eman ta zabal zazu

Universidad del País Vasco — Euskal Herriko Unibertsitatea

Implicit Semantic Roles in Discourse

Egoitz Laparra Martín

German Rigau Claramunt

Director

PhD Thesis

informatika fakultatea — facultad de informática

Donostia, January, 2015

This work has been funded by a research grant (BES-2007-15271) related to the Spanish Government Project KNOW (TIN2006-15049-C03-01).

Acknowledgements

First of all, I would like to thank my advisor, Dr. German Rigau, for introducing me in the world of research on Computational Linguistics. Without his wise guidance I would not have been able to carry this work out.

I also want to thank to all my partners in the IXA group for supporting me in many ways. My group has provided me with all the resources that I have needed for my work and they have never hesitated to give me a helping hand. I hope to continue this collaboration in many future projects.

I would like to thank Dr. Piek Vossen and his group in the Vrije Universiteit Amsterdam for their hospitality during my stay in this centre in 2011. Their interest in my research was really encouraging. I also must thank Dr. Martha Palmer from the University of Colorado for reviewing this document and providing me with valuable comments and suggestions.

But, of course, I could not have maintained my eagerness and enthusiasm during the last years without the support of my enviroment in the *real life*, my friends, my flat-mates, and above all my parents, Iñaki and Pepi. Thank you for just everything.

Contents

List of Figures

List of Tables

INTRODUCTION

CHAPTER 1

Introduction

This introductory chapter of the document is organized as follows. Section 1.1 describes our research framework and presents some concepts to introduce the reader in the notion of Implicit Semantic Role Labelling (ISRL). After that, section 1.2 describes the main goals of our research. Section 1.3 presents the main contributions of our research related with their corresponding chapters. Finally, we describe the organization of the rest of the document in section 1.4.

1.1 Research framework

Natural Language Processing is a sub-field of Artificial Intelligence that faces the task of analysing and processing automatically natural human language. The development of systems capable of understanding textual expressions has been a challenge for linguistics and computer scientists for many years. As many other areas in Artificial Intelligence, first systems obtained very successful performances working in very closed domains. However, these early systems had to address really challenging problems when applied to real world tasks. Nowadays, computational linguistics is still a main focus of interest for scientists due to their relevance and intrinsic difficulties.

One of the main problems that researchers must face is that the texts need to be accurately analysed at many distinct levels for a full understanding. Furthermore, each of these levels are affected by ambiguous expressions that can not be interpreted in isolation. Solving the inherent ambiguity and vagueness of natural language is the main goal of several tasks in natural language processing.

For example, the correct *semantic interpretation* of the text demands for capturing the meaning of each word according to their context. This process is called Word Sense Disambiguation (WSD) (Agirre and Edmonds, 2007) and is the task of matching each word with its corresponding word sense in a lexical knowledge base, like WordNet (Fellbaum, 1998). This semantic analysis can be performed over any type of word, as nouns, verbs or adjectives, and also over named entities. In the later case, the task, called Named Entity Recognition (NER) (Nadeau and Sekine, 2007), focuses in labelling the entities with general semantic categories like *person*, *organization* or *place*. However the semantic interpretation of a sentence do not only depends on the meaning of the words. Semantic Role Labelling (SRL) (Gildea and Jurafsky, 2000, 2002) tries to discover the predicates and its semantic roles in a sentence. In other words, *who* make *what* in a sentence. Like WSD, it is also to label each element of the sentence with knowledge taken from a semantic source, like FrameNet (Baker et al., 1998), PropBank (Palmer et al., 2005) or NomBank (Meyers et al., 2004), that describes predicate structures including roles as *Agent*, *Patient* or *Location*.

Previous tasks limit their scope to expressions occurring within the same sentence boundaries, but full understanding of the discourse requires relating elements occurring in different parts of the document. This process requires to perform a *pragmatic analysis*. For instance, many times some components of the text can be only interpreted by matching those mentions that refer to the same entity or event. The task of capturing such mentions for every entity is called Coreference Resolution (CR) (Poesio et al., 2011) and, depending on the approach, can face both nominal entities and events (Humphreys et al., 1997). The clusters resulting of merging coreferent mentions gives a complete idea about all the events, and their participants, that appear in the discourse. Discovering and labelling the relationships between the different events have been the focus of a number of research lines. For instance, a very important task for language understanding is sorting the events in order to reconstruct narrative chains or story lines. This task can be approached by relating each pairs of events by temporal labels (Schilder et al., 2007), as *Before*, *After* or *Simultaneous*. This job can be also carried out by obtaining the complete *narrative schema* (Chambers and Jurafsky, 2009) of the events that share the same participants in the discourse.

But temporality is not the only type of event relations expressed in a text. For example, common narrations include descriptions about which events causes other events, which events are parts of more complex events, etc. In the literature, there exists some theories that try to formalize the relations that occur at the discourse level (Webber et al., 2012) and are the base for systems that guess this type of complex pragmatic parsing. However, although many advances have been performed for all these task independently, the join interpretation of the human language is far to be completely reached, and there are a bunch of aspects that require further investigation.

In short, natural language understanding demands deep semantic and pragmatic analysis to infer the ambiguous and inherent knowledge that remains inside the explicit realizations of the elements of the discourse. We can say that the correct interpretation of a written text is the correct labelling of the *actions* and the *actors*, and capturing the relations that connect all of them. For this, many analysis must be performed both at sentence and document level. This process should result not only in representing the explicit information but also in discovering the implicit information denoted by the text.

1.2 Main goals

As said in the previous section, even a joint application of a set of semantic and pragmatic analysis can be unable to provide a full interpretation of the text because there remains a great amount of information that lies behind the explicit realizations of the text. Specifically, Semantic Role Labelling (SRL) systems frequently returns incomplete role structures. This fact can give the misconception that the text describes insufficient information about the events described. However, human reader can overcome this apparent lack of information by inferring and recovering the missing roles from the discourse. Those roles, whose fillers do not share any syntactic relation with their predicates, are called **Implicit Semantic Roles** and according to Palmer et al. (1986) can be considered as a special case of anaphora.

[arg_0 The network] had been expected to have [np **losses**] [arg_1 of as much as $20 million] [$arg_3$ on baseball this year]. It isn't clear how much those [np **losses**] may widen because of the short Series.

Table 1.1: Traditional SRL misses the roles of the predicate **losses**.

Quest Medical Inc said it adopted [arg_1 a shareholders' rights] [np **plan**] in which rights to purchase shares of common stock will be distributed as a dividend to shareholders of record as of Oct 23.

Table 1.2: The arg_0 of the predicate **plan** is missed by traditional SRL systems.

In consequence, works that have faced the task of **Implicit Semantic Roles Labelling** (ISRL) combine SRL and entity Coreference Resolution (CR) techniques, because, traditionally, SRL systems have focused in searching the fillers of those explicit roles appearing within sentence boundaries (Gildea and Jurafsky, 2000, 2002; Carreras and Màrquez, 2005; Surdeanu et al., 2008; Hajič et al., 2009). These systems limited their search-space to the elements that share a syntactical relation with the predicate. However, when the participants of a predicate are implicit, this approach obtains incomplete predicative structures with null arguments.

The example in Table 1.1 contains the gold-standard annotations for a traditional SRL process. That analysis includes annotations for the nominal predicate **loss** based on the NomBank structure (Meyers et al., 2004). In this case the annotator identifies, in the first sentence, the arguments arg_0, the entity losing something, arg_1, the thing lost, and arg_3, the source of that loss. However, in the second sentence there is another instance of the same predicate, **loss**, but in this case no argument has been associated with it. Traditional SRL systems facing this type of examples are not able to fill the arguments of a predicate because their fillers are not in the same sentence of the predicate. Moreover, these systems also let unfilled arguments occurring in the same sentence, like the example in Table 1.2. For the predicate **plan** in that sentence, a traditional SRL process only returns the filler for the argument arg_1, the theme of the plan.

However, in both examples, a human reader could easily infer the missing arguments from the surrounding context of the predicate, and determine that

It isn't clear how much those [np **losses**] may widen because of the short Series. [$iarg_0$ The network] [$iarg_1$ of as much as \$20 million] [$iarg_3$ on baseball this year]
Quest Medical Inc said it adopted [arg_1 a shareholders' rights] [np **plan**] in which rights to purchase shares of common stock will be distributed as a dividend to shareholders of record as of Oct 23. [$iarg_0$ Quest Medical Inc]

Table 1.3: Implicit arguments ($iarg_n$) inferred from the context.

in the example in Table 1.1 both instances of the predicate share the same arguments and in the case in Table 1.2 the missing argument corresponds to the subject of the verb that dominates the predicate, *Quest Medical Inc.* The result of this inference process is shown in Table 1.3. Obviously, these additional annotations could contribute positively to its semantic analysis. In fact, Gerber and Chai (2010) pointed out that implicit arguments can increase the coverage of argument structures by **71%**. However, current automatic systems dealing with implicit semantic roles require large amounts of manually annotated training data for each predicate. In fact, the effort required for this manual annotation explains the absence of generally applicable tools for this task.

The research we present on this dissertation aims to label **Implicit Semantic Roles**, but stars with the hypothesis that this task has strong dependences with the different coreferent elements in a discourse. Specially, solving nominal and event coreference and the recovering of event-relations can contribute to capture explicit evidence about the actual fillers of the elided roles. In fact, our research has two main goals:

1. To study **Implicit Semantic Role Labelling** with respect to three different types of relations in the discourse. First, solving implicit roles as a special case of anaphora resolution. Second, taking into account the event coreference. Finally, including some entailment relations between roles and predicates.

2. To study and develop novel methods, tools and resources to overcome the lack of training data. First, applying lexically independent features that generalize for any predicate. Second, developing deterministic algorithms that do not require any training data. Finally, exploiting wide

coverage lexical knowledge resources to obtain rules that relate events and arguments semantically.

1.3 Main contributions

Implicit Semantic Roles Labelling is the task of finding the fillers of the roles of the predicates beyond syntactic dependencies and sentence boundaries. For this type of analysis, a system must combine techniques of both semantic and pragmatic processing. We propose a novel approach to solve implicit roles in a discourse as anaphoric expressions that depend on the coreferent information of the events. Furthermore, we also face the challenge of working with very sparse training sets that are very complex to develop. In order to cover these two aspects, we study and combine a set of methods, tools and resources in three different approaches.

Our first approach follows previous works that relate ISRL with entity coreference. We study some methods commonly used for anaphora and coreference resolution in order to describe a set of features that do not depend lexically on the predicates. Consequently, we train a model that can generalize better for predicates with no manual training set. This model is trained and evaluated in a corpus based on FrameNet and obtains the best results on this dataset.

For the second approach we introduce a new strategy that exploits event coreference. We follow the intuition that different mentions of the same event in a discourse tend to share their roles. As result, we develop **ImpAr**, a deterministic algorithm that does not require any training data. This algorithm is evaluated in a dataset based on PropBank and NomBank annotations. Its performance shows competitive results with fully supervised systems and, unlike them, it can be applied to any predicate.

Finally, we present a new approach that extends the basic configuration of **ImpAr**. In this case, we take advantage of the semantic relations between roles defined in FrameNet. These relations are included in **ImpAr** in order to discover entailments between the events. The evaluation proves that this kind of information can improve the results of ISRL systems.

Moreover, we have obtained other interesting results as a side-effect of the work carried out during this research. The main contributions presented in this document are the following:

- An exhaustive overview of the state of the art about implicit semantic roles and other types of natural language phenomena that are also related to this task. See chapter 2.

- A novel model for Implicit Semantic Role Labelling that focuses in relating different types of coreferential elements on the discourse, including entities and events. See chapter 3.

- The description and comparison of the two existing evaluation frameworks for ISRL. One based on FrameNet and another based on PropBank/NomBank. This study covers both the datasets and the scorers used. See chapter 3.

- Focusing on FrameNet, a new method to determine which missing roles must be recoverable from the context that outperforms previous approaches. See chapter 4.

- A set of lexically independent features to train a supervised system that, following previous related works, considers the implicit roles as a particular case of anaphora. The system is trained and evaluated in the FrameNet dataset, which contains a very sparse training set. This approach obtains the best results on this dataset. See chapter 5.

- A deterministic algorithm for ISRL that combines an adapted version of a well known method of anaphora resolution and a basic approach for event coreference. The algorithm, evaluated in the dataset based on ProbBank/NomBank, obtains results comparable to those obtained by fully supervised systems but, unlike them, it can be applied for any predicate, even when there is not any training data available. See chapter 6.

- A novel approach for exploiting semantic relations between predicates and semantic roles for ISRL. The results show that using relations such as those described in FrameNet improves the performance of an ISRL system. See chapter 7.

- A new automatic mapping between FrameNet and PropBank/NomBank. According to the evaluation results, the mapping provides a reliable set of links that outnumber existing manual mappings. See chapter 7.

- A new resource resulting of transferring the semantic relations of FrameNet to the predicates and roles of PropBank/NomBank. See chapter 7.

1.4 Organization of the document

This thesis presents chronologically the research we have carried out on Implicit Semantic Role Labelling. Starting from a general basic model, each chapter includes incrementally new approaches and methods. The rest of this document is organized as follows:

- **Chapter 2: State of the art**

  This chapter presents an in depth review of the state of the art on implicit semantic role labelling and related tasks. It shows a general overview of different types of analysis over the text for the interpretation of the discourse. Among them, we describe more deeply those that are closely related with the methods proposed in this dissertation. That is, Semantic Role Labelling and both entity and event coreference resolution. Finally, the chapter revises all state of the art approaches to Implicit Semantic Role Labelling, including the available evaluation frame-works.

- **Chapter 3: A framework for Implicit Semantic Role Labelling**

  This chapter presents a complete description of the framework of the research that will be developed in the rest of chapters of this work. It describes a new general model for Implicit Semantic Role Labelling that exploits different coreferential elements related along the discourse. It includes a brief summary of the techniques we propose to overcome the lack of annotated corpora for ISRL. We list how the subsequent chapters relate these techniques with our model. The chapter also describes the two available datasets and the scoring methodologies we use to evaluate our approaches.

- **Chapter 4: First steps for Implicit Semantic Role Labelling**

  This chapter briefly describes the general approach for ISRL consisting in two steps. The first one selects the missing roles of a predicate mention and the second one captures the actual fillers of these roles in the surrounding context of the predicate. The chapter focuses on the differences between FrameNet and PropBank when facing the first step. We emphasize the relevance of that step for FrameNet based labelling and include an initial system for detecting the missing roles for this schema.

- **Chapter 5: Elided roles as a particular case of anaphora**

  This chapter follows previous works that relate ISRL with zero anaphora. It contains a detailed study of a set of features that have been traditionally used to solve anaphora and entity coreference. Focusing on the FrameNet dataset, we propose an adaptation of those features for implicit roles trying to make them not lexically-dependant. We describe how they behave on this dataset, obtaining the best performance in the sate-of-the art. Finally, we discuss their benefits and drawbacks.

- **Chapter 6: Completing role labelling via coreferent predicates**

  This chapter presents a novel deterministic algorithm for ISRL that, in addition to entity coreference, introduces the notion of event coreference. This system do not require training data. Thus, in contrast to supervised approaches, our system can be applied to any predicate. We include an evaluation over the PropBank/NomBank dataset showing that our system obtains very competitive results without training data.

- **Chapter 7: Extending event relationships for a full role annotation**

  This chapter exploits semantic relations between predicates and roles acquired from existing resources. The chapter focuses on the relations described in FrameNet and contains the definition of some declarative rules for making them explicit. We propose a method to automatically map FrameNet to PropBank/NomBank roles in order to transfer the relations. We also describe how to include this knowledge in the system presented in the previous chapter. Finally, we prove that, although their limitations in coverage, these type of relations can improve the labelling of implicit roles.

- **Chapter 8: Conclusion and further work**

  Finally, in this chapter we draw the main conclusions of this research and present some ideas for future work.

CHAPTER 2

State of the art

This chapter presents a review of the state of the art of different research lines regarding Implicit Semantic Role Labeling. The chapter starts in Section 2.1 with an overview of the different tasks in Computational Linguistics that aim to understand automatically the natural human language. After this, we present a deeper review of those types of analysis that are more related to ISLR, including the available resources, datasets and tools. In Section 2.2, we include a description of the research on Semantic Role Labelling. Section 2.3 and Section 2.4 present the most relevant works on Entity Coreference and Event Coreference respectively. And finally, we detail the state of the art that focuses directly on ISRL in Section 2.5.

2.1 Towards natural language understanding

Real language understanding demands deep semantic and pragmatic analysis to obtain the large amount of implicit information that can be inferred from the explicit realizations of the elements of the discourse. A correct interpretation of a written text consists in getting the best possible description of the *events* and the *participants* that take part of the discourse. In order to obtain a complete representation of the events occurring in a text, several tasks are needed both at the sentence level and beyond the boundaries of

Armstrong World Industries Inc. agreed in principle to sell its carpet operations to Shaw Industries Inc. The price wasn't disclosed but one analyst estimated that it was $150 million. Moreover, such a sale could help Armstrong reassure its investors.

Table 2.1: Text example from Penn TreeBank(Marcus et al., 1993)

the sentence. In particular, the full understanding of a text can be viewed as the result of a semantic and pragmatic analysis combined with a grounded resolution of all coreferential relations among the different elements of the document. Different Natural Language Processing (NLP) subtasks can contribute to complete this analysis. These NLP tasks are described below in more detail. The example of Table 2.1 is used to clarify the contribution of each one.

