Algorithm 2. Lit. translation: "*The Council has demonstrated its wish for dialogue*"

In Fig. 4.8, for example, the main verb "*hear*", which is the lexical unit of the Hear frame, is followed directly by the subordinate clause, while in the Italian translation the lexical unit "*Sentiamo*" is first followed by "*affermare*" (*claim*), which is in turn followed by the subordinate clause. The target transfer is correct because the two lexical units are correctly aligned. The *Hearer* FE, instead, is not overtly expressed in Italian, so the FE transfer cannot be carried out. As for the *Message*, both algorithms identify the VP "*affermare che l'Europa è lontana dai suoi cittadini*" as the best candidate, whereas the correct FE would include just the SBAR "*che l'Europa è lontana dai suoi cittadini*". The problem with algorithm 1 is that, according to the compatibility rules, an Italian VP can correspond to an English SBAR and, since it is the maximum projection of the verbal semantic head "*è*", the clause starting with "*affermare*" is selected as the best FE candidate. In algorithm 2, both the SBAR and the VP including "*affermare*" show four aligned words with the English FE. The problem is that the higher VP is the direct syntactic dependent of the lexical unit "*Sentiamo*", so it is chosen as the best FE candidate for bearing the *Message* label.

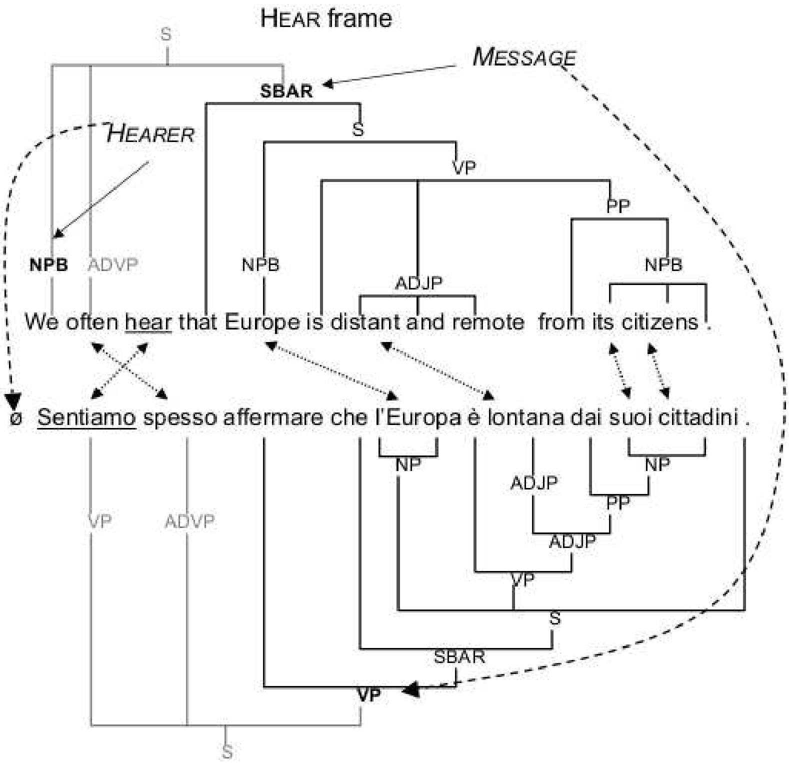

Figure 4.8: Example of wrong FE transfer. Lit. translation: "*We often hear claim that Europe is distant from its citizens*"

4.5 The gold standards

In order to investigate the influence of corpus characteristics on transfer quality, and meanwhile to develop a small manually-annotated corpus, we took into account two different parallel corpora. The resources will be used for different evaluations described in Section 4.6.

4.5.1 EUROPARL

The first corpus was an excerpt of 987 English and Italian sentences taken from the EUROPARL multilanguage parallel corpus (Koehn, 2005). EUROPARL includes the proceedings of the European Parliament in 11 European languages, namely French, Italian, Spanish, Portuguese, English, Dutch, German, Danish, Swedish, Greek and Finnish. The texts, aligned at sentence level, were first collected and made available to train statistical machine translation systems. For each language, the corpus contains about 1 million sentences and 30 million tokens, which makes it the largest multilingual corpus available for research purposes (it can be dowwnloaded at `http://www.statmt.org/europarl/`).

Two crucial considerations have to be pointed out about EUROPARL. The first one concerns its specificity in terms of genre: EUROPARL contains only transcriptions of political speeches, so it is strongly characterized by the use of the first and the second person singular and frequent personal addresses. Besides, the discussion topics include different aspects of social and political activity, but they show the predominance of a formal political vocabulary. Another issue about the use of EUROPARL for annotation transfer is the degree of semantic correspondence in the bitexts. Each speaker in the European Parliament addresses the assembly in his or her native language, and then the speech is translated into all other languages. This means that we cannot rely on the fact that all Italian sentences have been *directly* translated from English, which impacts on the number of free translations in the bitext. As an example, we report in Fig. 4.5.1 two parallel sentences extracted with the online EUROPARL browser (`http://urd.let.rug.nl/tiedeman/OPUS/cwb/Europarl/frames-cqp.html`). The sentences belong to the English - Italian bitext, but they are both translated from German, since the speaker is labeled with (DE). This might explain the different discourse structure, depending on the diverging interpretation of the speaker's opinion given by the translator.

In order to carry out the evaluation of the two algorithms, we extracted a subcorpus of the English-Italian bitext comprising 987 sentences. The same English

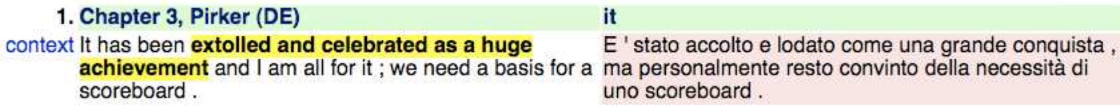

Figure 4.9: Parallel sentence displayed in the EUROPARL browser. Lit. translation: "*It has been welcomed and extolled as a big achievement, but I am personally convinced of the need for a scoreboard*"

sentences had been previously used to build an English-German and an English-French gold standard[4] and to evaluate some transfer experiments with these two language pairs (Padó and Lapata, 2005, Padó and Pitel, 2007). The use of this subcorpus offers several advantages:

- The English side has already been parsed and manually enriched with frame information in the context of past experiments described above, which would reduce the effort to create a gold standard for our language pair.
- With the annotation of frame information on the Italian part of the subcorpus, we would contribute to the creation of a multilingual parallel corpus including English, German, French and Italian. This would be very useful to carry out comparative studies about the applicability of the FrameNet paradigm to different languages.
- Since we use the same dataset employed for past transfer experiments, it is easier to compare the results and to highlight the differences between language pairs.

While the English side of the gold standard was already available, the Italian side was manually annotated with frame information, after some pre-processing (see Section 4.5.4). Note that the annotator did the work without knowing the information annotated for English.

In Table 4.2, we report a comparison between the subcorpora annotated with frame information in the four languages. The German, English and Italian corpus all comprise 987 sentences, each one with one frame annotation. The French one, instead, includes 1076 sentences, but only 951 are annotated with frame information. The others are commented as problematic cases, mainly because a suitable frame label is not available.

[4]The data are available at `http://www.nlpado.de/~sebastian/srl_data.html`

	Eng	Ita	Ger	Fr
N. of frames	83	158	74	121
N. of FE annotations	1938	1733	2141	1669
N. of unique FEs	97	172	91	121

Table 4.2: Comparison among the 4 annotated subcorpora

The figures show that there is a discrepancy between the annotated data in the four languages. This may depend on the fact that the gold standard for English, German and French had been annotated with FrameNet 1.1., containing about 520 frames, while the annotation for Italian was based on FrameNet 1.3 and, if necessary, on the online version of the database. This choice was motivated by the intention to build not only a gold standard for evaluation, but also an annotated corpus that may become part of a future Italian FrameNet and that should comply with the latest version of the English database.

The table shows also a parallelism between Romance languages on one side and Germanic languages on the other. In particular, Italian and French show a higher variability of frames and FE labels than English and German. This is particularly evident for French, since the corpus comprises fewer annotated sentences than the others. Despite this, Italian and French texts contain less FE annotations. On the French side, this may depend on the corpus size, while for Italian it is strongly influenced by the presence of null-subject pronouns, since the subject of a sentence in Italian can be left unexpressed. This means that every time we find a role-bearing subject pronoun such as *I*, *you*, *they*, *we*, *he* or *she* in the English corpus, we can expect that no corresponding overt lexical item is found in the Italian translation, as shown in Example 4.1. This phenomenon is very frequent in the EUROPARL corpus because of the wide usage of the first person. As a matter of fact, about 15% of all English FEs correspond to a null-subject pronoun in the Italian gold standard.

(4.1) JUDGMENT_DIRECT_ADDRESS frame:
[I]$_{\text{Speaker}}$ <u>thank</u> [you]$_{\text{Addressee}}$ [for your report]$_{\text{Reason}}$.
[La]$_{\text{Addressee}}$ <u>ringrazio</u> $\emptyset_{\text{Speaker}}$ [per la relazione]$_{\text{Reason}}$.

The same can be observed for expletive *it*, that is never expressed in Italian.

In order to understand if, in spite of the different amount of annotated data, the four gold standards contain the same frame labels, we extracted the 10 most frequent frames from every subcorpus. Results are reported in Table 4.3. The number between parenthesis indicates the number of occurrences.

English	Italian	German	French
Awareness (142)	Awareness (112)	Awareness (154)	Awareness (145)
Statement (68)	Statement (72)	Statement (78)	Statement (73)
Questioning (47)	Opinion (30)	Questioning (48)	Questioning (54)
Communicate_cat. (40)	Removing (28)	Evidence (40)	Hear (31)
Arriving (37)	Handling (28)	Categorization (39)	Arriving (31)
Removing (30)	Questioning (27)	Hear (36)	Removing (28)
Giving (29)	Arriving (27)	Judgment_comm. (33)	Endangering (27)
Event (29)	Killing (24)	Endangering (32)	Judgment (24)
Evidence (25)	Evidence (22)	Removing (31)	Judgment_dir._add. (23)
Perception_active (21)	Hear (21)	Giving (29)	Giving (21)

Table 4.3: The 10 most frequent frames in the 4 subcorpora

The list shows that the most frequent frames are mainly connected to the communication and the political scenarios. AWARENESS is by far the most frequent frame, even if it is less recurrent in Italian, followed by STATEMENT. In the Italian corpus the third most frequent frame is OPINION, which is not present in the other gold standards because it was not defined in FrameNet 1.1.

In order to classify 28 sentences having *trattare.v* (treat) as target, we decided to introduce a new frame, which we called HANDLING, defined as follows: "An *Agent* behaves towards an *Affected_party* in a certain way or *Manner*". This frame resembles the CONDUCT frame, but it is more focused on the *Affected_party* than on the *Agent*. As a matter of fact, *Affected_party* is extra-thematic in CONDUCT and core in HANDLING. If we look at the frame label assigned to the translation of the Italian sentences classified as HANDLING, we notice that there is not a homogeneous annotation, which means that it could be a case of missing or incomplete frame definition in FrameNet. We report in 4.2 the translation of the same sentence in the four languages of the gold standard, with each sentence bearing a different frame label, even if the translation is quite faithful. In French no label was assigned because the annotators could not find any satisfactory frame definition in the database.

(4.2) ITA: In altri Parlamenti, gli uscieri sono trattati con il rispetto che è loro dovuto (HANDLING)
ENG: In other parliaments ushers are treated with the respect that they deserve (COMMUNICATE_CATEGORIZATION)
GER: In anderen Parlamenten ist es ueblich, Saaldiener mit dem gebuehrenden Respekt zu behandeln (CONDUCT)
FRA: Dans les autres parlements, les huissiers sont traités avec le respect auquel ils ont droit (No frame assignment)

Finally, we compared the degree of frame and FE parallelism for every bitext considered. Results are reported in Table 4.4. Frame parallelism measures the percentage of English sentences having the same frame label in another language, and FE parallelism was computed on this subset of parallel sentences.

	Eng-Ita	Eng-Ger	Eng-Fr
Frame parallelism	0.61	0.71	0.69
FE parallelism	0.82	0.91	0.88

Table 4.4: Comparison of frame and FE parallelism

The English-Italian bitext presents the lowest parallelism in both cases. Apart from free translations, missing parallelism can be negatively affected also by different interpretations of the sentences given by the annotators. This involves in particular frame elements which are semantically similar, such as *Topic/Message* in the STATEMENT frame, *Agent/Cause* in the CAUSE_HARM frame or *Area/Path* in the MOTION frame. Another cause of missing parallelism is the different version of FrameNet used in the annotation of the bitexts, as we mentioned before. In version 1.1, for example, the SCRUTINY frame had the *Standard* frame element, which was called *Enabled_situation* in version 1.3. The same happened to the LIKELIHOOD frame, where the *Event* frame element in version 1.1 was newly changed into *Hypothetical_event*. The complete list of all frames and LUs present in the Italian part of the gold standard is reported in Appendix B.1.

4.5.2 MULTIBERKELEY

In order to highlight the impact of different corpora on evaluation, we took into account also a second parallel corpus which we called MULTIBERKELEY. In this case, the corpus was built by manually translating in a controlled way a number of sentences from the Berkeley FrameNet corpus. The selection of sentences was guided by the desire to include in the resulting Italian corpus frames that were not already present in EUROPARL. While most LUs annotated in EUROPARL were verbs, in MULTIBERKELEY also frames with LUs of different categories were included, for example nominal LUs as in the CLOTHING frame or adjectival LUs as in COLOR. In this way, it was possible to account for lexical units of different categories. Besides, as past experiments on annotation transfer have shown (see Bentivogli and Pianta, 2005), the automatic projection of annotation between two parallel corpora in different languages can benefit from a translation that minimizes syntactic differences from source and target language.

The corpus creation comprises seven steps:

1. Select a set of frames F that are not present in the *Europarl* gold standard
2. $\forall f \in F$, choose the lexical unit $l_n \in f$ with the largest set of example sentences S in the English FrameNet database
3. $\forall s \in S$, compute the number of tokens n_s
4. Pick $s_n \in S$ having the lowest n_{s_n}
5. Manually translate s_n into Italian
6. Pre-process the Italian sentence s_{ita} (see Section 4.5.4)
7. Manually annotate s_{ita} with frame information

The selection carried out at step 2 was aimed at choosing the *most representative* LUs for a frame. Even if the information in the FrameNet database is not statistically significant w.r.t. the frequency of the occurrence of the different targets, we assumed that a target with several attestations in the Berkeley corpus and a complete annotation should be considered significant of the frame it belongs to. In fact, as described in Baker et al. (1998), at least the initial set of frames was defined starting from a skeletal description of each frame and an intuitive choice of the major lexical units, which were then annotated on a selected subcorpus of examples. For this reason, the lexical units with a complete annotation and a rich set of example sentences in the FrameNet database could be considered typical and relevant to the frame.

As for Step 4, we selected the shortest sentences but we also discarded the instances where all frame elements are expressed by a personal pronoun (e.g. "*He took it*"). This process was carried out semi-automatically, with the automatic selection of the 5 shortest example sentences ordered by length and a manual check of the first sentence in the list, which was eventually discarded in favor of the second one and so on.

MULTIBERKELEY comprises 391 sentences with one example per frame. The sentences are taken from the English FrameNet database, thus they are PoS tagged and annotated with frame information. All frame elements are also labeled with phrase type (NP, PP, VP, etc.) and grammatical function (Ext, Dep, Head, etc.). We manually translated the English corpus into Italian trying to limit "free" translations in order to enhance the correspondence between source and target texts. If possible, we preferred Italian translations minimizing divergences with English. However, priority was always given to good Italian prose.

The complete list of frames and LUs annotated in the Italian part of MULTIBERKELEY is reported in Appendix B.2.

4.5.3 Gold standard comparison

As a preliminary step to the evaluation of the two algorithms illustrated in Section 4.6, we compared the gold standards in two ways: first we focus on the difference between the Italian side of EUROPARL and that of MULTIBERKELEY, and then we take into account the bitexts. The first comparison is reported in Table 4.5:

	Europarl	*MBerk.*
Avg. sent. length (tokens)	23±9	10±4
N. of frames	158	387
N. of unique FEs	172	256
N. of unique LUs	413	390
Avg. annotated FEs per target	1.75	1.59
LU category		
Nouns	66 (15.98%)	130 (33.33%)
Verbs	319 (77.24%)	200 (51.28%)
Adjectives	25 (6.05%)	57 (14.61%)
Adverbs	2 (0.48%)	2 (0.51%)
Prepositions	1 (0.24%)	1 (0.26%)

Table 4.5: Comparison of the gold standards (Italian)

The average sentence length in the EUROPARL corpus is more than double than that in the MULTIBERKELEY corpus due to the different selection strategy of the sentences. The sentence length impacts also on the average number of FEs annotated for each target, which is higher in EUROPARL because longer sentences may include more FEs realizations, especially the peripheral ones (i.e. *Time*, *Place*, etc.). The selection strategy maximizes also frame and FE variability, as shown by the number of frames, unique FEs and LUs represented in the two corpora. As for the frames, their number is much higher in MULTIBERKELEY because almost every sentence belongs to a different frame[5]. The lower number of unique FEs in EUROPARL than in MULTIBERKELEY, instead, depends on the fact that in EUROPARL there are a lot of frames that are semantically related to each other and that share the same frame elements. For example, AWARENESS and OPINION, which have many occurrences in EUROPARL, have the *Cognizer* FE in common, while

[5] In the English part of the corpus, every sentence belongs to a different frame, so that there are 391 frames for 391 sentences. With the translation into Italian, the frame label changed for 4 sentences, so that we obtained 387 frame attestations for 391 sentences.

STATEMENT and QUESTIONING share the *Speaker*, *Addressee*, *Message*, *Medium* and *Topic* frame elements. As for the lexical units, in EUROPARL there are 2.6 LUs for every frame on average, which means that the LU sets that can be derived from the corpus are richer than in MULTIBERKELEY. Besides, every LU is instantiated in 2.4 sentences on average. On the other hand, the LU distribution according to the PoS is unbalanced w.r.t. the FrameNet database (see Table 2.4): in EUROPARL 77.24% of all LUs are verbs, while in the FrameNet database both verbal and nominal LUs represent about 40% of all lexical units. From this point of view, MULTIBERKELEY shows a similar LU distribution to the English database.

In both corpora, the average LU polysemy is low, being 1.04 in MULTIBERKELEY and 1.11 in EUROPARL, and the great majority of LUs has just one occurrence. The most polysemous LU in the MULTIBERKELEY is "*urlare.v*" (*to cry*), belonging to COMMUNICATION_MANNER, COMMUNICATION_NOISE and MAKE_NOISE. In EUROPARL, instead, the most polysemous LU, "*trattare.v*" (*to treat*), occurs in 4 frames: HANDLING, CATEGORIZATION, COMMUNICATE_CATEGORIZATION and SPEAK_ON_TOPIC.

A second comparison was carried out taking into account the two bitexts. In particular, we measured the frame and FE parallelism between the gold standards. Results are reported in Table 4.6. Note that FE parallelism is computed on the set of sentences annotated with the same frame.

	Europarl	*MBerk.*
Frame parallelism	0.61	0.98
FE parallelism	0.82	0.91

Table 4.6: Corpus comparison

As expected, the two bitexts present very different degrees of parallelism. This depends on the fact that the two sides of the EUROPARL gold standard were annotated using different FrameNet versions, but also on free translations, which have been minimized in MULTIBERKELEY. In such corpus, the few cases of missing frame parallelism depend on *lexical gaps* rather than on *translation shifts*. As an example, we report in (4.3) two sentences from MULTIBERKELEY which present different frame labels (between parenthesis) because of a lexical gap, while (4.4) is extracted from EUROPARL and was annotated with two different frame labels due to a translation shift.

(4.3) Didn't it <u>smell</u> odd? [APPEARANCE]
Non aveva un <u>odore</u> strano? [SENSATION]

Didn't it have an odd smell?

(4.4) We <u>hear</u> now that other national associations are expressing concern. [PERCEPTION_EXPERIENCE]
E adesso ci viene <u>detto</u> che altre associazioni nazionali hanno espresso una certa preoccupazione. [STATEMENT]
And now we are told that national associations have expressed a certain concern.

The comparison between the bitexts contributed to point out the advantages and the weaknesses of using them for evaluation. EUROPARL contains extremely free translations, the sentences are usually quite long and complex and the topics dealt with in the corpus are quite homogeneous. For this reason, we expect annotation transfer algorithms to perform poorly on such datasets, even if the restricted domain could represent an advantage if the corpus were used to evaluate systems for automatic frame identification. MULTIBERKELEY, on the contrary, contains simplified sentences aimed at maximizing cross-lingual parallelism. In this way, we can investigate the real performance of transfer algorithms, reducing noise coming from wrong parsing or bad word alignment. Because of these characteristics, though, it is far from being an example of "real-word text". Instead, it should be seen as a *seed* corpus, a starting point for the automatic extension of FrameNet for Italian.

4.5.4 Gold standard development

In order to build the gold standards described in the previous sections, we follow 2 directions. On the one hand, we align the parallel corpora at word level using KNOWA. On the other hand, we add the annotation layer to the Italian part of the corpus in five steps, as shown in Fig. 4.10.

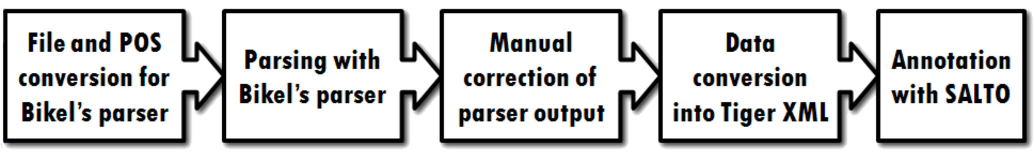

Figure 4.10: Steps for the development of the Italian gold standard

The five phases leading to the annotation of the Italian part of the corpus with frame information are the following:

Pre-processing: The Italian sentences are first tokenized, tagged and converted for the parsing. In this pre-processing step, we use TagPro (Pianta and Zanoli, 2007), the PoS tagger included in the TextPro suite for morphological analysis of Italian[6] (Pianta et al., 2008). The PoS labels and the format are then converted for the parser. We report in Table 4.7 the conversion table.

TagPro Label	Parser Label	Description
XPS, XPW, XPO	Punctuation as it is	Punctuation
XPB	-LRB- or -RRB-	Brackets
N	NUMR	Number
RS, RP	ART	Article
AS, AP, AN	ADJ	Qualitative adjective
DS, DP, DN	ADJ	Determinative adjective
E, ES, EP	PREP	Preposition
B	ADVB	Adverb
C, CCHE, CCHI, CADV	CONJ	Conjunction
PS, PP, PN	PRO	Pronoun
SS, SP, SN, SPN	NOU	Noun
QNS, QNP	PRO	Relative pronoun
YA, YF	NOU	Acronym, foreign term
I	INTERJ	Interjection
#, %	SPECIAL	Special characters
VI, VF, VSP, VPP, VG, VM	VMA	Main verb
VIY, VFY, VSPY, VPPY, VGY, VMY	VAU	Auxiliary verb

Table 4.7: Conversion table TagPro - Bikel's parser

Parsing: The Italian sentences are parsed with Bikel's phrase-based statistical parser trained for Italian (Corazza et al., 2007), used also in the pre-processing step of algorithm 1 and 2.

Manual correction of parse trees: In order to annotate constituents with the correct label and the right span, we carried out a shallow correction of parse trees. The correction was manual and was realized with the help of an online visualization tool called *phpSyntaxTree* (`http://ironcreek.net/phpsyntaxtree/`), that takes the parenthesized format of the sentence as input and displays a parse tree in output. To our knowledge, no visualization tool that allows to directly correct the parse trees in the graphical environment is available. For this reason, the correction step could not be carried out directly on the displayed tree, but had to be manually accomplished in the output file of the parser.

[6]The tagger, which is available at `http://textpro.fbk.eu/`, performed the best in the task of Italian PoS tagging at the evaluation campaign EVALITA 2007

We did not correct all trees because it would be too time-consuming. We only corrected those nodes that we assume to be good candidates for bearing frame information. For example, the sentence "*Indossava occhiali dalla montatura leggera*" (*He wore thin-rimmed spectacles*) was parsed as:
(S (VP (VMA Indossava) (NP (NOU occhiali)) (PP (PREP dalla) (NP (NOU montatura) (ADJ leggera)))) (. .)) (Fig. 4.11).
The parsing delivered an error in the PP-attachment. The correct version is:
(S (VP (VMA Indossava) (NP (NP (NOU occhiali)) (PP (PREP dalla) (NP (NOU montatura) (ADJ leggera))))) (. .)) (Fig. 4.12).

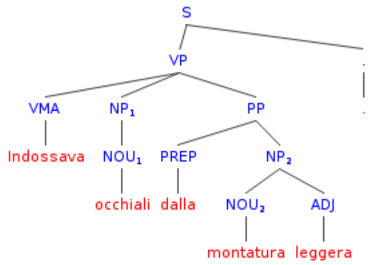

Figure 4.11: Wrong parse tree

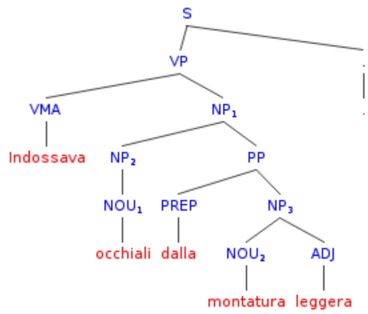

Figure 4.12: Correct parse tree

Format conversion: In order to annotate the dialogs with the SALTO tool (Burchardt et al., 2006), data must be first converted into Tiger-XML format (Mengel and Lezius, 2000) using the freely available TIGERRegistry tool (`http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERSearch/doc/html/TIGERRegistry.html`). This application supports many popular treebank and parser output formats, e.g. Penn Treebank, SWITCHBOARD, Susanne and Negra, and converts them into the Tiger-XML format required by the tool for frame annotation. In such format, which represents a standard for XML-based annotation of syntactic information, every sentence is seen as a `<graph>` consisting of `<terminals>` and `<nonterminals>`. The `<terminals>` element is a list of `<t>`erminals with PoS information reported as attribute. Instead, `<nonterminals>` include a list of syntactic nodes `<nt>`. Within each node, the `<edge>` label links the node to its direct constituents (`<t>`s or `<nt>`s). An example sentence in XML-Tiger format is reported in Fig. 4.13, where the sentence of Fig. 4.12 is displayed.

The sentence number is 3430-459771, which is repeated in every terminal id (e.g. 3430-459771_1 etc.) and in every node id. Words are numbered in increasing order and are described by the PoS feature. Non-terminal nodes

```
<s id="3430-459771">
    <graph root="3430-459771_500">
        <terminals>
            <t id="3430-459771_1" word="Indossava" pos="VMA"/>
            <t id="3430-459771_2" word="occhiali" pos="NOU"/>
            <t id="3430-459771_3" word="dalla" pos="PREP"/>
            <t id="3430-459771_4" word="montatura" pos="NOU"/>
            <t id="3430-459771_5" word="leggera" pos="ADJ"/>
            <t id="3430-459771_6" word="." pos="."/>
        </terminals>
        <nonterminals>
            <nt id="3430-459771_503" cat="NP">
                <edge label="--" idref="3430-459771_2"/>
            </nt>
            <nt id="3430-459771_505" cat="NP">
                <edge label="--" idref="3430-459771_4"/>
                <edge label="--" idref="3430-459771_5"/>
            </nt>
            <nt id="3430-459771_504" cat="PP">
                <edge label="--" idref="3430-459771_3"/>
                <edge label="--" idref="3430-459771_505"/>
            </nt>
            <nt id="3430-459771_502" cat="NP">
                <edge label="--" idref="3430-459771_503"/>
                <edge label="--" idref="3430-459771_504"/>
            </nt>
            <nt id="3430-459771_501" cat="VP">
                <edge label="--" idref="3430-459771_1"/>
                <edge label="--" idref="3430-459771_502"/>
            </nt>
            <nt id="3430-459771_500" cat="S">
                <edge label="--" idref="3430-459771_501"/>
                <edge label="--" idref="3430-459771_6"/>
            </nt>
        </nonterminals>
    </graph>
</s>
```

Figure 4.13: Example of Tiger-XML format

are listed separately with the category label and the edges that they include. Nodes are numbered starting from 500 and follow a top-down order, from the root node down to pre-terminals.

Manual annotation: Manual annotation of frame information was carried out using SALTO (Burchardt et al., 2006), a freely available Java application that can be downloaded at `http://www.coli.uni-saarland.de/projects/salsa/page.php?id=software`. The tool can load parsed sentences in Tiger-XML format, displays them as parse trees and gives the possibility to add frame and FE labels pointing to the tree nodes. An example sentence displayed with SALTO is reported in Fig. 4.14.

We take as lexical unit the Italian word that is the translation equivalent of the English LU. In case the bitext translation is so free that no translation equivalent is found in Italian, we choose the main verb, if present, or the main nominal head, if the Italian sentence consists of a nominal phrase.

After selecting the target, the annotator has to identify the conceptual situation

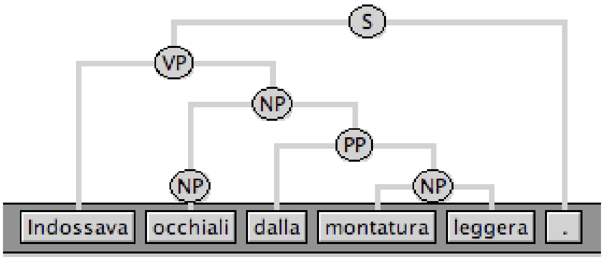

Figure 4.14: Parse tree displayed with SALTO

evoked by it and to assign a frame label. SALTO allows to import the frame list from the FrameNet database and to assign a frame label to a target just by clicking on one item in the list. Otherwise, it is possible to add newly created frame descriptions.

The **frame assignment** task can be problematic in case of ambiguous targets or when no frame seems to be suitable for the assignment. In general, annotators should first look at the textual definition of the candidate frames and at the list of available frame elements on the FrameNet website. They should check if this definition can be applied also to Italian sentences, and see if the candidate frame(s) contain an English translation equivalent of the Italian target. In case of doubt, it is recommended to look at the example sentences available for the English target in order to compare the usage of the English and corresponding Italian LUs. Annotators should also check if some annotated sentences in Italian containing the given LU are already available. If the Italian example sentence to annotate contains an ambiguous target, it is also recommended to try and paraphrase it using a possibly unambiguous LU, so as to clarify which the evoked frame is.

Some ambiguities that are present in English can occur also in Italian, for example the polysemous verb *ask.v* evokes both QUESTIONING and REQUEST, as does the Italian translation equivalent *chiedere.v*. In other cases, frame assignment in Italian is more straightforward than in English because the alternation between reflexive and non-reflexive forms is captured by different frame types, as in *svegliarsi.v* (*to wake up*, intrans.), belonging to WAKING_UP, and *svegliare.v* (*to wake up*, trans.) in CAUSE_TO_WAKE_UP. In English, instead, the verbs *wake.v*, *wake_up.v* and *got_up.v* appear in both frames and there is no distinction between the reflexive and the causative form.

If frame assignment is still problematic after looking at the frame definitions and at the English examples, annotators should try and match the FEs provided for every candidate frame to the subcategorisation pattern of the current Italian target. If the target in the example sentence has some (realized or unrealized) arguments that do not correspond to any FE of the candidate frame, then the frame has to be discarded. For example, the target *collegare.v* (*to connect*, trans.) in the sentence "*Il tecnico collega la stampante alla rete*" (*The technician connects the printer to the net*) could in principle be assigned to the following candidate frames:

INCHOATIVE_ATTACHING: An Item comes to be attached to a Goal, with a Connector forming a bond between the Handle of the Item and the Goal.

ATTACHING: An Agent attaches an Item to a Goal by manipulating a Connector, creating an asymmetric relationship between the Item and the Goal.

The INCHOATIVE_ATTACHING frame is clearly not appropriate to the sentence because its definition does not include the *Agent* frame element, which is the role of "*Il tecnico*" (*The technician*) in the Italian sentence. On the contrary, ATTACHING is a suitable frame because it includes the roles *Agent*, *Item* and *Goal* needed to annotate all the constituents in the example sentences.

Despite these suggestions, there are still cases in which it is very difficult to make a decision between two frames, because they express related meaning components. For example, the sentence "*Credo che X*" (*I believe that X*) without further context could belong both to the AWARENESS frame (to have a fact in his/her mental representation, as a belief or knowledge) and to CERTAINTY (to be certain of a fact). The assignment decision should depend on which of the meaning components is dominant in the example at hand. If it focuses on expressing the content of the belief/knowledge, like in "*So che X*" (*I know that X*), AWARENESS is more appropriate; if the main information is about the degree of certainty of the belief, like in "*Sono sicuro che X*" (*I am sure that X*), it is a case of CERTAINTY. Again, frame assignment could benefit from paraphrasing the example in order to stress the role of its meaning components. Anyhow, such examples involve case-to-case decisions which can be influenced by the annotator's interpretation.

After a frame has been assigned to the target in a sentence, annotators have to **identify the frame elements**. If a frame label has been chosen for the target, SALTO displays automatically the core FEs available for the given frame, as shown in Fig. 4.15 for the ADDICTION frame pointing to *dipendenza.n* (*dependency*) as a target. A FE label is assigned by dragging it onto a syntactic constituent. Other labels for peripheral and extra-thematic FEs can be added by hand.

The annotated information is internally recorded in Tiger/SALSA XML format (Erk and Padó, 2004), a modular extension of Tiger-XML where syntax and se-

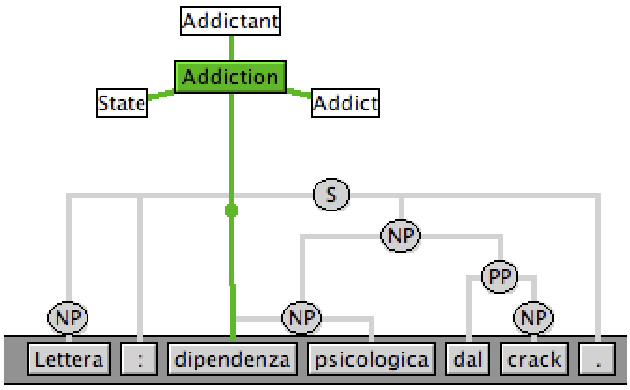

Figure 4.15: Core FEs available with SALTO

mantics are stored independently. The semantic information layer is contained in the additional element <sem>, while the syntactic representation is still contained in <graph>, as represented in Fig. 4.16. <sem> contains the semantic information for the current sentence, with unique identifiers for all semantic nodes and edges. In our annotation, the only semantic information encoded is about frames, and is contained in the <frames> element. <frames> can potentially include the annotation of different frames in the same sentence (between <frame> tags), although only one frame per sentence is annotated in our gold standard. For each <frame>, <target> and <fe> can be encoded and connected to the tree nodes. Even if all information about a sentence is included within one <s> element, the different annotation levels, namely <graph> and <sem>, are kept in separate blocks. However, they can be straightforwardly related through pointers from semantic labels to syntactic nodes.

Annotators are instructed to annotate all frame elements which can be recognised with certainty. Sometimes, a FE can be annotated considering different extensions and it can be difficult to choose one over the other. In such cases, we adopted the *maximality principle* described in Padó (2007, pp. 188-189), which we summarize as:

(4.5) If possible, the complete lexical material describing a frame element should be annotated. Ideally, this material is located below one single node, the so-called maximal constituent. If the lexical material of a FE is distributed over several syntactic constituents, it is allowed to annotate *discontinuous* frame elements.

```xml
<s id="3430-459771">
    <graph root="3430-459771_500">
        <terminals>
            <t id="3430-459771_1" word="Indossava" pos="VMA"/>
            <t id="3430-459771_2" word="occhiali" pos="NOU"/>
            <t id="3430-459771_3" word="dalla" pos="PREP"/>
            <t id="3430-459771_4" word="montatura" pos="NOU"/>
            <t id="3430-459771_5" word="leggera" pos="ADJ"/>
            <t id="3430-459771_6" word="." pos="."/>
        </terminals>
        <nonterminals>
            <nt id="3430-459771_503" cat="NP">
                <edge label="--" idref="3430-459771_2"/>
            </nt>
            <nt id="3430-459771_505" cat="NP">
                <edge label="--" idref="3430-459771_4"/>
                <edge label="--" idref="3430-459771_5"/>
            </nt>
            <nt id="3430-459771_504" cat="PP">
                <edge label="--" idref="3430-459771_3"/>
                <edge label="--" idref="3430-459771_505"/>
            </nt>
            <nt id="3430-459771_502" cat="NP">
                <edge label="--" idref="3430-459771_503"/>
                <edge label="--" idref="3430-459771_504"/>
            </nt>
            <nt id="3430-459771_501" cat="VP">
                <edge label="--" idref="3430-459771_1"/>
                <edge label="--" idref="3430-459771_502"/>
            </nt>
            <nt id="3430-459771_500" cat="S">
                <edge label="--" idref="3430-459771_501"/>
                <edge label="--" idref="3430-459771_6"/>
            </nt>
        </nonterminals>
    </graph>
    <matches>
    </matches>
    <sem>
        <globals>
        </globals>
        <frames>
            <frame name="Accoutrements" id="3430-459771_f1">
                <target>
                    <fenode idref="3430-459771_2"/>
                </target>
                <fe name="Descriptor" id="3430-459771_f1_e1">
                    <fenode idref="3430-459771_504"/>
                </fe>
            </frame>
        </frames>
        <usp>
            <uspframes>
            </uspframes>
            <uspfes>
            </uspfes>
        </usp>
        <wordtags>
        </wordtags>
    </sem>
</s>
```

Figure 4.16: Example of Tiger/SALSA-XML format

SALTO gives the possibility to assign a FE label to part of a word. This is useful in case of verbal targets with role-bearing clitics, as shown in Fig. 4.17: the clitic *li* was split from the target word *inzuppare.v* (*to soak*), so that it can bear the *Theme* FE label.

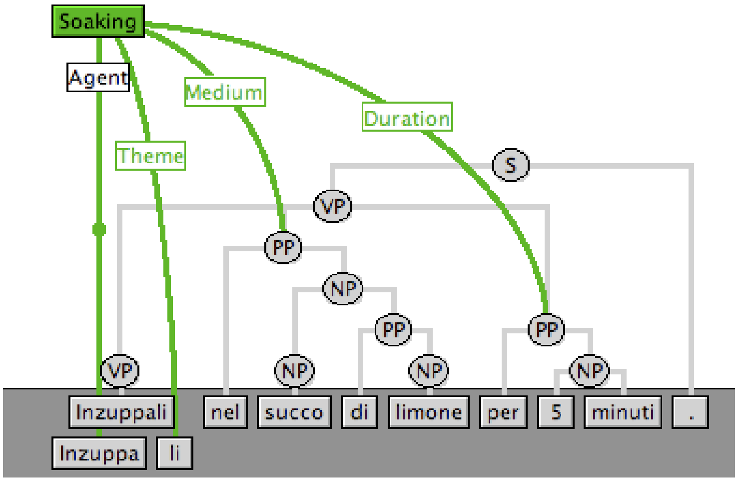

Figure 4.17: Splitting words with SALTO. Transl.: "*Soak them in lemon juice for 5 minutes*"

Since the syntactic structure was created automatically and only partially corrected, the parse trees can contain some errors, for instance a constituent could be wrongly split in two phrases. Even if we cannot modify the constituent structure, SALTO gives the possibility to duplicate a FE label. In this way, the FE can cover all terminals expressing the corresponding role, even if the nodes are wrong.

With the SALTO interface, it is possible to associate attributes to words. We exploit this mechanism to annotate the fact that one of the unlinked FEs could be assigned to an empty subject whose presence is revealed by the verb morphology. In practice, we associate the attribute *Empty_subj: FE label* to the verb word. In this way, we can cope with the problem of unexpressed roles, since Italian verbs can have an empty subject whose person and number is conveyed by the verb conjugation, whereas English and German verbs all require a mandatory explicit subject. An example annotation is reported in Fig. 4.18.

In this sentence, the verbal target *rassicurò* (*reassured*), which evokes the REASSURING frame, implies the presence of an implicit subject in the third person singular. This subject would bear the *Speaker* role.

In general, we add the empty subject label only when the target is a finite verb bearing explicit subject agreement information. This excludes for instance

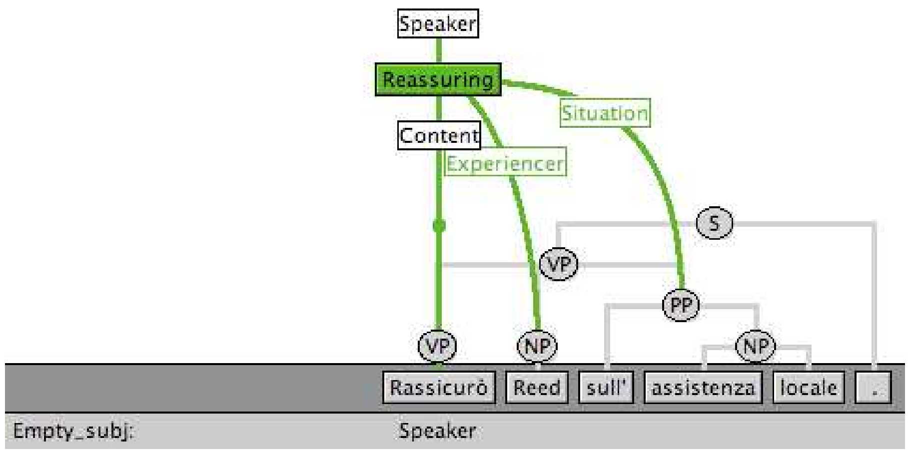

Figure 4.18: Empty subject annotated with SALTO.
Transl.: "∅ *reassured Reed on local assistance.*"

the *passato prossimo* tense (present perfect). In this case, the target is the past participle which in some cases bears subject gender and number agreement but in no case person agreement. Instead, most subject agreement information is conveyed by the auxiliary. For this reason, if the target is a past participle, we do not add the *Empty_subj* attribute. The same applies to cases of *gerundio* (gerund).

It is very important to distinguish the cases labeled as *Empty_subj* from the *null instantiations* encoded in the English FrameNet (Ruppenhofer et al., 2006, pp. 33-36). In the Berkeley database, FEs that do not appear in a given sentence as lexical or phrasal material are annotated to convey lexicographically relevant information about *omissibility conditions*. In particular, null instantiations defined as Constructional (CNI) are usually motivated by particular grammatical constructions that license such omissions such as imperative sentences with missing subject or passive sentences with missing agent. We report in example 4.6 a CNI annotation from the FrameNet database, where the core FE *Communicator* is omitted in the passive construction:

(4.6) But [anyone]$_{\text{Evaluee}}$ could be arrested and <u>accused</u>$_{\text{JUDGMENT\_COMMUNICATION}}$ [of communist sympathies]$_{\text{Reason}}$ (CNI=Communicator).

While we do not annotate cases of null instantiations in our gold standard, we identify empty subjects because they are morphologically expressed through verb agreement.

4.6 Evaluation framework

In different research works about frame annotation transfer, several evaluation criteria have been applied. The common feature among them is the choice to include in the testset only sentences that present a certain degree of semantic parallelism in the parallel gold standards. We believe that this approach is not suitable for our goal: since we aim at creating an annotated corpus with near manual annotation quality, we need to evaluate all the annotations resulting from the transfer, reproducing the complete task under real-world conditions.

In the following subsections, we will illustrate two existing evaluation approaches and add our proposal for a more general and effective one. Moreover, we will evaluate the output of our algorithms applying the presented metrics.

In order to carry out the evaluation, we divided both corpora into a development and test set. The former was used to tune the transfer algorithms, while the latter was employed to run the algorithms and carry out evaluation, comparing the output to the Italian gold standard. The EUROPARL corpus was split into a devset of 300 sentences and a testset of 687 sentences. The MULTIBERKELEY corpus comprised a development set of 91 sentences and a testset of 300 sentences.

The basic evaluation metrics adopted are precision, recall and F1. In the context of frame information transfer, *precision* is the number of correct transfers divided by the total number of transfers carried out by the algorithm. A perfect precision score of 1.0 means that every transfer delivered was correct, but says nothing about whether all semantic information has been transferred.

Recall is the number of correct transfers divided by the number of all semantic elements that have to be transferred. A perfect recall score of 1.0 means that all semantic information in the source text has been correctly transferred to the target text, but says nothing about how many incorrect transfers were also carried out by the system. For example, if 10 annotations out of 1000 were transferred, and all of them were correct, precision scores 1.0 (n. of correct transfers / n. of transferred elements) but the general performance of the transfer system is not good because precision is very low (10/1000 = n. of correct transfers / n. of elements to be transferred).

Usually, precision and recall are combined into a single measure, such as the *F-measure*, also called *F1*, which is the weighted harmonic mean of precision and recall and is computed as follows:

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

4.6.1 Evaluation 1

In the evaluation of frame information transfer between English and German and English and French, Padó and Lapata (2009) and Padó and Pitel (2007) proposed to evaluate the task following three main criteria: first, they do not consider target transfer because they focus only on FE transfer. Second, they consider for evaluation only the subset of parallel sentences in the source and target gold standard having the same frame, in order to focus on the alignment and transfer quality and exclude free translations from evaluation. Third, they propose to measure performance only on frame elements using the "exact match condition", i.e. both the label and the span of the projected role have to match the gold standard annotation for the target language to count as a true positive. We first apply the same evaluation framework and compare the results obtained with algorithm 1 and 2 on EUROPARL to the results obtained by Padó and Pitel (2007) for the English-French pair, given that they worked on the same subset of sentences taken from EUROPARL and used the same English gold standard. Since Italian and French are both romance languages, we assume that they should show the same degree of syntactic and semantic similarity to English. Results are reported in Table 4.8.

Europarl	**Precision**	**Recall**	**F1**
Algorithm 1	0.48	0.39	0.43
Algorithm 2	0.66	0.40	0.50

Table 4.8: FE transfer evaluation (1) on Europarl

The second algorithm improves on the first for every measure. The constituent alignment strategy based on word overlap outperforms the head alignment approach, especially in precision, while recall seems to remain a weak point of both approaches. Padó and Pitel (2007) report that the best full constituent-based model on the French testset, with filters for non-aligned words and arguments, achieves 63.1 as best f-measure (0.66 precision, 0.60 recall). Our best results on EUROPARL scored the same precision but a lower recall. This discrepancy may depend on different factors. First of all, frame instance parallelism between English and French gold standards is higher than between English and Italian, with 0.69 target parallelism and 0.88 FE parallelism for the English-French couple against 0.61 and 0.82 on English-Italian EUROPARL (see Section 4.5). Besides, the French parser used in the pre-processing phase scores 76.3 f-measure, whereas the Bikel parser trained on Italian reaches 70.79 f-measure on a text with gold standard PoS[7].

[7] In our case the parser performance may be worse because PoS are automatically annotated

In order to verify the impact of wrong parse trees, we compared the performance of the two algorithms when starting from automatically generated parse trees and from the same trees after manual correction. The corresponding evaluation on EUROPARL highlighted that for algorithm 1 the correction step enhances precision of 0.14 and recall of 0.12. With the second algorithm, the values improved respectively of 0.14 and 0.9. This proves that parsing problems are a relevant source of error.

Results for MULTIBERKELEY are reported in Table 4.9.

MultiBerkeley	Precision	Recall	F1
Algorithm 2	0.75	0.49	0.59

Table 4.9: FE transfer evaluation (1) on MultiBerkeley

In this case, we could not apply algorithm 1 because it requires the source sentences to be represented as syntactic trees, whereas the English FrameNet corpus used to build the gold standard has annotation pointing to flat chunks without parsing information. Also for this second corpus, we evaluated the improvement of the algorithm on manually corrected parse trees on the Italian side. Precision scores an enhancement of 0.16, and recall of 0.11. The improvement via correction step is greater for MULTIBERKELEY than for EUROPARL. This means that in MULTIBERKELEY parsing problems are the main source of error, whereas in the EUROPARL corpus also other factors have a significant impact on the algorithm performance, for instance free translations. In general, we notice that the transfer approach performs better on a corpus like MULTIBERKELEY, where syntactic complexity is limited by the sentence length and the faithful translation of the parallel sentences enhances the performance of the aligner.

4.6.2 Evaluation 2

Basili et al. (2009) presented a fully automatic transfer process based on alignment with *Moses* (Koehn et al., 2007) at chunk level between English and Italian parallel sentences and a selection of the best candidate segment for semantic transfer according to some ranking and post-processing criteria. The algorithm was evaluated on the same subset of EUROPARL corpus that we used. However, they apply an evaluation framework that is different from that of Padó and Lapata (2009) presented in the previous section. In fact, they consider each FE and target annotation as independent and include in the testset only those FEs having the same label both in the Italian and in the English gold standard. In order to compare this approach to ours, we decided to adopt the same evaluation measures.

Accuracy is evaluated in two alternative ways, either on all semantic elements of the target language (both *targets* and *frame elements* together) or only on FEs. The transfer of target annotations was considered correct if the alignment was correct, independently of the actual label used in the Italian gold standard.

As for FEs, two kinds of match were computed: Perfect Matching (the projected segments in the target language exactly match with the gold standard ones) and Partial Matching (the intersection between the target projected segments and the ones in the gold standard is not empty). Moreover, in order to measure the gap between perfect and partial matching, evaluation included also token precision, recall and f-measure computed over all transferred labels (micro-average). We apply the same measures separately to the EUROPARL gold standard and to the MULTIBERKELEY corpus. In this way we can compare our algorithm run on EUROPARL to the evaluation by Basili et al. (2009), and compare the different algorithm performance obtained with EUROPARL and MULTIBERKELEY.

In Table 4.10 we report the evaluation of our annotation transfer with algorithm 2, which performs better than algorithm 1, run on the EUROPARL gold standard following the above mentioned criteria. We show the results of perfect and partial match applied to all semantic elements (targets + FEs), while the values for FEs only are reported between parenthesis.

Europarl	**PerfMatch**	**PartialMatch**
	LUs+FEs (FEs only)	LUs+FEs (FEs only)
	0.77 (0.66)	0.90 (0.89)

Token	**Precision**	**Recall**	**F1**
	0.83 (0.82)	0.75 (0.78)	0.79 (0.80)

Table 4.10: Evaluation 2 of Alg. 2 on Europarl

The best model reported in Basili et al. (2009) on the same gold standard scored 0.73 PerfMatch and 0.90 PartialMatch on LUs+FEs, and 0.42 and 0.78 respectively as PerfMatch and PartialMatch on FEs only. This means that both approaches reach high accuracy on target words, whereas our model performs significantly better on FEs only. In general, the two results reflect the different goals of the two approaches: Basili et al. (2009) are interested in investigating and adopting unsupervised techniques with poor semantic and syntactic information to automatically annotate a large scale training set and exploit it for semantic role labelling. On the contrary, we are interested in developing annotated resources with nearly manual quality, so we consider particularly important FE transfer precision.

We report in Table 4.11 the evaluation of algorithm 2 on the *MultiBerkeley* corpus following the same criteria mentioned above.

MultiBerkeley		**PerfMatch** LUs+FEs (FEs only)		**PartialMatch** LUs+FEs (FEs only)
		0.84 (0.75)		0.92 (0.88)
Token	**Precision**	**Recall**	**F1**	
	0.88 (0.85)	0.84 (0.86)	0.86 (0.85)	

Table 4.11: Evaluation 2 of Alg. 2 on M.Berkeley

As expected, the algorithm behaves differently on the two corpora, and all values obtained on MULTIBERKELEY outperform those on EUROPARL, except for PartialMatch on FEs only (0.88 vs. 0.89). This may depend on the fact that the constituents in the MULTIBERKELEY corpus are generally quite short, so the annotation transfer tend to be either a perfect match or to fail. On the contrary, the constituents in the EUROPARL sentences tend to be more complex, thus it is likely that they have at least one aligned token with the English source FE that matches with the gold standard, but exact match is less probable.

4.6.3 Evaluation 3: a proposal

A common feature of the two evaluation frameworks presented in Section 4.6.1 and 4.6.2 is that they exclude from evaluation cases of missing parallelism between source and target sentences. We propose a third approach based on 3 main ideas: 1) We think that it is preferable to evaluate separately targets and frame elements, because of the different nature of the two tasks: target transfer is more influenced by word alignment quality and is generally more straightforward than FE projection. On the other hand, the latter requires a different strategy because it involves selection procedures at chunk or constituent level. While target projection is mainly based on single-word alignment, FE projection requires both role identification and boundary detection. 2) Since we are interested in the (semi) automatic creation of FrameNet for new languages, we want to evaluate the quality of the resulting corpus as a whole, so we consider all transferred annotation regardless of parallelism between the two gold standards. 3) As for the evaluation of FE transfer, we propose two different criteria for assessing the match between automatic annotation and gold standard that are looser than the exact match condition. In both cases, the automatically annotated FE matches the gold standard FE if they share at least the same semantic head.

In this way, we can reduce the impact of parsing error on evaluation because we consider a match as correct if at least the main semantic element in the target sentence has been identified, even if the constituent boundaries are not exactly matching. Besides, this strategy minimize differences in manual annotation criteria, for example as regards the inclusion of punctuation at the beginning and at the end of constituents. A further advantage is that it allows for a direct comparison of FE transfer based on constituent alignment to other models relying on dependency relations. Some proposals have been recently put forward about the use of dependency graphs for frame annotation (Fürstenau, 2008, Johansson and Nugues, 2007a), and since dependencies in sentences are usually based on the concept of semantic heads, an evaluation approach relying on semantic heads could be easily applied both to constituency and to dependency representations.

The first type of FE matching criterion is more strict in that it requires that also the annotation of the corresponding targets match. Type 2, instead, considers correct all matching frame elements between automatic and manually annotated sentences regardless of whether the target has been annotated with the right frame.

We report in Table 4.12 the evaluation of target transfer on the two corpora. We don't distinguish between algorithm 1 and algorithm 2 on the EUROPARL corpus because the alignment step for targets is the same and relies on word alignment.

	Precision	**Recall**	**F1**
Europarl	0.71	0.50	0.59
MultiBerkeley	0.93	0.81	0.86

Table 4.12: Target transfer evaluation

Europarl	**Precision**	**Recall**	**F1**
Algorithm1			
Type 1	0.46	0.30	0.37
Type 2	0.64	0.41	0.49
Algorithm2			
Type 1	0.55	0.28	0.37
Type 2	0.64	0.32	0.43

Table 4.13: FE transfer evaluation 3 on Europarl

In Table 4.13 we report the evaluation of FE transfer on the EUROPARL corpus according to the two criteria we have proposed, using both algorithm 1 and algorithm 2. The results reflect different features of the two algorithms that had not been highlighted in the previous evaluations. In particular, algorithm 2 achieves a better performance on precision for evaluation Type 1, but the overall recall value are worse for both types. Since FE transfer in algorithm 2 depends on a correct target transfer, it is clear that missing target alignments influence in turn also the FE transfer performance. The evaluation shows that it is probably better to make the two transfer steps independent, like in algorithm1, so that one can try and align FEs even if no target has been transferred.

In Table 4.14 we report the evaluation of FE transfer on the MULTIBERKELEY corpus according to the two criteria we have proposed and applying algorithm 2.

MultiBerkeley	**Precision**	**Recall**	**F1**
Type 1	0.68	0.54	0.60
Type 2	0.69	0.55	0.61

Table 4.14: FE transfer evaluation 3 on MBerk.

All results on MULTIBERKELEY generally achieve an improvement w.r.t. EUROPARL, particularly on recall. This can be explained by the nature of the corpus, that maximizes word alignment, so that less constituents are left out in the alignment step. Moreover, we noticed in the EUROPARL corpus a greater difference between type 1 and type 2 than in MULTIBERKELEY. In fact, in the former there are a lot of frames that are semantically related and share the same frame elements (for example *Cognizer* is a core FE of several frames in the corpus such as AWARENESS, CERTAINTY, COMING_TO_BELIEVE, JUDGMENT, OPINION, etc.). For this reason, the set of all matching frame elements between automatic and manually annotated sentences regardless of the frame identity (type 2) is bigger than that considering also the corresponding target match (type 1). In MULTIBERKELEY, instead, the two sets almost coincide because the frame variability is much higher, thus it is less likely that two frame elements of different sentences are the same even if the frame is different.
Error analysis shows that transfer quality of the EUROPARL corpus is crucially affected by syntactic complexity and free translation of the target corpus, which in turn impact on alignment quality. See for example the sentences reported at 4.7:

(4.7) EN: 85% of Mexico's exports go north.
ITA: L'85 percento delle esportazioni messicane è destinato all'America del nord.
85 percent of Mexican exports are destined to North America.

In order to determine the parallelism between the two sentences, we need to make the inference that North America is north of Mexico, which is out of the current capability of any word-alignment tool. Furthermore, "*go*" and "*essere destinato*" (*to be destined*)" do not exactly express the same predicate and it is likely that they won't be aligned.

Other problems involve both corpora and arise from different interpretations given by the annotators to the aligned sentences, which may depend also on inherent ambiguity of FrameNet definitions. For example, in the STATEMENT frame, English annotators tend to prefer to label as *Topic* the content of the communication, whereas in Italian it is mostly annotated as *Message*. Probably the difference between the two frame elements is not clear enough, especially if not applied to English. Other minor problems depend on the recognition and alignment of multiwords in Italian. In general, both algorithms fail to find the correct constituent for frame element transfer in case of complex tree nodes, where different terminals and nodes dominated by the same parent bear different FE labels.

4.7 Summary

In this Chapter, we have focused on the projection of frame annotation from English to Italian and we have investigated to what extent similar approaches developed for other language pairs can be applied to this language pair. The research activity has dealt with different issues involving the development of the projection algorithm, the choice of the gold standard for evaluation and the selection of the evaluation metrics. In particular, we tried to answer the 3 following questions: (1) What is the best annotation transfer algorithm for the English-Italian couple? (2) What kind of parallel corpus is best suitable to the annotation transfer task? (3) How should the annotation transfer be evaluated, given the final aim of the transfer?

In order to answer the first question, we gave an overview of the existing projection approaches for automatic frame annotation and we presented two algorithms developed for the English-Italian pair. While the target transfer step was based in both cases on word alignment, the two algorithms are consistently different in the procedure for aligning English and Italian constituents, which is the preliminary step to FE transfer. In particular, one algorithm based such alignment on the semantic head of the constituents, while the second employed a similarity function relying on the maximum word alignment. A comparison between the two approaches in the light of some indicative examples helped us to highlight the pros and cons of each approach. We came to the conclusion that the methodology exploiting the word overlap measure as similarity function can be better generalized and is significantly simpler to implement because it requires less compatibility rules dependent on the language pair. On the other hand, an approach that carries out target and FE transfer independently can contribute to achieving a better recall.

Another main concern of our investigation was to understand to what extent different types of corpora can influence the transfer process. For this reason, we developed two different gold standards for evaluation: one was extracted from the English-Italian bitext in the EUROPARL corpus, while the other was created by manually translating into Italian a selection of English sentences from the Berkeley FrameNet database, which we called MULTIBERKELEY. The former includes parallel sentences with high syntactic complexity but low topic variability, since it deals mainly with political and social matters. Besides, a lot of translational divergences are reported. MULTIBERKELEY, instead, has been created in order to maximize the semantic and syntactic parallelism of the English-Italian sentence pairs, but also to include a wide range of topics, i.e. of frames.

While evaluation results on the EUROPARL subcorpus were still unsatisfactory because they did not allow for a completely automatic development of FrameNet-like resources, we noticed that with MULTIBERKELEY we could optimize algorithm performance and minimize alignment errors. Evaluation results show that the translation effort to produce the corpus is repaid by the remarkable reduction of correction work.

As for the third issue, i.e. the best evaluation framework for the projection task, we took into account three methodologies. Two of them have already been applied in previous evaluations of the projection task, while the third one represents our proposal. We showed that the evaluation results can considerably change according to the framework applied and we suggested that the evaluation approach should depend on the goal of the transfer task. In particular, experiments aimed at assessing the performance of automatic semantic role labeling via annotation projection should apply evaluation metrics focusing only on FE transfer, as in Padó and Lapata (2009). On the contrary, if the primary investigation concerns the development of methodologies for the automatic creation of new FrameNets, as in our case, evaluation should be more general and include the whole transfer task under real-world conditions, i.e. it should comprise both frame and FE transfer.

At the moment, the annotation transfer strategy cannot produce new FrameNets in a fully automatic and reliable way, so it should be seen as a starting point in this direction. For example, even if a good transfer performance is achieved on MULTIBERKELEY, it requires a *controlled* translation. Besides, transferring only one sentence per frame allows for covering only one of the possible valence patterns of the frame. For this reason, a simplified corpus like MULTIBERKELEY we can be seen as a *seed* database, that represents the starting point for investigating procedures to automatically acquire new and more complex example sentences.

Chapter 5

Using WordNet to populate Italian frames

5.1 Introduction

In this Chapter, we show that the semi-automatic development of FrameNet-like resources can go beyond the automatic annotation of example sentences. Indeed, it can be carried out also at frame level, through the automatic extraction of new LUs. In particular, we will propose a methodology to link FrameNet frames and WordNet synsets (Fellbaum, 1998) in English and then to acquire new Italian lexical units using MultiWordNet (Pianta et al., 2002) as a bridge. Besides, the mapping between the two resources can be exploited to add frame labels to the MultiSemCor corpus (Bentivogli and Pianta, 2005).

Our approach makes use of a supervised learning framework for the mapping of FrameNet lexical units onto WordNet synsets based on a reduced set of novel and semantically rich features. The proposed approach addresses some of the limitations of previous works on the same task (see for example DeCao et al. (2008) and Johansson and Nugues (2007b)). Most notably, as we do not train the classifier on a per-frame basis, our model is able to cope also with those frames that have little or no annotated sentences to support the frame description. After learning a very fast model on a small set of annotated lexical unit-synset pairs, we can automatically establish new mappings in never-seen-before pairs and use them for our applications.

The discussion is structured as follows: in Section 5.2 we describe the main characteristics of WordNet and MultiWordNet; in Section 5.3 we motivate the mapping task comparing the information delivered by FrameNet and WordNet and describing the advantages of the mapping. In Section 5.4 we discuss previous works dealing with the task; in Section 5.5 we formalize the mapping task and we describe our supervised approach to link lexical units with synsets; Section 5.6 details the dataset that we employed for our experiments; Section 5.7 describes the novel features that we used to characterize the mapping; in Section 5.8 we discuss the results of our experiments and the contribution of the different features; in Section 5.9 we apply the mapping to three different tasks: the automatic induction of new LUs for English, the identification of LUs for Italian FrameNet and the annotation of MultiSemCor with frame labels; finally, in Section 5.10 we give an overview of the chapter.