A crucial step for text understanding is the correct interpretation of the meaning of each word depending on their context. A single lemma can represent different senses when it is applied on texts of different domains and usually the only clue that helps to infer its proper meaning are the rest of words that takes part of the same discourse. The process of labelling every word in a text with its appropriate meaning or sense is called **Word Sense Disambiguation** (WSD) (Agirre and Edmonds, 2007). This task can be faced adopting different levels of detail. For example WSD can consist on identifying just the semantic class of the words, **coarse grain**, or on capturing a much more close-fitted meaning, **fine grain**. The latter approach gives a more detailed knowledge about the text but, obviously, it is more difficult. State of the art systems obtain around 60-70% precision for fine-grained senses while 80-90% for coarser distinctions (Izquierdo et al., 2009). As in many other NLP tasks, either supervised and unsupervised learning methods are commonly used for WSD. But in the last year graph algorithms (Agirre and Soroa, 2009; Laparra and Rigau, 2009) have gained much attention because they obtain very good results and do not require a costly annotation effort to develop the training data. Independently of the chosen technique, any WSD system must be based on a knowledge source that provides the set of senses to be associated to the words of the text. The most popular structured knowledge source is WordNet (Miller et al., 1990; Fellbaum, 1998), a computational lexicon of English that encodes in a huge graph sets of synonyms (synsets). In this graph the synsets are connected by semantic relations such as antonymy, hyperonymy, metonymy and so on. The version 3.0 of WordNet contains about 155,000 words organized in over 117,000 synsets. On the other hand, a widely used unstructured data source is SemCor (Miller et al., 1993), a corpus that contains 352 texts manually annotated with part-of-speech tags, lemmas, and word senses from the WordNet inventory. Similarly, OntoNotes(Pradhan et al., 2007) is a large-scale corpus of multiple levels of the shallow semantic annotation in text for three different languages, English, Arabic and Chinese. For English, this corpus contains more than 1,300,000 words from different sources and its semantic representation includes word sense disambiguation for nouns and verbs. There are a total of 264,622 words in the combined corpora tagged with word sense information. These cover 1,338 noun and 2,011 verb types. Related with WSD, **Named Entity Recognition and Classification** (NERC) is the task of detecting and labelling words that refer to specific entities and their classification into general categories like *persons*, *organization* or *places* (Nadeau and Sekine, 2007). A further step of this task is the Entity Linking or **Named Entity Disambiguation** (NED) where the recognized entities are not just linked to their categories but to their references in a knowledge base. In this case the semantic resources used in WSD are replaced by catalogues that contain sets of names for each entity. Wikipedia has became the standard catalogue for NED, specially its structured representation DBpedia.

The example presented previously could be annotated by these processes as Table 2.2 shows. Specifically, this analysis is given by the *SuperSenseTagger*[1] (Ciaramita and Altun, 2006).

This annotation includes NERC, second column, using CoNLL-2003 (Tjong Kim Sang and De Meulder, 2003) entity label. The third column contains supersenses, overall categories that groups the senses of WordNet following its lexicographical files[2]. That is, this example presents an example of coarse grained WSD.

However, word senses by themselves are not sufficient to represent the actions described in a sentence and their participants. In order to correctly interpret the semantic information at a sentence level, it is necessary to know which relations connects the different elements of the sentence. The syntax of the sentence, either dependencies and constituents, shows how words are

[1] http://supersensetag.sourceforge.net/
[2] http://wordnet.princeton.edu/wordnet/man/lexnames.5WN.html

word	named-entity	super-sense
Armstrong	ORGANIZATION:CORPORATION	noun.person
World	ORGANIZATION:CORPORATION	noun.group
Industries	ORGANIZATION:CORPORATION	noun.group
Inc.	ORGANIZATION:CORPORATION	noun.group
agreed	-	verb.communication
in	-	adv.all
principle	-	adv.all
to	-	-
sell	-	verb.possession
its	-	-
carpet	-	noun.artifact
operations	-	noun.act
to	-	-
Shaw	ORGANIZATION:CORPORATION	noun.group
Industries	ORGANIZATION:CORPORATION	noun.group
Inc	ORGANIZATION:CORPORATION	noun.group
.	ORGANIZATION:CORPORATION	-
The	-	-
price	-	noun.attribute
was	-	verb.stative
n't	-	adv.all
disclosed	-	adj.all
but	-	-
one	CARDINAL	adj.all
analyst	PER_DESC	noun.person
estimated	-	verb.cognition
that	-	-
it	-	-
was	-	verb.stative
$	MONEY	-
150	MONEY	-
million	MONEY	noun.quantity
.	-	-
Moreover	-	adv.all
,	-	-
such	-	adj.all
a	-	-
sale	-	noun.act
could	-	-
help	-	verb.social
Armstrong	ORGANIZATION:CORPORATION	noun.person
reassure	-	verb.emotion
its	-	-
investors	PER_DESC	noun.person
.	-	-

Table 2.2: WSD and NER performed by the SuperSenseTagger.

related to each other but these syntactic relations do not provide any explicit semantic relation. In consequence, another analysis layer is required over the syntactic parsing. In particular, for recognizing the predicates of the sentence and their semantic arguments or roles, like *Agent*, *Patient*, *Instrument* or *Location*. This task is called **Semantic Role Labelling** (SRL) (Gildea and Jurafsky, 2000, 2002) and it interfaces the syntax with the semantics of the sentence. In the last years several lexical-semantic resources have emerged including descriptions of predicative structures. Depending on the approach, these descriptions can be closer to the syntax, like the predicate-argument relations of PropBank (Palmer et al., 2005) and NomBank (Gerber and Chai, 2010), or more abstract, like VerbNet (Kipper, 2005) or FrameNet (Baker et al., 1998). The formalisms provided by these lexical databases, along with the corpora annotated with them, allowed a great boost for developping many SRL systems. The most successful SRL systems, as Björkelund et al. (2009), rely on supervised learning.

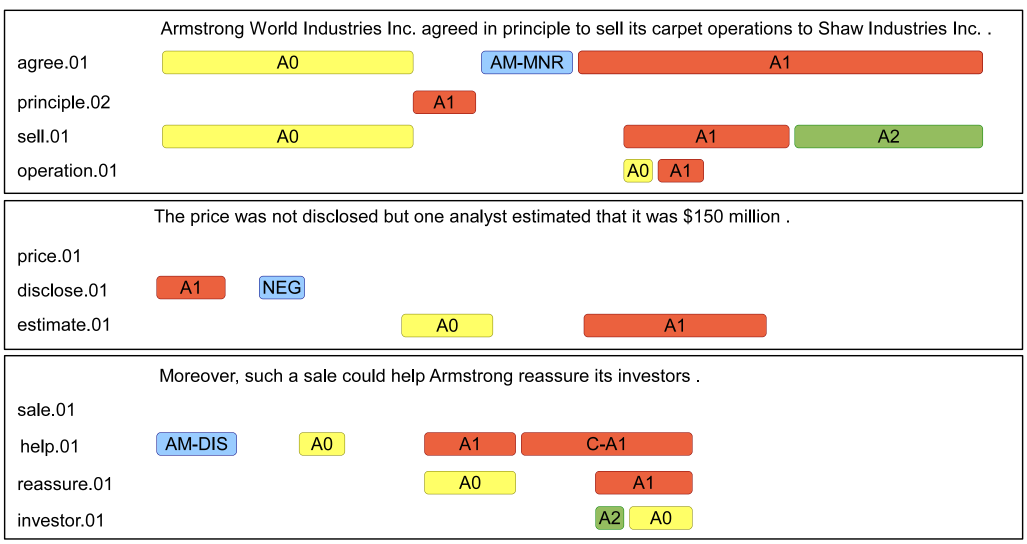

Figure 2.1: SRL annotations as provided by mate-tools SRL system.

Figure 2.1 presents the example of Table 2.1 including Semantic Role Labelling annotations following the argument structures of PropBank and NomBank as provided by mate-tools[3] (Björkelund et al., 2009). The first column shows the proper sense of the predicate, represented by the number

[3]http://code.google.com/p/mate-tools

with a lemma. In fact, SRL includes a WSD process as it also selects the prediction of these senses. In colors yellow, red, blue and green appear the different participants of the events mentioned in the texts. *A0* (in yellow) stands for the argument arg_0 which usually corresponds to the *Agent* of the event. *A1* and *C-A1* (in red) stands for the argument arg_1 which usually corresponds to the *Theme* of the event. *AM-MNR*, *NEG* and *AM-DIS* (in blue) corresponds to different modifiers of the event. Finally, *A2* (in green) stands for the argument arg_2 of the event. For the **sell.01** predicate, the argument arg_2 corresponds to the *Recipient* while for the predicate **investor.01**, the same argument corresponds to the participant that receives the investment.

The cases presented above show how this semantic analysis provides an interpretation about what is expressed within the sentence boundaries. But full understanding of the text demands capturing the relations that connect the elements of different sentences on a discourse level, that is, beyond sentence boundaries. This kind of text processing requires matching the different elements that refer to the same entity or event, reconstructing the discourse and recognizing rhetorical relations such as *Antithesis*, *Condition* or *Purpose*. The first of these tasks is called **Coreference Resolution** (CR) and deals with the identification of those expressions in the text that are actually mentions of the same individual. One of the main difficulties for CR is that coreference arises in many distinct ways. For instance, one of the most frequent occurrences of coreference is the *anaphora* that takes place when an expression can not be interpreted without the participation of its referent, such as pronouns. Moreover, although the majority of the research performed in this area has focused on entities, coreference also affects to events, that can be expressed many times in the same text by either nominal and verbal forms. Currently, the performance of the sate of the art for entity coreference achieves around 80% (Lee et al., 2011) and from 80% to 90% on event coreference (Bejan and Harabagiu, 2010).

Armstrong World Industries Inc. agreed in principle to **sell** its carpet operations to Shaw Industries Inc.

The price wasnt disclosed but one analyst estimated that it was $150 million.

Moreover, such a **sale** could help Armstrong reassure its investors.

Figure 2.2: Example of entity and event coreference.

Figure 2.2 makes evident that entities and events occurs along the discourse generating chains of coreferent mentions of events and entities. For instance, in this example, a coreference chain if the one formed by the three mentions of the entity *Armstrong World Industries Inc.* which are marked by blue boxes. The example also contains another anaphoric mention of the *price* marked in yellow and two mentions of the same *selling* event highligthed in red.

Clustering all the mentions in their corresponding coreferent chains provides a more complete vision about the events and participants in a discourse. However, in order to achieve a complete understanting of the text, it is still necesary to explicitly represent all the relations between the events and participants occurring in or implied by the discourse.

In this sense, a crucial task for language understanding is reconstructing the time lines occurring in a document[4]. In other words, ordering cronologically the events with respect to one another. In order to perform this kind of arrangement, every event must be anchored in time and connected with other events through temporal relations. The main approaches that face this analysis can be distinguished in two groups.

On the one hand, there are system that apply supervised algorithms learning from annotated corpora (Mani et al., 2006).For English, TimeBank (Pustejovsky et al., 2003a) contains 183 annotated news articles following TimeML specifications (Pustejovsky et al., 2003b). Figure 2.3 presents the TimeML annotation of the example of Table 2.1 as obtained by using the Tarsqi tools[5] (Mani et al., 2006). This annotation presents three different temporal relations between events, when they occur simultaneously, like *agree* and *sell* and when one of the events occurs after or before another event, as *help* that occurs after the event *sell*. Element *t0* stands for the time point when the discourse takes place. This time reference is considered for establishing relative temporal relations. For example, the event *sell* is supposed to occur before the writing of the text.

Another approaches use unsupervised techniques to learn sequences of events or narrative schemas (Chambers and Jurafsky, 2009, 2010). For example, a schema of a criminal act would include events like *the commitment*

[4] `http://alt.qcri.org/semeval2015/task4`
[5] `http://www.timeml.org/site/tarsqi/toolkit`

```xml
<TimeML>

  <DCT>
    <TIMEX3 tid="t0" type="DATE" value="2014-03-09" temporalFunction="false"
functionInDocument="CREATION_TIME"> 2014-03-09 </TIMEX3>
  </DCT>

  <TEXT>
    Armstrong World Industries Inc.
    <EVENT class="I_ACTION" eid="e1">agreed</EVENT>
    in principle to
    <EVENT class="OCCURRENCE" eid="e2">sell</EVENT>
    its carpet operations to Shaw Industries Inc. The price wasn't
    <EVENT class="REPORTING" eid="e3">disclosed</EVENT>
    but one analyst
    <EVENT class="I_ACTION" eid="e4">estimated</EVENT>
    that it was $150 million. Moreover, such a
    <EVENT class="OCCURRENCE" eid="e5">sale</EVENT>
    could
    <EVENT class="I_ACTION" eid="e6">help</EVENT>
    Armstrong
    <EVENT class="OCCURRENCE" eid="e7">reassure</EVENT>
    its investors.
  </TEXT>

  <MAKEINSTANCE eiid="ei1"  eventID="e1" pos="VERB" tense="PAST" aspect="NONE" polarity="POS" />
  <MAKEINSTANCE eiid="ei2"  eventID="e2" pos="VERB" tense="PAST" aspect="NONE" polarity="POS" />
  <MAKEINSTANCE eiid="ei3"  eventID="e3" pos="VERB" tense="PAST" aspect="NONE" polarity="NEG" />
  <MAKEINSTANCE eiid="ei4"  eventID="e4" pos="VERB" tense="PAST" aspect="NONE" polarity="POS" />
  <MAKEINSTANCE eiid="ei5"  eventID="e5" pos="NOUN" tense="NONE" aspect="NONE" polarity="POS" />
  <MAKEINSTANCE eiid="ei6"  eventID="e6" pos="VERB" tense="NONE" aspect="NONE" polarity="POS" />
  <MAKEINSTANCE eiid="ei7"  eventID="e7" pos="VERB" tense="NONE" aspect="NONE" polarity="POS" />

  <TLINK lid="l1" relType="BEFORE" eventInstanceID="ei3" relatedToTime="t0" />
  <TLINK lid="l2" relType="BEFORE" eventInstanceID="ei1" relatedToTime="t0" />
  <TLINK lid="l3" relType="SIMULTANEOUS" eventInstanceID="ei2" relatedToEventInstance="ei1" />
  <TLINK lid="l4" relType="AFTER" eventInstanceID="ei3" relatedToEventInstance="ei1" />
  <TLINK lid="l5" relType="AFTER" eventInstanceID="ei4" relatedToEventInstance="ei1" />
  <TLINK lid="l6" relType="AFTER" eventInstanceID="ei5" relatedToTime="t0" />
  <TLINK lid="l7" relType="SIMULTANEOUS" eventInstanceID="ei1" relatedToEventInstance="ei5" />
  <TLINK lid="l8" relType="BEFORE" eventInstanceID="ei6" relatedToEventInstance="ei5" />
  <TLINK lid="l9" relType="BEFORE" eventInstanceID="ei7" relatedToEventInstance="ei5" />

</TimeML>
```


Figure 2.3: TimeML annotations obtained by Tarsqi tools.

of the crime, *the arrest of the criminal*, *the trial*, and so on. The roles or participants of those events are also taken into account.

Besides temporal relations, there exist many other types of discursive connections between the events. In the literature, many frameworks have been proposed on modelling this kind of relations to support the **discourse analysis**. One of the most prominent is the Rhetorical Structure Theory (RST) proposed by Mann and Thompson (1988). For RST the text is a a sequence of non-overlapping units related to each other by a set of defined relations. For each relation, RST declares one of the discursive units as the nucleus of the relationship, while the other one, called satellite, is considered as a piece of supporting information for the nucleus. However, in some of the relations both units are marked as nucleus. These relations can be embedded recursively giving as a result a tree structure of the discourse. A different approach is followed in the Penn Discourse Treebank (PDTB) (Prasad et al., 2008). This corpus, created using the Penn Treebank (Marcus et al., 1994), follows a novel annotation framework based on the predicate-argument structures. The annotation of PDTB include explicit and implicit relations. The span of the arguments and the sense of the relation are also present.

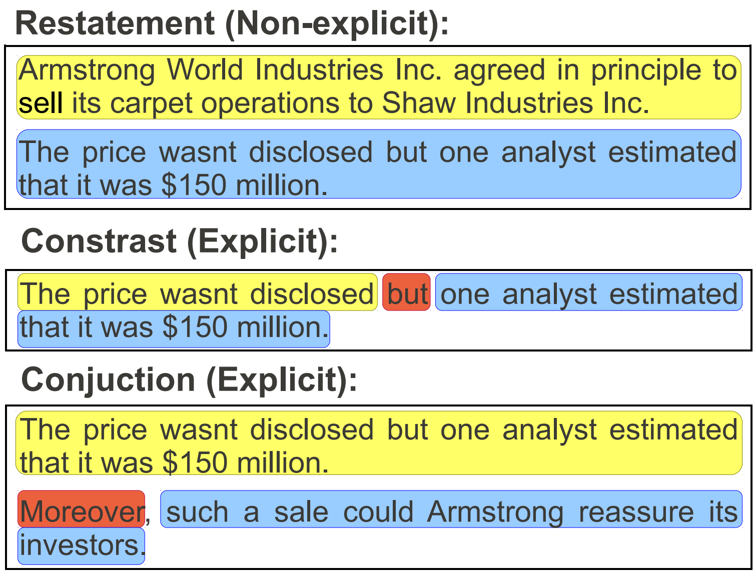

Figure 2.4: Output provided by the PDTB tagger.

Figure 2.4 shows a the discursive relations between the events of the example of Table 2.1 as obtained using PDTB tagger[6] (Lin et al., 2014). In this case, there are three different relations connecting some discourse units. As shown, the *Restatement* relation is not marked by any explicit connective, while the *Contrast* and *Conjuction* relations are pointed by the connectives *but* and *moreover* respectively.

The set of language analysis described up to this point establish a comprehensive set of contributions towards natural language understanding. However, even the joint application of all these processes does not provides a full interpretation of the text (Glavaš and Šnajder, 2014). There is a considerable amount of information that lies behind the phenomena treated by those

[6] http://wing.comp.nus.edu.sg/~linzihen/parser

processes. For example, going back to the traditional SRL annotation (see Figure 2.1), it is clear that, in many cases, the role structures are incomplete. In fact, some of the predicates, like *price*, has not assigned even a single argument. However, for a human reader the text contains enough knowledge to fill those missing roles just looking beyond direct syntactic dependencies and sentence boundaries.