5.2 WordNet and MultiWordNet

WordNet (Fellbaum, 1998) is a lexical resource for English based on psycholinguistics principles and developed at Princeton University. It has been conceived as a computational resource aimed at improving some drawbacks of traditional dictionaries such as the circularity of definitions and the ambiguity of sense references. At present, version 3.0 contains 210,000 entries that cover the majority of nouns, verbs, adjectives and adverbs in the English language, organized in synonym sets called *synsets*, which correspond to lexical concepts. WordNet also includes a rich set of semantic relations across concepts, such as hyponymy, entailment, antonymy, similar-to, etc. Each synset is encoded as a set of synomyms having the same part of speech and described by a definition or *gloss*. In some cases, one or more example sentences may also be reported. For example, the lemma *bottleneck* occurs in 4 synsets of WordNet v.3 (`http://wordnetweb.princeton.edu/perl/webwn`), two for *bottleneck* as a noun and two as a verb, in the following form:

(5.1) (n) constriction#1, **bottleneck#1**, checkpoint#1 (a narrowing that reduces the flow through a channel)
(n) **bottleneck#2** (the narrow part of a bottle near the top)
(v) **bottleneck#1** (slow down or impede by creating an obstruction)
"*His laziness has bottlenecked our efforts to reform the system*"
(v) **bottleneck#2** (become narrow, like a bottleneck)
"*Right by the bridge, the road bottlenecks*"

Given a set of synsets containing the same lemma, as above, an intuitive way to understand the meaning of the different senses and to disambiguate them is to consider the lemmas included in the same synset or to look at the semantic relations linking each synset to other synsets. In (5.1), for example, the direct hypernym of *bottleneck.n#1* is *narrowing.n#1* (an instance of becoming narrow), while the hypernym of *bottleneck.n#2* is formed by *part.n#2* and *portion.n#2* (something less than the whole of a human artifact).

The Princeton English WordNet (PWN) has been augmented with domain labels (Magnini and Cavaglià, 2000) that group synsets into homogeneous clusters. Domains can include synsets of different syntactic categories, given that they share a strong semantic relation. For example, the Medicine domain includes nouns such as *doctor.n#1* and *emergency_room.n#1* but also verbs such as *operate.v#7*.
Domain labels include 164 categories manually selected from subject field codes used in current dictionaries and from the categories of the Dewey Decimal Classification (DDC), the most widely used taxonomy for library organization. The labels are organized in five main trees reaching a maximum depth of four, with an increasing degree of specialization. For example, the Doctrines tree is divided into four *basic domains*, i.e. Art, Psychology, Philosophy and Religion. Art encompasses six more specialized classes, which are Theatre, Dance, Drawing, Music, Photography and Plastic arts. The latter, in turn, contains Sculpture, Numismatics and Jewellery, and so on (Bentivogli et al., 2004).

Domain information can contribute to the disambiguation of a lemma belonging to two or more synsets. The verb *coin.v*, for example, occurs in two synsets, one belonging to the Literature domain and the other to the Money domain. If we can determine the overall domain dealt with by the text containing the ambiguous word, this may be sufficient for choosing the right sense.

MultiWordNet (Pianta et al., 2002) is a multilingual lexical database where the synsets of different languages are strictly aligned with the Princeton WordNet 1.6. The basic assumption is that two synsets in PWN being connected by a certain relation keep that relation also in another language. For example, if *vertebrate#1* is the hypernym of *bird.n#1*, and if they are aligned respectively to *vertebrato.v#* and *uccello.n#1* in the Italian part of MultiWordNet, then we can assume that *vertebrato.v#* is a hypernym of *uccello.n#1*.
MultiWordNet differs from EuroWordNet (Vossen, 1998), another project aimed at the development of WordNets for the main European languages, in the procedure followed for the database development. In EuroWordNet, in most cases the language-specific WordNets have been first created independently and then there was an attempt to find correspondences among them. On the contrary, MultiWordNet has been created trying to preserve the semantic relations in the PWN and building the new synsets starting, if possible, from the corresponding English synsets. For this reason, MultiWordNet is particularly suitable for experimenting algorithms that exploit the information available for one language to extract new

data for other languages, like the one we present in the following sections.

The MultiWordNet database that can be browsed via the online interface (`http://multiwordnet.fbk.eu/online/multiwordnet.php`) is available in English, Italian, Spanish, Portuguese, Hebrew, Romanian and Latin. The distributed version, instead, contains only the Italian database aligned with the Princeton WordNet.

5.3 FrameNet and WordNet

From a theoretical point of view, a mapping between FrameNet and WordNet would be an interesting issue because it would help investigate to what extent these two different approaches can be integrated. Baker and Fellbaum (2009) have presented a case study in which a short text passage is annotated with the two frameworks, showing that both paradigms contribute to text understanding in a complementary way. Besides, they reported an ongoing project about the semi-automatic annotation of lemmas from the American National Corpus with frame and synset information, in order to help align the word senses of the two resources and to provide a gold standard for automatic frame and synset detection.

As highlighted by Boas (2005a), the key differences between FrameNet and WordNet mainly depend on the different theoretical approaches underlying the two resources, i.e. frame semantics vs. more "traditional" lexical semantic relations and psycholinguistic principles. Besides, FrameNet was originally conceived and organized as a resource for computational lexicography, while WordNet was created primarily as a semantic database. As such, the former is provided with a rich list of lexico-syntactic patterns and a set of corpus examples for every LU, whereas the latter has a richer and more structured hierarchy of semantic relations such as synonymy, antonymy, polysemy, etc. and a more systematic treatment of polysemy. Furthermore, WordNet provides also frequency information, has a better coverage especially for adjectives and nouns and is more fine-grained w.r.t. frame information. In the light of these considerations, we believe that mapping FrameNet LUs to WordNet synsets would be very useful for the following reasons:

- Most of *WordNet* information is of paradigmatic nature. Almost no, or very shallow, valency information is represented. The link between synsets and frames adds information of syntagmatic nature about how concepts combine to describe situations in texts.

- For the *FrameNet* side, a mapping would automatically increase the number of LUs for frame by importing all synonyms from the mapped synset(s), and would allow to exploit the semantic and lexical relations in WordNet to enrich the information encoded in FrameNet. This would help coping with coverage problems and disambiguating the LU senses[1].

- Such mapping would allow us to model frame-based resources *for new languages* using minimal supervision, because we could first link the frames in the Berkeley database with the English synsets and then automatically populate frame sets for new languages via MultiWordNet by importing all lemmas from the mapped synsets.

In Figure 5.1, we show as an example the information that could be merged from FrameNet and WordNet by linking the *court.n* LU in the JUDICIAL_BODY frame with the synset n#06176884.

First, it would be possible to import the *judicature.n* lemma in JUDICIAL_BODY, and to acquire it as a new lexical unit. Then, the mapping would establish a connection between the frame and the Law domain. Third, the frame element description could be used also for the lemmas in the synset, adding some information about the semantics of their valence. Then, the annotated sentences in the FrameNet database containing the *court.n* and the *tribunal.n* LUs for the JUDICIAL_BODY frame could be associated to the corresponding lemmas in the synset, providing more examples for this sense.

As for the automatic development of multilingual FrameNets, the figure includes also the content of the synset in Italian, Spanish, Rumanian and Portuguese as reported in MultiWordNet. By linking the English synset with the frame, we could assume that also the lemmas listed for the different languages in synset n#06176884 are lexical units of the JUDICIAL_BODY frame: if no FrameNet is available for these languages, this could be a first step towards the automatic population of frames with lexical units.

5.4 Previous mapping approaches

Several experiments have been carried out to develop a FrameNet-WordNet mapping. Shi and Mihalcea (2005) described a semi-automatic approach to exploit VerbNet as a bridge between FrameNet and WordNet for verbs, using synonym and hyponym relations and similarity between Levin's verb classes and FrameNet frames. Their mapping was used to develop a rule-based semantic parser (Shi and

[1]This would require a preliminary investigation about the compatibility of WordNet domains and the frame hierarchy, which may be worth including in our future research

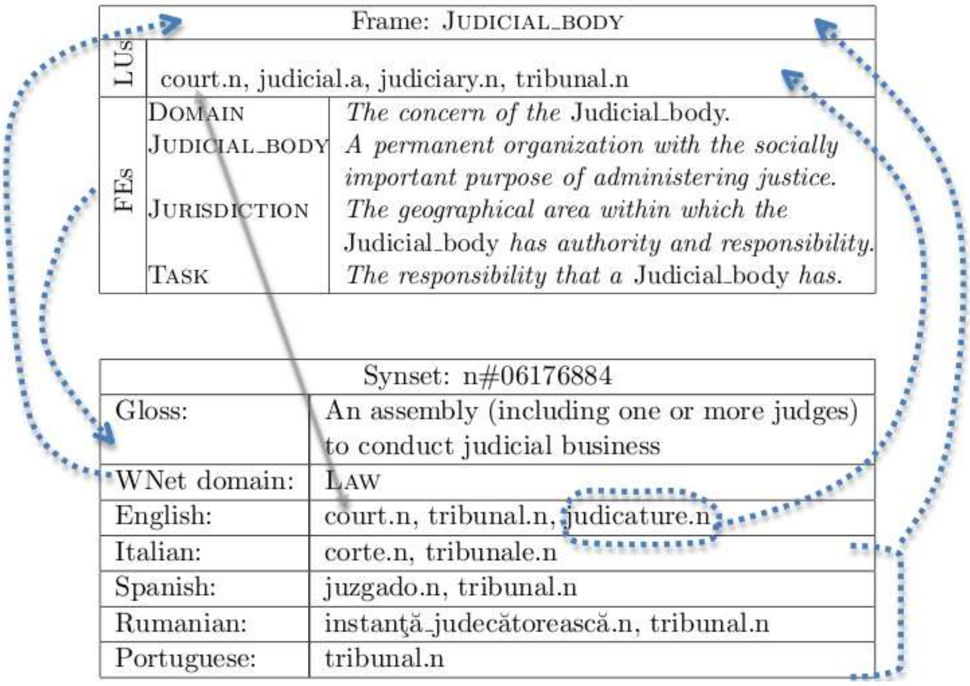

Figure 5.1: FrameNet - WordNet mapping of *court.n*

Mihalcea, 2004) as well as to detect target words and assign frames for verbs in an open text (Honnibal and Hawker, 2005).

Burchardt et al. (2005) presented a rule-based system for the assignment of FrameNet frames by way of a "detour via WordNet". They applied a WordNet-based WSD system to annotate lexical units in unseen texts with their contextually determined WordNet synsets and then exploited synonyms and hypernyms information to assign the best frame to the lexical units. The system was integrated into the SALSA RTE system for textual entailment (Burchardt et al., 2007) to cope with sparse-data problems in the automatic assignment of frame labels.

Johansson and Nugues (2007b) created a feature representation for every WordNet lemma and used it to train an SVM classifier for each frame that tells whether a lemma belongs to the frame or not. The best-performing feature representation was built using for each synset the sequence of its hypernyms and a weight based on their relative frequency in the SemCor corpus[2]. They used the mapping in the Semeval-2007 task on frame-semantic structure extraction (Baker et al., 2007) in

[2]SemCor (Landes et al., 1998) is a collection of 352 texts extracted from the Brown corpus, syntactically tagged with Brill's part of speech tagger and manually annotated with WordNet synsets.

order to find target words in open text and assign frames.

Crespo and Buitelaar (2008) carried out an automatic mapping of medical-oriented frames to WordNet synsets applying a *Statistical Hypothesis Testing* to select synsets attached to a lexical unit that were statistically significant using a given reference corpus. The mapping obtained was used to expand Spanish FrameNet using *EuroWordNet* and evaluation was carried out on the Spanish lexical units obtained after mapping.

Given a set of lexical units, DeCao et al. (2008) proposed a method to detect the set of suitable WordNet senses able to evoke a frame by applying a similarity function that exploits different WordNet information, namely conceptual density for nouns, synonymy and co-hyponymy for verbs and synonymy for adjectives. The mapping approach was applied also to LU induction for the English FrameNet and for Italian frames via MultiWordNet.

More recently, Laparra and Rigau (2009) presented a methodology to integrate FrameNet and WordNet which is based on a knowledge-based word sense disambiguation algorithm called SSI-Dijkstra. The basic idea is to order all LUs in a frame by polysemy degree and to link the less ambiguous ones to the most probable sense in WordNet. Then, the algorithm is applied to disambiguate the remaining ambiguous LUs selecting the synsets that are closer to the senses of the already disambiguated words. This process is carried out by computing path distances on a very large connected graph containing almost 100,000 nodes (synsets) and 636,077 edges (the relations between synsets). The approach outperforms past mapping algorithms, but it can provide answers only for those frames having at least two LUs, because one or more context words are needed to disambiguate every LU.

5.5 Problem formulation

Our objective is to be able to assign to every lexical unit l, belonging to a frame F_i defined in the FrameNet database, one or more WordNet senses that express in the best way the meaning of l. More specifically, for every $l \in F_i$, we consider the set of all WordNet senses where l appears, *CandSet*, and then find the best WordNet sense(s) $best_s \subset CandSet$ that express the meaning of l.

For example, the lexical unit *rouse.v* belonging to the CAUSE_TO_WAKE frame, is defined in FrameNet as "bring out of sleep; awaken". Its *CandSet* comprises 4 WordNet senses[3]: 1# *bestir, rouse* (become active); 2# *rout_out, drive_out, force_out, rouse* (force or drive_out); #3 *agitate, rouse, turn_on, charge, commove, excite,*

[3]The gloss is reported between parenthesis

charge_up (cause to be agitated, excited or roused); #4 *awaken, wake, waken, rouse, wake_up, arouse* (cause to become awake or conscious). In this example, $best_s = \{\#4\}$ for *rouse.v* in CAUSE_TO_WAKE.

Our aim is to develop a mapping system that can achieve a good accuracy also with poorly-documented lexical units and frames. In fact, we believe that under real-usage conditions, the automatic induction of LUs is typically required for frames with a smaller LU set, especially for those with only one element. In the FrameNet database (v. 1.3), 33 frames out of 720 are described only by one lexical unit, and 63 are described by two. Furthermore, more than 3,000 lexical units are characterized only by the lexicographic definition and are not provided with example sentences. For this reason, we suggest an approach that makes also use of usually unexploited information in the FrameNet database, namely the *definition* associated to every lexical unit, and disregards example sentences.

This is the main point of difference between our and other similar works, e.g. Johansson and Nugues (2007b), DeCao et al. (2008) and Laparra and Rigau (2009), where unsupervised approaches are proposed which strongly rely either on the number of lexical units in a frame or on the example sentences available for l in the FrameNet corpus. We claim that the relative short time necessary to annotate a small dataset of frame-synset pairs will result in a more reliable mapping system and, as a consequence, in consistent time savings when we actually try to use the mappings for some tasks. The ability to cope with different cases while retaining a good accuracy makes it possible to bootstrap the mapping process in many cases where other approaches would have failed due to lack of training data.

To this end, we can train a binary classifier that, given l and *CandSet*, for each pair $\langle l, s\rangle$, $s \in CandSet$, delivers a positive answer if $s \in best_s$, and a negative one otherwise. To follow on the previous example, for *rouse.v* we would have 4 classifier examples, i.e. the pairs ⟨rouse.v,#1⟩, ⟨rouse.v,#2⟩, ⟨rouse.v,#3⟩ and ⟨rouse.v,#4⟩. Of these, only the last would be considered a positive instance. As a learning framework, we decided to use Support Vector Machines (SVMs) due to their classification accuracy and robustness to noisy data (Vapnik, 1998).

5.6 Dataset description

In order to train and test the classifier, we created a gold standard by manually annotating 2,158 LU-synset pairs as positive or negative examples. We don't have data about inter-annotator agreement because the dataset was developed only by one annotator, but DeCao et al. (2008) report 0.90 as Cohen's Kappa computed over 192 LU-synset pairs for the same mapping task. This confirms that senses and lexical units are highly correlated and that the mapping is semantically motivated.

The annotation process can be carried out in reasonable time. It took approximately two work days to an expert annotator to manually annotate the 2,158 pairs that make up our gold standard. The lexical units were randomly selected from the FrameNet database regardless of their part of speech or amount of annotated data in the FrameNet database. For each lexical unit, we extracted from WordNet the synsets where the LU appears. Then, we assigned a positive label in case the LU-synset pairs share the same meaning, and a negative label otherwise. Statistics about the dataset are reported in Table 5.1.

N. of LU-synset pairs	2,158
N. of lexical units	617
Verbal lexical units	39%
Nominal lexical units	51%
Adjectival lexical units	9%
Adverbial lexical units	<1%
Targeted frames	375
Pairs annotated as positive	32%
Pairs annotated as negative	68%
Average polysemy	3.49
LUs with one candidate synset	204
LUs with 10 or more cand. synsets	32

Table 5.1: Statistics on the dataset

The 375 frames that are present in the dataset represent about one half of all lexicalized frames in the FrameNet database. This proves that, despite the limited size of the dataset, it is well representative of FrameNet characteristics. This is confirmed by the distribution of the part of speech. In fact, in the FrameNet database about 41% of the LUs are nouns, 40% are verbs, 17% are adjectives and <1% are adverbs (the rest are prepositions, which are not included in our experiment because they are not present in WordNet). In our dataset, the percentage of nouns is higher, but the PoS ranking by frequency is the same, with nouns being the most frequent PoS and adverbs the less represented. The average polysemy corresponds to the average number of candidate synsets for every LU in the dataset. Note that the high number of lexical units with only one candidate does not imply a more straightforward mapping, because in some cases the only candidate represents a negative example. In fact, a LU could be encoded in a frame that does not correspond to the sense expressed by the synset.

5.7 Feature description

For every LU-synset pair in the gold standard, we extracted a set of features that characterize different aspects of the mapping. In the remainder, we detail the meaning as well as the feature extraction procedure of each of them.

Stem overlap (2 real features): Both WordNet glosses and LU definitions in FrameNet are manually written by lexicographers. We noticed that when they share the same sense, they show high similarity, and sometimes are even identical. For example, the definition of *thicken* in the *Change_of_consistency* frame is *"become thick or thicker"*, which is identical to the WordNet gloss of synset n. v#00300319. The *thicken* lemma occurs in three WordNet synsets, and in each of them it is the only lemma available, so no other information could be exploited for the sense disambiguation.

We believe that this information could help in the choice of the best candidate synset, so we stemmed all the words in the synset gloss and in the lexical unit definition and measured their overlap. As features, we used the ratio between the number of overlapping stems and the number of stems resp. in the gloss and in the LU description. In a more formal way, we can define these two features as follows:

Let X be the set of (non-stop-word) stems in the definition of l, and Y the stems in the gloss of s. The two features are defined as $S_{WN} = \frac{|X \bigcap Y|}{|Y|}$ and $S_{FN} = \frac{|X \bigcap Y|}{|X|}$.

For example, if we consider $l =$ *bejewelled.a* in $F_i =$ Abounding_with and $s =$ *a#00057580*, we extract the following definitions[4]:

WNet gloss of *a#00057580* (4 stems): "**covered** with beads or **jewels** or sequins"

Definition of *bejewelled.a* (2 stems): "**covered** with **jewels**"

Then we compute the following feature values:

$S_{WN} = 2/4 = 0.5$

$S_{FN} = 2/2 = 1$

[4]The words in bold contained in the definitions are overlapping. The number of stems reported between parenthesis is computed discarding stop words.

Prevalent Domain and Prevalent Synset (boolean): Since a frame represents a prototypical situation evoked by a set of lexical units, our intuition is that it should be possible to assign a frame to a WordNet domain that groups homogeneous clusters of semantically similar synsets (see Section 5.2).

Given the LU-synset pair $\langle l, s\rangle$, $l \in F_i$, $s \in CandSet$, we extract all the lexical units in F_i and then build a set *AllCandSet* of pairs $\langle s_j, c_j\rangle$, where s_j is a synset in which at least one $l_i \in F_i$ appears, and c_j is the count of all lexical units of F_i that are found in s_j.

We exploit the information conveyed by *AllCandSet* in two ways: 1) if there is a prevalent WordNet domain that characterizes the majority of the synsets in *AllCandSet*, and $s \in CandSet$ belongs to that same domain, we add a boolean feature to the feature vector representing $\langle l, s\rangle$. 2) if s is the synset with the highest count in *AllCandSet*, i.e. if $s = s_j$ and $c_j > c_i \forall \langle s_j, c_j\rangle \in AllCandSet$, $i \neq j$, then we add another boolean feature to encode this information.

The basic assumption about the concept of prevalent domain is that, if the majority of LUs in a frame can be linked to a WordNet domain, then it is very likely that the domain is representative of the semantics of the given frame and that the synsets included in that domain have a connection with the LUs in the frame. For example, 149 LUs in the CLOTHING frame occur also in synsets that belong to the Fashion domain, so we can assess a relationship between the frame and the domain and between the LUs and the synsets $\subset$ Fashion. In order to assign this feature, we first computed the mapping between frames and their prevalent domains. We were able to assign one domain label to 85 frames[5], 10 of which are reported in Table 5.2:

Frame	Domain	N. of common LUs/lemmas
CLOTHING	Fashion	149
OBSERVABLE_BODYPARTS	Anatomy	81
BUILDING_SUBPARTS	Buildings	64
FOOD	Gastronomy	58
MEDICAL_CONDITIONS	Medicine	56
BUILDINGS	Buildings	54
CALENDRIC_UNIT	Time_period	53
ACCOUTREMENTS	Fashion	42
INTOXICANTS	Pharmacy	31
WEAPON	Military	31

Table 5.2: The 10 WordNet domains most frequently assigned to a frame

The information extracted about the frame-domain pairings is then exploited to

[5]The general domain Factotum was not taken into account for the assignment

assign a positive binary feature to all synsets that could be a good candidate for a given LU. For example, if we want to find the synset that most likely expresses the meaning of *regalia.n* in ACCOUTREMENTS, we see that it occurs in two WordNet synsets, namely n#03225941 "paraphernalia indicative of royalty (or other high office)" in the Politics domain and n#02212047 "especially fine or decorative clothing" belonging to the Fashion domain. According to Table 5.2, we give our preference to the latter because a connection between ACCOUTREMENTS and Fashion is more probable.

As for the most frequent synset, the basic idea is similar to that of the most frequent domain, i.e. we assume that if a frame and a synset share a high number of lemmas, than it is very likely that their semantics is similar, or that the meaning of the synset is included in the semantics of the frame. In the light of the mapping task, for example, we assume that if we want to find the synset that best corresponds to the *mess_up.v* LU in the BUNGLING frame, the synset v#01723558 "make a mess of, destroy or ruin" is a better candidate than v#00951547 "disturb the smoothness of" because the former synset shares 17 lemmas with BUNGLING , while the latter has no common elements with it. For this reason, we assign a positive binary feature to all LUs in BUNGLING that appear also in v#01723558.

Cross-lingual parallelism (boolean): Our idea is that, if an English lexical unit and its Italian translation equivalent belong to the same frame, they are likely to appear in an English and an Italian synset that are aligned in MultiWordNet, and vice versa. More formally, given an English lexical unit l belonging to a frame F_i and an English synset s, if the Italian translation equivalent of l occurs in an Italian synset that is aligned to s in MultiWordNet, then it is likely that the meaning of l corresponds to the meaning of s.

We used the two gold standards developed for cross-language transfer experiments and described in Section 4.5. Such corpora contain about 1,300 English - Italian parallel sentences altogether, enriched with manually annotated frame information on both sides.

Given a pair $\langle l, s\rangle$, we check if l appears in the English side of the parallel corpus with the frame label F_i and extract its Italian translation l_{it}. If l_{it} appears also in the Italian synset aligned to s in MultiWordNet, we consider s as a good candidate for the mapping of l and encode this information as a binary feature.

For example, given F_i = IMPRISONMENT, l = put_away.v and s = v#01699803, we identified in the parallel corpus the following sentences (the lexical units are underlined):

S_{en} = *Nelson Mandela was put away in 1962.*

S_{it} = *Nelson Mandela fu incarcerato nel 1962.*

In *MultiWordNet*, *put_away.v* appears in the three synsets below:

v#01699803: {put_away} - {incarcerare}

v#01520955: {put_away} - {gettare}

v#00921958: {put_away} - {riporre}

Since only v#01699803 contains also *incarcerare.v*, its pairing with *put_awat.v* in IMPRISONMENT is seen as a very likely mapping and is assigned a positive binary feature.

Simple synset-frame overlap (real feature): Intuitively, the more lemmas a frame and a synset have in common, the more semantically similar they are. In order to take into account this similarity in our feature vector, given the pair $\langle l, s \rangle$, $l \in F_i$, we extract all lexical units in F_i and all lemmas in s and we compute the number of overlapping elements. Then we divide the value by the number of LUs in F_i excluding l . In a more formal definition, if L_i is the set of LUs in F_i besides l, the value of this feature is $\frac{|L_i \bigcap s|}{|L_i|}$, i.e. the number of lemmas in s overlapping those in L_i, divided by the number of LUs in L_i. For example, given F_i = IMPRISONMENT, l = incarcerate.v and s = v#01699803:

L_i = {**imprison.v**, imprisonment.n, incarceration.n, **jail.v**, **put_away.v**}

s = {**imprison**, incarcerate, lag, immure, **jail**, jug, **put_away**, ...}

Overlap = 0.6 (3/5)

Extended synset-frame overlap (real feature): This feature is a generalization of overlapping value described above because it considers also the hypernym information in WordNet to disambiguate the synsets. In other words, we take into account not only the overlaps according to the previous criterion, but also the number of overlapping words between the lexical units in a frame and the hypernyms of a synset. For example, the *party.n* lexical unit in the AGGREGATE frame has 5 senses in WordNet. According to the previous criterion, there is no overlap between the LUs in the frame and the lemmas in any of the five synsets. Instead, if we look at the direct hypernym relation of *party*, we find that sense #3 is described as *set, circle, band*, that are also lexical units of AGGREGATE.

In the cases where the hypernym relation is not defined, e.g. adjectives, we used the *similar-to* relation.

5.8 Experimental setup and evaluation

To evaluate our methodology we carried out a 10-fold cross validation using the available data, splitting them in 10 non-overlapping sets. For each iteration, 70% of the data was used for training, 30% for testing. All the splits were generated so as to maintain a balance between positive and negative examples in the training and test sets.

We used the SVM optimizer SVM-Light[6] Joachims (1999), and applied polynomial kernels (*poly*) of different degrees (i.e. 1 through 4) in order to select the configuration with the best generalization capabilities. The accuracy is measured in terms of Precision, Recall and F_1 measure, i.e. the harmonic average between Precision and Recall. For the sake of annotation, it is important that an automatic system be very precise, thus not producing wrong annotations. On the other hand, the higher the recall, the larger the amount of data that the system will be able to annotate.

The macro-average of the classifier accuracy for the different configurations is shown in Table 5.3. We report results for linear kernel (i.e. poly 1), maximizing recall and f-measure, and for polynomial kernel of degree 2 (i.e. poly 2), scoring the highest precision. In general, we notice that all our models have a higher precision than recall, but overall are quite balanced. Different polynomial kernels (i.e. conjunction of features) do not produce very relevant differences in the results, suggesting that the features that we employed encode significant information and have a relevance if considered independently.

As a comparison, we also carried out the same evaluation by setting a manual threshold and considering a LU-synset pair as a positive example if the sum of the feature values was above the threshold. We chose two different threshold values: the first (Row 1 in Table 5.3) was selected so as to have comparable precision with the most precise SVM model (i.e. poly2), while the second (Row 2) was set in order to have recall comparable with poly1, i.e. the SVM model with highest recall. In the first case, the model has a recall that is less than half than poly2, i.e. 0.214 vs. 0.569, meaning that such model would establish a half of the mappings while making the same percentage of mistakes. In the second, the precision of the SVM classifier is 0.114 points higher, i.e. 0.794 vs. 0.680, meaning the SVM can retrieve

[6]Available at http://svmlight.joachims.org/

as many mappings but making 15% less errors.

In order to investigate the impact of different features on the classifier performance, we also considered three different groups of features separately: the ones based on stem overlap, those computed for prevalent domain and synset, and the features for simple and extended frame - synset overlap. We did not take into account cross-lingual parallelism because it is one single feature whose coverage strongly relies on the parallel corpus available. As a consequence, it is not possible to test the feature in isolation due to data sparseness.

Results are shown in Table 5.3, in the second group of rows. Also in this case, we carried out a 10-fold cross validation using a polynomial kernel of degree 2. The stem overlap features, which to our best knowledge are an original contribution of our approach, score the highest recall among the three groups. This confirms our intuition that LU definitions and WordNet glosses can help extending the number of mapped LUs, including those that are poorly annotated. For instance, if we consider the KNOT_CREATION frame, having only *tie.v* as LU, the features about prevalent domain & synset and about synset-frame overlap would hardly be informative, while stem overlap generally achieves a consistent performance regardless of the LU set. In fact, *tie.v* is correctly mapped to synset v#00095054 based on their similar definition (respectively *"to form a knot"* and *"form a knot or bow in"*). Best precision was scored by the feature group considering prevalent domain & synset, which are also new features introduced by our approach. The positive effect of combining all features is clearly shown by comparing the results obtained with individual feature groups against the figures in the row labeled *poly2*.

	Prec.	Recall	F1
Man. thresh. (P)	0.789	0.214	0.337
Man. thresh. (F1)	0.680	0.662	0.671
Stem Overlap	0.679	0.487	0.567
Prev.Dom.& Syn.	0.756	0.434	0.551
Syn.- Frame Overlap	0.717	0.388	0.504
poly1	0.761	**0.613**	**0.679**
poly2	**0.794**	0.569	0.663

Table 5.3: Mapping evaluation

5.9 MapNet and its applications

Since we aim at assigning at least one synset to every lexical unit in FrameNet, we considered all the frames and for every LU in the database we created a list of LU-synset pairs. We re-trained the classifier using the whole annotated gold standard and classified all the candidate pairs. The mapping produced between the two resources, that we call *MapNet*, comprises 5,162 pairs. Statistics on MapNet are reported in table 5.4.

N. of LUs with at least one syn.cand.	9,120 (89.45%)
N. of LU-synset candidate pairs	33,698
N. of mapped pairs	5,162

Table 5.4: Statistics on the mapping

About one thousand lexical units in FrameNet have no candidate synsets because the lemma is not present in WordNet. The remaining LUs have 3.69 candidate synsets each on average, similarly to the average polysemy reported for the gold standard (see Table 5.1). This confirms our hypothesis that the data used for training are well representative of the characteristics of the whole resource. We expect about 80% of these mappings to be correct, i.e. in line with the precision of the classifier.

5.9.1 Automatic FrameNet extension

MapNet can be easily exploited to automatically extend FrameNet coverage, in particular to extend the set of lexical units for each frame. In fact, we can assume that all lemmas in the mapped synsets have the same meaning of the LUs in the corresponding frames. *MapNet* can be exploited to extract from WordNet the lemmas in the mapped synsets and add them to the frames.

For English FrameNet, we can acquire 4,265 new lexical units for 521 frames. In this way, we would extend FrameNet size by almost 42%. In the random evaluation of 100 newly acquired LUs belonging to 100 different frames, we assessed a precision of 78%. For the Italian side, we extract 6,429 lexical units for 561 frames. Since no Italian FrameNet has been developed yet, this would represent a first attempt to create this resource by automatically populating the frames. We evaluate the content of 15 complete frames containing 191 Italian LUs. The assigned LUs are correct in 88% of the considered cases, which represent a promising result w.r.t. the unsupervised creation of Italian FrameNet.

The difference in the evaluation for the two languages most likely lies in the smaller number of synsets on the Italian side of MultiWordNet if compared to the English, which results in less ambiguity. Furthermore, we should consider that the task for Italian is easier than for English, since in the former case we are building a resource from scratch, while in the latter we are extending an already existing resource with lexical units which are most likely peripheral with respect to those already present in the database.

Although the mapping performance achieves promising results, it can only partially contribute to the creation of Italian FrameNet, because it populates frames with lexical units but does not provide example sentences. Since one of the main properties of FrameNet is the presence of an annotated corpus, it is important to provide also example sentences that instantiate the acquired lexical units. To this purpose, we devise a methodology that exploits MapNet to annotate sentences from the MultiSemCor corpus with frame labels. We describe it in detail in the following section.