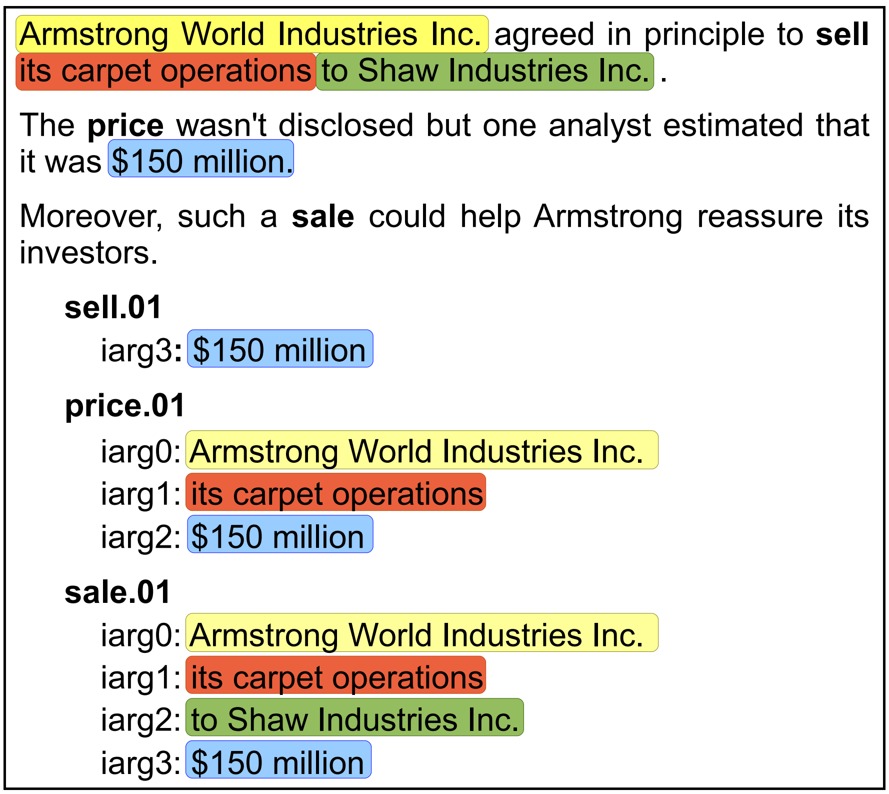

Figure 2.5: Example of an Implicit Semantic Role Labelling.

Figure 2.5 shows how frequent these type of *implicit* roles are. A traditional SRL system applied to these three predicates does not obtain a complete argument structure for any of the three predicates. However, in this case, the context contains enough information to recover, implicitly, all the participants of the events described.

Palmer et al. (1986) called these missing arguments **Implicit Semantic Roles** and argued that they can be considered as a special case of anaphora.

Since then, just a handful of works have been approached to the task. Moreover, all these efforts shared similar strategies trying to combine entity coreference resolution and traditional semantic role labelling.

Our research stars with the hypothesis that implicit roles have strong implications with other characteristics of the discourse. Since our research focuses on detecting such Implicit Semantic Roles, we also reviewed the state of the art of the tasks briefly introduced in this section that are most related with the work carried out in this thesis. That is, Semantic Role Labelling, Nominal Coreference resolution, Event Coreference and Implicit Semantic Role Labelling. The interested reader are referred to extensive surveys, summaries and books of additional related work. For instance, Named Entity Recognition and Classification (Nadeau and Sekine, 2007), Word Sense Disambiguation (Agirre and Edmonds, 2007), extracting timelines and temporal relations (Schilder et al., 2007) and discourse structure (Webber et al., 2012).

2.2 Semantic Role Labeling

2.2.1 Early SRL systems

As many other early works on NLP tasks, the research in automatic semantic role annotation started by manually defining large sets of rules that tried to encode world knowledge. Along this line, Hirst (1987) proposed a compositional building of semantic representation linking syntactic constituents to corresponding manually defined frame positions. In a similar fashion, the PUNDIT (Prolog Understanding of Integrated Text) (Hirschman et al., 1989; Dahl, 1992) system included, among many others, a set of hand-coded logic rules to map syntactic and semantic constituents in order to process military and medical domain texts. Many later works followed this approach and built several hand-crafted lexicons, grammars and sets of logic rules as Richardson et al. (1998), even for nominalizations, like the work by Hull and Gomez (1996) and Meyers et al. (1998). Remarkably, PUNDIT (Dahl et al., 1987) also designed discourse heuristics for the identification of the role structures of pronouns referring to nominal predicates.

The example in Table 2.3, taken from Hirst (1987), provides an overview of the kind of mappings performed by the systems described.

[Nadia$_{subject}$] **bought** [the book$_{object}$] from a store in the mall.

Word or Phrase	Semantic Object
SUBJECT	*AGENT*
Nadia	
buy	
OBJECT	*PATIENT*
the book	
from	*SOURCE*
a store	
in	*LOCATION*
the mall	
.	

Table 2.3: SRL taken from Hirst (1987)

Obviously, the rules and predicate frames required in order to build such systems must be very specific and require large sets of hand-crafted rules. For that reason, the systems explained previously were developed in a very domain-specific manner and display a weak performance when they are applied to novel domains that do not share the same predicates and roles. This handicap made statistical method gain relevance and interest with respect to rule based systems. However, if the rules are well defined the predictions obtained by them are likely to be correct, so they are still a proper approach when high precision is needed. Moreover, the rules by themselves provide a proper explanation of why the predictions matches the correct answers. Obviously, this explanation is much more difficult to obtain from statistic models.

2.2.2 Resources and corpora

In the last few years, several efforts have been developing lexical-semantic resources that describe the predicate and role structures. These efforts have been also producing a set of manually annotated corpora based on different theories and paradigms. Predicate structures and the role descriptions contained in these resources vary from syntactic arguments to thematic roles, including some semantic and ontological knowledge. Although early work on

Predicate **sell**, roleset 1:
arg_0 : the seller arg_1 : the thing sold arg_2 : the buyer arg_3 : the price paid arg_4 : the benefactive
[arg_0 Maxwell] agreed to **sell** [arg_1 its U.S. printing unit] [arg_2 to Quebecor] [arg_3 for $500 million].

Table 2.4: PropBank roleset and annotated sentence for sense 1 of verb **sell**.

the description of the role structures started from verbs some of the existing resources also take into account other forms of the predicates and particularly their nominalizations. Finally, as the resulting corpora and predicate structures highly vary from one resource to another, it has been also invested some effort in their harmonization and mapping.

2.2.2.1 PropBank

Palmer et al. (2005) provided a semantic annotation over the syntactic structures of the Penn TreeBank(Marcus et al., 1993). The result of this work is the Proposition Bank (PropBank)[7], a wide corpus with more than 112,000 semantic role analyses that includes relations between verbal predicates and their roles or arguments. PropBank also contains a description of the frame structures, called **rolesets**, of each sense of every verb that belong to its lexicon. This lexicon contains up to 3,256 different verbs. Unlike other similar resources, PropBank defines the arguments, or roles, of each verb individually. Consider the frame in Table 2.4 taken from PropBank for the sense 1 of the verbal predicate **sell** and an annotated sentence included in the corpus.

In the example of Table 2.4, the *Agent* and the *Theme* of the predicate **sell** correspond to the argument arg_0 (*the seller*) and the argument arg_1 (*the thing sold*) respectively. Unfortunately, PropBank does not encode explicit relations between arguments of different predicates. For instance, consider the example shown in Table 2.5 representing the frame for the sense 1 of

[7] `http://verbs.colorado.edu/~mpalmer/projects/ace.html`

Predicate **buy**, roleset 1:
arg_0 : the buyer
arg_1 : the thing bought
arg_2 : the seller
arg_3 : the price paid
arg_4 : the benefactive

Table 2.5: PropBank roleset for sense 1 of verb **buy**.

Predicate **sale**, roleset 1:
arg_0 : the seller
arg_1 : the thing sold
arg_2 : the buyer
arg_3 : the price paid
arg_4 : the benefactive

Table 2.6: NomBank roleset for sense 1 of the nominal predicate **sale**

the predicate **buy**. Although both predicates share *the buyer* and *the seller* arguments, are not explicitly related since the descriptions of the roles are not systematic.

2.2.2.2 NomBank

NomBank (Meyers et al., 2004) follows the annotation effort of PropBank focusing on the argument structures of instances of nominal predicates over the same Penn TreeBank corpus (Marcus et al., 1993). NomBank also bases its lexicon on PropBank. In fact, most of the predicates of NomBank are nominalizations of verbs existing in PropBank and inherit the role sets and argument definitions from their corresponding verbal forms. For instance, consider the example from Table 2.6 that contains the frame of the sense 1 of the nominal predicate **sale** included in NomBank. It can be observed that it matches the same frame of its verbal form **sell** shown in Table 2.4.

This concordance between the frames of the nominal and verbal forms of the same predicates allows to maintain the coherence among the verbal

[arg_0 Maxwell] agreed to the **sale** of [arg_1 its U.S. printing unit] [arg_2 to Quebecor] [arg_3 for $500 million].

Table 2.7: NomBank annotations for the predicate **sale**.

and nominal annotations of the same predicates. In the example shown in Table 2.7, all the arguments of the nominal predicate **sale** have the same interpretation as the verbal predicate **sell**.

The NomBank corpus shows that nominal predicates are as frequent as verbal predicates. In total, the corpus contains argument annotations for 114,574 instances of 4,704 distinct nominal predicates.

2.2.2.3 VerbNet

Inspired by the work of Levin (1993) on verb classes, Kipper et al. (2000); Kipper (2005) developed hierarchical domain-independent broad-coverage verb lexicon for English. This resource, called **VerbNet**[8], is organized into coherent verb classes and subclasses. The most recent version of the resource groups 5,257 verb senses into 274 classes. Each verbal class in VerbNet is completely described by thematic-roles, selectional restrictions on the arguments, and frames consisting of a syntactic description and semantic predicates. For instance, the VerbNet class 13.1, shown in Table 2.8, groups verbs related to *giving* acts.

Table 2.8 shows how the members of the subclass 13.1-1 inherit the semantics from its parent class 13.1. Subclass 13.1-1 includes a more specific set of events grouped together because they share an additional thematic-role, *Asset*. Thus, instances of the members 13.1-1 like **sell** can be annotated as the sentence shown in Table 2.9.

Although VerbNet does not provide any corpus with annotations, this resource is very rich since it encodes semantic descriptions of the events that represent each class and also includes as selectional preferences a set of semantic types, hierarchically classified, for some of the roles of the classes.

Table 2.10 shows the information encoded in the VerbNet class 13.1. Its semantics describe the exchanging process of the **selling** event of the example

[8] `http://verbs.colorado.edu/~mpalmer/projects/verbnet.html`

Class 13.1:	
Members:	**deal, lend, loan, pass, peddle, refund, render**
Thematic-roles:	*Agent, Theme, Recipient*
Subclass 13.1-1:	
Members:	**give, hock, hawk, rent, sell, lease, pawn**
Thematic-roles:	*Asset*

Table 2.8: VerbNet class 13.1 description.

[*Agent* Maxwell] agreed to **sell** [*Theme* its U.S. printing unit] [*Recipient* to Quebecor] [*Asset* for $500 million].

Table 2.9: Example annotation for the verbal predicate **sell**

Selectional restrictions:

Agent:	animate or organization
Recipient:	animate or organization

Semantics :

has_possession(start(**Event**), *Agent*, *Theme*)
has_possession(start(**Event**), *Recipient*, *Asset*)
has_possession(end(**Event**), *Agent*, *Asset*)
has_possession(end(**Event**), *Recipient*, *Theme*)
transfer(during(**Event**), *Theme*)

Table 2.10: Information encoded in VerbNet for class 13.1

Commerce_sell:	
lexical-units:	**auction.n, auction.v, retail.v, retailer.n, sale.n, sell.v, seller.n, vend.v, vendor.n**
frame-elements:	
Core:	*Seller Goods Buyer*
Non-Core:	*Manner Means Money Period_of_iteration Place Purpose Purpose_of_Goods Rate Reason Rely Result Reversive Time Unit*

Table 2.11: Lexical-units and frame-elements in the frame **Commerce_sell**

in Table 2.9. In this case, an *animate* entity (*Maxwell*) plays the role *Agent* and an *organization* (*Quebecor*) plays the role *Recipient*. At the start of the **selling** event the *Agent* possesses a *Theme* (*U.S. printing unit*) and the *Recipient* an *Asset* (*$500 million*). At the end of the event, the *Agent* and the *Recipient* have exchanged their possessions.

2.2.2.4 FrameNet

Fillmore (1976) established a new paradigm to describe events called Frame Semantics. In this paradigm, a Frame corresponds to a scenario that involves the interaction of a set of typical participants, each one playing a particular role in the scenario. Some of those roles would be essentials for the scenario. Following this paradigm, Baker et al. (1998) built **FrameNet**[9], a very rich semantic resource that contains and groups sets of 12,940 English words into 1,019 coherent semantic classes or frames that describe a particular event or scenario. Each frame, that is further characterized by a list of participants, can be evoked by the set of words that belong to it. Different senses for a word are represented in FrameNet by assigning different frames. In FrameNet, these trigger words are called *lexical-units* (LU) and the participants or roles that describe the frame are called *frame-elements* (FE).

Table 2.11 shows the **Commerce_sell** frame. The *Core frame-elements* are those essential FEs of the frame that can define the frames by themselves. In the **Commerce_sell** frame, the core frame-elements are the *Seller*, *Goods* and *Buyer*, while the rest of the participants are considered less descriptive or

[9] `http://framenet.icsi.berkeley.edu/`

[*Seller* Maxwell] agreed to **sell** [*Goods* its U.S. printing unit] [*Buyer* to Quebecor] [*Money* for $500 million].

Table 2.12: Example annotation for the verbal predicate **sell** in FrameNet

too general. FrameNet contains a corpus of approximately 197,055 annotated frame instances as the example shown in Table 2.12.

However, not every *Core FEs* is always present in a sentence. These FEs are considered as *Null Instantiations* (NI). FrameNet classifies the NIs in three different sets:

1. If the FEs are grammatically omitted, NIs are called *Constructional Null Instantiations* (CNI).
2. If the omissions are licensed lexically and the fillers are inaccessible, the NIs are called *Indefinite Null Instantiations* (INI).
3. When the omissions are licensed lexically and the fillers are recoverable, the NIs are called *Definite Null Instantiations* (DNI).

FrameNet defines a complex semantic network linking the frames with twelve different relationships. These relations are of the type *subclass*, *causation* and *perspective* that also connect the FEs among the frames. The resulting network contains 10,076 direct relations between frames that allow to perform inferences involving the different events and their participants.

Figure 2.6 shows a small portion of the whole ontology involving 4 different frames. The ***Inheritance*** relation between the frame **Transfer** and **Commerce_goods_transfer** means that the latter describes a more specific case of the first one. In consequence, all the properties of the frame **Transfer** are inherited by the frame **Commerce_goods_transfer**, in particular, the corresponding frame-elements. However, the frame *Commerce_goods_transfer* describes an scenario from a neutral point of view. That is why this frame does not contain any lexical-unit to evoke the frame. The event described by this frame varies its interpretation depending on the perspective of the participants. This is expressed by the relation ***Perspective_on*** that connects the frame **Commerce_goods_transfer** to the frame **Commerce_sell**, if the perspective of the *seller* is assumed, and to the frame **Commerce_buy**,

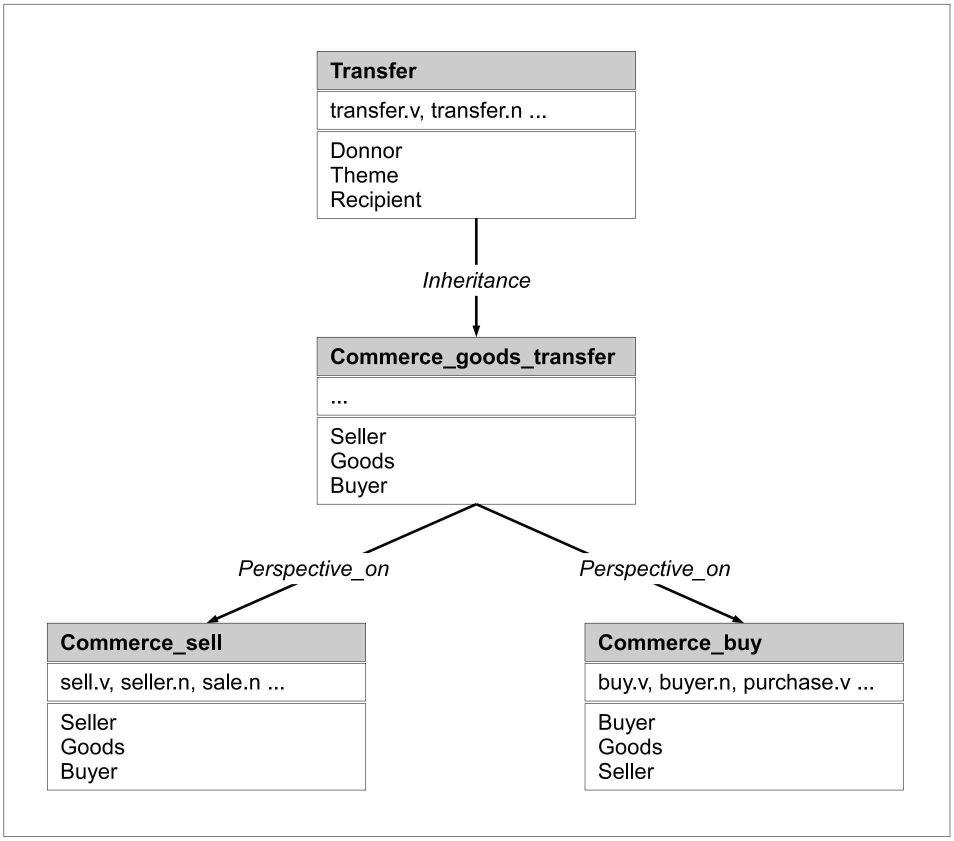

Figure 2.6: Examples of FN frames connected by *Inheritance* and *Perspective_on* relations.

in the case of the perspective of the **buyer**. Thus, the lexical-units such as **sell.v**, **seller.n**, **buy.v** and **buy.n** are actually evoking the same event but from different points of view.

2.2.2.5 SemLink

SemLink[10] (Palmer, 2009) is a project whose aim is to link together different predicate resources via a rich set of mappings. Currently, SemLink provides partial mappings between FrameNet (Baker et al., 1998), VerbNet (Kipper et al., 2000), PropBank (Palmer et al., 2005) and WordNet (Fellbaum, 1998).