5.9.2 Frame annotation of MultiSemCor

MultiSemCor (Bentivogli and Pianta, 2005) is an English/Italian parallel corpus, aligned at word level and annotated with PoS, lemma and WordNet synsets. The parallel corpus was created starting from the SemCor corpus (Landes et al., 1998), which is a subset of the English Brown corpus containing about 700,000 running words, 200,000 of which have been lemmatized and annotated with WordNet synsets. The MultiSemCor corpus comprises 116 English texts[7] which were first manually translated into Italian. Then, the procedure of transferring word sense annotations from English to Italian was carried out automatically after aligning the parallel texts at word level with KNOWA (Pianta and Bentivogli, 2004). More recently, the Romanian SemCor has been added (Lupu et al., 2005), which comprises the manual translation of 34 English SemCor texts, aligned and annotated similarly to the Italian corpus.

The whole MultiSemCor corpus in English, Italian and Romanian can be browsed at the site `http://multisemcor.itc.it/`. It is possible both to display the parallel texts in a sentence-by-sentence way (Fig. 5.2), or to look for concordances and browse all occurrences of a word in the corpus with the corresponding translations (Fig. 5.3). The English-Italian MultiSemCor can also be obtained for free for research purposes.

[7] About one third of the SemCor corpus

Figure 5.2: An excerpt of the MultiSemCor parallel text *br-r04*

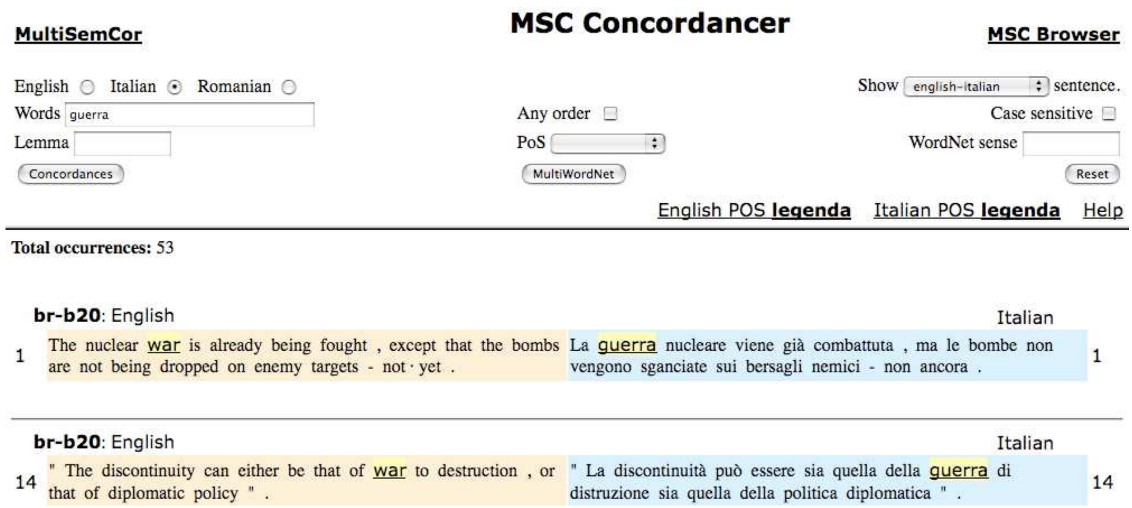

Figure 5.3: Occurrences of the word *guerra* - *war* in MultiSemCor

Since we are interested in the creation of FrameNet for Italian, we focused our research on the English-Italian part of the corpus, which is also more extensive. The basic idea is to exploit the mapping information contained in *MapNet* to straightforwardly assign a frame label to the synsets on both sides of the corpus. The automatic assignment of frame labels to the MultiSemCor sentences can be exploited from different point of views. Not only we enrich the resource with a new annotation layer, but we also automatically acquire a large set of English and Italian sentences having a lexical unit with a frame label. For the English side, it is a good solution to automatically extract a dataset with frame information and train, for example, a machine learning system for frame identification. For the Italian side, it represents a good starting point for the creation of a large annotated corpus with frame information, the base for a future Italian FrameNet.

MultiSemCor contains 12,843 parallel sentences. If we apply *MapNet* to the corpus, we produce 27,793 annotated instances in English and 23,872 in Italian, i.e. about two lexical units per sentence. The different amount of annotated sentences depends on the fact that in MultiSemCor some synset annotations have not been transferred from English to Italian. We acquired example sentences for 548 different frames in English and 533 frames in Italian. In Italian, we could acquire 3,370 new lexical units, while in English 3,149 LUs were automatically identified but only 861 are new, i.e are not present in the English FrameNet 1.6. Among these there are some LUs that introduce orthographical variations of the existing ones, for example *pajamas* in CLOTHING (*pyjama* is already present), while some others are completely new to the frame, such as *ceaseless*, *durable*, *endless*, *permanent* and *prolonged* in the DURATION frame.

Since the meaning of a LU can be really understood only in the context of a sentence, we carried out an evaluation on 200 randomly selected sentences labeled with 200 different frames, both in English and in Italian. As for the English corpus, 75% of the sentences were annotated with the correct frame label, while on the Italian side they were 70%. This result is in line with the expectations, since *MapNet* was developed with 0.79 precision. Besides, synset annotation on the English side of MultiSemCor was carried out by hand, while annotation in Italian was automatically acquired by transferring the information from the English corpus (precision 0.86). This explains why the resulting annotation for English is slightly better than for Italian.

In some cases, the wrongly annotated frame was strictly connected to the right one. For instance, the sentence reported in (5.2) from the MultiSemCor text "br-p09" was automatically annotated as an example of the MOVING_IN_PLACE frame, with *shake.v* as a LU. Instead, it should have been classified as belonging to Causing_to_move_in_place, i.e. in the causative sense of the predicate. Other classification errors include the assignment of the Apply_heat label instead of Cooking_creation and Attaching instead of Inchoative_attaching.

(5.2) She could not count the times Herman had rapped on the door , just a_couple_of bangs that *shook* (*lemma=shake, target=v#01291209*) the whole damned closet .

The same kind of mistakes are recorded also in the Italian annotation. For example, *scongelarsi.v* (*become thawed*) was classified as Cause_change_of_phase instead of Change_of_phase and *asciugarsi.v* (*dry up*) as Cause_to_be_dry instead of Becoming_dry.

In other cases, the frame definition in the Berkeley database seems to involve the *pragmatic* use of a word or a sentence and for this reason cannot be captured by our mapping model, which relies on lexically-motivated features. For example, the LUs in the Attention_getting frame include certain terms of address such as *boy.n*, *miss.n*, *sir.n* when they are used to get someone's attention. This sense distinction is not present in WordNet and for this reason all MultiSemCor sentences assigned to Attention_getting and containing such LUs are wrong.

The frame labels automatically assigned to MultiSemCor sentences are now part of the online version of the corpus and will be soon included in the corpus release. We report below a screenshot of the browser showing two parallel sentences extracted from the "br-a01" text with the annotation of *court* / *corte*. The information encoded includes the lemma, the WordNet synset, the PoS tag and the frame label.

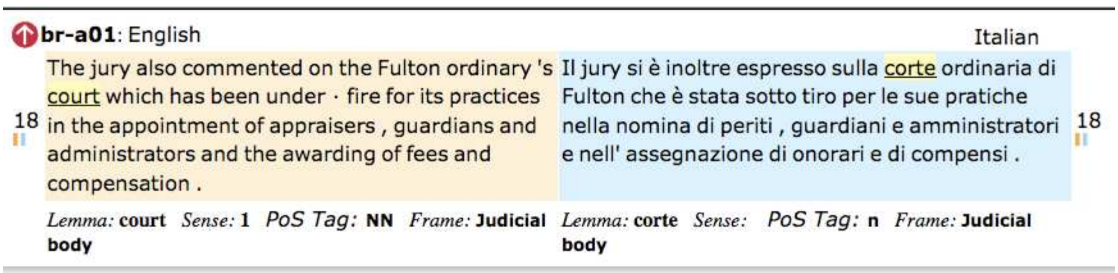

Figure 5.4: A parallel sentence from MultiSemCor with frame information

For the moment, we have not applied the mapping to the Romanian side of the corpus because we are focusing on the development of FrameNet for Italian. Anyhow, the mapping is language-independent and can be applied to every text with synset information, so we plan to deliver an automatic annotation also of the Romanian texts in MultiSemCor. Since the corpus creation and the sense annotation were carried out following the same procedure applied for Italian, we expect the frame assignment to score a similar precision.

5.10 Summary

In this chapter we have proposed a method to automatically acquire new LUs for English FrameNet and to induce also LUs for Italian FrameNet applying the same procedure, i.e. through the mapping of FrameNet LUs to WordNet synsets. To this purpose, we used SVM with minimal supervision effort.

To our best knowledge, this is the only approach to the task that exploits features based on stem overlap between LU definition and synset gloss and that makes use of information about WordNet domains. Differently from other models, the SVM is not trained on a per-frame basis and we do not rely on the number of annotated sentences for a LU in the FrameNet corpus, thus our mapping algorithm performs well also with poorly-annotated LUs.

After creating *MapNet*, the mapping between FrameNet and WordNet, we applied it to three tasks: the automatic induction of new LUs for English FrameNet, the population of frames for Italian FrameNet and the annotation of the parallel MultiSemCor corpus with frame information. The evaluation showed that the mapping can significantly reduce the manual effort for the development and the extension of FrameNet-like resources, both in the phase of corpus annotation and of frame population. Besides, adding a new annotation layer to the MultiSemCor can also be relevant from a theoretical point of view. Indeed, it can contribute to a new research direction (Baker and Fellbaum, 2009) aimed at investigating the relationship between FrameNet and WordNet with a corpus-based approach.

Our algorithm proved to cope with the different granularity of the two resources by allowing the assignment of several synsets to the same LU in a frame. On the other hand, evaluation highlighted that some frame definitions are not compatible with the sense differentiation in WordNet because they are rather pragmatically-based (for example ATTENTION_GETTING).

In the framework of a multi-language FrameNet, similar to MultiWordNet, our approach could be easily applied to acquire frame information for several new languages, at least those that are already encoded in MultiWordNet such as Spanish, Portuguese, Hebrew and Romanian. Also assignment of frame labels to sentences can be extended to the Romanian part of MultiSemCor without much effort.

In general, our proposal could give a contribution to projects aimed at the creation of multilingual FrameNets, such as the ROMANCE FrameNet initiative[8], by delivering a set of multilanguage annotations which can represent a good starting point and an initial reference for future validation.

[8] A joint initiative launched in 2005 in order to create a multilingual FrameNet resource for romance languages, namely French, Spanish, Italian, Romanian, Portuguese and Catalan, by manually translating the sentences in the original FrameNet database. See `http://www.icsi.berkeley.edu/~vincenzo/rfn/index.html`

Chapter 6

Wikipedia as frame example repository

6.1 Introduction

In this Chapter, we investigate a further research direction for the automatic annotation of frame information for Italian. In particular, we focus on the extraction of usage examples for various frames. We present an explorative approach that for the first time exploits Wikipedia to this purpose. We focused on Wikipedia because it offers several advantages: it is freely available and downloadable in the form of the so-called "dumps" i.e. copies of the available content in a given moment for offline use. Second, it covers a wide range of topics, which guarantees that the frames instantiated in the database are almost all frames of the original FrameNet database. Third, it is available in many languages, which can all benefit from the extraction methodology applied for English through a straightforward mapping. The task we have investigated can be formulated as follows:

(6.1) Given a lexical unit l belonging to a frame F, devise a strategy to link l to the Wikipedia article that best captures the sense of l in F.

This is basically a word disambiguation (WSD) problem (Erk, 2004) and to this purpose we employ a state-of-the-art WSD system (Gliozzo et al., 2005).

The mapping between (l, F) pairs and Wikipedia pages is then exploited for three further subtasks:

- Automatic extraction from Wikipedia of all sentences pointing to the Wikipage mapped with (l, F) to assign them the F label

- Automatic expansion of the the lexical units sets in the English FrameNet by exploiting the redirecting and linking strategy of Wikipedia
- Since Wikipedia is available in Italian among other languages, use the English Wikipedia article linked to (l, F) as a bridge to carry out sentence and lexical unit retrieval in Italian.

Since the aim of this work is the semi-automatic development of FrameNet for Italian, we place particular emphasis on the latter subtask. Indeed, the set of automatically collected data would give an important contribution to the creation of an annotated corpus for Italian FrameNet. In fact, having a repository of sentences extracted from Wikipedia which have already been divided by frame would significantly speed up the annotation process. The annotators would not need to extract all sentences in a corpus containing l and classify them by sense, they should simply validate the given sentences and assign the correct frame elements.

In the following, we start by providing a description of the main characteristics of Wikipedia, its structure and organization. Next, we present the task of sentence extraction giving a general motivation (Section 6.3). In 6.4, we describe the algorithm for mapping lexical units and Wikipedia articles (the so-called *Wikipages*) and the word sense disambiguation approach employed. In Section 6.5 we describe the dataset used in the experiments and report evaluation results of the mapping between (l, F) pairs and Wikipedia senses. In Section 6.6 we further describe and evaluate the quality of the extracted data to be included in the English FrameNet, while in Section 6.7 we describe and evaluate the data acquired for Italian FrameNet using Italian Wikipedia. Finally, we summarize our contribution and we draw some conclusions.

6.2 Wikipedia

Wikipedia[1] is one of the largest online repositories of encyclopedic knowledge, with millions of articles available for a large number of languages (>3,100,000 for English at the moment of writing). Such resource steadily becomes larger, because anyone is free to edit it and add content or correct the existing entries. This constant updating guarantees Wikipedia accuracy and makes it a reliable source of knowledge both for simple Internet users and for researchers.

The *article* (or *page*) is the basic entry in Wikipedia. Every article has a unique reference, i.e., one or more words that identify the page and are present in its

[1] `http://en.wikipedia.org`

URL. For example, `Ball_(dance)` identifies the page that describes several types of ball intended as formal dance, while `Dance_(musical_form)` describes the dance as musical genre. Every Wikipedia article is linked to others, and in the body of every page there are plenty of links or *anchors* that connect the most relevant terms to other pages. Such connections are manually added by Wikipedia contributors following the available *Manual of Style*[2] and should be used to increase the reader's understanding of the topic and to find related information.

Another important characteristic of Wikipedia is the presence of about 3,000,000 redirection pages. Given a keyword that is not the identifier of any Wikipedia article, redirection pages automatically display the article with the most semantically similar identifier (for example `Killing` is redirected to the `Murder` page). Wikipedia contains also more than 100,000 *disambiguation pages* listing all senses (pages) for an ambiguous entity. For example, the disambiguation page of the concept `Book` lists 9 senses, which correspond to 9 different articles.

Wikipedia structure and quality make this resource particularly suitable for a number of NLP tasks, for example coreference resolution (Ponzetto and Strube, 2006) and metonymy resolution (Vivi and Strube, 2009). In general, such tasks all require an intermediate step, the so-called *wikification*, in which the most important words and phrases are extracted from a document and are linked to a Wikipedia article. In other words, wikification allows to carry out automatically what Wikipedia contributors do by hand, i.e. to add links to Wikipages for the most important concepts in a never-before-seen document.

Some well-known approaches to wikification include the work presented by Csomai and Mihalcea (2008) and by Milne and Witten (2008). The former proposed to divide the task into two steps, namely the extraction of keywords from the document that has to be wikified and the disambiguation of such keywords in order to link them to the Wikipages that best express their meaning. For the first step, they exploit statistical-based filtering techniques to obtain the words or expressions that are most likely to be keywords of the given document. In the second step, they experiment two different WSD algorithms, one knowledge-based and one data-driven. The first algorithm relies on a measure of word overlap between the document paragraph where the expression appears and the candidate Wikipages, while the second integrates local and topical features, including contextual words and part-of-speech into a naive-Bayes classifier (Mihalcea, 2007). The machine-learning approach achieves better results, with a precision of 93% and a recall of 83%.

As for the approach by Milne and Witten (2008), it is to some extent similar

[2] `http://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style`

to Csomai and Mihalcea (2008) in that it decomposes the wikification task into the *detection* step, i.e. the identification of relevant terms within unstructured text, and the *disambiguation* process, which links the detected phrases to the appropriate Wikipedia article. However, the order of such steps has been inverted, because first all words in a document are linked to the appropriate Wikipedia pages and then the most relevant terms are selected. In this way, the term selection process can rely on link probability measures as in Csomai and Mihalcea (2008), but also on features obtained from the Wikipedia articles that have been previously connected to the words in the document. As for the disambiguation process, Milne and Witten employ a machine-learning approach based on two main features, i.e. the commonness of each sense in Wikipedia and its relatedness to the surrounding context. The best-model performance achieves 98.4% precision and 95.7% recall on the disambiguation task and 77.3% recall and 72.9% precision on the link detection task.

6.3 Motivation of the sentence extraction task

One of the peculiar features of FrameNet-like databases is that they rely on corpus annotation and that LUs and FEs are usually instantiated by example sentences. For this reason, it is very important that a large corpus is available for the given language and that it deals with different topics in order to cover a wide range of frames. While the English FrameNet project could rely on the British National Corpus (BNC), such a rich resource is missing for most languages. As shown in Section 4.5, the EUROPARL corpus (Koehn, 2005), which is often employed in NLP experiments having about 30 million tokens in 11 languages, would not be suitable to covering all frame variability: despite its dimensions, it focuses on sociopolitical issues and has less topic variability than the BNC. Besides, having a large multilingual corpus for the development of new FrameNets would be very important because it would be possible to integrate and re-use information available for the English FrameNet in order to extract "good" example sentences for the new language (see Section 6.7).

We devised a methodology that exploits Wikipedia for the automatic selection and frame labeling of example sentences in all languages available for Wikipedia, with a focus on Italian. This is the first time that Wikipedia is used for the task and also that Wikipedia texts are seen as a corpus for frame annotation. Our approach is based on one hand on the internal linking strategy of Wikipedia and on the other hand on its multilingual nature, because the pages in different languages corresponding to the same sense are (or should be) linked together.

We exemplify our intuition with the help of Figg. 6.1 – 6.3. Let's say that we want to enrich the existing Word_relations frame in English with new example sentences and that we want also to extract example sentences in Italian for the same frame, without having any LUs or Italian examples available. First, we start from the English frame and find the Wikipedia pages that best express the meaning of the English LUs in Word_relations. For example, we identify the `Collocation`, `Homonym` and `Opposite(semantics)` Wikipages by pairing them with *collocate.n*, *homonym.n* and *antonym.n* in Word_relations (Fig. 6.1).

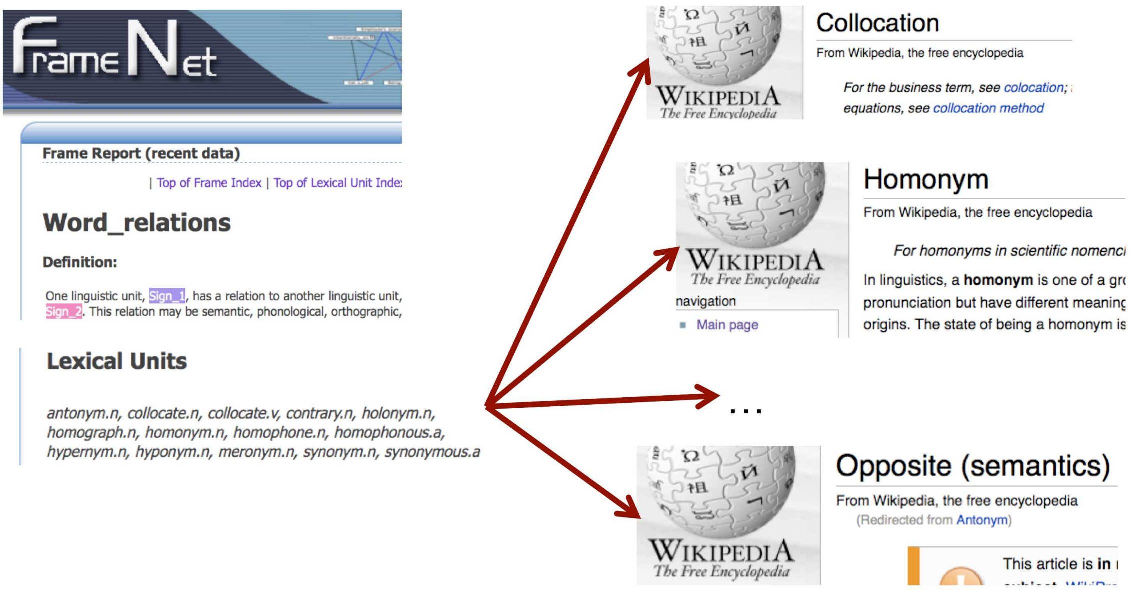

Figure 6.1: Linking the LUs in Word_relations to the corresponding Wikipages

Then, we extract from the English Wikipedia dump all sentences with an anchor pointing to the mapped Wikipages. In Fig. 6.2 we show three sentences with a word (in bold) anchored to the `Homonym` page, while the total number of extracted sentences pointing to `Homonym` is 186. Such sentences can be seen as corpus attestations of the Word_relations frame and the anchored words as lexical units.

The lemmas linked to the `Homonym` page are *homonym.n*, *homonymy.n*, *homophone.n*, *homograph.n*, *homophonous.n*, *homonymic.n*, *heteronym.n* and *same.a*. Among them, only the last is not appropriate for the Word_relation frame, even if the sentence where it occurs is semantically related to it, as shown in Example 6.2:

(6.2) In Hebrew the word 'thus' has the **same** [→ link to `Homonym`] triconsonantal root.

The lemmas *homonymic.a* and *heteronym.a* can be acquired as new lexical units

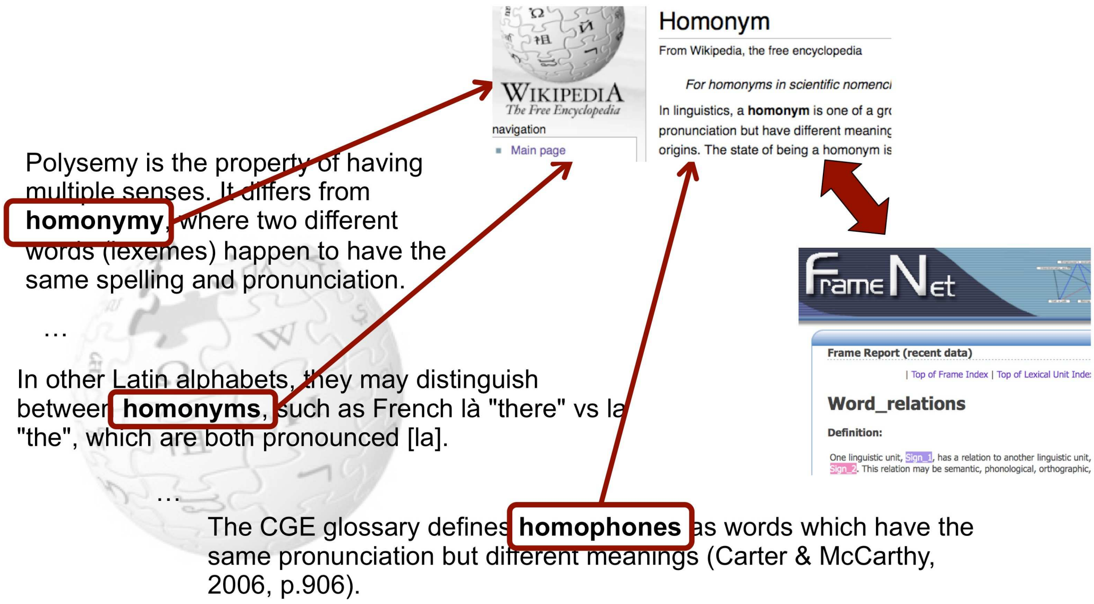

Figure 6.2: Wikipedia sentences anchored to the *Homonym* page

for WORD_RELATIONS, and *homograph.n*, for which no example sentence is provided in FrameNet, can be automatically instantiated by a set of examples.

In this way, we accomplish the first subtask, i.e. the automatic extraction of examples sentences to enrich the WORD_RELATIONS frame, and also the identification of new LUs. As for the automatic extraction of sentences belonging to the WORD_RELATIONS frame in new languages, we take all English Wikipages mapped to the given frame and extract their version in other languages. For example, `Homonym` is linked to the `Homonym` page in the German Wikipedia, `Homonimia` in Spanish and `Omonimia` in Italian, among others (Fig. 6.3).

Then, we repeat the step shown in Fig. 6.2, extracting all sentences in German, Spanish or Italian that point to the given Wikipage. With the same procedure, we obtain a large set of classified sentences assigned to a frame and also the corresponding list of LUs. For Italian, the extracted sentences would be like those reported in Examples 6.3 and 6.4. The words in bold contain an anchor to the Wikipage `Omonimia` and would be included in the LU list for WORD_RELATIONS. We report at the end of the sentence the Wikipage containing the example, while the anchored page is included between square brackets right after the word pointing to it.

(6.3) In poesia, la rima è l'**omofonia** [→ `Omonimia`], ovvero l'identità dei suoni,

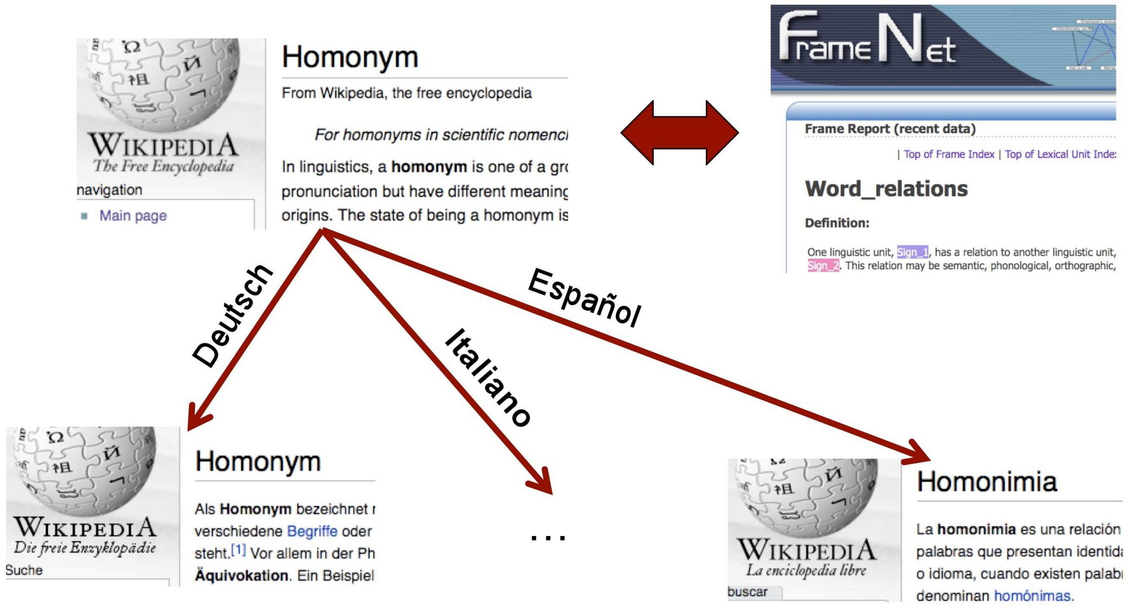

Figure 6.3: Mapping extension to new languages

tra due o più parole a partire dall'ultima vocale accentata (`Rima_(linguistica)`)
In poetry, rhyme is homophony, i.e. the sound identity between two or more words starting from the last accented vowel.

(6.4) Il calembour un termine preso in prestito dalla lingua francese che indica un particolare gioco di parole, basato sull'**omofonia** [→ `Omonimia`] di parole che si scrivono in maniera identica o simile ma hanno significato diverso (`Calembour`).
Calembour is a term borrowed from French which indicates a particular word game based on the homophony of words that are written in the same or similar way but have different meanings.

In this way, we accomplish also the second subtask, i.e. the multilingual extraction of example sentences as frame attestations. The whole procedure relies on a mapping algorithm, which in turn is based on a well-known word sense disambiguation algorithm. Further details are given in the following Section.

6.4 The Mapping Algorithm

In this section, we describe how to map a lexical unit-frame pair (l, F) into the Wikipedia article that best captures the sense of l as defined in F. The mapping problem is casted as a supervised WSD problem, in which information about l extracted from the FrameNet database are used to provide the test data and Wikipedia is exploited to provide the sense inventory and the training data. Even if the idea of using Wikipedia links for disambiguation is not novel (Cucerzan, 2007), it is applied for the first time to FrameNet lexical units, considering a frame as a sense definition. The proposed mapping relies on the following steps:

Step 1: Creation of the training set For each lexical unit l, we collect from the English Wikipedia dump[3] all contexts[4] where l is the anchor of an internal link. The set of linked pages represents the senses of l in Wikipedia and the contexts are used as labelled training examples. For example, given that we want to map the lexical unit *building.n* in the frame BUILDINGS to the Wikipedia page that best expresses its meaning, we first collect all paragraphs in the English Wikipedia dump that contain the word *"building"* with an embedded anchor (link) to some Wikipedia page. In this way, we collect 708 different paragraphs (or contexts) that point to 42 different Wikipedia pages (senses), such as `Civil_engineering` and `Building` (see Fig. 6.4). From a WSD point of view, we assume that there are 42 senses for the word *"building"* and that the paragraphs extracted can be divided into 42 groups, one for each sense, that can serve as training set for WSD system.

Step 2: Training of the WSD system The set of contexts with their corresponding senses is used to train the WSD system by Giuliano et al. (2009), in order to obtain a WSD system where each *sense* is expressed by a Wikipedia page. For example, the context "The 2008 budget was released on May 13, 2008, with a particular emphasis on family welfare and *building* funds." is a training example for the sense defined by the Wikipedia page `Civil_engineering` because *"building"* has an embedded anchor to the `Civil_engineering` page (see Fig.6.4).

The WSD system employs a kernel-based semi-supervised algorithm that exploits knowledge from Wikipedia acquired in an unsupervised way. In this approach, the linguistic phenomena are first represented independently to capture different *domain* and *syntagmatic* aspects of sense distinction, and then are combined together in a

[3] `http://download.wikimedia.org/enwiki/20090306`

[4] A context corresponds to a line of text in the Wikipedia dump and it is represented as a paragraph in a Wikipedia article.

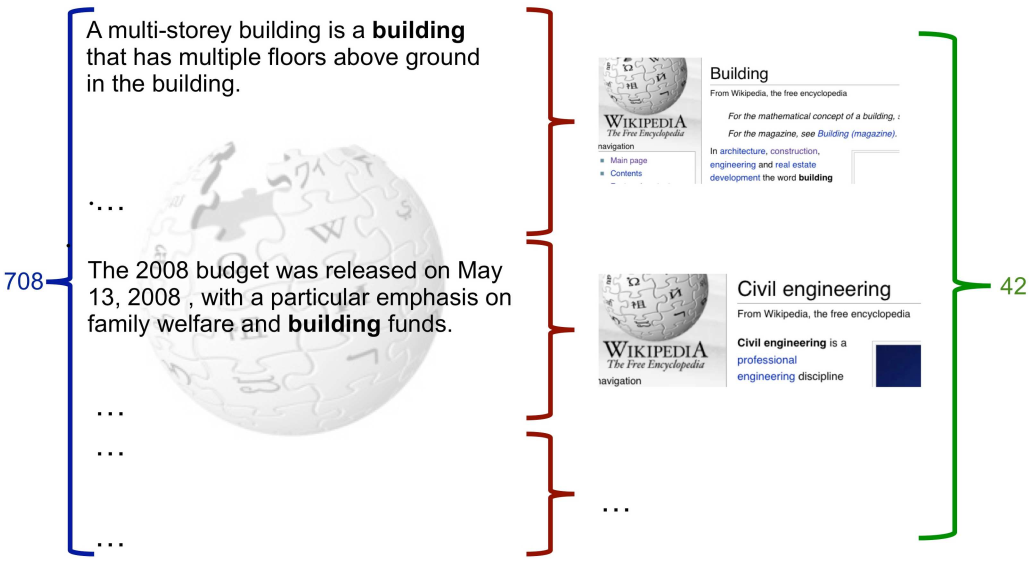

Figure 6.4: Context extraction and clustering

so-called *composite kernel*. In our case, the domain-oriented kernels considered for the training are basically two: one is based on *domain models* (DM), i.e. clusters of terms representing the words that tend to co-occur in texts with the same topic. This kernel type was estimated using the 200,000 most visited Wikipedia pages. The second domain kernel is the more standard *bag-of-words* kernel, relying on bag-of-words features extracted from a wide window of text around the word to be disambiguated. To represent also syntagmatic aspects, we integrated in the system a *collocation kernel*, which uses the local context of the word to be disambiguated. In this formulation, non-contiguous collocations are also taken into account. Details about the kernel functions and their combination are given in Giuliano et al. (2009).

At the end of the training phase, a disambiguation model is created that, given a word in a new context, can assign it to the most appropriate sense represented by a Wikipedia page[5].

Step 3: LU disambiguation This step represents the original part of our contribution, i.e. the idea to use an existing WSD system trained to recognize the Wikipedia pages as word senses and to apply it to FrameNet lexical units, in order to obtain a

[5]We thank Claudio Giuliano for sharing his WSD system and giving insights into Wikipedia and Kateryna Tymoshenko for the technical support and the help with the system

mapping between (l, F) pairs and Wikipedia articles.

The disambiguation model learned in the previous step is used to accomplish such mapping. The couple (l, F) is modeled by a representative text which is selected in two ways: in a first experiment using a *definition-based approach*, we build a *pseudo-text* by merging the frame definition and the lexical units in such frame. In a second experiment using an *example-based approach*, the (l, F) pair is modeled by the example sentences available in FrameNet.

For example, according to the first model, if we want to classify the lexical unit *building.n* in the BUILDINGS frame, we take the frame definition "This frame contains words which name permanent fixed structures forming an enclosure and providing protection from the elements" and the set of LUs contained in it comprising *acropolis, arena, auditorium, bar, barn, barracks, basilica, blockhouse, bungalow, bunker, cabin, campanile, caravan, caravanserai, castle, chalet*, etc. The pseudo-text modeling the couple (*building.n*, BUILDINGS) is created by using the frame definition as left context and the LU set as right context of *"building"*.

As for the second model, we collect all example sentences associated to the frame F in the FrameNet database containing l and we use each of them as a representative text. Every sentence modeling $l \in F$ is assigned to a Wikipedia page that expresses the sense of l in the given example. If several sentences are available for a single l, the Wikipedia page that is most frequently assigned to the examples is chosen as the best sense for (l, F). For example, if we want to classify the *church.n* lexical unit in the BUILDINGS frame, we take into account all 47 example sentences in FrameNet with the *church.n* LU belonging to BUILDINGS and use each of them as a representative sentence for the WSD process. The system delivers for every sentence a Wikipedia page that should correspond to the meaning of *church.n* in such sentence. In this particular case, 45 sentences are associated to `Church_(building)` while 2 are assigned to `Christian_Church`[6]. The first option is selected because it is the most frequent one. In this way, we can assess a relationship between the BUILDINGS frame and the `Church_(building)` page. In the first model, instead, only one test sentence per (l, F) pair is available, so that the mapped Wikipedia page is the only option.

The different performances of the two types of representative texts are discussed in the evaluation section.

[6]In order to display the page assigned, it is enough to open in a browser the page `http://en.wikipedia.org/wiki/+AssignedPage`, for example `http://en.wikipedia.org/wiki/Christian_Church`

6.5 The mapping experiment

6.5.1 Experimental Setup

In order to compare the two types of data available, i.e. the *definition-based* and the *example-based* texts, we take into account all lexical units having at least one example attestation in the FrameNet database. Besides, since Wikipedia is basically a resource organized by concepts that are generally expressed by nouns, we restrict our experiment to nominal lexical units. Furthermore, many verbal and adjectival concepts in Wikipedia are redirected to nominal identifiers.

After discarding all lexical units reported in the FrameNet database without example sentences and extracting only the nominal ones, we obtain 2,698 nominal lexical units, including 61 multiwords.

For every LU, the WSD system is run twice, one with the definition-based text and one with the example-based type. For example, the pair (*living.n*, DEAD_OR_ALIVE) is disambiguated using two different test data: in the first experiment, the pseudo-text reported in (6.5) is used, while in the second experiment we employ the three example sentences in (6.6) taken from the FrameNet database.

(6.5) A Protagonist is in the dynamic, maintained state of being alive or has exited this state. *living* alive , dead , deceased , lifeless , nonliving , undead .

(6.6) (1) The patron of the *living* asked the architect Samuel Saunders Teulon to make a report on it .

(2) The shapes , the voices that throng his mind , for there are days when the *living* have no substance and the dead are active .

(3) The Archdeacon did not , diplomatically , point out that the Bishop had not consulted Lord Dersingham when the *living* had been in his gift .

The domain model built for WSD (i.e. sets of term clusters, see Step 2 in previous section) is based on the 200,000 most visited Wikipedia articles. After removing terms that occur less than 5 times, the resulting dictionaries contain about 300,000 terms. The experiments are performed using the SVM package *LIBSVM* (Chang and Lin, 2001) customized to embed the kernels described in Section 6.4.

6.5.2 WSD statistics and analysis

We report in Table 6.1 some statistics about the coverage of the WSD system using the two test data. We compared the system output to a baseline computed considering the *most frequent sense* of every lexical unit in Wikipedia. The most frequent sense for a lexical unit l is represented by the page to which an occurrence of l is most frequently linked in Wikipedia. For example, for the pair (*bonnet.n*, ACCOUTREMENTS), the WSD system delivers in both experiments the Wikipedia page `Bonnet_(headgear)` as the best mapping for the given sense of "bonnet", which is correct. The baseline, instead, is the page `Hood_(vehicle)`, because most occurrences of "bonnet" in Wikipedia are internally linked to that page.

	Definition-based	*Example-based*
N. of mapped LUs	2,487	2,505
N. of missing mappings	211	193
Different mappings w.r.t. the baseline	154	163
Different mappings between the two models	256	

Table 6.1: Output of Wikipedia mapping

Since our final goal is the automatic annotation of sentences extracted from Wikipedia with frame labels, we consider this mapping procedure as an intermediate step of the task, so we do not evaluate its accuracy independently. Instead, we prefer to focus on the evaluation of the final resource. However, it is interesting to look at the WSD task and discuss some issues about the compatibility of FrameNet and Wikipedia.

In a previous study (Tonelli and Giuliano, 2009), we reported that 14% of the lexical units contained in a sample of 250 (l, F) pairs were not present in Wikipedia as a concept. On one hand, this confirms our intuition that FrameNet and Wikipedia are linkable resources to a large extent, and on the other hand it can partly explain why almost 8% of the (l, F) pairs considered in this experiment could not be mapped to any Wikipedia page, regardless of the type of test data (the system delivered the answer "concept_not_found"). For example, the system could not deliver any mapping for (*bafflement.n*, EMOTION_DIRECTED), (*tress.n*, HAIR_CONFIGURATION) and (*clink.n*, SOUNDS), which are indeed missing as senses in Wikipedia. As expected, the number of missing mappings is slightly lower for the *example-based* test set than for the *definition-based* version, showing that the use of several example sentences as test data for a given LU can increase the system recall.

In 256 cases, the system delivered a different mapping for the two kinds of test data. In 63 cases, the "concept_not_found" label was assigned to one of the two outputs. In particular, the *definition-based* test data produced 45 missing mappings, whereas the *example-based* setting generated only 18 mapping failures: as expected, a mapping is more likely to be found using a set of examples as test data instead of one pseudo-text.

If we look at the other diverging mappings, we notice that the pages mapped with the example-based methodology are generally more specific than those obtained with the definition-based test set. This is due to the fact that examples present a LU in a specific context of usage while a frame definition is more generic. We report some examples in Table 6.2:

	(l, F) pairs	*Definition-based*	*Example-based*
1	*plain.n*, BIOLOGICAL_AREA	`Plain`	`Great_Plains`
2	*lace.n*, CLOTHING_PARTS	`Lace`	`Shoelaces`
3	*queen.n*, LEADERSHIP	`Queen_regnant`	`Elizabeth_II_of_the_Unit._King.`
4	*inquiry.n*, QUESTIONING	`Inquiry`	`Public_inquiry`
5	*pro.n*, EXPERTISE	`Promagistrate`	`National_Football_League`

Table 6.2: Example of diverging mappings

Notice that, since we are not evaluating the mapping accuracy per se, we cannot assume that the definition-based model performs better. For the moment, we just point out that the definition-based approach tends to achieve mappings between (l, F) and Wikipages such that the meaning of l and the concept illustrated by the page are semantically equivalent. Instead, the example-based methodology usually delivers a Wikipedia page that is more specific than l.

For example, as reported in Table 6.2, line 1, the definition-based mapping of (*plain.n*, BIOLOGICAL_AREA) is the Wikipedia page `Plain`, which describes exactly a type of land, while the example-based output is `Great_Plains`, the article about a prairie lying west of the Mississippi River and east of the Rocky Mountains. Even if the (l, F) sense and the Wikipedia page obtained with the first approach are equivalent, this does not imply that the sentences extracted with this method are more appropriate as example sentences than those extracted with the example-based mapping. Further details will be given in Section 6.6.3.

In some cases, wrong assignments are not always directly connected to errors of the WSD algorithm. Instead, it seems that frame definitions are sometimes inconsistent and it is very difficult to discriminate between two frames even for a human annotator. For example, the *squabble.n* LU in the HOSTILE_ENCOUNTER frame and the *dispute.n* LU in the QUARRELING frame are both mapped to the `Controversy` Wikipage with every system configuration. This seems to contradict our initial assumption, i.e. that every LU in a frame represents a word sense that has no equivalence in other frames. However, the definition of HOSTILE_ENCOUNTER and QUARRELING are conceptually similar, and the LU sets for each frame are in some cases overlapping. In particular, the *squabble.n* LU should rather belong to QUARRELING than to HOSTILE_ENCOUNTER, since it denotes a particular type of quarrel[7]. This proves that part of the semantics of QUARRELING is shared by HOSTILE_ENCOUNTER, and that the assignment of the `Controversy` Wikipage to both frames is to some extent correct.

6.6 English FrameNet expansion

6.6.1 English data extraction

The core goal of our methodology is to investigate to what extent the FrameNet - Wikipedia mapping can be effectively applied to automatically expand the FrameNet database with new example sentences, and eventually to acquire new lexical units. This should be applied both to existing FrameNets, in order to enrich them, and to new ones, in order to collect an initial set of pre-annotated data. In this section we focus on the expansion of the English FrameNet, while in the next one we investigate the algorithm applicability to Italian.

For every (l, F) pair, we consider the linked Wikipedia sense s and extract all sentences C_s in Wikipedia which are linked to s. In this way, we can assume that, if (l, F) was mapped to s, then C_s can be included in the example sentences of F. This repository of sentences is already grouped by sense and can significantly speed-up manual annotation. On the other hand, the extracted sentences could enrich the training set of machine learning systems for frame annotation to improve the frame identification step. In fact, this task has raised growing interest in the NLP community, with a devoted challenge at the last SemEval campaign (Baker et al., 2007).

This retrieval process allows also to extract from C_s all words W_s that have an embedded reference to s in the form `<ahref="/wiki/Wiki_Sense"...>word</a>`. In this way, W_s are automatically included in F as new lexical units. In this phase, also redirecting links are very useful because they automatically connect a word or expression to its nearest sense in case there is no specific page for this word. The information about redirecting allows also to account for orthographic variations of the same lexical unit, for example *collectible* is redirected to *collectable*.

[7] Indeed, *squabble.n* has been recently added to QUARRELING in the online version of the FrameNet database, even if no annotated sentences are present yet. See `http://framenet.icsi.berkeley.edu/index.php?option=com_wrapper&Itemid=118&le=8941&source=frame&sourcevar=Quarreling&`

6.6.2 Output statistics

In Table 6.3 some statistics about the sentences extracted from the English Wikipedia are reported. Again, we compare the output obtained with the two different test data.

	Definition-based	*Example-based*
N. of mapped (l, F) pairs	2,487	2,505
N. of extracted sents	304,011	303,297
Pairs with at least 1 sentence	2,110	2,164
Avg. sents per (l, F)	144	140
LU candidates	18,305	19,915

Table 6.3: Extracted data from English Wikipedia

The extraction algorithm is followed by a post-processing step which filters the sentences in order to eliminate data that are not strictly related to the content of the articles. For example, the extraction algorithm retrieves also internal Wikipedia pages, or links about users, but since they are generally introduced by standard patterns such as "*Template:*", "*Portal:*" or "*Wikipedia:*", they can be automatically discarded with a simple rule-based filtering. Another filtering rule copes with a main problem that involves the linking of nationalities and nations in Wikipedia. The mapping algorithm usually connects a LU belonging to the PEOPLE_BY_ORIGIN frame, which collects the names describing humans w.r.t. their nationality, to the Wikipedia article about the corresponding nation. For example, the *Brit.n* lexical unit is linked to the `United_Kingdom` page. Even if this can be considered generally correct, it represents a relevant problem for our task, because nationalities and nations are not considered as belonging to the same frame in the FrameNet database. This means that, if a sentence like (6.7) is extracted for *Brit.n* in PEOPLE_BY_ORIGIN, it would be a wrong assignment.

(6.7) In the *United Kingdom* [→ `United_Kingdom`] and the Republic of Ireland, the majority of state secondary schools adopt a uniform for a more formal look.

Since many of the extracted sentences come from Wikipedia articles about nations, probably because such pages are very much linked and anchored by other pages, we decided to introduce a post-processing rule that filters all the cases where a PEOPLE_BY_ORIGIN frame is linked to the page of a country. This is carried out automatically in a very straightforward way, and it represents the only step in the whole algorithm which is language-dependent. Anyhow, this single rule eliminates about one third of the extracted sentences and is supposed to increase precision very much. The extracted sentences before post-processing amount to 527,156 for the definition-based approach and 641,068 for the example-based one, and after filtering they are reduced resp. to 304,011 and 303,297.

The number of mapped (l, F) pairs, for which the algorithm finds a corresponding Wikipedia article, is around 2,500 with both approaches, but the number of mappings for which at least an example sentence is extracted is lower, i.e. 2,110 with the definition-based and 2,164 with the example-based approach. This means that in some cases, the Wikipedia article that was mapped to (l, F) was not linked by any sentence in Wikipedia or had just "bad" links which were filtered in the post-processing step. It seems also that the filtering has higher impact on the example-based dataset. This can be explained by the fact that, as we mentioned in the previous section, the example-based mapping tends to link Wikipedia articles that are more specific and that are likely to be less edited and corrected than the general ones. Anyhow, even after the filtering, every (l, F) pair is represented by a good amount of example sentences (144 definition-based and 140 example-based), given that every LU in the FrameNet database is instantiated by 14 examples on average.

In every extracted sentence, the system highlights the word or expression with an embedded anchor to the mapped Wikipedia article, which can be seen as a good candidate for being a LU. For example, the lexical unit *junkie.n* in the ADDICTION frame was mapped to the Wikipage `Drug_addiction`. The set of sentences extracted which were pointing to `Drug_addiction` comprise some instances reported in Example (6.8). The words in italics have an embedded anchor to the `Drug_addiction` page and are candidate LUs for the ADDICTION frame.

(6.8) (1) LAAM is indicated as a second-line treatment for the treatment and management of opioid *dependence* [→ `Drug_addiction`] if patients fail to respond to drugs like methadone or buprenorphine.

(2) In many industrialized countries, nicotine is among the most significant *addictive* [→ `Drug_addiction`] substances and a cause for medical concern.

(3) His major international breakthrough soon followed with the role of heroin *addict* [→ `Drug_addiction`] Mark Renton in Boyle's film version of Irvine Welsh's Trainspotting (1996).

It is interesting to note that, while *addict.n* is already present in the ADDICTION frame, *addictive.a* and *dependence.n* would be newly introduced in a completely automatic way. In particular the latter is present in FrameNet as LU of the CONTINGENCY and of the RELIANCE frame, but the meaning of "*dependence*" as "*addiction*" is missing. With our extraction algorithm we would both provide a new LU and a set of example sentences that instantiate its meaning in context.

6.6.3 Evaluation of the English example sentences

The dimension of the extracted corpus makes it impossible to carry out a comprehensive evaluation. For this reason, we manually evaluated 1,000 sentences, i.e. we considered 50 (l, F) pairs, and for each of them we evaluated 20 sentences extracted from our large repository. Both (l, F) pairs and the assigned sentences were randomly selected, regardless of the mapping quality of the assigned Wikipage. The same evaluation procedure was applied both to the definition-based and to the example-based dataset. In particular, we considered the same (l, F) pairs in both datasets. Our evaluation was aimed at checking if an extracted sentence was appropriately assigned to a (l, F) pair and if the word or expression highlighted by the system as a LU candidate can be included in F. In other words, we aim at estimating the percentage of LU candidates which were assigned to the correct frame given the context where they appear.

We observed that 71% of the sentences extracted with the definition-based methodology were correctly linked to (l, F) pairs, while with the example-based approach the value increased up to 74%[8]. This proves that, even if the Wikipage assigned to (l, F) is not the article that best corresponds to the meaning of l in F, some sentences pointing to it may be appropriate to express l. In particular, the pages mapped using the frame definition tend to be more general and they may have been linked by sentences dealing with a variety of topics, which can limit the accuracy of the extracted sentences. Instead, the pages mapped with the example-based approach are generally more specific, which reduces the topic variability of the sentences containing outgoing links.

As for the error analysis, we can identify different aspects reducing accuracy:

Errors in the WSD algorithm: Such errors occur when the system maps a (l, F) pair with a wrong Wikipedia page, which impacts on the quality of the extracted sentences. For example *improbability.n* in the LIKELIHOOD frame is

[8] In a previous study (Tonelli and Giuliano, 2009), we reported an accuracy of 78% on a similar task. The different measure may depend on the different structure of the evaluated dataset, which comprises 20 (l, F) pairs with 50 sentences. A further difference involves the WSD system configuration, because in the previous experiments we filtered out mapped articles corresponding to disambiguation pages, while here we retain all delivered articles.

mapped to the `Probability_theory` article, and all sentences extracted pointing to it deal with mathematical theories and problems.

Linking strategy of the Wikipedia editors: This error class involves cases in which the sentence extraction algorithm cannot cope with the linking strategy of Wikipedia. For example, there are sentences about the "Gossip" music band that are linked to the `Gossip` page (in the sense of rumors), and sentences about the film "Assassination" that are anchored to `Murder` article. This means that, even if the *gossip.n* LU in the CHATTING frame is correctly mapped to the `Gossip` Wikipage, the extracted sentences will considered incorrect because "Gossip" denotes the proper name of a band.

Linking errors by the Wikipedia editors: While the previous point depends on a linking strategy which cannot be seen as wrong, there are also some cases of proper linking mistakes done by Wikipedia editors. For example, "ignorant" was linked to the `Gossip` page, and "infancy" to the `Infanticide` article, even if the `Ignorance` and the `Infancy` pages are available[9].

Different Wikipedia-FrameNet granularity: As mentioned above, if a (l, F) pair is mapped to a Wikipedia page which is more specific than the sense of the LU in the frame, this does not represent a problem because the sentences extracted for (l, F) may be appropriate anyway. However, the contrary is more likely to be a problematic case, that is the Wikipedia page should not be more general than the frame definition.

The major problem is that in the FrameNet database, frames describing an activity or a state are usually different from frames describing the people who perform such activities. This kind of distinction is not explicit in Wikipedia entries. For example, MEDICAL_SPECIALTIES and MEDICAL_PROFESSIONS are two distinct frames. If the `Gynaecology` page is correctly mapped to *gynaecology.n* in MEDICAL_SPECIALTIES, and a sentence is extracted where the word "*gynaecologist*" is linked to the `Gynaecology`, this will result in a wrong assignment because it should be be rather assigned to the MEDICAL_PROFESSIONS frame. The same problems arise with frames such as CAUSE_CHANGE and UNDERGO_CHANGE, because in Wikipedia no distinction is made for the causative use of a concept (for example, the "*conversion*" concept is described in one single Wikipedia article, while FrameNet distinguishes between "*conversion*" as the act of converting one's beliefs and the persuading of someone else).

[9]A funny editor linked the expression "married couples" to the `Penitentiary` page. This looks more like a humorous linking rather than an error.

Since the example-based dataset proved to be more accurate, we focused on it in the further phases of our evaluation, which concerned the analysis of the candidate LUs proposed by the system.

The 1,000 sentences evaluated contain 218 LU candidates. Among them, 47 would have been *new* LUs correctly assigned to existing frames, for example "*proprietor.n*" to POSSESSION and "*tollway.n*" to ROADWAYS. Exploiting redirections and anchoring strategies, our induction method can account for orthographical variations, for example it acquires both "*behavior.n*" and "*behaviour.n*". On the other hand, also misspelled words may be collected, for instance "*gynaecologial*" instead of "*gynaecological*". Some of the candidate LUs extracted are verbal targets, for example "*collaborate.v*" for COLLABORATION, "*behave.v*" for CONDUCT or "*withdraw.v*" for QUITTING_A_PLACE. This corroborates our intuition that, even if our Wikipedia - FrameNet mapping is noun-based, it can account, at least partially, also for verbal lexical units, providing correct examples for verbal targets.

As for the example-based methodology, we also wanted to see if the number of example sentences used as test data for the WSD algorithm can influence the mapping performance. Our hypothesis was that a high number of test sentences per (l, F) pair would lead to more precise mappings. So, we extracted for every (l, F) in the evaluation set the number of examples used for the mapping in order to see if there was some correlation between precision and dimensions of the test set. The data proved that our hypothesis was wrong: there is no correlation at all between the two factors. We run the WSD algorithm starting from test sets of different dimensions, from 2 to 69 example sentences per (l, F) pair, depending on the sentences available in the FrameNet database. Surprisingly, the performance achieved for many (l, F) pairs with a lot of example sentences was bad. For example, *function.n* in the CONTINGENCY frame is instantiated by 47 examples, which were used as test set, leading to 19 "bad" extracted sentences out of 20. On the contrary, *mass.n* in the RITE frame had just 2 test sentences, but 19 of the 20 sentences evaluated were correct. This shows that several other factors influence the mapping and sentence extraction quality rather than the number of test sentences available. In the case of *mass.n*, for instance, the sense is so specific that few examples are enough to portray the domain in which such meaning is valid. The same happens with *expressway.n* in ROADWAYS: the term is not ambiguous and, even if only 5 example sentences are available as test set, all extracted sentences that we have evaluated are correct. On the contrary, *sheet* in the SHAPES frame is used to describe "a flat, frequently rectangular portion of a substance" and is instantiated by 30 example sentences. However, this sense of "*sheet*" is completely missing in Wikipedia, where none of the listed senses includes the idea of a “rectangular portion” of something. For this reason, despite the high amount of test data, all extracted sentences are wrong.

Another hypothesis we wanted to verify was the idea that the quality of the extracted sentences may depend also on the reliability of the Wikipedia article where such sentences occur. Our intuition was that an article presenting many incoming links must have been edited and checked a lot of times, so the sentences contained in it are likely to be very accurate. For example, if we map the pair (*apple.n*, FRUIT) to the Wikipage `Apple` describing the fruit and then we extract all sentences pointing to the `Apple` article, we notice that they come from different source pages. We report two examples below, and we specify the source page at the end of the line between round parenthesis:

(6.9) (1) *Apples* [→ `Apple`] from New England include the original varieties, Baldwin, Lady, Mother, Pomme Grise, Porter, Roxbury Russet, Wright, Sops of Wine, Peck’s Pleasant, Titus Pippin, Westfield-Seek-No-Further, and Duchess of Oldenburg. `(Cuisine_of_the_United_States)`

(2) The building has a *Apple* [→ `Apple`] (Wireless) and two PC (IBM compatible) labs. `(Louisa-Muscatine_School_District)`

If we compute the number of incoming links to the page about `Cuisine_of_the_United_States`, which expresses the number of sentences in Wikipedia pointing to it, we can see that it is quite high, i.e. 59, while the incoming links of `Louisa-Muscatine_School_District` are 2. Following our hypothesis, we rank the extracted sentences according to the number of such links and check if the top-ranked sentences can be considered more reliable. In Example 6.9 the intuition would be confirmed, because the first sentence is a good example of the (*apple.n*, FRUIT) pair, while the second is not because the word *Apple* has been wrongly anchored to the `Apple` page instead of to `Apple_Inc`.

In other words, we want to test if a Wikipage with a lot of incoming links can be considered more reliable than a less linked one, and to what extent this can influence the sentence extraction task. To this purpose, for every (l, F) pair in the evaluation dataset, we ranked the 20 example sentences according to the number of incoming links of the Wikipage they were extracted from. The top-ranked sentences were those coming from the Wikipedia articles having the highest number of links, and that, according to our hypothesis, were more likely to be correct. Then, we computed the *mean average precision* (MAP), a measure widely used by researchers to evaluate information retrieval systems and algorithms (Wu and McClean, 2007). This measure captures not only single-value metrics such as precision and recall, but also the order in which the ranked sentences are listed, emphasizing the relevance of top-ranked sentences. Given that R is the total number of relevant documents in the whole collection of extracted sentences and p_i is the ranking position of the i-th relevant documents in the resulting list, the MAP is computed as follows:

$$MAP = \frac{1}{R}\sum_{i=1}^{R}\frac{i}{p_i}$$

The measure corresponds to the sum of the precision at each relevant position in the sentence list divided by the total number of relevant sentences in the sample of 20 examples considered for every (l, F) pair.

MAP is calculated for each of the 50 (l, F) pairs in the evaluation dataset and then an average for all pairs is computed. The obtained measure is 0.781. In order to compare it with a baseline, we computed MAP on the same dataset taking the 20 extracted sentences for every (l, F) pair in a random order. The sentences were randomly re-ranked 20 times, and every time the MAP was computed. Then, an average was calculated. The baseline obtained was 0.770. The comparison between the two values shows that the sentence re-ranking is effective in few cases and that mistakes in the linking strategy have a low impact on the sentence extraction task. This implies also that links in Wikipedia are generally correct and that the structure of Wikipedia and the linking methodology can be reliably exploited in WSD systems or similar. Besides, we are aware that a better evaluation based on MAP should have taken into account all sentences extracted for every (l, F) pair and not just 20 randomly chosen examples in order to have a complete representation of the ranking.

6.7 Multilingual FrameNet expansion

One of the great advantages of Wikipedia is its availability in several languages. The English version is by far the most extended, but a considerable repository of pages is available also for other languages, especially European ones. In general, articles about the same topic in different languages are edited independently and do not have to be translations of one another, but are linked to each other by their authors. In this way, the multilingual versions of Wikipedia can be easily exploited to build comparable corpora, with connected Wikipages in different languages dealing with the same contents.

In this research step, we focus on this aspect of Wikipedia and propose a methodology that, using the English Wikipages as a bridge, automatically acquires new lexical units and example sentences also for other languages.

We apply our extraction algorithm to the Italian Wikipedia, because this work is focused on the development of FrameNet for Italian. Anyhow, the approach can be exploited in principle for every language available in Wikipedia.

Similarly to the data extraction process described in Section 6.6, we consider for every (l, F) pair in English the linked Wikipedia sense s, in English as well. Then, we retrieve the Italian Wikipedia sense (= page) s_i linked to s and extract all sentences C_i in the Italian Wikipedia dump[10] with a reference to s_i. In this way, we can assume that C_i are example sentences of F and that the words or expressions W_i in C_i containing an embedded reference to s_i are good candidate lexical units of F in the Italian FrameNet. For example, if we link `http://en.wikipedia.org/wiki/Court` to the JUDICIAL_BODY frame, we first retrieve the Italian version of the site `http://it.wikipedia.org/wiki/Tribunale`. Then, with a top-down strategy, we further extract all Italian sentences pointing to the `Tribunale` page and acquire as lexical units all words with an embedded reference to this concept, for example *tribunale* and *corte*. In this way, we can include the extracted lexical units and the sentences where they occur in the JUDICIAL_BODY frame for Italian.

6.7.1 Italian data extraction

In order to carry out the sentence extraction task for Italian, we relied on the results obtained using the *example-based* methodology in English, because it achieved better results than the *definition-based* strategy. Given the 2,505 (l, F) pairs mapped to Wikipedia articles using the example-based test set (see Section 6.6.2), we first extracted the Italian Wikipages that are linked to the English ones. Then for every linked Wikipage in Italian, we retrieved all sentences with a reference pointing to that page in the Italian Wikipedia dump. Statistics about the extracted data are reported in Table 6.4.

	Italian Wikipedia
Linked Wikipages in Italian	1,197
N. of extracted sents	32,539
(l, F) pairs with at least 1 sentence	740
Avg. sents per Italian sense	44
LU candidates	3,605

Table 6.4: Extracted data from Italian Wikipedia

We applied to the Italian sentences the same filtering routines developed for

[10] `http://download.wikimedia.org/itwiki/20090203`

English, after having modified the language-dependent rules. Only about one third of the 104,845 extracted sentences were selected. Besides, since the Italian Wikipedia is about one fifth of the English one, it was not possible to map every English Wikipage with an Italian article. Indeed, only 1,197 senses out of 2,505 in English were linked to an Italian page. Furthermore, only for 740 (l, F) pairs mapped to a Wikipedia article it was possible to extract at least one example sentence from the Italian Wikipedia dump. Also the average number of sentences extracted for every sense is much smaller than in English (44 vs. 140). Anyhow, this does not represent a problem because in the English FrameNet, the lexical units are usually instantiated by a set of 14 annotated sentences on average. So, according to the FrameNet standard, 44 sentences are more than enough to represent the meaning of a lexical unit in a frame.

6.7.2 Evaluation of the Italian sentences

In this evaluation part, we took into account 1,000 sentences, in order to have a comparable dataset w.r.t. the evaluation for English. However, the sets of Italian sentences extracted for every (l, F), i.e. for every Wikipedia article, were much smaller, so we increased the number of randomly chosen (l, F) pairs to 200, taking into account 5 extracted examples for every pair. Our evaluation is focused on the quality of the sentences and aims at assessing if the given sentences are correctly assigned to the (l, F) pairs. We report 69% accuracy, which is 5% lower than for the best English model, even if the measures are not directly comparable because the evaluated datasets have different structures.

If we look at the most common mistakes, we notice the same problems recorded for English (Section 6.6.3). Errors can originate from different granularity between FrameNet and Wikipedia, especially when Wikipedia is more specific, as well as from errors by the WSD system. For example, if the WSD algorithm maps the *peanut.n* LU in the FOOD frame to the `Peanuts` article in the English Wikipedia, dealing with the famous comic strip, the corresponding article in Italian is about the comic strip as well, thus the error is propagated also to the sentence extraction step for Italian.

Besides, other problems arise from the fact that we rely on the Italian Wikipage linked to the English article that was previously mapped to (l, F) and we assume that an English page and its Italian version describe the same concept. If we look at the evaluated data, we can see that it is not always the case: since Italian Wikipedia contains less articles than the English version, there must be either some gaps, or several fine-grained sense distinctions in English must have been merged into a more general article in Italian. For example, in the English Wikipedia there are two different articles describing the sense of *prescription* in the medical and in the law domain, whereas in the Italian version only the legal sense is present. Another problem that affects the accuracy of the extracted sentences is the quality of the links. This element has already been noticed in the English dataset, but it is even more relevant in the Italian Wikipedia, since its articles are edited and checked by a smaller number of users. In general, redirections and internal links tend to be less precise than in English. For example, the word "*arma*" in Italian is ambiguous because it can describe both a weapon and an army. Even if both articles are present in the Italian Wikipedia, namely `Arma` and `Armi_e_Servizi_dell'Esercito_Italiano`, some sentences containing "*arma*" meaning weapon are linked to the wrong page, as the one reported below:

(6.10) Capitano di nobile famiglia siciliana, originaria di Catania, entrato, per avverse fortune, nell' *Arma* [→ link to `Arma`] dei Carabinieri.
He was a captain from a noble Sicilian family, originally from Catania, who entered the Army of Carabinieri through unfavorable fortune.