[10] `http://verbs.colorado.edu/semlink/`

PropBank	VerbNet	FrameNet
sell.01	13.1-1	Commerce_sell
arg_0	*Agent*	*Seller*
arg_1	*Theme*	*Goods*
arg_2	*Recipient*	-

Table 2.13: SemLink role mappings for the verbal predicate **sell**

These mappings make it possible to combine their information for tasks such as inferencing, consistency cheking, interoperable SRL, or automatic extending its current overlapping coverage.

Palmer (2009) states that PropBank, VerbNet and FrameNet are indeed complementary and the redundancy caused by grouping the three resources together can be useful. This assertion is justified because PropBank provides the best coverage and the largest corpus, VerbNet contains the clearest links between syntax and semantics and FrameNet provides the richest semantics and ontological knowledge. Thus, all together can provide a much richer and more complete resource. The example shown in Table 2.13 shows the full mapping between the different role structures of the same sense of the predicate **sell**.

Manually mapping these resources is a very difficult task. In the example, the argument arg_2 of the predicate **sell.01** of PropBank is only connected to its corresponding thematic role in VerbNet, but not to its corresponding frame-element of FrameNet which is *Buyer*. López de Lacalle et al. (2014a) provides a complete description of the partial coverage of SemLink across the different predicate resources.

2.2.3 Statistical SRL

As seen previously, early works on automatic semantic role annotation based on hand-coded rules required so much effort in development that the resulting systems only could be effective on specific domains. However, the emergence of new corpora and linguistic resources that described role structures for more general domains and genres allowed the application of machine learning techniques. This research line was introduced by Gildea and Jurafsky (2002). In this work, the authors modelled a set of features in order to discover the relation between the roles and the lexical and syntactic realizations. They trained their model on the FrameNet corpus employing simple maximum likelihood and got quite encouraging results reaching a 63% F1 in the identification and labelling of frame-elements. The analysis of these results showed that finding the proper syntax of the frame-elements, or roles, of each predicate is crucial in order to obtain their correct semantic interpretation. This conclusion matches the intuition behind early works on semantic role labelling. The proposal by Gildea and Jurafsky (2002) established a kind of architecture that have been followed by many other later works. Their model used separated classifiers in a two-steps process. The first step, for the identification of the frame-elements and the second one, for labelling them. However, in their experiments, Gildea and Jurafsky (2002) assumed ground-truth lexical-units and frames, letting out the important step of the predicate disambiguation.

The task gained relevance with the shared challenges of Senseval-2004 (Litkowski, 2004), CoNLL-2004 (Carreras and Màrquez, 2004), CoNLL-2005 (Carreras and Màrquez, 2005), CoNLL-2008 (Surdeanu et al., 2008) and CoNLL-2009 (Hajič et al., 2009). While Senseval-2004 maintained FrameNet as the reference resource, the CoNLL tasks distributed the corpus based on PropBank. Due to the success of these challenges, soon, PropBank emerged as the most popular paradigm for Semantic Role Labelling. The progress of the challenges also uncovered that the syntactic representation used to portray the text can highly vary the performance of the systems. Constituency trees was the first formalism used to represent the syntactic properties for SRL. This representation has a long tradition in computational linguistics and provides a wide range a of syntactic properties that are exploitable as features for a statistical model. However, since Surdeanu et al. (2008) the approaches to the task have turned to dependency trees. According to Johansson and Nugues (2008), the latter representation rely less on lexicalized features, which makes them more flexible for generalization and domain changes and in consecuence, more efficient with same amount of training data. But still, SRL systems based on consituents are more usefull for the proper identification of the fillers of the roles. The most successful systems presented in the mentioned tasks used Maximum Entropy models or Support Vector Machines, like Ngai et al. (2004),Hacioglu et al. (2004) or Zhao et al. (2009), or a combination of several classifiers, like Koomen et al. (2005).

Although the corpus of FrameNet(Baker et al., 1998) includes annotations for nominal predicates, statistical models for nominalizations have not been very frequent until the release of NomBank(Meyers et al., 2004). Previously, Lapata (2000) developed a statistical model for the semantic interpretation of compound nouns based on their cooccurrence with verb-argument tuples on a corpus. This preliminary approach only labelled the arguments corresponding to the subject and the object. The first use of NomBank for Semantic Role Labelling corresponds to the work by Jiang and Ng (2006) who adapted a system based on PropBank with some features extracted from a pre-release version of NomBank. Another initial study on NomBank was presented by Liu and Ng (2007). In this work, the authors applied a machine learning technique called alternating structure optimization (Ando and Zhang, 2005) using an improved version of NomBank. Following these works, from Surdeanu et al. (2008), CoNLL tasks have also included nominal predicates from NomBank in their challenges and, in consequence, the number of approaches facing role labelling for nominalizations has increased significantly.

Recently, after the success obtained in tasks such as topic modelling, bayesian inference has became one of the most applied techniques in NLP. This tendency has also influenced several works on unsupervised learning for semantic role labelling. Some of the first approaches in this line include the proposal by Swier and Stevenson (2004), who used the VerbNet lexicon together with a bootstrapping algorithm to unsupervisely learn the semantic roles that label the arguments of the verbs. In another early work, Grenager and Manning (2006) obtained prior distributions over syntactic-semantic links to perform unsupervised learning in a generative model with an Expectation-Maximization algorithm. Lang and Lapata (2010, 2011a, b) involve a significant progress towards the unsupervised induction of semantic roles from unannotated text. Lang and Lapata (2010) formulated the problem as one of detecting alternations and finding a canonical syntactic form for them. In a later work, Lang and Lapata (2011a) described an algorithm that iteratively splits and merges clusters representing semantic roles, improving the quality of the clusters in each iteration. Finally, Lang and Lapata (2011b) conceptualized the role induction as a graph partitioning problem where vertices in a graph represent argument instances of a verb and the edges quantify their role-semantic similarity. The graph is partitioned iteratively by an algorithm that assigns vertices to clusters based on the assignments of the neighbour vertices. This research line has been also followed by Poon and Domingos (2009) who transformed syntactic relations into logical forms for the application of Markov logic. In a very related manner, Titov and Klementiev (2011) proposed a non-parametric Bayesian model to apply a modification of the Metropolis-Hastings sampling. This model was later improved by including the Chinese Restaurant Process (CRP) as a prior. In this case the authors could evaluate the model on PropBank. More recently Modi et al. (2012) applied the same model to induce frame-semantic representations.

Thanks to the research works detailed above, there are currently available a wide set of tools that include SRL systems. The Mate tools[11](Björkelund et al., 2009) provide a pipeline of modules that carry out lemmatization, part-of-speech tagging, dependency parsing, and PropBank based semantic role labelling. The tools are language independent, provide a very high accuracy. SwiRL[12](Surdeanu and Turmo, 2005) is a PropBank based SRL system for English constructed on top of full syntactic analysis. It trains an AdaBoost classifier for each argument label. SENNA[13](Collobert et al., 2011) is a software package with a full NLP pipeline including part-of-speech tagging, chunking, name entity recognition, and syntactic parsing. The package also contains a SRL module based on PropBank. SEMAFOR [14](Das et al., 2010) is a tool for automatic analysis of the Frame Semantic structure of English text that performs a set of steps in order to find words that evoke the semantic frames and label each role of the frame. Shalmaneser[15](Erk and Pado, 2006) is a supervised learning toolbox for shallow semantic parsing that was developed for Frame Semantics, but its architecture is reasonably general and can be adapted for other paradigms like PropBank.

2.3 Entity coreference resolution

2.3.1 Early models

The first computational models for coreference and anaphora resolution implemented approaches relying heavily on linguistic and domain knowledge. Depending on their theoretical and formal assumptions these works can be

[11]http://code.google.com/p/mate-tools/
[12]http://surdeanu.info/mihai/swirl/
[13]http://ml.nec-labs.com/senna/
[14]http://www.ark.cs.cmu.edu/SEMAFOR/
[15]http://www.coli.uni-saarland.de/projects/salsa/shal

classified in three main groups. Some of these models assume that coreference is only affected by the syntax, others that it is a matter of common-sense knowledge and finally, there is a set of models that were pragmatically oriented.

Within the first group, the earliest and most well-known approach was the Hobbs' algorithm (Hobbs, 1977). The algorithm solves pronouns performing a breadth-first left-to-right search for antecedents in a constituency tree representation of the sentence. If no proper antecedent is found, the search continues in the preceding sentences starting from the closest one. Due to its nature the algorithm prefers noun phrases located to the left and in higher positions. This promotes subjects against other syntactic components. It is also remarkable that Hobbs (1977) presented a formal evaluation of the method, even though the testing set had just 100 examples. The algorithm achieved a hight accuracy, around 88%, and more recent studies have showed in larger evaluations that its performance is still quite competitive.

The second group, using common-sense knowledge, was introduced into coreference resolution in the very first works of computational linguistics. Charniak (1972), Winograd (1972) and Wilks (1975, 1986) presented basic systems that dispensed with syntax and proposed knowledge based approaches, like semantic preferences, causal relations and so on. This research line was later widely followed. For instance, Alshawi (1987) exploited a semantic knowledge base built of two types of relations: specializations and correspondences. Similar approaches were presented by Poesio et al. (1997) and Harabagiu and Moldovan (1998) who took those kind of semantic relations from WordNet. Hobbs et al. (1993) proposed a semantic interpretation theory based on abduction, a reasoning mechanism that provides the most plausible causes of a particular observation. Hobbs et al. (1993) applied large sets of axioms for several NLP tasks, including anaphora resolution. Jurafsky and Martin (2009) presented a modern review of this approach.

A third group focuses on theories about the salience of the text elements. A very early work in this research line is the one by Sidner (1978). Sidner's algorithm, besides the use of a semantic network inspired by Charniak (1972), includes the notion of **focus**, one of the most influential theories in subsequent research on coreference resolution. The theory establishes that in each position of a discourse there is an element that have the highest salience, the **focus**, that can be expressed around three information structures: the discourse topic, also named as **discourse focus**, the **actor focus** that points the syntactic subjects, and a ranked list of the entities mentioned in the last sentence. These three sources can expose the most salient entity in a discourse. Thus, the **local focus**, according to Sidner's theory, is the most likely entity to be referred. Carter (1987) combined Sidner's and Hobbs' algorithms in the the SPAR system leading interesting observations on the interaction of common-sense knowledge and focusing. Extending the previous ideas Grosz et al. (1995) presented the **centering** theory that studies how the **local focus** switches along the discourse. The theory establishes a rank with the possible transitions according to their occurrence probability. The most important implementation of centering theory were developed by Brennan et al. (1987), who related the transitions rank with grammatical functions, and Strube and Hahn (1999) who opted to set the ranking based on the taxonomy proposed by Prince (1981). Although posterior empirical studies (Poesio et al., 2004) seems to confirm that Strube and Hahn (1999) implementation fits better real text and performs more accurately, both algorithms have been extremely influential and some of their features are grounded in solid empirical evidence.

Another family of approaches based on salience effects make use of the notion of **activation**. These models assume that every entity in a discourse has a value, graded in a scale, that sets its activation level in each point of the discourse. In other words, the salience of all the entities can be determined by ranking them according to some linguistic features. First examples of activation-based models can be found in Lockman and Klappholz (1980) and Alshawi (1987), but the best known algorithm is RAP by Lappin and Leass (1994). RAP, that stands for Resolution of Anaphora Procedure, exploits syntactic information, taken from a full parser, to filter possible candidates to be referents of a pronoun. The algorithm filters out antecedents of non-reflexives[16] when the pronoun occurs in the argument, adjunct or NP domain of the potential antecedent, and non-pronominal antecedents that are contained in the governing phrase of the pronoun. After the filtering process from all the remaining candidates that are number and gender compatible with the pronoun, the closest one with the highest salience weight is selected. The salience weight is obtained adding a set of values pending on the syntactic realizations of the candidate. For example, a *sentence recency* weight is given to every candidate belonging to the same sentence of the pronoun. Other examples of salience weights used by RAP are *head noun emphasis*, *head*

[16] Reflexive pronouns are *itself, himself, herself, etc.*

object emphasis or *non-adverbial emphasis*. Kennedy and Boguraev (1996) presented a modification of the RAP algorithm using just morphological tags and grammatical functions.

2.3.2 Towards a corpus-based approach

After the works presented above, the research on coreference and anaphora resolution moved to more corpus-based approaches that rely in huge volumes of annotated corpora and in the formalization of the evaluation metrics. The biggest milestone was supplied by the Message Understanding Conferences (MUC), that hosted two evaluations of coreference resolution systems, MUC-6 (Grishman and Sundheim, 1995) and MUC-7 (Chinchor, 1998). These campaigns provided, besides annotated corpora, guidelines for the annotation of coreference and a common evaluation procedure. These tools made possible the comparison of different systems facing the same dataset, involving a capital advance for the research on coreference resolution. MUC-6 used 30 text documents with 4,381 mentions for training, and another 30 documents with 4,565 mentions for testing. MUC-7 consisted of 30 text documents with 5,270 mentions for training, and 20 documents with 3,558 mentions for testing.

Taking over from MUC, the National Institute of Standards and Technology (NIST) promoted the Automatic Content Extraction (ACE) program (Doddington et al., 2004) for the automatic processing of human language in text form. Originally oriented to the three source types of news-wires, broadcast news and newspapers, ACE has been enriched with different sources in each new version released, including texts for languages different to English, like Chinese and Arabic. Along with other kind of annotations, the documents contained in ACE include semantic information about the entities and their referential relation. The size of this corpus varies depending on the source of the documents. For example broadcast news and newspaper news have between 200 and 300 entity mentions on average, while documents from news-wires are commonly smaller with around 100 mentions.

In a more ambitious way, the OntoNotes(Pradhan et al., 2007) project has created a much larger corpus that integrates multiple levels of shallow semantic annotations. The texts of this corpus includes syntactic and semantic analysis, word sense disambiguation, named entity recognition and also coreferential links within the nominal entities. OntoNotes describes all the annotation layers as a relational database that captures inter and intra-layer dependencies and provides an object-oriented API to interact with the data. Thus, OntoNotes has become a very widely used resource, being included in some relevant shared task on coreference resolution, like SemEval-2010 (Recasens et al., 2010) and CoNLL-2011 (Pradhan et al., 2011). The English corpora is considerably larger than MUC and ACE. It contains about 1.3M words, comprising 450,000 words from news-wires, 150,000 from magazine articles, 200,000 from broadcast news, 200,000 from broadcast conversations, and 200,000 web data. As ACE corpora the newest versions of OntoNotes also contains documents for Arabic and Chinese.

Another relevant resource, in Catalan and Spanish, is AnCora-CO (Recasens and Martí, 2010). This corpora contains coreference annotations of pronouns and full noun phrases, including named entities, and takes into account the existence of singletons as one element coreference chains. The resource also contains several annotation layers including lemmas, parts-of-speech, morphological features, dependency parsing, named entities, predicates, and semantic roles. Most of these annotation layers are provided both as gold standard and predicted annotations. AnCora-CO was used in SemEval-2010 (Recasens et al., 2010) along with OntoNotes. The size of AnCora-CO is about 350,000 words of Catalan and a similar quantity in Spanish.

2.3.3 Machine learning approaches

In the same way that many other NLP taks, the availability of increasingly larger quantities of real annotated corpora have allowed to apply multiple machine learning approaches to automatic coreference resolution. Researches in this area argued that the early methods relied in linguistic and ontological information too difficult to obtain and too dependant on the domain. In consequence, they had to be adapted to other domains analysing and encoding relevant facts and assertions. Conversely, researchers in this area argue that machine learning approaches can rely on generalizable features.

The earliest machine learning models were based on decision trees. For instance, Aone and Bennett (1994) developed a system for persons and organizations in Japanese. Mccarthy and Lehnert (1995) developed a similar resolver for the MUC-5 information extraction task. This system exploits features that are domain independent (mention type, name sub-string, being in a common noun phrase, being in the same sentence) along with others that are domain specific (either or both mentions referring to a company created in a joint venture). A later version (Fisher et al., 1995) was evaluated in the MUC-6 coreference task including features such as having the same semantic type or being a role description for a person. The approach by Poesio and Vieira (1998); Vieira and Poesio (2000) focused on one of the hardest type of anaphora resolution, noun phrases in unrestricted domain. These cases are more general that named entities that usually belong to a restricted domain. Poesio and Vieira (1998); Vieira and Poesio (2000) developed two version of the same system. One using hand-coded rules and another one applying a decision tree. In both cases, they combined lexical and common-sense knowledge. The comparison of both versions of the system, using for the evaluation 14 texts from the Penn TreeBank, showed that although the hand-crafted version performed slightly better both versions obtained very similar results.

The inclusion of the probabilistic perspective has provided substantial advance on automatic coreference resolution. For instance, Ge et al. (1998) used a generative probabilistic model to include on the Hobbs' algorithm statistical gender identification, selectional preferences and a saliency measure based on counting mentions. They used texts from the Penn TreeBank to test the model, achieving 84.2% accuracy, while using just the Hobbs' distance only obtained 65.3%. Another early statistical model was proposed by Kehler (1997) who built a maximum entropy classifier to determine the probability that two mentions corefer. The model was applied following two different approaches. A first one using pairwise classification of all pairs of mentions, and another one creating sets of coreferents and deciding if a new mention belongs to an existing set or to a new one. For the evaluation, Kehler (1997) calculated cross-entropies on the test data as well as the number of exact matches. While both approaches overcome the baseline, the pairwise version showed to be superior to the second one.

Shortly after these previous models, Soon et al. (1999, 2001), (Morton, 2000) and Ng and Cardie (2002) established a new standard of statistical anaphora and coreference resolution called **mention-pair** model. Soon et al. (1999, 2001) proposed that coreference resolution can be viewed as a binary classification problem where for every mention and its possible antecedent the algorithm decides if they are coreferent or not. As previous approaches, they presented a system based on decision trees and evaluated it on the MUC-6 dataset reporting a F1 score of 62.6% F1 and MUC-7 getting a 60.4%. On the other hand, Ng and Cardie (2002) improved the system by Soon et al.

(1999, 2001) using a a much larger feature set. They evaluated their system in the same corpus obtaining a F1 score of 70.4% on the MUC-6 and 63.4% on MUC-7. The influence of these two algorithms, and the **mention-pair** model, is similar to the Hobbs' algorithm. In fact both approaches have become the standard benchmarking baselines for many years. Yang et al. (2003, 2005) extended this idea including in the training a negative mention-antecedent pair for each positive case. They proved that this strategy learns better to discriminate between true and false cases than the original models.

Another important standard of statistical coreference resolution is the clustering approach, also called **entity** model. Contrary to the mention-pair method, the coreference chains are obtained as partitions of the set of all the mentions in the text. The first known proposal of this approach was presented by Luo et al. (2004) who carried out the clustering process by structuring the search space as a Bell tree (Bell, 1934) where each leaf represents a possible partition. Their algorithm computes for each mention in a document the probability of belonging to a previously generated partial coreference chain. Following this approach, Daumé and Marcu (2005) presented a model based on online learning where a beam search is applied to avoid some problems derived by non-optimal local decisions. Their method keeps multiple partial solutions and discards those partial solutions that show to be inconsistent. More recently, Rahman and Ng (2009) developed a cluster algorithm that ranks the possible coreference chains for a mention and chooses the highest one. From a totally different point of view, Haghighi and Klein (2007, 2010) presented a non-parametric bayesian model for mention clustering. This model produces generatively each mention from a combination of global and local properties. Its main contribution is that the model is fully unsupervised and achieves 70,3% F1 on the MUC-6. A result that is quite competitive compared to supervised systems.

The success of the statistical approaches for anaphora and coreference resolution has induced a large interest in this particular task and many novel works have been performed in order to enhance the performance of these systems, either following a supervised approach (Culotta et al., 2007; Bengtson and Roth, 2008; Finkel and Manning, 2008; Sapena et al., 2010) or an unsupervised one (Poon and Domingos, 2008).

However, very recently, a new family of algorithms have recovered the early deterministic approaches obtaining, in many cases, better performances than the state-of-the-art machine learning approaches. Haghighi and Klein

Joe wants to sell the house. Sue has **offered** one million.

Table 2.14: Example of Zero anaphora.

(2010); Raghunathan et al. (2010); Lee et al. (2011, 2013) presented modular systems that select the referent of a mention applying sets of constraints that check syntactic and semantic compatibilities.

2.3.4 Zero anaphora

In many cases, entities are not expressed explicitly in the text. However, as Shopen (1973) pointed out, some of these missing elements have an anaphoric interpretation. Consider the example in Table 2.14 taken from Saeboe (1996). In the second sentence, *Sue* offers *one million* for a commodity that is omitted. The interpretation of that sentence depends on capturing that this commodity is *the house* that *Joe wants to sell*. These kind of mentions are called **Zero anaphora** (Mitkov, 2002).

Shopen (1973) made a distinction between those missing elements that do not refer to any entity, called **indefinite**, and those **zero-anaphoric** mentions that can be interpreted by signalising their context. For the latter, called **definite**, two different problems must be solved. On the one hand, deciding when there really exists a non-expressed entity. On the other hand, the reconstruction of the elided fragments. In other words, finding the antecedent of the particular zero-anaphoric entity. First works in this line described strategies based on syntactic evidence (Weischedel and Sondheimer, 1982), on more conceptual approaches like **focus** exploration (Sidner, 1986), or combinations of both techniques (Díaz de Ilarraza Sanchez et al., 1990).

Since then, research on **Zero anaphora** resolution has focused in so called *pronoun-dropping* or *pro-drop* languages, where omitted pronouns are frequent, like Chinesse (Zhao and Ng, 2007; Yeh and Chen, 2007) and Japanesse (Iida et al., 2007; Iida and Poesio, 2011), and some romance languages as Italian (Iida and Poesio, 2011) and Spanish (Ferrández and Peral, 2000; Rello and Ilisei, 2009).

2.4 Event coreference resolution

Resolving the coreference of event mentions is a quite different task from its counterpart for entity mentions. For instance, entity mentions usually are named entities (e.g. *Steve Jobs*) which are referred by a fixed set of names (e.g. *The founder of Apple*, etc.) that are available in several knowledge sources containing properties about the entity (e.g. *male*, etc.) which helps coreference resolution. Furtheremore, events can be referred by nouns or verbs. For example, the verb **sell** and noun **sale** can refer to the same event instance.

But entity and event coreference do not only differ in the manner that the mentions of the instances are expressed, but also in their inner semantics and pragmatics. Previous to any computational approach, the similarity between two events have been studied from a philosophical perspective. Davidson (1969) proposed the first relevant theory on event identity. He argued that two events are the same if they have the same causes and effects. This theory was refuted by Quine (1985) who considered that, as each event can be well defined in space and time, two events are identical if they share the same participants and spatio-temporal location. The suggestion by Quine (1985) become the most accepted proposal on event identity after Davidson (1985) also agreed with it. In accordance with the Quine's theory, we consider that two event mentions are coreferential if they have the same event properties and share the same event participants. Figure 2.7, taken from Bejan and Harabagiu (2010), shows sentences including different mentions of two events of the general type **buy**. The first event, ***b1***, is referred in **Document 1** by the predicates *[buy]*$_{b11}$ and *[acquisition]*$_{b12}$. The other event, ***b2***, is mentioned in both **Document 2** and **Document 3** by the predicates *[buy]*$_{b21}$, *[purchase]*$_{b22}$ and *[acquire]*$_{b23}$.

The first approaches to automatic event coreference resolution present pairwise models. For example, Humphreys et al. (1997) compute a similarity measure between two events based on the semantic classes and ontological features of those events. In another pioneer work Bagga and Baldwin (1999) presented a simple method for solving cross-document event coreference matching two event as identical if they share the same lexeme or they are synonyms. However, these approaches require the design of domain specific constraints. More recently, the ACE 2005 task included a dataset for the evaluation of the automatic event coreference resolution restricted to a

Document 1:	AMD agreed to **[buy]**$_{b11}$ Markham, Ontario-based ATI for around \$5.4 billion in cash and stock, the companies announced Monday. The **[acquisition]**$_{b12}$ would turn AMD into one of the world's largest providers of graphics chips.
Document 2:	Hewlett-Packard is negotiating to **[buy]**$_{b21}$ technology services provider Electronic Data Systems. With a market value of about \$115 billion, HP could easily use its own stock to finance the **[purchase]**$_{b22}$.
Document 3:	Industry sources have confirmed to eWEEK that Hewlett-Packard will **[acquire]**$_{b23}$ Electronic Data Systems for about \$13 billion.

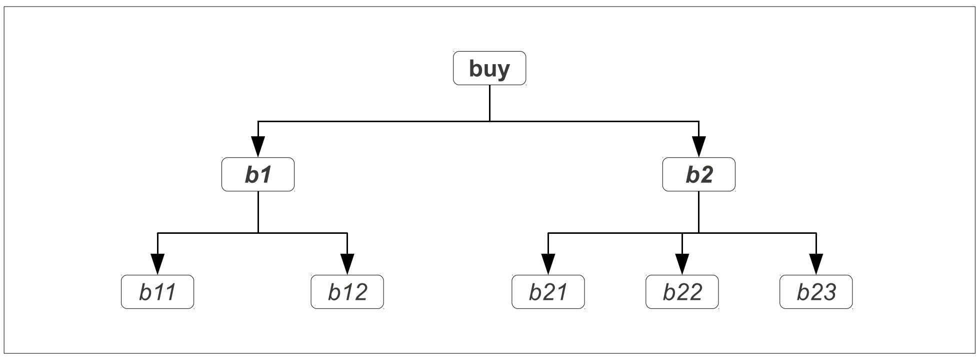

Figure 2.7: Intra and cross document event coreference.

small set of event types. The task proposed applied the usual metrics for entity coreference, as MUC (Vilain et al., 1995), B3 (Bagga and Baldwin, 1998) or CEAF (Luo, 2005). However, although the ACE dataset allows the formalization of the evaluation of the experiments not many works on event coreference have been developed. Some of the few works in this line are Ahn (2006) and Chen and Ji (2009).

In recent years, event-coreference has started receiving more attention. Bejan et al. (2009); Bejan and Harabagiu (2010) defined non-parametric bayesian models based on those proposed by Haghighi and Klein (2007) for entity coreference within and cross-document. They employed a combination of lexical, VerbNet class and WordNet features as well as semantic role structures. They evaluated their models on the ACE dataset and they obtained the results of 83.8% B3 and 76.7% CEAF F-score. But, due to the limited set of event types on that corpus and the fact that it does not include cross-document coreference annotations, they created a novel dataset called EventCorefBank annotating articles on 43 different topics from the GoogleNews archive. On their own dataset they reached a 90% B3 and 86.5% CEAF F-score for within document coreference and 80.4% B3 and 70.8% CEAF F-score for cross-document coreference. Chen et al. (2011) proposed a combination of three resolvers that are applied after an entity coreference resolution process. The chain of solvers works as follows. First, the results of the entity resolution solvers are used to reduce the false positive cases. Second, a mention-pair model is trained using a set of lexical, PoS, semantic and syntactic features. Finally, a coreference chain is formed using a globally optimized graph partitioning model. The strategy was evaluated on the OntoNotes 2.0 corpus obtaining a 46.91% B3 F-score.

At this point none of the approaches presented deal explicitly with partial-coreferent events. That is, when an event mention refers to only one part of another event (e.g. *pay* and *buy*). Bejan and Harabagiu (2010) noticed that not accounting for these cases is the reason for one of the most common errors in their output. Lee et al. (2012) incorporated partial coreference using distributional similarity as one of the features for cluster comparison. This approach performs a joint event and entity cross-document coreference resolution asserting that they influence one to another. In this work a SVM classifier is trained using a set of features similar to the set from Bejan et al. (2009); Bejan and Harabagiu (2010). The model achieved 62.7% MUC / 67.7% B3 / 33.9% CEAF / 71.7% BLANC F-score on an extended version of the ECB corpus. Cybulska and Vossen (2013) also faced partial-coreferent events matching hyponymic relations and granularity shifts. They combine granularity with similarity to model fine and coarse-grained matches across event descriptions that are likely to happen across different documents and sources. Their approach, that only acounts within document coreference, obtains on the ECB corpus 70% MUC / 72% B3 / 62% CEAF / 69% BLANC F-score. But hyponymy is not the only relation that can indicate partial event coreference, there exist more kinds of implications between the events and their participants. For example, LexPar (Coyne and Rambow, 2009) and FRED (Aharon et al., 2010) contain entailment relations derived from FrameNet such as ***Cause*** and ***Perspective_on***. These resources have proved their utility in datasets like ACE.

Disk drive was down at 11/16-2305. Has select lock.

Table 2.15: The *theme* of the predicate **has** is missing.

2.5 Implicit Semantic Role Labelling

Although recent works have shown that solving the implicit roles can extend significantly the semantic annotation, there is a remarkable lack of previous research in the literature. The first attempt for the automatic recovery of implicit semantic information was presented by Palmer et al. (1986). This paper contains a description of how the different components of the PUNDIT system interact to make explicit the implicit information. PUNDIT considers that some empty syntactic constituents and unfilled semantic roles may be implicit entities. So, once syntax and semantics capture these missing entities, PUNDIT lets its pragmatic component solve them as a coreference resolution task. Consider the example in Table 2.15 taken from Palmer et al. (1986).

PUNDIT analyses these sentences one by one. For the first one the system performs syntactic and semantic processing obtaining, for instance, for the predicate **be** the role *theme* filled by the entity *disk drive*. PUNDIT continues with the second sentence but in this case the semantic module identifies that the role *theme* for the predicate **has** is unfilled. Then the pragmatic module treats the empty subject as a pronoun and finds that the only possible referent according with some semantic selectional constraints is *disk drive*. In order to perform all this process Palmer et al. (1986) manually defined a set of logic rules that work along with some selectional preferences taken from a manually created domain specific knowledge repository.

In a very similar fashion, Whittemore et al. (1991) made use of an extended version of the Discourse Representation (Kamp, 1981) to treat missing roles, called by the author as *open roles*, as a particular case of anaphora. In the example in Table 2.16, included in Whittemore et al. (1991), the event **buy** of the first sentence has no entity associated with its corresponding thematic role *seller*.

However, the missing role is considered as an anaphoric mention whose proper referent entity is introduced in the next sentence. As *the salesman* satisfies the semantic properties of the thematic role *seller* it can be anchored

Pete bought a car. The salesman was a real jerk.

Table 2.16: The *seller* of the predicate **buy** is not explicit.

to the anaphora hypothesized in the previous sentence and, consequently, it can be annotated as the filler of the missing role. Whittemore et al. (1991) proposed some rules to cover different cases of implicit roles and, like Palmer et al. (1986), defined semantic properties for a specific set of thematic roles that can be used as selectional restrictions.

More recently, Tetreault (2002) described another automatic method that follows closely the previous approaches. In this case the proposal relies on a notion of focus, the approach used traditionally for anaphora resolution. The most relevant fact of this work is that it is the first attempt to evaluate empirically an implicit semantic role labelling algorithm. The author used a subset of the TRAINS-93 Corpus with coreference annotations and manually extends the corpus with implicit reference roles. Unfortunately, this was a very preliminary work that was not taken up again. The resulting corpus consists of only 86 short sentences with less than 10 words each and only a very reduced set of verbs and thematic roles (*Instrument*, *Theme*, *From-Location* and *To-Location*) were taken into account.

At this point all the systems developed for implicit semantic role labelling are very similar and are built following the same assumption. That is, the implicit roles are indeed a special case of anaphora or pronoun resolution. According to this hypothesis the approaches proposed are basically a combination of coreference systems that make use of some semantic knowledge that is exploited to set selectional preferences and logic rules for each thematic role. Nevertheless, in the previous works the construction of these constraints is conducted by manual effort over very specific domains or a very small set of events and thematic roles, resulting in quite limited systems.

Due to its complexity and the lack of knowledge sources or annotated corpora, the task of implicit semantic role labelling has not received enough interest for long time. However, over the last few years it has been taken up again around two different proposals. On the one hand, Ruppenhofer et al. (2009) presented a task in SemEval-2010 on *Linking Events and Their Participants in Discourse* that, besides the traditional semantic role labelling, included an implicit role identification challenge based on FrameNet (Baker et al., 1998). For developing the corpus for this task some chapters were chosen from Arthur Conan Doyle's novels and annotated with semantic information following the FrameNet paradigm. Specifically, the annotation of implicit roles covered a wide variety of nominal and verbal predicates, each one having only a small number of instances. Three systems were presented but only two faced the implicit semantic role labelling subtask obtaining quite poor results (F1 below 0,02) according to Ruppenhofer et al. (2010). VENSES++ (Tonelli and Delmonte, 2010) applied a rule based anaphora resolution procedure and semantic similarity between candidates and thematic roles using WordNet (Fellbaum, 1998). The system was tuned in Tonelli and Delmonte (2011) improving slightly its performance. SEMAFOR (Chen et al., 2010) is a supervised system that extends an existing semantic role labeler to enlarge the search window to other sentences, replacing the features defined for regular arguments with two new semantic features. Although this system obtained the best performance in the task, data sparseness strongly affected the results. Besides the two systems presented to the task, some other systems have used the same dataset and evaluation metrics. Ruppenhofer et al. (2011) focused on the correct identification of those missing roles that are actually implicit. They proved that improvements in this step can provide significant overall gain even when the procedure of finding the proper filler of the role is simple. However, the authors did not proposed any novel approach for this latter step. Gorinski et al. (2013) built four different resolvers based on semantic role labelling and coreference resolution, each trying to draw different aspects. The approach was weakly supervised and obtained quite competitive results. Although the significant gain obtained by the last works on the SemEval-2010 corpus, all attempts performed over this dataset do not reach even 0,3 F1. Besides the mentioned data sparseness, these poor results can also be caused by the nature of the selected documents. Being chapters from novels several dialogues are included, increasing the complexity of the task.

On the other hand, Gerber and Chai (2010, 2012) presented and studied a novel corpus extending the semantic role annotation of the documents from Penn TreeBank(Marcus et al., 1993) that were already annotated for PropBank and NomBank. Because the manual annotation of implicit roles requires a huge effort, the authors decided to focus just on a small set of ten predicates[17] that satisfy a set of criteria. All of them have an unambiguous sense, derive from a verb, have a high frequency in the corpus and express many implicit arguments. For each of the chosen predicates the number of instances annotated is much larger than the dataset from SemEval-2010 task. This allowed to avoid the data sparseness problem. The annotation process revealed that the addition of implicit arguments can increase **71%** the number of roles across all the instances. Gerber and Chai (2010, 2012) also proposed a fully supervised model for automatic implicit semantic role labelling. In particular, they used a huge set of syntactic, semantic and coreferential features[18] to train a logistic regression classifier. The results of this system were far better than those obtained by the systems using the SemEval-2010 dataset. However, many of the most important features are lexically dependent on the predicate and cannot been generalized. Thus, the applicability of the model is limited to the small set of predicates analysed in these works.

The previous work inspired LIAR$_c$ (Peris et al., 2013), a system for the automatic annotation of the implicit arguments of deverbal nominalizations in Spanish. The model was successfully trained and evaluated on a subset of the IARG-AnCora (Taulé et al., 2012) with an overall F-Measure of 89.9%. Taulé et al. (2012); Peris et al. (2013) pointed out very interesting differences between implicit roles in English and Spanish, like, for instance, that the frequency of elided arguments is much higher in Spanish.

The model by Gerber and Chai (2010, 2012) has been also adapted and applied for the dataset of the task of SemEval-2010 by Silberer and Frank (2012). Due to the lack of training data the authors also proposed a heuristic-based methodology to extend automatically the training corpus. Exploiting the additional data obtained, their system was able to improve the previous results but once again the performance of this model remains below 0,3 F1. Finally, following the same approach, Moor et al. (2013) presented a corpus of predicate-specific annotations for verbs in the FrameNet paradigm that are aligned with PropBank and VerbNet and could help to improve the automatic annotation in future works. In this line, Feizabadi and Padó (2014) described a study to prove that that implicit roles could be annotated as well as locally realized roles in a crowd-sourcing set-up.