A particularly interesting application of the multilingual sentence extraction is the straightforward induction of new LUs. Indeed, Italian FrameNet does not exist yet, so every lexical unit in an extracted sentence that is correct can directly populate the frames in the Italian version. For example, the 691 sentences in the evaluation dataset which were considered correct contain 387 unique LUs, which can be assigned to 72 different frames without any manual effort. The advantage is that not only we identify and classify new LUs, but we also provide them with example sentences that express the particular meaning of the LUs in the different frames. For example, the WEAPON frame for Italian was automatically created and populated with "*AK-47.n*", "*lanciafiamme.n*" (*flame-thrower*), "*fucile.n*" (*rifle*), "*arma.n*" (*weapon*), "*pistola*" (*gun*), "*revolver.n*" and "*rivoltella.n*" (*six-shooter*). The three last LUs were obtained starting from a single English LU, "*six-shooter.n*", which was then mapped to the English Wikipage `Revolver`. The Italian version of the page was `Rivoltella`, and the extracted sentences which were pointing to it contained three equivalent translations of "*six-shooter.n*", namely "*pistola*", "*revolver.n*" and "*rivoltella.n.*

6.8 Summary

In this chapter, we have investigated the use of Wikipedia for the automatic expansion of English FrameNet and for the automatic population of frames for Italian. We proposed to apply a word sense disambiguation system to a new task, namely the linking between LU senses and Wikipedia pages. Then, we have exploited such mapping to automatically extract a number of sentences with a pre-assigned frame label, both in English and in Italian.

This methodology offers several advantages. First of all, it is possible to extend the sentence extraction algorithm to every language available in Wikipedia. The amount of retrieved datasets would be of course smaller than for English, but the evaluation on Italian shows that accuracy could still be around 70%, at least for Wikipedia versions of similar dimensions, namely in German, French, Spanish and Portuguese. It would be interesting to check if the same algorithm can be applied also to languages that are typologically different from European languages, for example to Japanese.

A second advantage offered by Wikipedia is the encyclopedic knowledge it represents, meaning that it deals with a number of different topics. This guarantees that all frames, even domain-specific ones, could be potentially instantiated in the resource. On the contrary, other multilingual corpora, for example EUROPARL, deal with a more restricted set of topics and it would be impossible to retrieve the variety of word senses represented in FrameNet. Besides, Wikipedia is continuously updated and edited. In theory, this would allow us to study word usage in a dynamic way, even to acquire a new LU for a frame before it is identified and classified by FrameNet lexicographers. As for new FrameNets, it is worth noticing that Wikipedia is the only extensive resource available for many lesser-used languages, which makes it a valuable reference for corpus-based studies.

A drawback of our methodology is that the WSD system relies exclusively on nominal lexical units and verbal lexical units are not taken into account in the disambiguation step. Even if verbal LUs can be acquired in the sentence extraction process, our approach tends to retrieve them less frequently than nominal LUs. As a first solution, we could just try to run the WSD algorithm using test sentences with verbal LUs, and evaluate them.

Once the pre-classified sentences have been extracted, they can be exploited in different ways. On the one hand, the retrieved examples can speed up human annotation, requiring only a manual validation. On the other hand, the extracted sentences could provide enough training data to machine learning systems for frame assignment, since insufficient frame attestations in the FrameNet database are a major problem for such systems.

Chapter 7

Conclusions and perspectives for future research

7.1 Summary

In this dissertation, we presented different approaches aimed at developing in a semi-automatic way FrameNets for new languages, with a focus on Italian.

We have described the theoretical basis of frame semantics and construction grammar and we have presented in detail the structure of the English FrameNet database. Then, we have discussed the main issues related to the creation of FrameNet for Italian. We have shown that the original database structure has a language-independent component, which can be preserved in new FrameNets with minor adjustments, and a language-dependent part, which includes corpus-based attestations of lexical units. Even if our work was not aimed at obtaining a systematic analysis of the similarities and differences between English FrameNet and Italian FrameNet, the manual annotation of some sample Italian texts and the analysis of errors made by our automatic algorithms allowed us to point out a number of problematic issues such as lexical gaps or missing frames in the English database. Besides, we described some new frames introduced for Italian and, in Appendix A, we briefly presented some strategies adopted for the frame annotation of spontaneous speech.

From an NLP perspective, we investigated three research directions for the semi-automatic development of FrameNet for Italian, which can be potentially applied also to other languages. We call our approach *semi-automatic* and not *fully* automatic because the final objective of this investigation is to look for new ways to induce annotated data and lexical resources with near-manual quality, in order to avoid expensive and time-consuming annotation work.

The first approach investigated was the automatic projection of frame annotation from English to Italian using parallel texts. We demonstrated that this methodology can be applied to Italian without major changes with regard to other European languages, and on the other hand we showed that the transfer task can greatly benefit from the use of a *controlled* parallel corpus, minimizing syntactic complexity and free translations. We also proved that the evaluation framework adopted can lead to very different performance measures and we proposed an evaluation methodology which focuses on the quality of the annotated corpus resulting from annotation transfer. We also differentiated between target and FE evaluation and we prioritized semantic head match over exact constituent match.

The second research direction was the automatic acquisition of Italian lexical units through the FrameNet - WordNet mapping. We have shown that the introduction of features based on definition overlap and WordNet domains and the use of a machine-learning framework can achieve good mapping accuracy. It also allows for including in the mapping poorly-annotated LUs, which were usually discarded in previous works on the same task. Besides, we have empirically verified that the mapping can be exploited to populate Italian frames with LUs by importing lemmas from the mapped synsets with good accuracy. We have also shown that the mapping allows for the straightforward inclusion of a new annotation layer to the MultiSemCor corpus because every synset can be enriched with a frame label, both in the English and in the Italian version of the corpus. Results are browsable also from the MultiSemCor site.

The third approach taken into account was based on the idea that Wikipedia can be exploited as a repository of example sentences for different frames. This initial intuition was supported by the implementation of a system that, after mapping a LU to a Wikipedia article, extracts a set of example sentences for the given LU. The task was first tested on English Wikipedia and then a multilingual extension was added, in order to extract frame instantiations also from Italian Wikipedia. A set of evaluations involving both English and Italian extracted sentences proved that the approach is well substantiated. This extraction methodology can be easily extended to all languages available in Wikipedia with an accuracy that, we guess, may be similar to the measure obtained for Italian.

7.2 The Final Resource

The work presented in this thesis has required both implementation and annotation effort. Since the proposed methodologies are quite new, no benchmarks for evaluation were available, and we had to manually create the gold standards and verify the extracted data.

In the remainder we summarize the systems and the data produced during this work. The **algorithms / systems** are:

(a) Two transfer algorithms for cross-lingual projection of frame information, implemented as rule-based transfer systems in Prolog.

(b) A WordNet – FrameNet mapping system that can be exported for every language available in MultiWordNet and has been implemented using SVM with minimal supervision effort

(c) A sentence extraction system based on Wikipedia that can be exported for every language available in Wikipedia. The application relies on an existing state-of-the-art WSD system.

Our **annotated data** comprise both automatically extracted and manually annotated material, such as:

(a) EUROPARL gold standard with 1,000 parallel sentences in English and Italian, parsed, aligned at word level, manually annotated with frame information (see Appendix B.1).

(b) 400 sentences in English extracted from the Berkeley FrameNet database and translated into Italian, parsed (only Italian side), aligned at word level, manually annotated with frame information (see Appendix B.2).

(c) 2,158 LU-synset pairs manually annotated as positive or negative examples; 5,162 LU-synset pairs automatically annotated and available for download at `http://danielepighin.net/cms/research/MapNet`, with a precision around 80%.

(d) 27,793 English sentences and 23,872 Italian sentences from the *MultiSemCor* corpus, with PoS, lemma and synset information, automatically enriched with frame labels pointing to the synsets (`http://multisemcor.itc.it/`). 200 randomly extracted sentences in both languages have been manually evaluated, showing that precision achieves 75% on the English side and 70% on the Italian side.

(e) More than 300,000 sentences automatically extracted from English Wikipedia and annotated with a frame label. More than 32,000 sentences automatically retrieved from Italian Wikipedia and annotated with a frame label. 1,000 sentences have been randomly extracted from both datasets and manually verified, achieving 74% precision on English and 69% precision on Italian.

7.3 Future work

The explorative work presented in this thesis has the goal of investigating semi-automatic approaches for the development of Italian FrameNet. This implies that much work should still be consolidated in order to deliver a complete and fully-revised version of such resource. First of all, manual annotations should be checked and validated by at least another annotator in order to compute inter-annotator agreement. Another issue concerns adjudication of difficult cases that for the moment have been left unassigned. Besides, we sketched new definitions of frames which need to be integrated in the general frame ontology specifying frame-to-frame relations. In order to fully comply with the FrameNet format, we eventually need to add grammatical functions to the annotated FE labels, preferably in a semi-automatic way.

This work demonstrates that the addressed task involves several skills which can hardly be mastered by a single person. Indeed, a preliminary investigation of this kind should then give way to a more structured annotation and implementation work. At the time of writing, such change is taking place because different research teams with specific expertise are putting their efforts together and contributing to the creation of Italian FrameNet. This spontaneous initiative includes our contribution as well as the work carried out at University of Pisa, Istituto di Linguistica Computazionale in Pisa, University of Trento, University of Tor Vergata and IT company CELI. We will provide the consortium with the data and the software described in the present work, in order to create a shared repository of resources that serve as a starting point for future work and confrontation.

Since in this thesis we have often stressed the multilingual character of our approach and its applicability to new languages, it would be very interesting to actually extend it to other languages. In particular, the experiments relying on MultiWordNet and Wikipedia can be straightforwardly run on new languages, and the quality of the retrieved data could be evaluated and compared to the performance obtained on Italian. If results confirm the Italian measures, it may be worth creating a large multilingual repository of data and distribute them for research purposes. This would support research on FrameNet for new languages, in particular for less-resourced ones.

As for long-term activities, we believe that the NLP community could benefit from the development of domain-specific FrameNets. Since research about computational approaches to sentiment analysis and detection has deserved increasing attention among NLP researchers, an extension of FrameNet dealing with the affective domain would give an important contribution, adding information about situations and semantic roles in texts expressing attitudes and opinions. Some attempts have already been made in this direction, as reported by Kim and Hovy (2006), who propose to identify holder and topic of opinions in text using frame and FE labels. A 'FrameNet Affect' extension would offer additional information with regard to existing lexical resources for sentiment domain such as WordNet Affect (Strapparava and Valitutti, 2004).

Another approach we would like to explore is the use of emerging annotation strategies such as the one offered by Amazon's Mechanical Turk and their applicability to frame information. In particular, we plan to devise a methodology to decompose annotation into single simplified steps and to automatically merge them in order to obtain a complete and reliable annotation.

Appendix A

Frame semantics and dialogs

In this Appendix, we present a study which is not directly part of the semi-automatic approaches investigated for the creation of FrameNet for Italian. Instead, it aims at analyzing strategies and methodological issues for the annotation of spontaneous speech in Italian. To our knowledge, it is the first attempt to apply the FrameNet paradigm to dialogs, and it represents also an original contribution to the development of domain-specific FrameNets, since the spontaneous speech analyzed deals with the topic of software/hardware assistance.

So far, the application of frame semantics to specialized languages has been limited to few domains such as the biological, the patent and the legal domain. The biological domain was investigated by Dolbey et al. (2006), who successfully developed BioFrameNet, an extension of FrameNet to the molecular biology domain based on a corpus of scientific writings. The FrameNet paradigm has also been applied to patent processing in the PatExpert project[1], providing a model for identifying frame-based concepts to be processed by a Patent Upper Level Ontology (Codina et al., 2008). As for Legal FrameNet (Venturi et al., 2009), research is still ongoing and is currently dealing with the analysis of a corpus of Italian law texts in environmental and consumer rights domain.

With respect to such approaches, our study is not aimed at building a complete ontology for the technical domain. Its goal is rather to cover all annotated instances in the dialogs with suitable frame information, adding if necessary new frame definitions to the original FrameNet. In this respect, our approach is more corpus-driven and less theoretically-driven, because the annotation work was carried out to study the relationship between different annotation layers in the corpus and to apply them to the development of spoken dialog systems.

[1] `http://recerca.upf.edu/patexpert/`

The Appendix is structured as follows: in Section A.1 an overview of the LUNA project is given, in particular of the annotation framework involving the Italian dialogs. In Section A.2 we discuss the main issues of the annotation of frame information in dialogs and we describe how the standard annotation procedure was changed in order to face such issues. Then, the 20 newly introduced frames are reported in Section A.3. In Section A.4 we give a quantitative description of the annotation and compare statistical differences between human-machine and human-human dialogs. Then, we identify and analyze the relationship between frame and dialog act labels (Section A.5).

A.1 The LUNA project

The LUNA project (Language UNderstanding in multilinguAl communication systems)[2] was a three-year project funded under the Sixth Research Framework Programme of the European Union, whose main goal was to enhance real-time understanding of spontaneous speech in advanced telecom services. The project, which ended in August 2009 and operated over Italian, French and Polish, focused on different objectives, namely the language and semantic modeling of speech, the automatic learning and the multilingual portability of spoken language understanding components.

In this framework, a considerable part of the work about semantic modeling of dialogs consisted in the multi-layered annotation of a corpus of Italian spontaneous speech recorded in the help-desk facility of the Consortium for Information Systems of Piemonte Region. The corpus contains 1000 equally partitioned Human-Human (HH) and Human-Machine (HM) dialogs. The former are real conversations about software/hardware troubleshooting, while the latter are dialogs where an operator acting as Wizard of Oz reacts to the caller's requests following one of ten possible scenarios.

The annotation workflow is displayed in Fig. A.1: the dialogs are first recorded as audio files and then segmented at turn level and semi-automatically transcribed. Then, they are further segmented by hand at utterance level[3] and are annotated at three parallel semantic levels:

- The domain attribute annotation is based on a pre-definite domain ontology

[2] `http://www.ist-luna.eu/`

[3] The interval when the speaker is active is defined as a *turn*, which is included between two pauses in the speech flow. *Utterances* are complex semantic entities that usually represent the annotation unit for dialog acts. Their relation to speaker turns is not one-to-one, because in most cases a single turn contains multiple utterances, and sometimes utterances can span more than one turn.

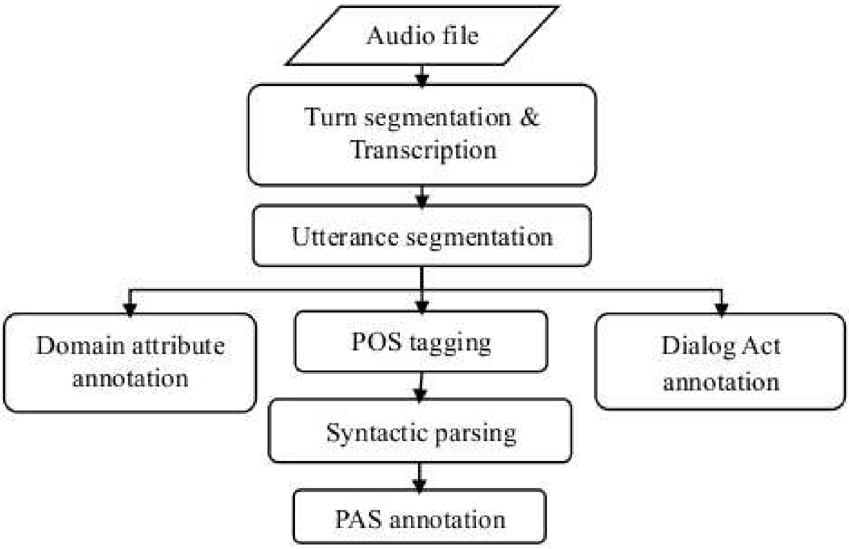

Figure A.1: The annotation process

and specifies concepts and their relations.

- Dialog acts (DAs), which describe the meaning of an utterance at the level of its illocutionary force (Austin, 1962), are annotated following a taxonomy that includes three main dialog act groups (Quarteroni et al., 2008): *core* acts, including offering or requesting to perform an action, asking or answering a question; *conventional* acts, such as greeting and quitting the conversation, and *feedback* acts representing clarification requests and acknowledgments of previous utterances. *Other* is used to annotate non interpretable/non-classifiable utterances. This partition is present in a number of well-known state-of-the-art DA taxonomies, for example the DAMSL (Core and Allen, 1997) and the TRAINS Traum (1996) coding scheme. The complete DA description is reported in Table A.1.

- Predicate-argument structure is annotated following the FrameNet paradigm. As shown in Fig. A.1, this step requires POS-tagging and syntactic parsing (via Bikel's parser trained for Italian (Corazza et al., 2007)). Then, a shallow manual correction is carried out to make sure that the tree nodes that may carry semantic information have correct constituent boundaries. The annotation workflow is the same as the one adopted for the gold standard development in the projection experiments (Chapter 4) and is described in Section 4.5.4.

	Core dialog acts
Info-request	Speaker wants information from addressee
Action-request	Speaker wants addressee to perform an action
Yes-answer	Affirmative answer
No-answer	Negative answer
Answer	Other kinds of answer
Offer	Speaker offers or commits to perform an action
ReportOnAction	Speaker notifies an action is being/has been performed
Inform	Speaker provides addressee with information not explicitly required (via an Info-request)
	Conventional dialog acts
Greet	Conversation opening
Quit	Conversation closing
Apology	Apology
Thank	Thanking (and down-playing)
	Feedback/turn management dialog acts
Clarif-request	Speaker asks addressee for confirmation/repetition of previous utterance for clarification.
Ack	Speaker expresses agreement with previous utterance, or provides feedback to signal understanding of what the addressee said
Filler	Utterance whose main goal is to manage conversational time (i.e. dpeaker taking time while keeping the turn)
	Non-interpretable/non-classifiable dialog acts
Other	Default tag for non-interpretable and non-classifiable utterances

Table A.1: DA taxonomy applied to the LUNA corpus

The multi-level annotation protocol was specifically studied within the project in order to investigate statistical relations between the layers, in particular between semantic and discourse features used in spontaneous conversations.

Figure A.2 displays four turns annotated according to the three-layered protocol. Domain attribute labels are reported above the utterances in capitals. In particular, the first and the last utterance don't present domain-attribute annotation, while the second and the third one bear several concept labels such ass ACTION, PART-OF-DAY, HARDWARE-COMPONENT, etc.

DA labels are placed before every utterance and correspond to *Info-request* in case of questions and *Info* when the speaker describes the problem he has with the printer.

Frame information is reported below the utterances. All lexical units are underlined and the frame is written in capitals, while the other labels refer to frame elements. In particular, ASSISTANCE is evoked by the lexical unit *aiutare* and has one attested frame element (*Benefitted_party*), GREETING has no frame element, and PROBLEM_DESCRIPTION and TELLING have two frame elements each.

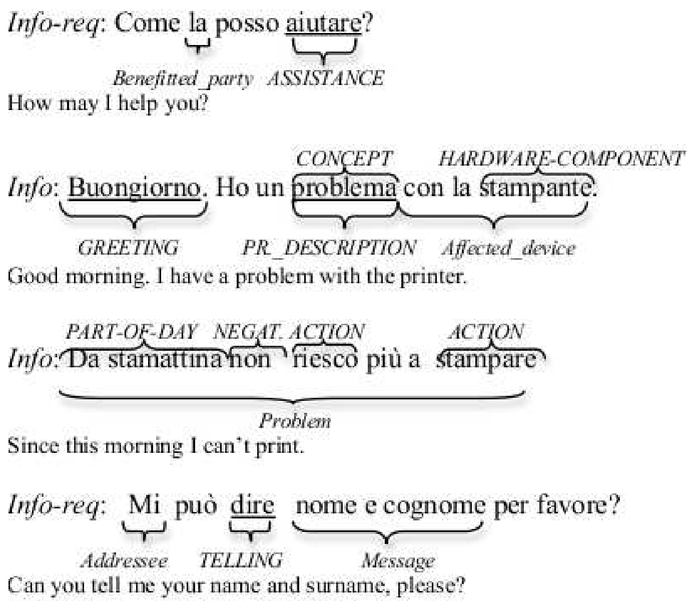

Figure A.2: Example of multi-layer annotation

While this study was part of a broad investigation involving several researchers, we focus here on the annotation of frame information, which represents the original part of our contribution.

A.2 Frame annotation of the LUNA corpus

We annotated 10% of the LUNA corpus with frame information, divided into 50 HM and 50 HH dialogs. The task was carried out with the aim of contributing to the study of spoken language understanding systems. For this reason, the annotation strategy was neither the one used in the original FrameNet project (i.e. one frame per sentence) nor the continuous-text annotation. Instead, we identified all lexical units corresponding to *semantically relevant* verbs, nouns and adjectives with a syntactic subcategorization pattern, possibly skipping the utterances with empty or fragmentary semantics (e.g. disfluencies). In particular, we annotated all lexical units that imply an action, introduce the speaker's opinion or describe the office environment, in other words all the concepts necessary to understand what is going on during the conversation. As for the annotation strategy, we followed the same workflow described in Section 4.5.4 for the annotation of the gold standards in the annotation transfer experiments. In particular, the dialogs were first parsed, then manually checked and finally annotated using SALTO (Burchardt et al., 2006). As in the previous annotation, the *Empty_subj* label was used to characterize verbal target words whose subject, bearing a FE label, is not overtly expressed but is conveyed by the morphological features of the verb (person and number). Besides, with SALTO we could easily introduce and annotate in the corpus 20 new frames (see following Section A.3).

The annotation of dialogs presented some problematic issues and it was not always possible to apply the annotation guidelines adopted for the current FrameNet projects. In particular, the main issues were:

- Repetitions are very frequent in dialogs, and it can happen that part of an utterance corresponding to a FE is repeated several times. In those cases, the FE label was assigned only to the closest candidate FE w.r.t. the target word.

- It was not always possible to annotate every utterance, particularly in case of disfluencies and semantically empty expressions. The general approach was to annotate utterances, if at least part of their meaning was expressed. For example, if a speaker could utter part of the turn and then was interrupted, annotation involved the tokens that were understandable, as shown in Fig. A.3 (translation: "No, well, in the sense that if you put the plug in the yes"). According to the guidelines, the *Goal* FE is assigned to the last PP in the sentence, even if it is not complete.

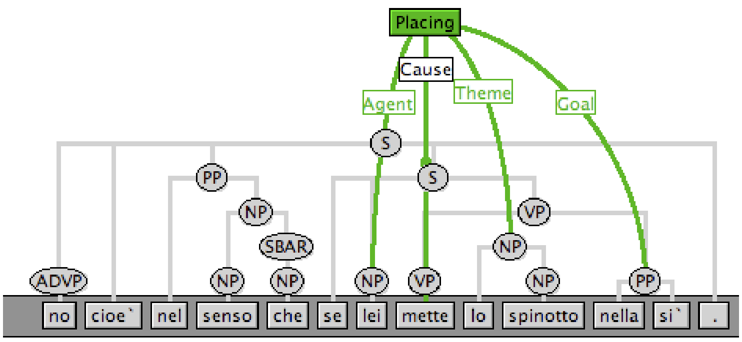

Figure A.3: Assignment of a FE label to an incomplete constituent

This approach represented a good solution to identify and annotate as much semantic information as possible, but we are aware that it cannot be easily generalized because the idea of "understandability" strongly depends on the annotator's choice and intuition.

- According to the same "understandability" principle, we introduced the *Corrected* flag for words which were clearly misspelled, due both to transcription errors and to speaker's mistakes. The flag allows to introduce the corrected version of misspelled word, which is displayed below the corresponding token. We report an example in Fig. A.4 (the sentence means "No, well, but I am sorry to make you go there"). From the context, it is clear that *mandare* is the misspelling of *andare* (*go*), which is manually added by the annotator. In this case, the correction was particularly relevant because it involved the target word of the ARRIVING frame.

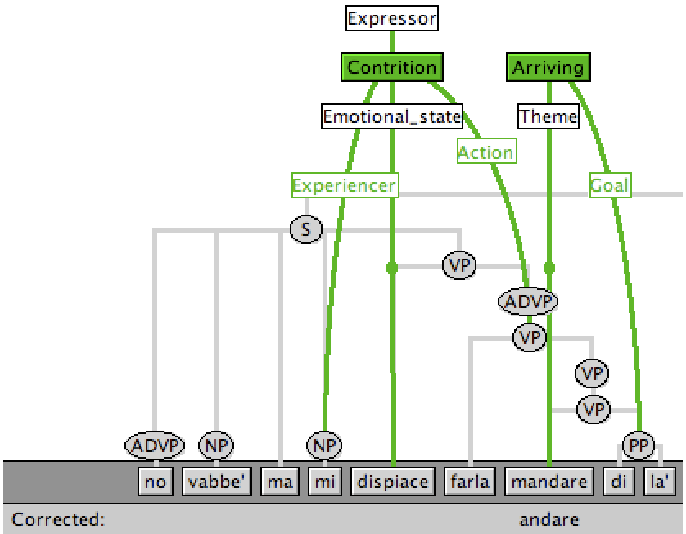

Figure A.4: Correction of misspelled *mandare*

- Annotation with SALTO is carried out sentence-by-sentence, while in dialogs semantic elements bearing frame information (LU and FEs) can span different turns because of interruptions and overlaps. For the moment, we limited annotation to the utterance level and to frame elements that are explicitly expressed by some lexical or phrasal material. In order to take into account inter-sentential relationships and the discourse context, however, it would be important to annotate also null instantiated FEs (Ruppenhofer et al., 2006). In particular, the so-called Definite Null Instantiations (DNI), which characterize lexically null instantiations of FEs that are already known from the context, could contribute to the identification of anaphoric relations between utterances. For example, the annotated sentence reported in Fig. A.5 ("Ok, just a minute and I connect myself") would allow the introduction of a DNI label for the *Goal* FE, which is not expressed but can be understood from the previous turns as being the computer. A further annotation step would then involve the identification of the connection between DNI label and the referent (Ruppenhofer et al., 2009).

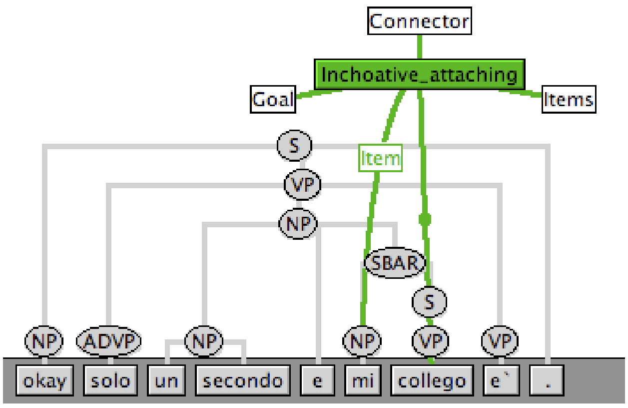

Figure A.5: Case of Definite Null Instantiation

- For all target words whose literal meaning does not correspond to the figurative one, we assign a frame label according to the figurative reading. This does not involve only idiomatic expressions, which are very frequent in dialogs, but also verbs with a generic meaning such as *"fare"* (*do/make*) and *"mettere"* (*put*), that in spontaneous conversations are used very often instead of more specific verbs. We prefer to annotate them with the *intended* meaning, if it can be unambiguously understood from the context.

As for the domain-specificity of the language and the influence of conversational style, several new frames were introduced, which will be described in the following Section.

A.3 Newly introduced frames

We introduced 20 new frames out of the 174 identified in the corpus. The newly introduced frames can be grouped into three main classes:

1. Some frames were created because there was a gap in the English FrameNet hierarchy, for example RENDER_NONFUNCTIONAL is in the FrameNet database, but BECOMING_NONFUNCTIONAL is missing. The newly introduced frames of this kind have the same level of specificity of the existing ones.

2. Some frames were created to cover the domain-specific topics discussed, since the original definition of frames related to hardware/software, data-handling and customer assistance was sometimes too coarse-grained. Some domain-oriented adaptations of existing frames were needed to describe specific situations, for example to distinguish between "real" and "virtual" movement (ARRIVING vs. NAVIGATION).

3. The GREETING frame was introduced because the FrameNet database so far does not take into account frames related to oral communication, apart from ATTENTION_GETTING. Few new frame elements were introduced as well, mostly expressing syntactic realizations that are typical of spoken Italian.

The list of new frames with the corresponding definition is reported in Table A.2. Notice that this list was developed for internal use during the annotation of the LUNA corpus and is not meant to be definitive.

Frame	**Definition**
ACQUIRE_DATA	A *User* moves some *Data* from a *Source* into an *Application* or a *Goal*. Ex. Hai <u>importato</u> [la password]$_{\text{Data}}$ [dalla vecchia versione]$_{\text{Source}}$? (Have you imported the password from the older version?).
ASSIGN	An *Item* is assigned to a *Receiver* so that he carries out or is in charge of a particular job or *Task*. Ex. [La richiesta]$_{\text{Item}}$ è <u>in carico</u> [al gruppo fonia]$_{\text{Receiver}}$. (Transl: "The telephonic group is in charge of the request". In English the order of the frame elements is inverted).
BECOMING_NONFUNCTIONAL	An *Artifact* becomes no longer capable of performing its inherent function. Ex. [La stampante]$_{\text{Artifact}}$ <u>si rompe</u> facilmente. ("The printer breaks easily"). Notice that "La stampante è rotta" (The printer is broken) would be BEING_OPERATIONAL. Precedes BEING_OPERATIONAL.
CHANGE_DATA	A *User* changes the content of a document so that *New_data* replace *Old_data*. Ex. Devi <u>aggiornare</u> [la password]$_{\text{Old\_data}}$. ("You must update your password"). This is the domain-specific version of REPLACING.
COME_TO_SIGHT	A *Graphical_element* becomes visible to a *Perceiver*. A *Duration* and *Manner* may also be specified

Frame	**Definition**
	Ex. Appare [per qualche istante]$_{\text{Duration}}$ [una schermata nera]$_{\text{Graphical\_element}}$. ("A black screen appears for some seconds").
CREATE_DATA	A *User* newly creates *Data*, information or a document containing them. He can assign a name or a *Label* to the data. Ex. [La richiesta]$_{\text{Data}}$ è stata aperta [come reset password]$_{\text{Label}}$. ("The request was opened as password reset").
DISPLAY_DATA	A *Device* shows to a *Perceiver* some *Data* or a *Message* through a *Support* so that the data become visible. Ex. [Il PC]$_{\text{Device}}$ [ti]$_{\text{Perceiver}}$ ripropone [delle altre lettere]$_{\text{Message}}$ [sullo schermo]$_{\text{Support}}$. ("The PC shows you some other letters on the screen").
CREATE_SPACE	This frame was introduced for the lexical units *liberare.v* (to free) and *libero.a* (free.a) when they are used to refer to *Empty_space* available on a *Device* Ex. Cerca di liberare [un po' di spazio]$_{\text{Empty\_space}}$ [sul PC]$_{\text{Device}}$. (Try to create some space on the PC).
GREETING	This frame includes all words and expressions used to give a sign of welcome or recognition to an *Addressee*. Ex. Buongiorno [a lei]$_{\text{Addressee}}$. (Good morning to you).
HANDLE_DATA	An *Operator* handles some *Data* or documents he is in charge of in order to carry out a task or a particular job. The *Application* used is optional Adesso gestisco [io]$_{\text{Operator}}$ [questa richiesta]$_{\text{Data}}$. (Now I handle this request).
INSERT_DATA	A *User* inserts some *Data* in a *Document* or a *Device*. The *Category* of the inserted data, the *Purpose* as well as the *Manner* in which the data are inserted may also be specified. Ex. [Il suo responsabile]$_{\text{User}}$ deve compilare [il documento]$_{\text{Document}}$ [per la richiesta]$_{\text{Purpose}}$. (Your boss has to fill out the document for the request).
LEND	A *Lender* grants to a *Borrower* the use of an *Object* on the understanding that it shall be returned. The *Duration* may also be specified. [Mi]$_{\text{Borrower}}$ puoi prestare [il tuo PC]$_{\text{Object}}$ [solo un secondo]$_{\text{Duration}}$? (Can you lend me your PC for a second?).
LOSE_DATA	Some *Data* or documents are unwillingly lost at the expenses of an *Affected_user* Ex. Se non lo risolvi potrei perder[mi]$_{\text{Affected\_user}}$ [qualche lavoro]$_{\text{Data}}$. (If you don't solve this, I could lose some work).