Finally, recent works have widen the scope and the targets of ISRL. For instance, Blanco and Moldovan (2014) focused on missing semantic roles

[17] *bid*, *sale*, *loan*, *cost*, *plan*, *investor*, *price*, *loss*, *investment*, *fund*

[18] Gerber and Chai (2012) includes a set of 81 different features.

within the same sentence of the predicate. Instead of core arguments, that were the goal of previous works, they run their approach to modifiers, including *TIME*, *LOCATION*, *MANNER*, *PURPOSE* and *CAUSE*. On the other hand, Stern and Dagan (2014) investigated implicit semantic roles in the the context of textual inference scenarios. For this purpose, they developed a novel dataset and propose some methods for a task that differs substantially the labelling of roles. In fact, it consisted in answering, *yes* or *not*, if a proposed candidate could be the implicit filler of a predicate according to some facilitated prior information that, according to the authors, eases the task.

As show in this section, the most recent proposals to the implicit semantic role labelling still follow strategies similar to those proposed by early works. The introduction of new corpora and the use of statistical and machine learning methods have allowed a deeper study of the task including more precise empirical evaluations. However, the hypothesis behind both early and recent approaches has remained substantially the same, namely, the implicit roles consist in a particular case of anaphora and can be recovered applying methods that combine entity coreference resolution and traditional semantic role labelling. Moreover, these approaches have shown that they need large amount of manually annotated training-data, otherwise they offer poor performances.

IMPLICIT SEMANTIC ROLES IN DISCOURSE

CHAPTER 3

A framework for Implicit Semantic Role Labelling

This chapter presents a summary of the complete framework of the research developed in this work. Section 3.1 presents a new general model for Implicit Semantic Role Labelling (ISRL). This is the model we unfold in subsequent chapters of the dissertation. In particular, the new model exploits different correferential elements which appear in the context of the predicate mentions. Section 3.2 briefly introduces the techniques we propose to overcome the need of developing large and costly annotated corpora for ISRL. In Section 3.3 we relate the new general model for ISRL described in this chapter with respect to the particular approaches for ISRL that are explained in the rest of chapters of this dissertation. Section 3.4 presents two existing datasets we use to evaluate our systems. The first one is based on FrameNet and the second on PropBank/NomBank. Finally, Section 3.5 introduces the scoring metrics used to evaluate ISRL systems

3.1 A new model for ISRL

Implicit roles are those participants of the event mentions that do not appear in the near context of their predicates. These *implicit* roles are not captured by traditional SRL systems Gildea and Jurafsky (2002) which only focus on the *explicit* ones. That is, those that have a direct syntactic relationship with their predicates. However, implicit roles usually appear expressed in the surrounding context of the predicate mentions and can be identified by interpreting the discourse. This anaphoric phenomena is very frequent in natural language because human writers tend avoid repetitions of the same information and rely on the inference capabilities of the readers for understanding the text.

In order to ilustrate our new model for Implicit Semantic Role Labelling (ISRL), Figure 3.1 presents a set of predicates (P_n) and their roles (R_n) in a discourse, some of which are *implicit* (marked by dotted lines). For instance, the predicate P_2 has an *explicit* role R_2 and an *implicit* one R_1. The figure shows that these missing roles may refer to anaphoric roles associated to other predicates in the discourse. For instance, in Figure 3.1, the *implicit* role R_3 of the predicate P_3 corresponds to the *explicit* role R_3 of predicate P_4 which appear later in the discourse. Moreover, the predicates sharing the same roles can also be related because they refer to the same event or to different but closely related events. For instance, the former case (i.e. same event) is represented by the *equal* $=$ relation between predicates P_3 and P_4. The later case (i.e. closely related event) is represented by the $\sim$ relation between predicates P_2 and P_3.

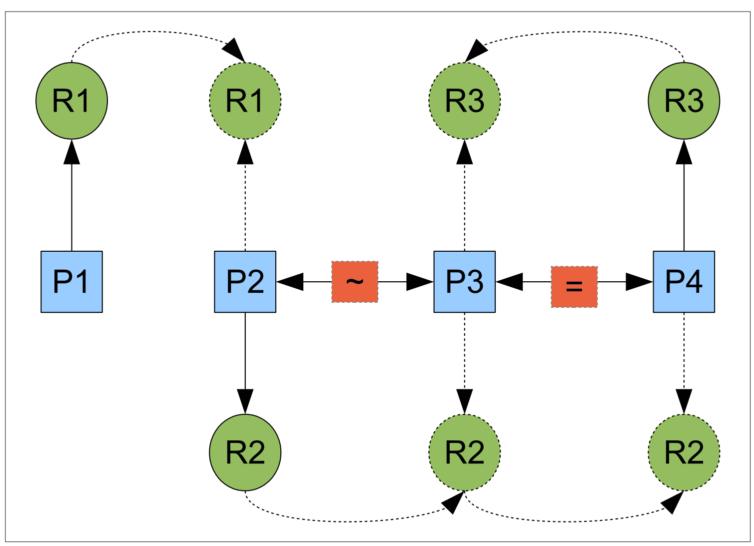

Figure 3.1: Our general model for ISRL.

Quest Medical Inc said it adopted [arg_1 a shareholders' rights] [np **plan**] in which rights to purchase shares of common stock will be distributed as a dividend to shareholders of record as of Oct 23.

Table 3.1: In this sentence the arg_0 of the predicate **plan** is missing.

Thus, the correct identification of the fillers of the implicit roles of a predicate consists on locating the proper elements in the context of the predicate that allows to complete the semantic interpretation of the event triggered by the predicate. As many other tasks in NLP, ISRL relies on the proper identification of the reference information expressed explicitly in other parts of the discourse.

In general, implicit roles have been considered as a particular case of anaphora (Palmer et al., 1986; Whittemore et al., 1991; Tetreault, 2002). For that reason, previous works on ISRL have commonly faced this task as an hybridization of traditional SRL and correference resolution (Silberer and Frank, 2012; Gerber and Chai, 2010). In these approaches, a missing role is considered as a not expressed mention of a referent entity. This assumption relates closely the implicit role resolution with zero-anaphora. The example in Table 3.1 shows a case where the filler of the implicit arg_0 of the predicate **plan**, *Quest Medical Inc*, could be recovered by this kind of techniques. The Chapter 5 of this document contains a system that follows this idea.

However, we believe ISRL can be also faced focusing on the predicates that have missing arguments. Figure 3.1 shows how in some cases the role of a predicate is not expressed because it has been already explicitly introduced associated to other related predicates. For example, this can be the case of role R_3 of the predicate P_3 which refer to the same event described by the predicate P_4 having the *explicit* role R_3. Dahl et al. (1987) suggested that pronouns that refer to predicates could acquire the same argument structures. This setting changes the approach from previous proposals because, the anaphora resolution is focused on the predicates, not on the roles. In other words, this strategy is based on event coreference and it can be extended not only to transfer annotations to pronouns but to solve missing roles of those predicates that refer to the same events. The suitability of applying event coreference for ISRL can be seen in the example of the Table 3.2. In this case, both mentions of the predicate **loss** refer to the same event and ob-

[arg_0 The network] had been expected to have [np **losses**] [arg_1 of as much as $20 million] [$arg_3$ on baseball this year]. It isn't clear how much those [np **losses**] may widen because of the short Series.

Table 3.2: In the second sentence the predicate **losses** does not have eny explicit role.

viously, they share the same arguments. Chapter 6 describes an algorithm that includes a basic event coreference strategy for Implicit Semantic Role Labeling.

The relation with equal sign between predicates P_3 and P_4 in Figure 3.1 means that these predicates are referring exactly to the same event. Thus, they possibly share their arguments, being these expressed (explicit) or not (implicit). That is, if predicate P_3 and predicate P_4 refer to the same event, their roles should be also the same[1]. Thus, role R_2 of predicate P_3 and role R_2 of predicate P_4 should contain the same fillers. Similarly, role R_3 of predicate P_3 and role R_3 of predicate P_4 should be associated to the same fillers.

Moreover, the same event can be expressed from different perspectives. This could be the case of predicates such as **sell** and **buy**. In Figure 3.1, this corresponds to the relation with the $\sim$ sign between predicates P_2 and P_3. In this case, these predicates can share some roles but not all of them. For instance, both predicates share role R_2 but they do not share neither R_1 nor R_3.

Furthermore, event coreference chains can also cover predicate mentions that refer to parts of a complex event or to other semantically related events. This could be the case of predicates such as **pay** and **buy**. Although in these cases the semantic roles could be different, their participants should be the same. Covering these relations between predicates and roles is another challenge for ISRL. The example in Table 3.3 illustrate this type of scenario. In this example, all the arguments of the predicate **price** are implicit. This case can be solved knowing that for every **buy** event there is implied a **price** event and knowing how their roles are interrelated. The influence of these kind of relations for ISRL is studied in Chapter 7.

Our research covers the general model descibed in this chapter, with spe-

[1]When the discourse does not contain contradictory statements.

[arg_0 He] has [arg_1 four stocks] in mind to [vp **buy**] if the [np **prices**] drop the level he wants.

Table 3.3: In this sentence all the roles of the predicate **price** are omitted.

cial emphasis on studiying existing relations between predicates. Of course, there exists even more complex situations. In Figure 3.2, the non-explicit role R_2 of the predicate P_2 refers to R_1 that belongs to the predicate P_1. But in this case the predicate P_1 is also not expressed.

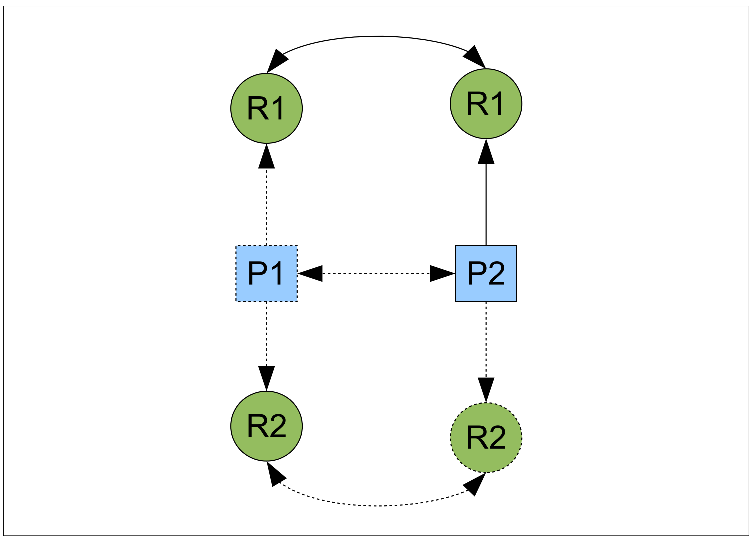

Figure 3.2: Both role *R2* and predicate **P1** are elided.

For instance, in the example of Table 3.4, the filler of the argument arg_0 of the nominal predicate **sale** is the *seller* of the event. However, the filler of arg_1, the *thing sold*, is not explicitly expressed. However, the noun phrase *computer companies* implies that there exist an implicit predicate **sell** where the argument arg_0 is *companies* and the argument arg_1 is *computers*. Thus, *computer*, the filler of argument arg_1 of predicate **sales** can be recovered. This kind of advanced inferencing can provide the clue to solve many implicit roles, but their study exceeds the scope of the current dissertation.

..., and [arg_0 computer companies] in general are experiencing slower [np **sales**].

Table 3.4: The arg_1 of the predicate **sales** in this sentence is not explicit.

3.2 Dealing with data sparsity

The most difficult subtask for any ISRL system and, consequently, the focus of the main body of our research, is the identification of the actual fillers of each inplicit role. On this subtask, the most successful approaches have been the supervised machine learning systems of Gerber and Chai (2010, 2012) and Silberer and Frank (2012). These methods, which are very similar each other but adapted to different datasets, rely on a set of features that combines entity coreference and SRL. But many of these features, specifically those related with SRL, strongly depend on the predicate lemma, in the case of PropBank/NomBank based systems, or in the name of the frame or the frame-element, in the case of FrameNet based ones. This point would no be really an inconvenience if there was available sufficient amount of corpora annotated with implicit roles. Unfortunately, this is not the case. On the one hand, available training dataset based on FrameNet is very limited and contains a few number of implicit roles. On the other hand, the training set developed for PropBank/NomBank only contains annotations for just a set of ten different nominal predicates.

Obviously, the limited coverage of available training corpora makes the supervised ISRL systems useless for real world applications. The features that those system exploit have proved to be really effective for ISRL but their use can not be easily extended to other predicates without additional training data. Moreover, the annotation effort of this data for all predicates and languages could be totally prohibitive. For that reason, in order to include implicit roles in an automatic natural language processing chain it is desirable to explore more generic approaches, avoiding lexical-dependant features, or approaches that do not require training data. This is why our research focuses on unsupervised or deterministic systems.

Thus, our research proposes three different main strategies:

- **Generalizable features**. As already said, lexical-dependant features need a large quantity of annotated instances for each predicate. Therefore, we focus on describing features that can capture the implicit roles characteristics in a more general way. Moreover, we also devised ways to exploit available annotations for explicit roles to learn useful information for ISRL.

- **Knowledge based systems**. SRL and ISRL systems both rely on existing large-scale semantic resources such as PropBank/NomBank or FrameNet. Furthermore, these semantic resources are also interconected to other semantic knowledge bases such as VerbNet (Kipper et al., 2000; Kipper, 2005) or WordNet(Fellbaum, 1998) through SemLink (Palmer, 2009) or the Predicate Matrix (López de Lacalle et al., 2014b, a). We can devise new ways to exploit these resources to derive a very rich set of semantic relations between predicates and roles. Although taking advantage of this information is not always trivial, it can be of utility for existing ISRL systems.

- **Deterministic algorithms**. In recent years the interest on unsupervised approaches has been increasing in many NLP tasks. A great number of works present models trained using huge amounts of raw data. Some other works develop algorithms that recover deterministic proposals based on simple but robust linguistic principles. As this kind of rule-based systems have proved to obtain very successful results for corefence resolution (Raghunathan et al., 2010), we expect also good performances for ISRL.

3.3 ISRL approaches

The combination of our model on ISRL explained in Section 3.1 and the strategies proposed to overcome the need of developing large and costly annotated corpora for ISRL descibred in Section 3.2 produces as a result the different approaches for ISRL of the current dissertation:

1. According to previous works that relate ISRL with entity coreference, we study a set of features and methods commonly used for anaphora

and coreference resolution in order to adapt them for training a lexical independent model for ISRL. This approach is described in Chapter 5.

2. Since focusing just on the characteristics of the coreferent entities may be insufficient to match the referents of elided roles, we present a novel deterministic method that includes a simple strategy for event coreference and does not require training data. Chapter 6 describes the resulting algorithm and experiments.

3. Obviously, more complex semantic event relations connect the participants in a discourse. Discovering these relations predicates and roles in the text can help to discover the fillers of the implicit arguments. Chapter 7 describes an extension of the previous approach using a new lexical knowledge base derived from FrameNet that relates different PropBank/NomBank predicates with their roles.

3.4 Datasets

The experiments presented in this research use two different datasets. On the one hand, Ruppenhofer et al. (2010) presents a task in SemEval-2010 that includes an implicit argument identification challenge based on FrameNet (Baker et al., 1998). The corpus for this task consists of some novel chapters. They cover a wide variety of **nominal** and **verbal** predicates, each one having only a small number of instances. On the other hand, Gerber and Chai (2010, 2012) study the implicit argument resolution on NomBank (Meyers et al., 2004). Unlike the dataset from SemEval-2010 (Ruppenhofer et al., 2010), in this work the authors focus just on a small set of ten predicates. But for these ten nominal predicates, they annotate a large amount of instances in the documents from the Wall Street Journal that were already annotated for PropBank (Palmer et al., 2005) and NomBank.

3.4.1 SEMEVAL-2010 dataset

The corpus released in SemEval-2010 for Task 10 *Linking Events and their Participants in Discourse* contains some chapters extracted from two Arthur

Conan Doyle's stories. *The Tiger of San Pedro* chapter from *The Adventure of Wisteria Lodge* was selected for training, while chapters 13 and 14 from *The Hound of the Baskervilles* were selected for testing. The text is written in first person and, as is common in novels, contains lots of quoted sentences that are usually part of dialogues between different characters. Table 3.5 presents an excerpt of chapter 13 of *The Hound of the Baskervilles*.

The texts are annotated using the frame-semantic structure of FrameNet 1.3. Those *Core frame-elements* not present in the sentences are annotated as *Null Instantiations* (NI). The annotation also includes the type of the NI and the recoverable fillers for each *Definite Null Instantiations* (DNI). Table 3.6 shows the number of sentences, DNI and explicit frame-elements (FE) in the dataset. The annotators of this dataset did not stablish any kind of restriction, resulting a very diverse set of nominal and verbal predicates, but having a low number of instances per predicate, resulting in a very sparse set of NIs. Moreover, the Null Instantiations in the training set is even smaller than the number of annotations in the testing set.

All the documents are enriched with a constituent-based parsing and for the training document there are also manual coreference annotations available. The dataset includes the annotation files for the lexical-units and the full-text annotated corpus from FrameNet. Although, the task was mainly focused on FrameNet, the dataset also contains annotations in PropBank style. However, due to the differences between those resources the annotations are not fully comparable.

The main problem of this corpus is its sparseness. There are only three documents in the training data and they do not cover a large number of Null Instantiations. Moreover, the nature of the documents (extracts of novels in first person with many dialogues) and its large size per document (changing the focus of the discourse quite frequently) makes this dataset quite difficult to process. However, it covers a wide number of different predicates including verbs and nouns, making possible to study them separately. In addition, the inclusion of a gold-standard coreference annotation in the training dataset allows to exploit some interesting features. The work in Chapter 4 and Chapter 5 use this dataset for evaluation.

3.4.2 Beyond NomBank

Gerber and Chai (2010, 2012) developed a different dataset (hereinafter BNB

Sir Henry was more pleased than surprised to see Sherlock Holmes, for he had for some days been expecting that recent events would bring him down from London. He did raise his eyebrows, however, when he found that my friend had neither any luggage nor any explanations for its absence. Between us we soon supplied his wants, and then over a belated supper we explained to the baronet as much of our experience as it seemed desirable that he should know. But first I had the unpleasant duty of breaking the news to Barrymore and his wife. To him it may have been an unmitigated relief, but she wept bitterly in her apron. To all the world he was the man of violence, half animal and half demon; but to her he always remained the little wilful boy of her own girlhood, the child who had clung to her hand. Evil indeed is the man who has not one woman to mourn him. "I've been moping in the house all day since Watson went off in the morning," said the baronet. "I guess I should have some credit, for I have kept my promise. If I hadn't sworn not to go about alone I might have had a more lively evening, for I had a message from Stapleton asking me over there." "I have no doubt that you would have had a more lively evening," said Holmes drily. "By the way , I don't suppose you appreciate that we have been mourning over you as having broken your neck?"

Table 3.5: A section of chapter 13 of *The Hound of the Baskervilles.*

which stands for *Beyond NomBank*) extending existing predicate annotations for NomBank and ProbBank. BNB presents the first annotation dataset of implicit arguments based on PropBank and NomBank frames. This annotation is an extension of the standard training, development and testing sections of the Penn TreeBank. The Penn TreeBank have been typically used for Semantic Role Labelling (SRL) evaluation and has been already annotated with PropBank and NomBank predicate structures. The documents are 500 articles (400 for the training set and 100 for the testing set) obtained from those of the Wall Street Journal corpus related to the domain of economy. They are written in a common journalistic style and tend to be quite concrete and brief, each document contains an average of around 21 sentences, and can include some quotations. Table 3.7 presents an example of the BNB corpus.

The authors selected a limited set of predicates. These predicates are

data-set	#sentences	DNIs (solved)	Explicit FE
train	438	303 (245)	2,726
test-13	249	158 (121)	1,545
test-14	276	191 (138)	1,688

Table 3.6: Number of DNI and Explicit FE annotations for the SemEval-10 Task-10 corpus.

Investors who bought stock with borrowed money – that is, "on margin" – may be more worried than most following Friday's market drop. That's because their brokers can require them to sell some shares or put up more cash to enhance the collateral backing their loans. In October 1987, these margin calls were thought to have contributed to the downward spiral of the stock market. Typically, a margin call occurs when the price of a stock falls below 75% of its original value. If the investor doesn't put up the extra cash to satisfy the call, the brokerage firm may begin liquidating the securities.
But some big brokerage firms said they don't expect major problems as a result of margin calls. Margin calls since Friday "have been higher than usual, but reasonable, " a spokesman for Shearson Lehman Hutton Inc said.

Table 3.7: A section of the document *wsj_2393* from the Wall Street Journal corpus.

all **nominalizations** of other verbal predicates, with no sense ambiguity, that appear frequently in the corpus and tend to have implicit arguments associated with their instances. These constraints allowed them to model enough occurrences of each implicit argument in order to cover adequately all the possible cases appearing in a test document. For each missing argument position they revised all the preceding sentences and annotated all mentions of the filler of that argument. Table 3.8 shows the list of predicates and the resulting figures of this annotation. From left to right, number of instances and implicit arguments per predicate in the whole dataset (including train and test) and in the test set.

As this corpus is the same that the one distributed in the CoNLL SRL tasks, it is also possible to integrate the information from these datasets. Specifically, the documents of the CoNLL-2008 task Surdeanu et al. (2008) include both manual and predicted annotations for syntactic dependencies,

	Full		Test	
Predicate	**Inst.**	**Imp.**	**Inst.**	**Imp.**
sale	184	181	64	60
price	216	138	121	53
bid	88	124	19	26
investor	160	108	78	35
cost	101	86	25	17
loan	84	82	11	9
plan	100	77	25	20
loss	104	62	30	12
fund	108	56	43	6
investment	102	52	21	8
Overall	1,247	966	437	246

Table 3.8: BeyondNomBank annotation figures. Columns 2 and 4 gives the number of predicate instances in the dataset. Columns 3 and 5 indicates the number of implicit arguments per predicate.

named entities and super-sense labels as semantic tags (Ciaramita and Altun, 2006).

This dataset covers a large number of documents. Although taking into account that just a short set of nominal predicates limits the analysis of the experiments, the number of instances of implicit arguments per each of these predicates can result on quite a good confidence on the final results. The style of the texts and, specially, their small size makes the information contained in one document much concise and precise. This feature allows to study new strategies for ISRL as those explained in Chapter 6 and Chapter 7.

3.5 Scorers

Unlike in traditional semantic role labelling tasks, the gold standard fillers of the implicit semantic roles are not always unique. In many cases, the elided arguments refers to entities that are mentioned more than once in the discourse. For this reason, any ISRL dataset must include full coreference chains for every entity being a gold-standard filler of an implicit role. That is, an answer provided by an ISRL system should be taken as correct if it matches any of the possible mentions of the actual filler.

The proposal by Ruppenhofer et al. (2010) defines the NI linking *Precision* as the number of all true positive links divided by the number of links made by a system. NI linking *Recall* is defined as the number of true positive links divided by the number of links between a NI and its equivalence set in the gold standard. NI linking $F-Score$ is then calculated as the harmonic mean of precision and recall.

However, since any prediction including the head of the correct filler is scored positively, selecting very large spans of text would obtain very good results[2]. For example, [*madam*] and [*no good will, madam*] would be evaluated as positive results for a [*madam*] gold-standard annotation. Therefore, the scorer also computes, for those cases when the head of the gold-standard has been correctly matched, the overlap (Dice coefficient) between the words in the predicted filler (P) of an NI and the words in the gold standard one (G):

$$\textbf{Dice coefficient} = \frac{2\,|P\ \cap\ G|}{|P|\ +\ |G|} \tag{3.1}$$

On the other hand, Gerber and Chai (2010, 2012) follow the same proposal but they perform the evaluation slightly differently. For every argument position in the gold-standard their scorer also expects a single predicted constituent to fill in. In order to evaluate the correct span of a prediction, the result is scored using the same Dice coefficient presented above over all the mentions of the gold-standard missing argument. Then, the highest value is chosen. If the predicted span does not cover the head of the annotated filler, the scorer returns zero. Then, *Precision* is calculated by the sum of all prediction scores divided by the number of attempts carried out by the system. *Recall* is equal to the sum of the prediction scores divided by the number of actual annotations in the gold-standard. As usual, $F - Score$ is calculated as the harmonic mean of recall and precision.

Table 3.9 shows how differently these two strategies score the same case. Suppose that the only annotation in the gold-standard is the [*madam*] from the previous example and a hypothetical system returns [*no good will, madam*] as its prediction. Assuming that the answer of the system has five words, including the *comma*, the results given by the two scoring methods would be as follows:

[2] In particular, returning the whole document would obtain perfect precision and recall.

	Precission	**Recall**	**F-Score**	**Overlap**
Ruppenhofer et al. (2010)	1.0	1.0	1.0	0.3
Gerber and Chai (2010)	0.3	0.3	0.3	-

Table 3.9: Diffences between the two scoring strategies. The method by Gerber and Chai (2010) does not compute the Overlap separately.

Obviously, the Gerber and Chai (2010) scorer is more strict than the one designed in Ruppenhofer et al. (2010) because it provides a better correlation with respect to the errors performed. However, in the following chapters, in order to obtain a proper comparison with respect state-of-the-art systems, we use the evaluation approach associated to its corresponding dataset. That is, we apply the scorer proposed by Ruppenhofer et al. (2010) in Chapter 4 and Chapter 5, and the scorer defined by Gerber and Chai (2010, 2012) in Chapter 6 and Chapter 7.

CHAPTER 4

First steps for Implicit Semantic Role Labelling

This chapter presents some first considerations on implicit semantic role labeling (ISRL). Section 4.1 describes our general approach for ISRL. That is, first selecting the missing roles of a predicate mention and second, discovering in its surrounding context the actual argument fillers of the missing roles. This section also describes the main differences between FrameNet and PropBank schemas when facing the first step of the general approach for ISRL. Section 4.2 presents an initial evaluation of a basic system on the FrameNet dataset for detecting the missing roles of a predicate mention. The chapter finishes with some concluding remarks in Section 4.3

4.1 Labelling implicit roles

In general, any system developed for Implicit Semantic Role Labeling (ISRL) must solve two essential subtasks:

1. Selecting the missing roles of a predicate mention that should be processed in the second step

For private-sector union workers, the cost of wages and benefits rose 0.9% in the third quarter. For non-union workers, the **costs** rose 1.4%.

Table 4.1: Example of the surrounding context of a mention of the predicate **cost**.

2. Discovering in its surrounding context the actual argument fillers of the implicit roles selected in the previous step.

In fact, the most difficult subtask of any ISRL system and, consequently, the focus of the main body of our research, is the second step. That is, the identification in their surrounding context of the actual fillers of each implicit roles. However, we need to address the first step before. The current chapter describes the main differences between FrameNet y PropBank schemas when facing this step.

4.1.1 Detecting the missing roles of a predicate mention

Depending whether we use FrameNet or PropBank schemas and datasets, the first step of the general approach for ISRL can become a much more complex task. Table 4.1 presents an example to illustrate the first step of a general ISRL process.

Table 4.2 presents the role structure of the predicate **cost.n** according to NomBank. Using this schema, the two steps designed to address the ISRL can be instantiated as:

1. Selecting the missing PropBank/NomBank arguments of a predicate mention that should be processed in the second step.

2. Discovering in its surrounding context the actual argument fillers of the implicit PropBank/NomBank arguments selected in the previos step.

arg_1	*Commodity*
arg_2	*Price*
arg_3	*Buyer*

Table 4.2: Role structure of the predicate **cost.01** in NomBank.

Checking its structure, we can easily discover that arg_1-*Commodity* and arg_2-*Price* have not been realized in its sentence boundaries. The argument arg_1-*Commodity* could be filled with [*the wages and benefits*] of the previous sentence and be considered as implicit. With respect the argument arg_2-*Price*, its filler is not mentioned at all.

When processing NomBank/PropBank, we follow a very simple strategy. We rely on the predicate-frame description to detect the missing roles of a predicate mention. That is, we perform the second ISRL subtask for all core arguments of the predicate structure of the mention which are not explicitly captured by a traditional SRL system. In our example, we perform the second step of the ISRL task for both arg_1-*Commodity* and arg_2-*Price*. That is, we select all PropBank/NomBank arguments of all predicate mentions to be considered by the second step of the general ISRL process.

On the other hand, using the FrameNet schema, the two steps designed to address the ISRL can be instantiated as:

1. Detecting which are the missing roles (or DNIs) of a predicate mention (or LU) that should be processed in the second step.

2. Discovering in its surrounding context the actual fillers of the DNIs detected in the previos step.

However, in FrameNet (Baker et al., 1998), the structures of the frames evoked by the predicates, or *lexical-units* (LU), are more complex. For example, the set of possible roles tend to be much larger. These roles or *frame-elements* (FE) involve both essential participants of the frame, *Core frame-elements*, and extra thematic pieces of information, *Non-Core frame-elements*. For the predicate **cost.n** of the previous example the role structure of the frame it evokes, **Expensiveness**, is presented in Table 4.3.

Expensiveness
Core:
Assert
Goods
Intended_event
Payer
Non-core:
Degree
Origin
Rate
Time

Table 4.3: Role structure of the predicate **cost.n** of the frame **Expensiveness** in FrameNet.

A large number of FE per frame introduces an extra level of difficulty. Not all of them will appear neither explicit nor implicit. Applying the **Expensiveness** frame to the previous case, the FE *Payer* would result filled explicitly by [*non-union workers*] and *Goods* implicitly by [*wages and benefits*]. This process leaves seven remaining FEs as unfilled. Four of them non-core. Obviously, trying to solve all of them would increase the probability of introducing a large number of mistakes. Moreover, FrameNet defines some relations between FEs to stablish their simultaneous compatibility/incompatibility. For instance, between the FEs *Goods* and *Intended_event* the relation ***Excludes*** is defined. This relation means that if one of the FEs is present for a LU the other one is not supposed to be present. Ruppenhofer et al. (2010) pointed out the convenience of including an additional step for selecting the roles that should be considered on the filling process and proposed some suggestions for exploiting the ontological knowledge of FrameNet. Ruppenhofer et al. (2011) applied these recommendations and proved empirically their utility.

However, the fixed set of rules that can be obtained from FrameNet are not always able to capture completely the interaction between the FEs. For example, applying the approach of Ruppenhofer et al. (2010) to the previous example, only the Core FEs would be taken into account. This would let just the roles*Goods*, *Intended_event* and *Assert* to be processed by the ISRL system. Furthermore, if the FE *Goods* is filled, it would exclude the FE *Intended_event*. Although the inclusion of these constraints could reduce the possibility of committing mistakes, we devised a new filtering process triying to overcome the limitations of the existing constraints.

This section presents a new strategy that relies on the hypothesis that the patterns of explicit realizations of roles can be learned and used to perform the filtering process. For instance, from the patterns of explicit occurrences of FEs of the frame **Expensiveness**, it can be learned that when the FE *Payer* is present, the most common pattern is the one that includes just the FEs *Payer* and *Goods*. According to this, in the case presented above, only the FE *Goods* should be considered in the search of implicit roles, reducing dramatically the risk of mistakes.

Following this approach, we present a system that performs, not only the DNI identification step but also a very simple DNI filling process. The approach just exploits the annotations of the explicit roles in both cases. However, this system is focused on the first step of the ISRL process.

4.2 A first model for DNI resolution

As said in Chapter 3, the sparseness of the training data in the SemEval-2010 dataset presents a very challenging problem. In previous works, as the one by Chen et al. (2010), the authors complain that the total number of annotations for DNIs is small to train supervised systems.

Thus, we propose that the explicit annotations can be exploited for the *Null Instantation* (NI) resolution. In particular, this preliminary version of our model learns the semantic knowledge associated to the heads of the participants that fill the explicit *frame-elements* occurring in the text. This knowledge is then used to capture the heads of the participants that should fill the *Definite Null Instantiations* (DNIs). For this preliminary model we leave aside the definition of the correct spans of the fillers.

First of all we perform a syntactic and semantic analysis of the dataset, for both training and testing parts. We use the Stanford parser[1] to obtain the named entities and coreference chains in order to process all the occurrences of the same participant as a unique item. We also perform a very simple Word Sense Disambiguation (WSD) process assigning to each word, when possible, the most frequent sense of WordNet (Fellbaum, 1998). This heuristic has been

[1] `http://nlp.stanford.edu/software/dcoref.shtml`

Frame#frame-element	Head	SemanticType
Expectation#Cognizer	Holmes [person]	Function#Human#Living#Object
Residence#Location	hotel	Artifact#Building#Object
Opinion#Opinion	that	that#IN

Table 4.4: Some examples of semantic types assigned to FEs. In brackets, the label of the named entity recognition

Apparently [the $\mathbf{tenants}_{\mathbf{Residence}}$]$_{Resident}$ had brought little or nothing with them. $\text{DNI}_{Location}$

Table 4.5: Example of annotation for the lexical-unit **tenant.n** of the frame **Residence**.

used frequently as a baseline in the evaluation of WSD systems and it seems to be very hard to beat (Gale et al., 1992). As the senses of WordNet have been mapped to several ontologies, this disambiguation allows us to label the documents with ontological features that can work as semantic types. In this experiment we use the Top Ontology (TO) (Álvez et al., 2008). For those cases where the words can not be labelled with any feature we define the pair lemma#part-of-speech as their semantic type (see the last example in Table 4.4). Our model assigns to each instantiated FE, the ontological feature of the syntactic head of its filler. Then, our model learns from the training data the probability distribution s of the semantic types of each FE. It also calculates the probability distribution p of the part-of-speech of the head of their fillers. The model includes in the training data the *explicit* annotations from the test document which is being analyzed. Table 4.4 contains three different examples of this assignment, the first one corresponds to a case where the TO feature has been assigned through the named entity label.

The first step of the ISRL process consist on deciding which not instantiated FEs should be filled. That is, which are *Definite Null Instantiations*.

Since our approach exploits the explicit annotations, our strategy differs notably from those presented in previous works. Following with the example shown in Table 4.5, suppose we are processing the lexical-unit **tenant.n** belonging to the frame *Residence* and the instantiated FE *Resident*:

First, our system collects from the training data the most common FE patterns of the corresponding frame of the LU under study. The patterns collected must contain the instantiated FEs of the LU. Table 4.6 shows the patterns collected for the previous example.

Pattern	Freq.
Resident Location	384
Resident Co_resident Location	34
Resident Co_resident	14
Resident	13
Resident Location Manner	1
Resident Location Time	1

Table 4.6: Most common patterns for the frame **Residence** containing the FE *Resident* ordered by frequency

Then, the system defines as DNIs all the *Core frame-elements* of the most common pattern that are missing for the lexical-unit that it is being processed. In the example above the most common pattern having the FE *Resident* is the one formed by the sequence of FEs *Resident* and *Location*. As the FE *Location* is indeed a *Core FE* of the frame **Residence**, it will be defined as a DNI of **tenant.n** and our system will try to find a filler for it.

Once the previous process has been applied for all the lexical-units in the document, our system can perform the DNI resolution. Gerber and Chai (2010) showed that the vast majority of the fillers of the implicit arguments can be found within the same sentence containing the predicate or in the two previous ones. They establish a window formed by these three sentences and considers as possible candidates the participants belonging to that window. We use the same criteria in our current model. Thus, our system selects the filler among the terminals that belong to the three sentences, the closest one that maximizes $P(s, p)$, the joint probability of s and p. Following with the example, our system calculates $P(s, p)$ for all terminals in the three sentence window of **tenant.n** as shown in Table 4.7.

Table 4.8 shows that, in this case, [*house*] obtains the higher joint probability. Consecuently, our model selects this participant as the filler for the NI *Location* for the predicate $\mathbf{tenant}_n$.

"Now, Mr. Holmes, with your permission, I will show you round the house." The various bedrooms and sitting-rooms had yielded nothing to a careful search. Apparently the **tenants** had brought little or nothing with them.

Table 4.7: Example of the three sentence window for the lexical-unit **tenant.n**.