Frame	Definition
Navigation	A *User* enters or leaves an *Application* or moves in a virtual space (ex. to a generic *Goal*). *Reason*, *Time* and *Means* may also be specified. The valence pattern of this frame is similar to the Arriving frame but it refers to the software environment. Ex. Devi andare [alla web-mail]$_{\text{Application}}$. (You have to go to the web-mail).
Open_data	A *User* opens an already existing *Document* with an *Application.* The frame is different from Create_data because in this case the document already exists. It is also different from Change_operational_state, which involves the opening of an application. Ex. Devi aprire [il file]$_{\text{Document}}$. (You must open the file).
Problem_description	This frame was created to annotate all the possible elements involved in the description of a problem in the technical domain. The lexical units in the frame are usually synonyms of *problema.n* (problem). It is a more specific version of Predicament. The problem can have a certain *Degree* and can involve a *Device* and an *Affected_person.* Sometimes it can hinder the execution of an *Affected_activity.* Ex. [I tuoi colleghi]$_{\text{Affected\_person}}$ hanno lo stesso problema [a collegare la stampante]$_{\text{Affected\_activity}}$. (Your colleagues have the same problem while connecting the printer).
Read_data	A *User* reads some *Data* or a *Device_with_data* using some *Reading_device.* Alternatively, a *Reading_device* reads some *Data* for a *User.* It is a domain-specific version of the Reading frame because it involves a device and some electronic data. Ex. Faccio fatica a leggere [il CD]$_{\text{Device\_with\_data}}$ [dal lettore]$_{\text{Reading\_device}}$. (I can hardly read the CD from the CD-player).
Run_operation	This frame refers specifically to the technical domain and describes a situation where an *Operator* runs or executes an *Operation* in a given *Environment* Ex. Hai provato ad eseguire [un ipconfig]$_{\text{Operation}}$ [da una sessione DOS]$_{\text{Environment}}$? (Have you tried to run an ipconfig from a DOS session?).
Select_data	An *Operator* selects some *Data*, optionally with a *Device.*

Frame	Definition
	Ex. <u>Clicca</u> [sulla password]$_{\text{Data}}$ [con il tasto destro del mouse]$_{\text{Device}}$. (Click on the password with the right mouse button).
UNDERGO_CHANGE_OF_OP._STATE	A *Device* or application goes in (or out of) service. The *Time* and the *Place* where the *Device* goes in or out of use may be specified. Precedes BEING_IN_OPERATION. It is similar to PROCESS_START but in that case an *Event* is involved, while here it is a *Device*. Ex. [Il computer]$_{\text{Device}}$ non <u>si accende</u>. (The computer does not start).

Table A.2: The 20 newly introduced frames

A.4 Statistics about the annotated corpus

Table A.3 shows some statistics about the corpus dimension and the results of our annotation. On average, HH dialogs are longer than HM, which concerns both the number of turns in a dialog and the number of tokens in a turn. However, HH dialogs contain less frame instances in average than the HM group, meaning that speech disfluencies, not present in turns uttered by the WOZ, negatively affect the semantic density of a turn. For the same reason, the percentage of turns in HH dialogs that were manually corrected in the pre-processing step (see Section 2.2) is lower than for HM turns, since HH dialogs have more turns that are semantically empty and that were skipped in the correction phase.

	HM	HH
Total number of turns	662	1,997
Mean dialog length (turns)	13.2	39.9
Mean turn length (tokens)	11.4	10.8
Mean nb of frame instances per dialog	18.5±5.1	39.0±17.2
Corrected turns (%)	50	39
Total number of annotations	923	1951
Mean nb of frame annotations per turn	1.4	1.0
Mean nb of FEs per frame annotation	1.6	1.7

Table A.3: Dialog statistics for the human-machine resp. human-human corpus

Table A.4 shows how many frames have a certain occurrence. These values are very important because they allow to extract statistically significant data about the corpus, which could be very useful in a machine learning system for automatic frame recognition. In the English FrameNet, instead, the lemma-by-lemma annotation style does not deliver useful measure about frame and LU frequency, which is a reason why the annotation of continuous text was introduced in the project.

The most frequent frame group comprises frames related to information exchange that is typical of the help-desk activity, including TELLING, GREETING, CONTACTING, STATEMENT, RECORDING, COMMUNICATION. Another relevant group encompasses frames related to the operational state of a device, for example BEING_OPERATIONAL, CHANGE_OPERATIONAL_STATE, OPERATIONAL_TESTING, BEING_IN_OPERATION.

The two groups also show high variability of lexical units (Table A.5). GREETING and TELLING have the richest lexical unit sets, resp. with 12 and 11 LUs each. ARRIVING, AWARENESS and CHANGE_OPERATIONAL_STATE are expressed by 10 different lexical units, while STATEMENT, BEING_OPERATIONAL, REMOVING and UNDERGO_CHANGE_OF_OPERATIONAL_STATE have 9 different lexical units each. Also in this case, the frames with large LU sets characterize the technical domain and the conversational context. Besides, the informal nature of the spoken dialogs influences the composition of the LU sets. In fact, they are rich in verbs and multiwords used only in colloquial contexts, for which there are generally few attestations in the English FrameNet database.

Occurrences	Nr. of frames
1	39
2-5	53
6-10	19
11-20	22
>20	41

Table A.4: Frame occurrences

LUs per frame	Nr. of frames
1	73
2-5	79
6-10	20
11	1 (*Telling*)
12	1 (*Greeting*)

Table A.5: Nr. of LUs per frame

Table A.6 reports the 10 most frequent frames occurring in the human-machine resp. human-human dialogs. The relative frame frequency in HH dialogs is more sparse than in HM dialogs, meaning that the task-solving strategy followed by the WOZ limits the number of digressions, whereas the semantics of HH dialogs is richer and more variable. As mentioned above, we had to introduce and define new frames which were not present in the original FrameNet database for English in order to capture all relevant situations described in the dialogs. A number of these frames appear in both tables, suggesting that they are indeed relevant to model the general semantics of the dialogs we are approaching.

HM corpus			HH corpus		
Frame	count	freq-%	Frame	count	freq-%
Greeting*	146	15.8	Telling	143	7.3
Telling	134	14.5	Greeting*	124	6.3
Recording	83	8.9	Awareness	74	3.8
Being_named	74	8.0	Contacting	63	3.2
Contacting	52	5.6	Giving	62	3.2
Usefulness	50	5.4	Navigation*	61	3.1
Being_operational	28	3.0	Change_operational_state	51	2.6
Problem_description*	24	2.6	Perception_experience	46	2.3
Inspecting	24	2.6	Insert_data*	46	2.3
Perception_experience	21	2.3	Come_to_sight*	38	1.9

Table A.6: 10 most frequent HM and HH frames (* = newly introduced frame)

With respect to the development of spoken dialog systems, it is crucial to identify recurring patterns of frames in order to model the semantics of the dialogs. For this reason, we analyzed the most frequent frame bigrams and trigrams in HM and HH dialogs. Results are reported in Table A.7. Both HH bigrams and trigrams show a sparser distribution and lower relative frequency than HM ones, implying that HH dialogs follow a more flexible structure with a richer set of topics, thus the sequence of themes is less predictable. In particular, 79% of HH bigrams and 97% of HH trigrams occur only once (vs. 68% HM bigrams and 82% HM trigrams). On the contrary, HM dialogs deal with a fix sequence of topics, driven by the turns uttered by the WOZ, which influences the sequence of annotated frames. For instance, the most frequent HM bigram and trigram both correspond to the opening utterance of the WOZ:

Help desk <u>buongiorno</u>GREETING, <u>sono</u>BEING_NAMED Paola, in cosa posso esserti <u>utile</u>USEFULNESS?
(Good morning, help-desk service, Paola speaking, how can I help you?)

As for HH dialogs, the most frequent patterns characterize the opening and the end of the dialog, for example the repeated greetings between caller and operator (bigram Greeting, Greeting and trigram Greeting, Greeting, Greeting), the greetings of the two and the presentation of the operator (trigram Greeting, Being_named, Greeting) or the final greetings with the operator's promise to call back as soon as the problem is solved (trigram Contacting, Greeting, Greeting).

Frame bigrams		Frame trigrams	
human-machine (HM)	freq-%	human-machine (HM)	freq-%
Greeting Being_named	17.1	Greeting Being_named Usefulness	9.5
Being_named Usefulness	15.3	Recording Contacting Greeting	5.7
Telling Recording	12.9	Being_named Usefulness Greeting	3.7
Recording Contacting	10.9	Telling Recording Contacting	3.5
Contacting Greeting	10.6	Telling Recording Recording	2.2
human-human (HH)	freq-%	human-human (HH)	freq-%
Greeting Greeting	4.7	Greeting Greeting Greeting	1.6
Navigation Navigation	1.2	Greeting Being_named Greeting	0.5
Telling Telling	1.0	Contacting Greeting Greeting	0.3
Change_op._state Change_op._state	0.9	Navigation Navigation Navigation	0.2
Telling Problem_description	0.8	Working_on Greeting Greeting	0.2

Table A.7: 10 most frequent frame bigrams and trigrams

A.5 DA-frame Relationship

A unique feature of the LUNA corpus is the availability of both a semantic and a dialog act annotation level: it is intuitive to seek relationships for the purpose of improving the recognition and understanding of each level by using features from the other. We considered a subset of 20 HH and 50 HM dialogs and computed an initial analysis of the co-occurrences of dialog acts and frames. We noticed that each frame tended to co-occur only with a limited subset of the available dialog act tags, and moreover in most cases the co-occurrence happened with only one dialog act. For a more thorough analysis, we computed the weighted mutual information (MI) between frames and dialog acts[4].

In the HM corpus, we noted some interesting associations between dialog acts and frames. First, *info-req* has the maximal MI with frames such as BEING_IN_OPERATION and BEING_ATTACHED, as requests are typically used by the operator to get information about the status of device. Several frames denote a high MI with the *info* dialog act, including ACTIVITY_RESUME, INFORMATION, BEING_NAMED, CONTACTING and RESOLVE_PROBLEM. CONTACTING refer to the description of the situation and

[4]Following Bechet et al. (2004), we define the weighted MI between two events x_i and y_j as:

$$wMI(x_i; y_j) = p(x_i; y_j) log \frac{p(x_i; y_j)}{p(x_i)p(y_j)},$$

where $p(x_i; y_j)$ is the probability of co-occurrence of x_i and y_j and $p(x_i)$ and $p(y_j)$ are the marginal probabilities of occurrence of x_i resp. y_j in the corpus. We approximate all probabilities using frequency of occurrence.

of the speaker's point of view (usually the caller). BEING_NAMED is primarily employed when the caller introduces himself, while ACTIVITY_RESUME usually refers to the operator's description of the scheduled interventions. As for the remaining acts, *clarif* has the highest MI with PERCEPTION_EXPERIENCE and STATEMENT, used to warn the addressee about understanding problems resp. asking him to repeat/rephrase an utterance. The *answer* tag is highly informative with frames referring to the exchange of information (READ_DATA) or to actions performed by the user after a suggestion of the system (CHANGE_OPERATIONAL_STATE). Action requests (*act-req*) seems to be correlated to REPLACING as it usually occurs when the operator requests the caller to carry out an action to solve a problem, typically to replace a component with another. Another frequent request may refer to some device that the operator has to test.

In the HH corpus, most of the frames are highly mutually informative with *info*: indeed, this is the most frequently occurring act in HH except for *ack*, i.e. speaker's feedback or agreement, which rarely contains verbs that can be annotated as lexical units. As for the remaining acts, there is an easily explainable high MI between *quit* and GREETING, because both characterize the end of dialogs; moreover, *info-req* denotes its highest MI with GIVING, as in requests to give information.

After analyzing the mutual information in both the HM and HH cases, we corroborated our initial observation that for most frames, the mutual information tends to be very high in correspondence of one dialog act type, and lower or null with the others. This suggests a high correlation between specific frames and dialog acts, and is an important result for which we can think of several applications. One of these is the beneficial effect of including shallow semantic information such as frames as features for dialog act classification. On the contrary, the correspondence between dialog acts and frames is less clear as the same dialog act can relate to a span of words covered by multiple frames and generally, several frame types co-occur with the same DA.

A.6 Summary

The annotation of frame information in the LUNA corpus has proved to be very interesting from a theoretical point of view for at least three reasons: 1) new data for Italian FrameNet were annotated and analyzed, 2) the frame paradigm was applied to a new domain and 3) for the first time it was extended to conversational speech. In this respect, we introduced 20 new frames, which mostly concerned domain-specific situations about the technical field of software/hardware assistance. Besides, a preliminary study about the correspondences between frames and dialog acts showed that there are complex cross-layer dependencies between semantic and discourse features which may be exploited for the training of semantic models accounting for predicate interpretation. Besides, the comparison between HM and HH dialogs highlighted more regularities and less topic variability in the former corpus type, suggesting that HM dialogs could be employed as a model to build simplified dialog systems, while HH dialogs would require a more complex modeling. As for the applicative side, the HH dialogs in the LUNA corpus were used to train a system for automatic FramNet-based annotation of Italian dialogs (Coppola et al., 2008), which achieved with the best model F1 0.76 on the FE detection and classification task, outperforming the result obtained with the same system on the English FrameNet dataset (F1 0.59) (Coppola et al., 2009). This proved that some typical features of conversational speech such as repetitions and disfluencies do not impact on system performance. On the contrary, a small set of manually annotated data is enough for achieving good performance of the FE identification task because a limited number of topics is dealt with in domain-specific dialogs and because the distribution of annotated data is statistically significant.

Appendix B

Italian LUs and frames in the gold standards

B.1 Europarl

We report the list of frames identified in the Europarl gold standard for Italian with the corresponding lexical units. If the LU is associated more than once to the same frame, we report the frequency between parenthesis. LUs in italics need a further revision.

Frame	**Lexical unit**
Activity_finish	concludere.v
Activity_start	iniziare.v, mettersi_a.v
Activity_stop	abbandonare.v (2), fermare.v
Agree_or_refuse_to_act	consenziente.a
Amassing	accumularsi.v (2)
Appearance	sembrare.v
Arriving	arrivare.v (5), giungere.v (5), pervenire.v, provenire.v (6), provenienza.n, raggiungere.v (2), venire.v (7)
Attack	aggredire.v (2), attaccare.v, assalire.v
Attempt	tentare.v, provare.v
Attention	attenzione.n (3)
Awareness	apprendere.v (2), capire.v (4), comprendere.v (4), conoscere.v (5), credere.v (24), essere_noto.a (2), pensare.v (4), rendersi_conto.v, ritenere.v (16), sapere.v (48), tenere_conto.v
Becoming	diventare.v
Becoming_aware	constatare.v, notare.v, registrare.v, rilevare.v, sapere.v, trovare.v, vedere.v
Being_at_risk	minacciato.a (2), pericolo.n

Frame	Lexical unit
Being_employed	lavorare.v (4)
Being_in_category	consistere.v, ricadere.v, rientrare.v
Being_named	chiamarsi.v (7), intitolarsi.v, noto_come.a
Being_necessary	necessario.a, necessità.n
Being_operational	funzionare.v
Body_movement	applaudire.v
Building	costruzione.n
Categorization	considerare.v (11), identificare.v, percepire.v, prendere_sul_serio.v, rappresentare.v, reputare.v, trattamento.n (2), trattare.v
Causation	indurre.v
Cause_change	cambiare.v
Cause_change_of_position_on_a_scale	aumentare.v, ridurre.v, riduzione.n
Cause_change_of_strength	consolidare.v
Cause_harm	colpire.v, ferire,v (5), picchiare.v (3)
Cause_to_fragment	rompere.v
Cause_to_make_progress	rifarsi.v
Certainty	confidare.v, dubitare.v (2), certo.a
Change_of_leadership	caduto.a, rovesciare.v (2)
Change_position_on_a_scale	calare.v, diminuzione.n, ridursi.v, scendere.v,
Choosing	scegliere.v (2)
Claim_ownership	rivendicazione.n
Cogitation	considerare.v (4), pensante.a, prendere_in_considerazione.v (5), riflessione.n, riflettere.v, soffermarsi.v, tenere_in_considerazione.v (3)
Collaboration	lavorare.v
Coming_to_be	diventare.v, emergere.v, levarsi.v, *porre.v*, presentarsi.v, sorgere.v, *andare.v*
Coming_to_believe	apprendere.v, rendersi_conto.v, sapere.v, trovare.v
Commerce_pay	compratore.n, *fare_le_spese.v*[1], finanziare.v, pagare.v (11), pagamento.n (2), romborsare.v, risarcire.v
Commerce_scenario	prezzo.n, acquirente.n
Commerce_sell	vendere.v
Commitment	minacciare.v (2), promettere.v
Communicate_categorization	trattamento.n (2), trattare.v (3)
Communication	diffondere.v
Communication_response	rispondere.v, risposta.n
Compliance	infrangere.v, rispettare.v (3), rompere.v, spezzare.v, violare.v (4), violazione.n
Conduct	approccio.n, comportamento.n, procedere.v
Contrition	dispiacersi.v (3), rammarico.n, spiacente.a

[1] Metaphorical use

Frame	Lexical unit
Convey_importance	evidenziare.v, ricordare.v, puntualizzare.v, sottolineare.v
Cure	somministrazione.v, curare.v (2)
Damaging	compromettere.v (2), pregiudicare (3)
Death	morire.v (8), morto.a, perdere_la_vita.v
Desiring	desiderare.v
Destiny	condannato.a, destinato.a
Destroying	distruggere.v
Discussion	dibattere.v (2), dibattito.n, discussione.n (5), discutere.v (5)
Dispersal	distribuire.v
Emotion_directed	rincrescersi.v, rincrescimento.n
Endangering	mettere_a_repentaglio.v (7), mettere_in_pericolo.v (2), mettere_a_rischio.v, minacciare.v (4), pregiudicare.v
Event	accadere.v (11), avvenire.v, realizzarsi.v, essere.v, succedere.v (3), verificarsi.v
Evidence	confermare.v, dimostrare.v (17), emergere.v, mostrare.v, provare.v, risultare.v
Execution	esecuzione.n (2), giustiziare.v (7), pena_capitale.n
Existence	condurre.v
Expectation	aspettarsi,v (3), attendere.v (2)
Expensiveness	costo.n
Experiencer_subj	rammarico.n
Experience_bodily_harm	ferito.n
Expressing_publicly	esprimere.v (3)
Fall_asleep	addormentarsi.v
Fame	noto.a
Feeling	nutrire.v, provare.v
Gathering_up	raccogliere.v
Getting	avere.v, conquistare.v
Giving	conferire.v (2), dare.v (10), fornire.v (2), lasciare.v, offrire.v (2)
Grasp	capire.v (2)
Guilt_or_innocence	responsabile.a
Handling[2]	trattamento.n (11), trattare.v (17)
Have_as_requirement	richiedere.v
Hear	avere_notizia.v, ascoltare.v (8), sentire.v (11), udire.v
Hindering	intralciare.v, frapporsi.v
Importance	stare_a_cuore.v
Import_export	importatore.n
Imprisonment	incarcerare.v

[2]Newly introduced frame. Our definition: "An *Agent* behaves towards an *Affected_party* in a certain way or *Manner*". Core FEs: *Agent*, *Affected_party* and *Manner*. Peripheral FEs: *Place* and *Time*.

Frame	Lexical unit
Inclusion	annettere.v
Institutionalization	ricoverare.v
Intentionally_act	intervenire.v (2)
Intentionally_create	fare.v (2)
Interrupt_process	rompere.v
Judgment	apprezzare.v (2), deplorare.v, gradire.v, lodare.v
Judgment_communication	accogliere_benevolmente.v, accusare.v (4), complimentarsi.v, condannare.v (4), contestare.v, criticare.v (2), denunciare.v, deprecare.v, rimproverare.v (2), prendere_posizione.v
Judgment_direct_address	ringraziare.v (19)
Killing	abbattere.v (3), assassinare.v (3), eliminare.v, eliminazione.n, giustiziare.v, macellare.v, massacrare.v, mietere.v, uccidere.v (11), uccisione.n
Likelihood	ipotizzabile.a, indubbiamente.adv, probabile.a, possibile.a, possibilità.n
Linguistic_meaning	significare.v
Locating	ritrovare.v
Manufacturing	produrre.v (6), produzione.n
Memory	ricordare.v, tenere_presente.v
Mental_property	lungimirante.a
Money	finanziamento.n
Motion	avanzare.v, inoltrarsi.v, muoversi.v, procedere.v, spingersi.v (3), spostarsi.v, andare.v (3)
Needing	necessitare.v
Notification_of_charges	accusare.v (5), sottoporre_a_giudizio.v
Opinion	avviso.n (3), credere.v (12), parere.n (3), pensare.v (5), ritenere.v (3), secondo.prep, sembrare.v (2), sentire.v
Perception_active	ascoltare.v (2), esaminare.v, guardare.v (8), occuparsi,v, osservare.v, rivolgersi.v, vedere.v (2)
Perception_experience	ascoltare.v, sentire.v (2)
Performing_arts	spettatore.n
Placing	aggiungere.v, frapporre.v, gettare.v, inserire.v (3), introdurre.v, nutrire.v, porre.v (2), riporre.v
Point_of_dispute	interrogazione.n, istanza.n, problema.n (2), questione.n (10)
Possession	possedere.v (3)
Posture	sedere.v
Predicting	prevedere.v (4)
Presence	presente.a
Process_resume	riprendere.v

Frame	**Lexical unit**
Purpose	destinare.v (2), obiettivo.n, perseguire.v, prefiggersi.v, scopo.n
Questioning	chiedere.v (3), domanda.n (21), domandare.v, interrogativo.n, interrogazione.n
Quitting_a_place	farsi_da_parte.v, ritirarsi.v (7)
Reasoning	dare_prova.v, dimostrare.v (2), mostrare.v (2), prova.n, provare.v (2)
Receiving	ricevere.v
Regard	apprezzamento.v, apprezzare.v (9), plaudere.v, riconoscere.v
Reliance	affidarsi.v, fare_affidamento.v
Remembering_information	dimenticare.v
Removing	disboscamento.n, eliminare.v (3), privare.v, revocare.v, ritirare.v (20), ritiro.n, togliere.v
Renunciation	rinunciare.v
Reporting	riportare.v
Request	chiedere.v (8), esigere.v (2), esortare.v, invitare.v, pretendere.v, richiesta.n (2)
Required_event	occorrere.v
Resolve_problem	affrontare.v (5), risolvere.v
Rewards_and_punishments	punire.v
Risky_situation	minacciare.v (3), minaccia.n (2), rischiare.v, pericolo.n
Run_risk	rischio.n
Scrutiny	analizzare.v, disamina.n, prendere_in_esame.v
Shoot_projectiles	sparare.v
Speak_on_topic	accennare.v, affrontare.v (2), passare_a.v, pronunciare.v (2), ricordare.v, soffermarsi.v, trattare.v (2)
Statement	affermare.v (4), ammettere.v, chiarire.v, confermare.v, consigliare.v, dire.v (26), dichiarare.v (3), evocare.v, *fare_un_esempio.v*, garantire.v, indicare.v, menzionare.v, osservazione.n (2), parlare.v (2), proporre.v (14), proposta.n (2), ribadire.v, ricordare,v, riferire.v, sostenere.v, spiegare.v (3), suggerire.v
State_continue	stare.v
Successful_action	fallire.v
Sufficiency	abbastanza.adv (3), bastare.v, eccessivo.a, sufficiente.a (3)
Supply	assicurare.v
Taking	assumere.v, levare.v
Taking_sides	approvare.v, puntare_su.v
Taking_time	lento.a
Telling	garantire.v, informare.v, precisare.v
Temporal_pattern	ritmo.n

Frame	Lexical unit
Temporary_stay	rimanere.v
Thwarting	impedire.v
Topic	idea.n, materia.n, oggetto.n, trattare.v (2)
Transfer	trasferimento.n
Travel	andare.v, viaggiare.v (3)
Traversing	travalicare.v
Trust	credere.v (2)
Undergoing	(essere_fatto_)oggetto.n, preda.n, vittima.n
Using	sfruttare.v
Verdict	condannare.v, condanna.n
Verification	confermare.v
Waiting	attendere.v
Willingness	disposto.a
Withdraw_from_participation	ritirarsi.v (2)
Working_on	seguire.v, lavorare.v
Total n. of frames: 157	Total n. of LUs: 412

B.2 MultiBerkeley

Frame	Lexical unit
Abounding_with	pieno.a
Absorb_heat	sfrigolare.v
Abundance	abbondare.v
Accompaniment	con.p
Accomplishment	portare_a_termine.v
Accoutrements	occhiali.n
Achieving_first	coniare.v
Active_substance	agente.n
Activity_done_state	finito.a
Activity_pause	finito.v
Addiction	dipendenza.n
Adding_up	sommare.v
Adducing	addurre.v
Adjusting	sistemare.v
Adopt_selection	adottare.v
Adorning	decorare.v
Age	anziano.a
Aggregate	serie.n
Altered_phase	surgelato.a
Amalgamation	unirsi.v
Amassing	accumulare.v

Frame	Lexical unit
Ambient_temperature	fresco.a
Amounting_to	ammontare.v
Appeal	appello.n
Apply_heat	friggere.v
Arraignment	imputazione.n
Arranging	sistemare.v
Arrest	arrestare.v
Arson	incendio.n
Artificiality	falso.a
Assistance	assistenza.n
Atonement	espiazione.n
Attaching	fissare.v
Attack	attaccare.v
Attempt	tentativo.n
Attempt_suasion	incitare.v
Attention	attenzione.n
Avoiding	evitare.v
Bail_setting	libertà.n
Be_in_agreement_on_assessment	accordo.n
Bearing_arms	armato.a
Becoming	diventare.v
Becoming_a_member	unirsi.v
Becoming_aware	notare.v
Becoming_detached	staccarsi.v
Behind_the_scenes	produttore.n
Being_attached	attaccare.v
Being_born	nascere.v
Being_dry	secco.a
Being_necessary	indispensabile.a
Being_obligated	obbligato.a
Being_obligatory	obbligatorio.a
Being_rotted	marcio.a
Being_wet	umido.a
Beyond_compare	impareggiabile.a
Biological_area	foresta.n
Biological_urge	stanco.a
Birth	nascita.n
Body_decoration	tatuaggio.n
Body_description_holistic	magro.a
Body_mark	cicatrice.n
Body_movement	calpestare.v
Bragging	vantarsi.v

Frame	Lexical unit
Breathing	respirare.v
Bringing	portare.v
Building	costruire.v
Building_subparts	stanza.n
Buildings	casa.n
Bungling	rovinare.v
Businesses	mulino.n
Calendric_unit	autunno.n
Candidness	sincero.a
Capability	capacità.n
Catastrophe	disgrazia.n
Categorization	interpretazione.n
Causation	provocare.v
Cause_change	modificare.v
Change_of_consistency	restringere.v
Cause_change_of_phase	sciogliere.v
Cause_change_of_position_on_a_scale	aumentare.v
Cause_expansion	estendere.v
Cause_fluidic_motion	aspergere.v
Cause_to_move_in_place	scuotere.v
Cause_temperature_change	scaldare.v
Cause_to_amalgamate	combinare.v
Cause_to_be_dry	asciugare.v
Cause_to_be_sharp	affilare.v
Cause_to_be_wet	inumidire.v
Cause_to_experience	terrorizzare.v
Cause_to_move_in_place	ruotare.v
Cause_to_start	generare.v
Cause_to_wake	svegliare.v
Certainty	dubbio.n
Change_direction	sterzare.v
Change_event_time	posticipare.v
Change_of_consistency	indurirsi.v
Change_of_phase	sciogliersi.v
Chatting	spettegolare.v
Clemency	clemenza.n
Clothing	pantaloni.n
Clothing_parts	manica.n
Cognitive_connection	correlato.a
Color	nero.a
Commerce_buy	acquistare.v
Commerce_collect	pagare.v

Frame	**Lexical unit**
Commitment	promettere.v
Committing_crime	commettere.v
Communicate_categorization	definire.v
Communication	comunicare.v
Communication_manner	urlare.v
Communication_means	telegrafare.v
Communication_noise	urlare.v
Competition	giocare.v
Complaining	lamentela.n
Congregating	incontrare.v
Connectors	catena.n
Containers	valigia.n
Contingency	dipendere.v
Cooking_creation	cucinare.v
Corporal_punishment	scudisciata.n
Corroding	arrugginire.v
Corroding_caused	corrodere.v
Cotheme	seguire.v
Court_examination	interrogare.v
Creating	creare.v
Criminal_investigation	investigare.v
Custom	tradizione.n
Cutting	tritare.v
Daring	osare.v
Dead_or_alive	vivo.a
Delivery	consegna.n
Departing	scomparire.v
Desirability	cattivo.a
Detaining	trattenere.v
Differentiation	differenziare.v
Difficulty	difficile.a
Dimension	profondo.a
Distinctiveness	tipico.a
Dodging	schivare.v
Dressing	vestire.v
Duplication	riprodurre.v
Duration	prolungato.a
Eclipse	nascondere.v
Emanating	irradiare.v
Emitting	secernere.v
Emotion_active	preoccuparsi.v
Emotion_directed	delusione.n

Frame	**Lexical unit**
Emotion_heat	ribollire.v
Employing	assumere.v
Emptying	disarmare.v
Encoding	formulare.v
Endangering	pericolo.n
Entering_of_plea	dichiarazione.n
Escaping	scappare.v
Estimating	stimare.v (2 occurrences)
Evading	eludere.v
Evaluative_comparison	eguagliare.v
Evoking	ricordare.v
Examination	test.n
Execution	esecuzione.n
Existence	esistere.v
Expansion	espandersi.v
Expectation	prevedere.v
Expensiveness	costare.v
Experience_bodily_harm	rompere.v
Expertise	esperto.n
Explaining_the_facts	spiegare.v
Exporting	esportazione.n
Extradition	estradizione.n
Facial_expression	sorriso.n
Fairness_evaluation	scorretto.a
Feeling	sentirsi.v
Feigning	fingere.v
Fighting_activity	rissa.n
Filling	pitturare.v
Fining	multare.v
Finish_competition	vincitore.n
Firing	licenziamento.n
First_rank	principale.a
Fleeing	fuggire.v
Fluidic_motion	scorrere.v
Food	minestra.n
Forging	contraffatto.a
Forgiveness	condonare.v
Forgoing	astenersi.v
Forming_relationships	corteggiare.v
Frequency	raro.a
Friction	grattare.v
Frugality	parsimonioso.a

Frame	**Lexical unit**
Gesture	segno.n
Gizmo	equipaggiamento.n
Grinding	macinare.v
Grooming	lavare.v
Ground_up	sminuzzato.a
Hair_configuration	ricciolo.n
Health_response	soggetto_a.a
Hiring	assumere.v
Hostile_encounter	lotta.n
Immobilization	ammanettare.v
Impact	colpire.v
Import_export	importazione.n
Inclination	propensione.n
Increment	aggiuntivo.a
Infrastructure	intrastruttura.n
Ingest_substance	fumare.v
Ingestion	sorseggiare.v
Ingredients	precursore.n
Inhibit_movement	confinato.a
Inspecting	esaminare.v
Instance	esemplare.n
Intentional_traversing	guadare.v
Intentionally_act	attività.n
Intercepting	intercettare.v
Intoxicants	tabacco.n
Intoxication	ubriaco.a
Invention	progettare.v
Judgment_communication	criticare.v
Judgment_direct_address	ringraziare.v
Jury_deliberation	discutere.v
Justifying	razionalizzare.v
Kidnapping	rapire.v
Killing	letale.a
Kinship	padre.n
Knot_creation	allacciare.v
Labeling	definire.v
Leadership	direttore.n
Legality	legittimo.a
Likelihood	potere.v
Linguistic_meaning	significato.n
Locale_by_use	laboratorio.n
Location_of_light	lucere.v

Frame	Lexical unit
Locative_relation	confinare.v
Make_agreement_on_action	accordo.n
Make_noise	urlare.v, strombazzare.v
Making_faces	sorridere.v
Manipulate_into_doing	abbindolare.v
Manipulation	afferrare.v
Manufacturing	fabbricare.v
Mass_motion	affollare.v
Measure_area	acro.n
Measure_duration	secolo.n
Measure_linear_extent	miglia.n
Measure_mass	tonnellata.n
Measure_volume	gallone.n
Membership	iscritto.n
Memorization	memorizzare.v
Mental_property	insensato.a
Morality_evaluation	malvagio.a
Motion_directional	cadere.v
Motion_noise	tintinnare.v
Moving_in_place	vibrare.v
Name_conferral	chiamare.v
Namesake	omonimo.a
Natural_features	lago.n
Objective_influence	influsso.n
Observable_bodyparts	occhio.n, barba.n
Offenses	stupro.n
Omen	presagio.n
Opinion	opinione.n
Part_edge	bordo.n
Part_inner_outer	mezzo.n
Part_ordered_segments	ripresa.n
Part_orientational	meridionale.a
Part_piece	pezzo.n
Part_whole	parte.n
Partiality	neutrale.a
Participation	partecipante.n
Path_shape	snodarsi.v
People	signora.n
People_by_age	ragazza.n
People_by_morality	delinquente.n
People_by_origin	straniero.n
People_by_religion	cristiano.n

Frame	**Lexical unit**
People_by_vocation	servitù .n
Perception_active	guardare.v
Perception_body	duolere.v
Perception_experience	vedere.v
Performers_and_roles	impersonare.v
Personal_relationship	moglie.n
Piracy	dirottare.v
Place_weight_on	enfasi.n
Political_locales	paese.n
Position_on_a_scale	elevato.a
Possession	proprietà.n
Practice	provare.v
Praiseworthiness	ammirevole.a
Precipitation	piovere.v
Predicting	predire.v
Preserving	affumicare.v
Prevarication	distorsione.n
Prison	prigione.n
Process	processo.n
Process_continue	durata.n
Processing_materials	arricchimento.n
Project	programma.n
Proliferating_in_number	proliferazione.n
Purpose	obiettivo.n
Quantity	mucchio.n
Quarreling	lite.n
Questioning	chiedere.v
Quitting	dimissione.n
Range	raggio.n
Rape	violentare.v
Reading	leggere.v
Reasoning	dimostrare.v
Reassuring	rassicurare.v
Recovery	guarire.v
Referring_by_name	chiamare.v
Reforming_a_system	ristrutturare.v
Relative_time	precedente.a
Reliance	affidamento.n
Remainder	resto.n
Remembering_experience	dimenticare.v
Remembering_information	dimenticarsi.v
Remembering_to_do	dimenticarsi.v

Frame	Lexical unit
Render_nonfunctional	disabilitare.v
Renting	affittare.v
Renting_out	affittare.v
Research	ricerca.n
Reshaping	piegare.v
Residence	abitare.v
Response	risposta.n
Reveal_secret	confessare.v
Revenge	contraccambiare.v
Rewards_and_punishments	punire.v
Ride_vehicle	volare.v
Rite	rito.n
Roadways	strada.n
Robbery	rapinare.v
Rope_manipulation	allacciare.v
Rotting	marcire.v
Safe_situation	rischio.n
Scouring	rovistare.v
Seeking_to_achieve	cercare.v
Self_motion	nuotare.v
Sensation	suono.n, odore.n
Sentencing	sentenza.n
Separation	dividere.v
Setting_fire	incendiare.v
Severity_of_offense	incriminabile.a
Shapes	tratto.n
Sharpness	affilato.a
Sign	segno.n
Sign_agreement	firmare.v
Silencing	zittire.v
Similarity	differenza.n
Sleep	dormire.v
Smuggling	smerciare.v
Soaking	inzuppare.v
Sociability	timido.a
Social_event	festival.n
Social_interaction_evaluation	cordiale.a
Sound_movement	riverberarsi
Sounds	tintinnio.n
Source_of_getting	fonte.n
Spanning_values	variare.v
Stinginess	generoso.a

Frame	**Lexical unit**
Strictness	indulgente.a
Suasion	persuadere.v
Submitting_documents	presentare.v
Success_or_failure	riuscire.v
Successfully_communicate_message	esprimere.v
Surpassing	superare.v
Suspiciousness	sospetto.a
Talking_into	istigare.v
Telling	dire.v
Text	poesia.n
Theft	fregare.v
Time_vector	fa.adv
Topic	argomento.n
Toxic_substance	tossico.a
Transfer	trasferimento.n
Translating	tradurre.v
Trial	processo.n
Try_defendant	processare.v
Type	varietà.n
Unattributed_information	apparentemente.adv
Undressing	sfilarsi.v
Vehicle	auto.n
Volubility	zitto.a
Waiting	aspettare.v
Waking_up	svegliarsi.v
Wealthiness	ricco.a
Weapon	pistola.n
Wearing	indossare.v
Weather	tempesta.n
Word_relations	sinonimo.n
Total n. of frames: 387	Total n. of LUs: 391

Bibliography

Atkins, S. (2008). *Then and Now: Competence and Performance in 35 Years of Lexicography*, chapter 17, pages 247–272. Practical Lexicography: a reader. Oxford University Press, Oxford.

Atkins, S. (2009). Using the FrameNet approach in a professional lexicography project (the DANTE database). In *Proceedings of the FrameNet Masterclass and Workshop, co-located with the 8^{th} International Workshop on Treebanks and Linguistic Theories (TLT8)*, Milan, Italy.

Austin, J. L. (1962). *How to do Things with Words*. Clarendon Press, Oxford.

Baker, C. F. (2008). FrameNet, present and future. In *Proceedings of the First International Conference on Global Interoperability for Language Resources*, pages 12–17, Hong Kong, China.

Baker, C. F., Ellsworth, M., and Erk, K. (2007). SemEval-2007 Task 10: Frame Semantic Structure Extraction. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 99–104, Prague, CZ.

Baker, C. F. and Fellbaum, C. (2009). WordNet and FrameNet as Complementary Resources for Annotation. In *Proceedings of the Third Linguistic Annotation Workshop*, pages 125–129, Singapore.

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The Berkeley FrameNet Project. In *Proceedings of the 36th ACL Meeting and 17th ICCL Conference*. Morgan Kaufmann.

Baker, C. F. and Sato, H. (2003). The FrameNet data and software. In *ACL '03: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, pages 161–164, Morristown, NJ, USA.

Basili, R., DeCao, D., Croce, D., Coppola, B., and Moschitti, A. (2009). Cross-language frame semantics transfer in bilingual corpora. In *Proceedings of CICLing*. Springer-Verlag.

Basili, R. and Zanzotto, F. M. (2002). Parsing Engineering and Empirical Robustness. *Journal of Natural Language Engineering*.

Bechet, F., Riccardi, G., and Hakkani-Tur, D. (2004). Mining Spoken dialogue Corpora for system Evaluation and Modeling. In Lin, D. and Wu, D., editors, *Proceedings of EMNLP 2004*, pages 134–141, Barcelona, Spain.

Bentivogli, L., Forner, P., Magnini, B., and Pianta, E. (2004). Revising WordNet Domains Hierarchy: Semantics, Coverage and Balancing. In *Proceedings of the COLING Workshop o Multilingual Linguistic Resources*, pages 101–108, Geneva, Switzerland.

Bentivogli, L. and Pianta, E. (2005). Exploiting Parallel Texts in the Creation of Multilingual Semantically Annotated Resources: The MultiSemCor Corpus. *Natural Language Engineering, Special Issue on Parallel Texts*, 11(03):247–261.

Bisazza, A., Dinarelli, M., Quarteroni, S., Tonelli, S., Moschitti, A., and Riccardi, G. (2008). Semantic Annotations for Conversational Speech: from Speech Transcriptions to Predicate Argument Structures. In *Proceedings of the 2nd IEEE Workshop on Spoken Language Technology*, Goa, India (to appear).

Boas, H. C. (2002). Bilingual framenet dictionaries for machine translation. In *Proceedings of the 3^{rd} International Conference on Languages Resources and Evaluations (LREC 2002)*, volume IV, pages 1364–1371, Las Palmas, Spain.

Boas, H. C. (2005a). *From Theory to Practice: Frame Semantics and the Design of FrameNet*, pages 129–160. Semantik im Lexikon, Tübingen edition.

Boas, H. C. (2005b). Semantic Frames as Interlingual Representations for Multilingual Lexical Databases. *International Journal of Lexicography*, 18(4):445–478.

Boas, H. C. (2006). *A frame-semantic approach to identifying syntactically relevant elements of meaning*, pages 119–149. Contrastive Studies and Valency. Studies in Honor of Hans Ulrich Boas.

Borin, L., Dannélls, D., Forsberg, M., Gronstaj, M. T., and Kokkinakis, D. (2009). Thinking Green: Toward Swedish FrameNet++. In *Proceedings of the FrameNet Masterclass and Workshop, co-located with the 8^{th} International Workshop on Treebanks and Linguistic Theories (TLT8)*, Milan, Italy.

Bos, J. and Nissim, M. (2008). Combining Discourse Representation Theory with FrameNet. In Favretti, R. R., editor, *Frames, Corpora and Knowledge Representation*, chapter 9, pages 169–183. Bononia University Press, Bologna, Italy.

Bosco, C. (2007). Multiple-step treebank conversion: from dependency to Penn format. In *Proceedings of the ACL Workshop on Linguistic Annotation (LAW)*, Prague, Czech Republic.

Brants, S., Dipper, S., Hansen, S., Lezius, W., and Smith, G. (2002). The TIGER Treebank. In *Proceedings of the 1st Workshop on Treebanks and Linguistic Theories*, Sozopol, Bulgaria.

Burchardt, A., Erk, K., and Frank, A. (2005). A WordNet detour to FrameNet. In Fisseni, B., Schmitz, H., Schröder, B., and Wagner, P., editors, *Sprachtechnologie, mobile Kommunikation und lingistische Resourcen*, Frankfurt am Main, Germany. Peter Lang.

Burchardt, A., Erk, K., Frank, A., Kowalski, A., Padó, S., and Pinkal, M. (2006). SALTO - A Versatile Multi-Level Annotation Tool. In *Proceedings of the 5th Language Resources and Evaluation Conference (LREC 06)*, pages 517–520, Genoa Italy.

Burchardt, A., Erk, K., Frank, A., Kowalski, A., Padó, S., and Pinkal, M. (2009a). FrameNet for the semantic analysis of German: Annotation, representation and automation. In Boas, H. C., editor, *Multilingual FrameNets in Computational Lexicography: Methods and Applications*, volume 200 of *Trends in Linguistics*, pages 209–244. Mouton de Gruyter.

Burchardt, A. and Frank, A. (2006). Approximating Textual Entailment with LFG and FrameNet Frames. In *Proceedings of the 2nd PASCAL RTE Workshop*, Venice, Italy.

Burchardt, A. and Pennacchiotti, M. (2008). FATE: a FrameNet-Annotated Corpus for Textual Entailment. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco.

Burchardt, A., Pennacchiotti, M., Thater, S., and Pinkal, M. (2009b). Assessing the impact of frame semantics on textual entailment. *Natural Language Engineering, Special Issue on Textual Entailment*, 15(04):527–550.

Burchardt, A., Reiter, N., Thater, S., and Frank, A. (2007). A semantic approach to textual entailment: System evaluation and task analysis. In *Proceedings of Pascal RTE-3 Challenge*, Prague, CZ.

Chang, C.-C. and Lin, C.-J. (2001). *LIBSVM: a library for support vector machines.* Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Chen, B. and Fung, P. (2004). Automatic Construction of an English-Chinese Bilingual FrameNet. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLT/NAACL)*, Boston, USA.

Chomsky, N. (1957). *Syntactic Structures.* Mouton.

Codina, J., Pianta, E., Vrochidis, S., and Papadopoulos, S. (2008). Integration of Semantic, Metadata and Image search engines with a text search engine for patent retrieval. In *Proceedings of SemSearch.*

Collins, M. (1999). *Head-Driven Statistical Models for Natural Language Parsing.* PhD thesis, University of Pennsylvania.

Coppola, B., Moschitti, A., and Riccardi, G. (2009). Shallow Semantic Parsing for Spoken Language Understanding. In *Proceedings of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT) 2009 conference*, Boulder, CO, USA.

Coppola, B., Moschitti, A., Tonelli, S., and Riccardi, G. (2008). Automatic FrameNet-Based Annotation of Conversational Speech. In *Proceedings of the 2nd IEEE Workshop on Spoken Language Technology*, Goa, India.

Corazza, A., Lavelli, A., and Satta, G. (2007). Analisi Sintattica-Statistica basata su Costituenti. *Intelligenza Artificiale - Numero speciale su Strumenti per l'elaborazione del linguaggio naturale per l'Italiano*, 4(2):38–39.

Core, M. G. and Allen, J. F. (1997). Coding dialogs with the DAMSL annotation scheme. In *Proceedings of the AAAI Fall Symposium on Communicative Actions in Humans and Machines.*

Crespo, M. and Buitelaar, P. (2008). Domain-specific English-to-Spanish Translation of FrameNet. In *Proceedings of the 6th Language Resources and Evaluation Conference (LREC 2008)*, Marrakech, Morocco.

Csomai, A. and Mihalcea, R. (2008). Linking Documents to Encyclopedic Knowledge. *IEEE Intelligent Systems, special issue on "Natural Language Processing for the Web"*, 23(5):34–41.

Cucerzan, S. (2007). Large-scale named entity disambiguation based on Wikipedia data. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 708–716, Prague, Czech Republic.

Dagan, I. and Glickman, O. (2004). Probabilistic textual entailment: Generic applied modeling of language variability. In *Proceedings of the PASCAL Workshop on Learning Methods for Text Understanding and Mining*, Grenoble, France.

Dagan, I., Glickman, O., and Magnini, B. (2006). The PASCAL Recognising Textual Entailment Challenge. In Quiñonero-Candela, J., Dagan, I., Magnini, B., and D'Alché-Buc, F., editors, *Machine Learning Challenges*, volume 3944, pages 177–190. Springer Verlag.

DeCao, D., Croce, D., Pennacchiotti, M., and Basili, R. (2008). Combining Word Sense and Usage for modeling Frame Semantics. In *Proceedings of STEP 2008*, Venice, Italy.

Delmonte, R. (2008). Inducing frames in the Italian lexicon. In Favretti, R. R., editor, *Frames, Corpora and Knowledge Representation.* Bononia University Press, Bologna, Italy.

Delmonte, R., Tonelli, S., Boniforti, M. A., and Bristot, A. (2005). Venses - A Linguistically-Based System for Semantic Evaluation. In Quiñonero-Candela, J., Dagan, I., Magnini, B., and D'Alché-Buc, F., editors, *First PASCAL Machine Learning Challenges Workshop*, Southampton, UK.

Dinarelli, M., Quarteroni, S., Tonelli, S., Moschitti, A., and Riccardi, G. (2009). Annotating Spoken Dialogs: from Speech Segments to Dialog Acts and Frame Semantics. In *Proceedings of the EACL Workshop on Semantic Representation of Spoken Language (SRSL)*, Athens, Greece.

Dolbey, A., Ellsworth, M., and Scheffczyk, J. (2006). BioFrameNet: A Domain-Specific FrameNet Extension with Links to Biomedical Ontologies. In Bodenreider, O., editor, *Proceedings of KR-MED*, pages 87–94.

Ellsworth, M. and Janin, A. (2007). Mutaphrase: Paraphrasing with FrameNet. In *Proceedings of the Workshop on Textual Entailment and Paraphrasing*, Prague, Czech Republic.

Erk, K. (2004). Frame assignment as Word Sense Disambiguation. In *Proceedings of IWCS-6*, Tilburg, NL.

Erk, K. and Padó, S. (2004). A powerful and versatile XML Format for representing role-semantic annotation. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 04)*, Lisbon, Portugal.

Erk, K. and Padó, S. (2006). Shalmaneser - a flexible toolbox for semantic role assignment. In ELRA, editor, *Proceedings of the 5th Language Resources and Evaluation Conference (LREC 06)*, Genoa Italy.

Fellbaum, C. (1998). *WordNet. An Electronic Lexical Database.* MIT Press.

Fillmore, C. J. (1968). The Case for Case. In Bach, E. and Harms, R. T., editors, *Universals in Linguistic Theory*, pages 1–88. Holt, Reinhart and Winston, New York.

Fillmore, C. J. (1976). Frame Semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language*, pages 20–32. Blackwell Publishing.

Fillmore, C. J. (1977). Topics in lexical semantics. In Cole, R. W., editor, *Current Issues in Linguistic Theory*. Indiana University Press.

Fillmore, C. J. (1985). Syntactic Intrusions and the Notion of Grammatical Constructions. In *Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistic Society*, volume 11.

Fillmore, C. J. (1988). The Mechanisms of "Construction Grammar". In *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistic Society*, pages 35–55, Berkeley, California.

Fillmore, C. J. and Atkins, B. T. (1992). Toward a Frame-Based Lexicon: The Semantics of RISK and its Neighbors. In Lehrer, A. and Kittay, E., editors, *Frames, fields and Contrast: New Essays in Semantics and Lexical Organization*, pages 75–102. Lawrence Erlbaum Associates.

Fillmore, C. J., Baker, C. F., and Sato, H. (2004a). FrameNet as a 'Net'. In *Proceedings of the 4th International Conference on Languagw Resources and Evaluation (LREC 04)*, pages 1091–1094, Lisbon, Portugal.

Fillmore, C. J., Ruppenhofer, J., and Baker, C. F. (2004b). Framenet and representing the link between semantic and syntactic relations. In Huang, C. and Lenders, W., editors, *Frontiers in Linguistics*, volume I of *Language and Linguistics Monograph Series*, pages 19–59. Academia Sinica.

Fürstenau, H. (2008). Enriching frame semantic resources with dependency graphs. In *Proceedings of the 6th Language Resources and Evaluation Conference (LREC 2008)*, Marrakech, Morocco.

Giuglea, A.-M. and Moschitti, A. (2006). Semantic role labeling via FrameNet, VerbNet and PropBank. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual ACL meeting*, pages 929–936, Morristown, US.

Giuliano, C., Gliozzo, A. M., and Strapparava, C. (2009). Kernel Methods for Minimally Supervised WSD. *Computational Linguistics*, 35(4):513–528.

Gliozzo, A., Giuliano, C., and Strapparava, C. (2005). Domain kernels for word sense disambiguation. In *Proceedings of the 43^{rd} annual meeting of the Association for Computational Linguistics (ACL-05)*, pages 403–410, Ann Arbor, Michigan.

Goldberg, A. (1992). *Constructions. A Construction Grammar Approach to Argument Structure*. Cognitive Theory of Language and Culture. The University of Chicago Press.

Honnibal, M. and Hawker, T. (2005). Identifying FrameNet frames for verbs from a real-text corpus. In *Proceedings of Australasian Language Technology Workshop 2005*.

Jelinek, F. (2005). Some of my Best Friends are Linguists. *Language Resources and Evaluation*, 39(1):25–34.

Joachims, T. (1999). Making large-scale support vector machine learning practical. In Schölkopf, B., Burges, C. J. C., and Smola, A. J., editors, *Advances in kernel methods: support vector learning*, pages 169–184. MIT Press, Cambridge, MA, USA.

Johansson, R. and Nugues, P. (2006). A FrameNet-based Semantic Role Labeler for Swedish. In *Proceedings of Coling/ACL 2006*.

Johansson, R. and Nugues, P. (2007a). Syntactic Representations Considered for Frame-Semantic Analysis. In *Proceedings of the Sixth International Workshop on Treebanks and Linguistic Theories*, Bergen, Norway.

Johansson, R. and Nugues, P. (2007b). Using WordNet to extend FrameNet coverage. In *Proc. of the Workshop on Building Frame-semantic Resources for Scandinavian and Baltic Languages, at NODALIDA*, Tartu.

Kaisser, M. and Webber, B. (2007). Question Answering based on Semantic Roles. In *Proceedings of the ACL Workshop on Deep Linguistic Processing*, Prague, Czech Republic.

Kay, P. and Fillmore, C. J. (1999). Grammatical constructions and linguistic generalizations: The what's X doing Y? construction. *Language*, 75:1–33.

Kim, S.-M. and Hovy, E. (2006). Extracting Opinions, Opinion Holders and Topics Expressed in Online News Media Text. In *Proceedings of the ACL Workshop on Sentiment and Subjectivity in Text*, pages 1–8, Sydney, Australia.

Koehn, P. (2005). Europarl: A Parallel Corpus for Statistical Machine Translation. In *Proceedings of the Machine Translation Summit X*, pages 79–86.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., and Herbst, E. (2007). Moses: Open source tool for statistical machine translation. In *Proceedings of ACL 2007*, Prague, CZ.

Landes, S., Leacock, C., and Tengi, R. I. (1998). Building semantic concordances. In Fellbaum, C., editor, *WordNet: An Electronic Lexical Database*. The MIT Press.

Laparra, E. and Rigau, G. (2009). Integrating WordNet and FrameNet using a knowledge-based Word Sense Disambiguation algorithm. In *Proceedings of Recent Advances in Natural Language Processing*, Borovets, Bulgaria.

Litowski, K. C. (2004). Senseval-3 task Automatic Labeling of Semantic Roles. In *Proceedings of SENSEVAL-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, pages 9–12, Barcelona, Spain.

Lönneker-Rodman, B. (2007). Multilinguality and FrameNet. Technical Report TR-07-001, International Computer Science Institute, Berkeley, California.

Lönneker-Rodman, B. and Baker, C. F. (2009). The FrameNet model and its applications. *Natural Language Engineering*, 15(03):415–453.

Lupu, M., Trandabat, D., and Husarciuc, M. (2005). A Romanian SemCor aligned to the English and Italian MultiSemCor. In *Proceedings of the ROMANCE FrameNet Workshop and Kick-off Meeting*, Cluj-Napoca, Romania.

Magnini, B. and Cavaglià, G. (2000). Integrating Subject Field Codes into WordNet. In *Proceedings of LREC 2000*, pages 1413–1418, Athens, Greece.

McConville, M. and Dzikovska, M. O. (2005). Extracting a verb lexicon for deep parsing with FrameNet. In *Proceedings of the 5th Workshop on Important Unresolved Matters*, pages 112–119, Ann Arbor, Michigan.

Mengel, A. and Lezius, W. (2000). An XML-based encoding format for syntactically annotated corpora. In *Proceedings of LREC-2000, Second International Conference on Language Resources and Evaluation*, Athens, Greece.

Mihalcea, R. (2007). Using Wikipedia for Automatic Word Sense Disambiguation. In *Proceedings of NAACL-HLT*, pages 196–203.

Milne, D. and Witten, I. H. (2008). Learning to link with Wikipedia. In *CIKM '08: Proceedings of the 17th ACM conference on Information and knowledge management*, pages 509–518, NY, USA. ACM.

Narayanan, S. and Harabagiu, S. M. (2004). Question Answering Based on Semantic Structure. In *Proceedings of the 20th international conference on Computational Linguistics (COLING04)*, pages 184–191, Geneva, Switzerland.

Och, F. J. and Ney, H. (2003). A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1):19–51.

Ohara, K. H. (2008). Lexicon, Grammar, and Multilinguality in the Japanese FrameNet. In *Proceedings of the 6th Language Resources and Evaluation Conference (LREC 2008)*, pages 3264–3268, Marrakech, Morocco.

Ohara, K. H., Fujii, S., Saito, H., Ishizaki, S., Ohori, T., and Suzuki, R. (2003). The Japanese FrameNet Project: A Preliminary Report. In *Proceedings of Pacific Association for Computational Linguistics (PACLING03)*, pages 249–254, Halifax, Canada.

Östman, J.-O. and Fried, M. (2004). Historical and intellectual background to construction grammar. In Fried, M. and Östman, J.-O., editors, *Construction grammar in a cross-language perspective*. John Benjamins.

Padó, S. (2007). *Cross-Lingual Annotation Projection Models for Role-Semantic Information*. PhD thesis, Universität des Saarlandes.

Padó, S. and Lapata, M. (2005). Cross-linguistic Projection of Role-Semantic Information. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 859–866, Vancouver, British Columbia, Canada.

Padó, S. and Lapata, M. (2009). Cross-lingual Annotation Projection of Semantic Roles. *Journal of Artificial Intelligence Research*, 36:307–340.

Padó, S. and Pitel, G. (2007). Annotation précise du français en sémantique de roles par projection cross-linguistique. In *Proceedings of TALN-07*, Toulouse, France.

Palmer, M., Gildea, D., and Kingsbury, P. (2005). The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31.

Petruck, M. R. (2009). Typological considerations in constructing a Hebrew FrameNet. In Boas, H. C., editor, *Multilingual FrameNets in Computational Lexicography: Methods and Applications*, Trends in Linguistics, pages 183–205. Mouton de Gruyter.

Pianta, E. and Bentivogli, L. (2004). KNOwledge Intensive Word Alignment with KNOWA. In *Proceedings of Coling 2004*, pages 1086 – 1092.

Pianta, E., Bentivogli, L., and Girardi, C. (2002). MultiWordNet: developing an aligned multilingual database. In *First International Conference on Global WordNet*, pages 292–302, Mysore, India.

Pianta, E., Girardi, C., and Zanoli, R. (2008). The TextPro tool suite. In *Proceedings of the 6th Language Resources and Evaluation Conference (LREC 2008)*, Marrakech, Morocco.

Pianta, E. and Zanoli, R. (2007). TagPro: a system for Italian Pos Tagging based on SVM. *Intelligenza Artificiale - Numero speciale su Strumenti per l'elaborazione del linguaggio naturale per l'Italiano*, 4(2):8–9.

Pitel, G. (2009). Cross-lingual labeling of semantic predicates and roles. In Boas, H. C., editor, *Multilingual FrameNets in Computational Lexicography: Methods and Applications*, Trends in Linguistics, pages 245–284. Mouton de Gruyter.

Ponzetto, S. P. and Strube, M. (2006). Exploiting semantic role labeling, WordNet and Wikipedia for coreference resolution. In *Proceedings of the Human Language*

Technology Conference of the North American Chapter of the Association of Computational Linguistics, pages 192–199.

Postolache, O., Cristea, D., and Orasan, C. (2006). Transferring coreference chains through word alignment. In *Proceedings of the 5th Language Resources and Evaluation Conference (LREC 06)*, Genoa, Italy.

Quarteroni, S., Riccardi, G., Varges, S., and Bisazza, A. (2008). An Open-Domain Dialog Act Taxonomy. Technical Report 08-032, DISI, University of Trento.

Ruppenhofer, J., Ellsworth, M., Petruck, M. R., Johnson, C. R., and Scheffczyk, J. (2006). *FrameNet II: Extended Theory and Practice.* Available at http://framenet.icsi.berkeley.edu/book/book.html.

Ruppenhofer, J., Sporleder, C., Morante, R., Baker, C. F., and Palmer, M. (2009). SemEval-2010 Task 10: Linking Events and Their Participants in Discourse. In *Proceedings of the NAACL-HLT Workshop on Semantic Evaluations: Recent Achievements and Future Directions*, pages 106–111, Boulder, CO, USA.

Sato, H. (2003). FrameSQL: A Software Tool for FrameNet. In *Proceedings of ASIALEX 03*, pages 251–258, Tokyo, Japan. Asian Association of Lexicography.

Sato, H. (2008). New Functions of FrameSQL for Multilingual FrameNets. In *Proceedings of the 6th Language Resources and Evaluation Conference (LREC 2008)*, pages 758–762, Marrakech, Morocco.

Shen, D. and Lapata, M. (2007). Using Semantic Roles to Improve Question Answering. In *Proceedings of EMNLP and CONLL*, pages 12–21, Prague, CZ.

Shi, L. and Mihalcea, R. (2004). Open Text Semantic Parsing Using FrameNet and WordNet. In *Proceedings of HLT-NAACL 2004*.

Shi, L. and Mihalcea, R. (2005). Putting Pieces Together: Combining FrameNet, VerbNet and WordNet for Robust Semantic Parsing. In *Proceedings of CICLing 2005*, pages 100–111. Springer.

Strapparava, C. and Valitutti, A. (2004). WordNet-Affect: an affective extension of WordNet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 04)*, pages 1083–1086, Lisbon, Portugal.

Subirats, C. (2009). Spanish FrameNet: A frame-semantic analysis of the Spanish lexicon. In Boas, H. C., editor, *Multilingual FrameNets in Computational Lexicography: Methods and Applications*, Trends in Linguistics, pages 135–162. Mouton de Gruyter.

Subirats-Rüggeberg, C. and Petruck, M. R. L. (2003). Surprise: Spanish FrameNet! In Hajicova, E., Kotesovcova, A., and Mirovsky, J., editors, *Proceedings of CIL 17*, Prague, CZ. Matfyzpress.

Tonelli, S. and Giuliano, C. (2009). Wikipedia as frame information repository. In *Proceedings of Empirical Methods in Natural Language Processing*, Singapore.

Tonelli, S. and Pianta, E. (2008). Frame Information Transfer from English to Italian. In *Proceedings of LREC 2008*, Marrakech, Morocco.

Tonelli, S. and Pianta, E. (2009a). A Novel Approach to Mapping FrameNet lexical units to WordNet synsets. In *Proceedings of IWCS-8*, Tilburg, NL.

Tonelli, S. and Pianta, E. (2009b). Three issues in cross-language frame information transfer. In *Proceedings of Recent Advances in Natural Language Processing*, Borovets, Bulgaria.

Tonelli, S. and Pighin, D. (2009). New features for FrameNet - WordNet Mapping. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, Boulder, CO, USA.

Traum, D. (1996). Conversational agency: The TRAINS-93 dialogue manager. In *Proceedings of the Twente Workshop on Language Technology 11:Dialogue Management in Natural Language Systems*, pages 1–11.

Vapnik, V. N. (1998). *Statistical Learning Theory*. Wiley-Interscience.

Venturi, G., Lenci, A., Montemagni, S., Vecchi, E. M., Sagri, M. T., Tiscornia, D., and Agnoloni, T. (2009). Towards a FrameNet Resource for the Legal Domain. In *Proceedings of the III Workshop on Legal Ontologies and Artificial Intelligence Techniques (LOAIT '09)*, Barcelona, Spain.

Vivi, N. and Strube, M. (2009). Combining Collocations, Lexical and Encyclopedic Knowledge for Metonymy Resolution. In *Proceedings of Empirical Methods in Natural Language Processing*, pages 1219–1224, Suntec, Singapore.

Vossen, P., editor (1998). *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Springer.

Wu, S. and McClean, S. (2007). Several methods of ranking retrieval systems with partial relevance judgment. In *Proceedings of the Second IEEE International Conference on Digital Information Management (ICDIM)*, pages 13–18, Lyon, France.

Yarowsky, D. and Ngai, G. (2001). Inducing multilingual POS tagger and NP bracketers via robust projection across aligned corpora. In *Proceedings of the 2^{nd} Annual Meeting of the North American Chapter of the Association of Computational Linguistics*, pages 200–207, Pittsburgh, PA.