Head	**PoS**	**SemanticType**	**P(s)**	**P(p)**	**P(s,p)**
house	N	Artifact#Building#Object	0.164	0.376	0.062
bedroom	N	Artifact#Building#Object#Part#Place	0.005	0.376	0.002
sitting-room	N	Artifact#Building#Object#Part#Place	0.005	0.376	0.002
Holmes [person]	N	Function#Human#Living#Object	0	0.376	0
show	V	Communication#Dynamic#Experience	0	0	0

Table 4.8: Resulting probabilities for some of the candidates in the context of $\textbf{tenant}_n$

4.2.1 Evaluation

As explained previously, the first step of the model consists on the correct identification of those missing FEs that are actually DNIs. Given that the final output of the system depends strongly on this first step, we evaluate the performance of our methodology in the DNI identification process. Table 4.9 shows how our system outperforms state of the art systems on this subtask.[2]

System	**P**	**R**	**F1**
Tonelli and Delmonte (2010)	-	-	-
Chen et al. (2010)	**0.57**	0.03	0.06
Tonelli and Delmonte (2011)	0.39	0.43	0.41
Our model	0.50	**0.66**	**0.57**

Table 4.9: Evaluation of DNI identification.

For the second subtask, we have used the scorer provided for NI subtask for the evaluation of the DNI resolution. This scorer works slightly different

[2] Values for the first version of VENSES++ were not reported. Silberer and Frank (2012) obtain a Recall of 0.4 in NI classification but they do not report results separately for DNI.

that the one for the traditional SRL subtask. Table 4.10 presents the results of the different systems.[3] It includes the performance of our system when learning either the lemmas or the semantic types of the head of the fillers. These results show that there is an additional gain exploiting the semantics of the fillers.

System	**P**	**R**	**F1**	**Over.**
Tonelli and Delmonte (2010)	-	-	0.01	-
Chen et al. (2010)	**0.25**	0.01	0.02	-
Tonelli and Delmonte (2011)	0.13	0.06	0.08	-
Silberer and Frank (2012)	0.09	0.11	0.10	-
This work (lemmas+pos)	0.13	0.23	0.17	0.54
This work (semantic-types+pos)	0.15	**0.25**	**0.19**	0.54

Table 4.10: Performance of our system compared with the systems using SemEval 2010 dataset.

Although our system obtains better results than alternative approaches, such a low figures clearly reflect its inherent difficulty. Our system clearly outperforms VENSES++ in terms of both precision and recall. SEMAFOR seems to solve much accurately a very limited number of cases. Finally, we also include the best results from (Silberer and Frank, 2012) obtained when using for training a larger corpus extended heuristically. In fact, their results are much lower when using only the training corpus provided for the task. It is worth mentioning that a window of three sentences around the predicate sets the upperbound recall to 76% for DNIs appearing in the test documents.

These results are just preliminary because, as it can be seen, the *Overlap* value is quite low. That means that spans of the predicted fillers are not accurately adjusted and they may contain too many tokens. Unfortunately, none of the rest of the proposals facing this task provides this score, so there is no possibility of performing a reliable comparison.

[3]The values of P and R for the first version of VENSES++ were not provided.

4.3 Conclusion

In this chapter, we explain that a general ISRL process consists of two different steps and we focus on the first one. We describe that for PropBank/NomBank we can apply a very simple strategy but for FrameNet we prefer to perform a more sophisticated strategy due to the complexity of the FE structures of the frames. This strategy is based on learning for each frame patterns of FEs that occur explicitly in the text. We present a preliminary system for the full ISRL process based on FrameNet that integrates this strategy. The results show that it obtains much better results than previous works.

The rest of the chapters of this document present approaches to face the second step of ISRL. That is, for detecting the fillers of the already selected implicit roles. However we also need to apply any of the strategies described in this section for solving the first step. For the model presented in Chapter 5 we use the strategy explained in Section 4.2 because this model is trained in the FrameNet based dataset. On the other hand, for the other two approaches described in Chapter 6 and Chapter 7, as they are focused on the PropBank/NomBank dataset, we apply the simple strategy showed in Section 4.1.1.

CHAPTER 5

Elided roles as a particular case of anaphora

In this chapter we present an approach for Implicit Semantic Role Labeling based on entity coreference. After a motivation of this work in Section 5.1, we study the adaptation of a set of theories and models that have been commonly used for anaphora resolution. This study is described in Section 5.2. We prove empirically the positive contribution of these features in Section 5.3 and discuss their advantages and drawbacks in Section 5.4. Finally, we present some concluding remarks about this approach in Section 5.5.

5.1 Introduction

Documents written in natural language commonly contain the same elements repeated many times along the discourse. However the different occurrences of the same elements tend to appear expressed in diverse forms even though they refer the same entities. Moreover, in many cases, these elements can not be interpreted by themselves in solitude and need some other coreferential elements to be understood (e.g. pronouns). Furthermore, sometimes the mentions of an entity do not even appear explicitly. This phenomena is also known as **zero anaphora**.

For instance, consider the example shown in Table 5.1. In this example, the word *road* should accompany the adjectives *straight*, *narrow*, *broad* and

There are two roads to eternity, a straight and narrow, and a broad and crooked.

Table 5.1: Example of **zero anaphora** where two mentions of *road* are elided.

John wants to sell his house. Sue has offered one million.

Table 5.2: Example of **zero anaphora**. In the second sentence *house* is omitted.

crooked but both mentions have been completely omitted. In order to fully understand this sentence it is necessary to infer that there exist references to each one of the *two roads* that appear previously.

As in zero anaphora the mention of the entity is not present, some of the evidences used in anaphora resolution such as the gender, number, etc. can not be applied. As this knowledge is not available in zero anaphoric cases, approaches for this task (Zhao and Ng, 2007; Yeh and Chen, 2007; Iida et al., 2007; Iida and Poesio, 2011) have to learn and exploit syntactic patterns in order to fulfil these gaps of information.

However, when the zero anaphora corresponds to an argument of a predicate, this phenomena can be faced as an extension of a traditional Semantic Role Labelling task.

Consider the example shown in Table 5.2 where the *benefactive* of the offer in the second sentence is missing. Once the predicate **offer.v** is identified as a specific member of a semantic resource, such as PropBank, it is possible to determine the missing arguments checking the role structure of the predicate in that resource. This fact solves some of the problems of zero anaphora resolution because for each semantic role it is possible to learn some useful information, like the most common semantic classes of the fillers, to be used as evidence instead of those traditionally used for anaphora resolution (gender, number, etc.). In the example, as a *benefactive* role is usually filled by a *Person* or *Organization*, the only potential candidate is *John*. Therefore, some of the approaches commonly applied for anaphora or coreference resolution can be adapted for **Implicit Semantic Role Labelling**. For that reason, early studies on implicit roles described this problem as a special case of anaphora or coreference resolution (Palmer et al., 1986; Whittemore et al., 1991; Tetreault, 2002) and treated elided arguments in similar ways as pronouns. Also recent works cast this problem as an anaphora resolution task (Silberer and Frank, 2012).

This chapter contains a detailed study of a set of features that have been traditionally used to model anaphora and coreference resolution tasks. We also describe how these features manifest in a FrameNet based corpus for modeling implicit argument resolution, including an analysis of their benefits and drawbacks.

5.2 Entity coreference based sources of evidence

Many sources of evidence have proved their utility in reference resolution (Burger and Connolly, 1992). This section presents how we adapt these sources of evidence to the specific characteristics of the DNI linking task and their behaviour over the Semeval-2010 training data[1]. Two main differences must be taken into account with respect to anaphora and coreference tasks. First, in anaphora and coreference tasks, mentions occur explicitly and they can be exploited to check particular constrains. Without an explicit argument, in some cases, the evidences can only be obtained from the predicate (that is, the lexical-unit) of the target DNI. Second, the referenced entities are not just nouns or pronouns but also verbs, adjectives, etc. Therefore, some features must be generalized. We adapted some of the sources of evidence studied. We group all of them in four different types:

- Syntactic
- Morpho-syntactic and Semantic Agreement
- Discoursive
- Coreference chains

Let's describe them in detail.

[1]See Section 3.4.1 for a complete description of the Semeval-2010 task

5.2.1 Syntactic

Some of the earliest theories studying pronoun resolution focused on the syntactic relations between the referenced entities. Here we present two syntactic features that also exploit this source of evidence. In both cases, an artificial node is included covering all document sentence trees in order to generalize its behaviour beyond sentence boundaries.

Command: C-command (Reinhart, 1976) is a syntactic relationship between nodes in a constituency tree. One node *N1* is said c-commanded by another *N2* if three requirements are satisfied:

- *N1* does not dominate *N2*
- *N2* does not dominate *N1*
- The first upper node that dominates *N1*, also dominates *N2*

This syntactic relation has proved to be useful to locate anaphoric references. Now, we study if this relationship can also be of utility for DNI resolution. We implemented this relation as a distance measure in the syntactic tree between the candidate filler node and the nearest common ancestor with respect the lexical-unit of the target DNI (see a simple example in Figure 5.1). Note that a value equal to zero means that either the filler dominates the target or the target dominates the filler. Besides, those fillers having a command value equal to one satisfy the c-command theory. Figure 5.2 presents the frequency distribution of our distance measure on the training data. It seems that most fillers have a command value equal or close to one.

Nearness: The constituency tree can also be exploited for anaphora resolution using breadth-first search techniques. A widely known algorithm based in this search is the Hobbs' algorithm (Hobbs, 1977). This algorithm follows a traversal search of the tree looking for a node that satisfies some constraints. Because of the nature of these constraints this algorithm cannot be directly applied to the implicit argument annotation task. Instead, we studied if the breadth distance can be an evidence through a measure we call **nearness**. We calculate **nearness** N as follows:

- P is the first upper node that dominates the lexical-unit T and the filler F
- B is the tree branch containing F whose parent is P
- If F precedes T, N is the number of following siblings of F in B

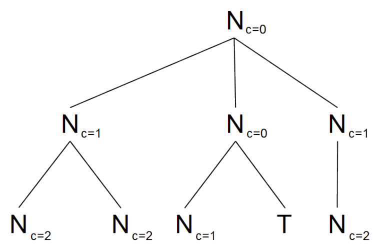

Figure 5.1: Sample values of **command** for different nodes in a constituency tree. T represents the lexical-unit of the target DNI

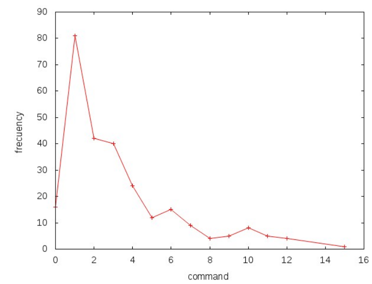

Figure 5.2: Frequency distribution of the different values of **command** in the training data

- If F follows T, N is the number of preceding siblings of F in B
- If T dominates F or F dominates T, N is equal to 0

Figure 5.3 presents some examples of values obtained using this measure. Figure 5.4 shows the frequency distribution of the different values of **nearness** in the training data. It also seems that most fillers prefer small **nearness** values.

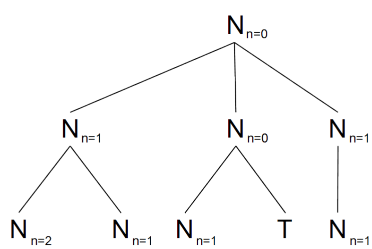

Figure 5.3: Sample values of **nearness** for different nodes in a constituency tree. T represents the lexical-unit of the target DNI

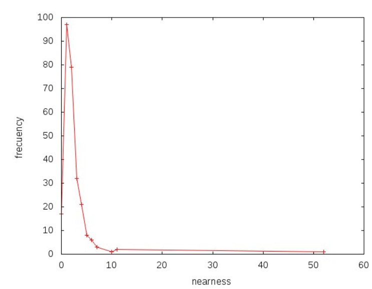

Figure 5.4: Frequency distribution of the different values of **nearness** in the training data

5.2.2 Morpho-syntactic and Semantic Agreement

Anaphora and coreference solvers usually apply morpho-syntactic and semantic agreement tests. These constraints check for the consistency between the properties of the target entities and the referents. Several agreement tests such as gender, number or semantic class can be applied. Since most of these tests cannot be applied to this task, in this work we only have studied the part of speech and semantic type agreement.

Semantic Type: To extract the semantic type of the filler of a frame-element, we first perform a very simple Word Sense Disambiguation (WSD) process assigning to each word, whenever possible, the most frequent sense of WordNet (Fellbaum, 1998). This heuristic has been used frequently as a baseline in the evaluation of WSD systems. As WordNet senses have been mapped to several ontologies, this disambiguation method allows us to label the documents with ontological features that can work as semantic types. In this work we have used the Top Ontology (TO) (Álvez et al., 2008). We assign to each filler the ontological features of its syntactic head. In this way, we can learn from the training data and for each frame-element a probability distribution of its semantic types. Table 5.3 contains some examples.

Frame#frame-element	SemanticType	Probability
Expectation#Cognizer	Human	0.93
	Group	0.07
Residence#Location	Building	0.77
	Place	0.33
Attempt#Goal	Purpose	0.41
	UnboundEvent	0.37
	Object	0.13
	Part	0.09

Table 5.3: Some examples of semantic types assigned to frame-elements.

Part of Speech: We also calculate the probability distribution of the part of speech (POS) of the head of the fillers similarly as for the semantic types.

5.2.3 Discoursive

Recency: While reading, recent entities are more likely to be a coreferent than more distant ones. This fact can be easily represented as the sentence distance between the lexical-unit of the target DNI and its referent. This feature has been used frequently not only in coreference and anaphora resolution but also in implicit argument resolution. Gerber and Chai (2010) noticed that the vast majority of the fillers of an implicit argument can be found either in the same sentence of the predicate or in the two preceding sentences. In our training data, this fact accounts for 70% of cases. Moreover, only around 2% of the fillers are located in posterior sentences. Figure 5.5 presents a frequency distribution of the different **recency** values.

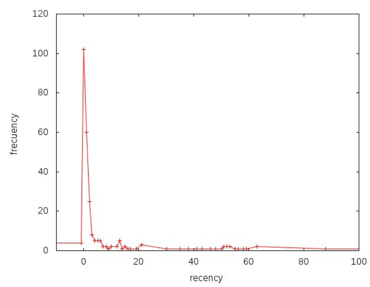

Figure 5.5: Frequency distribution of the different values of **recency** in the training data

filler \ LU	dialogue	monologue
dialogue	77.8%	5.4%
monologue	22.2%	94.6%

Table 5.4: Distribution of discourse level membership. The most common cases occur when the filler and the LU are both in the same level.

Dialogue: Since the corpus data consists of different chapters of a novel, it contains many dialogues inserted in a narrative monologue. The resolution of pronoun and coreference in dialogues dealing with a multi-party discourse have been largely studied (e.g. Byron and Stent (1998); Poesio et al. (2006); Stent and Bangalore (2010)). In our experiments, we just studied how referents are maintained with respect the two different levels of discourse. Table 5.4 shows that, in the vast majority of cases, both lexical-unit and filler belong to the same level of discourse[2]. Consequently, this fact can be used to promote those candidates that are at the same discourse level of the lexical-unit of the target DNI.

5.2.4 Coreference chains

An important source of evidence for anaphora resolution is the focus. The entity or topic which is salient in a particular part of the discourse is the

[2]Moreover, as expected, it is more frequent to refer from a monologue to a dialogue entity than the opposite.

most likely to be coreferred in the same part of the discourse. Thus, given a coreference annotation of a document it is possible to know how the focus varies along the discourse. As we explain in Section 3.4.1, the training data contains a full coreference annotation that we use to study three sources of evidence related to both focus and coreference chains.

Non singleton: Using the same training data, Silberer and Frank (2012) found that 72% of the DNIs are linked to referents that belong to non-singleton coreference chains. This means that candidate entities that are mentioned just once are less likely to be a referent filler of an implicit argument.

Focus: The **focus** refers to the entities that are most likely to be coreferred in a given point in the discourse (Sidner, 1978; Grosz and Sidner, 1986). Now, we study if this is also satisfied for DNI referents by checking if the filler of a DNI corresponds to the **focus** of a near context. We define the **focus** in a near context as follows. Consider the following definitions:

- F is the mention of an entity that is annotated as a filler of a target DNI.
- T is the lexical-unit of the target DNI.
- E is any entity between F and T.
- F_{-1} is the previous mention of F in the coreference chain.
- N_f is the number of mentions of F from F_{-1} to T.
- N_e is the number of mentions of E from F_{-1} to T.

If F_{-1} is the previous mention of F in the coreference chain, then N_f is equal to two. If there are no previous mentions of F, then F_{-1} is equal to F, and N_f is equal to one. F is the focus of the near context of T if and only if there is no E having $N_e{>}N_f$.

From our training data, we also observe that the **focus** matches the filler of a DNI in 72% of the cases.

Centering: **Centering** (Grosz et al., 1995; Brennan et al., 1987) is a theory that tracks the continuity of the focus to explain the coherence of the referential entities in a discourse. The theory establishes three different types of focus transition depending on the relation within the previous focus, $C_b(U_{n-1})$, the actual focus, $C_b(U_n)$, and the element that is most likely to be the focus, $C_p(U_n)$, according to its grammatical function. Figure 5.6 shows the three different kinds of **centering** transitions.

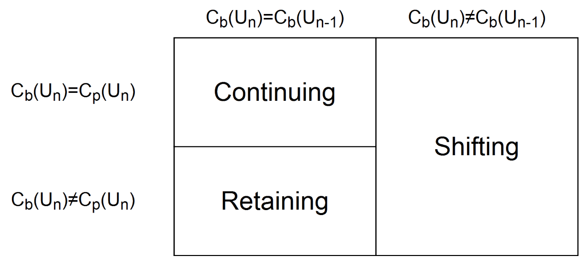

Figure 5.6: Types of **centering** transitions

The theory establishes that the most common transition is **continuing**. The second most common transition is **retaining** and the least common transition is **shifting**. Applying this schema to the training data, we found that the following probability distribution:

- Continuing: 41.0%
- Retaining: 25.2%
- Shifting: 18.8%
- Other: 15.0%

Since in the DNI filling task the referents can be of any kind of part-of-speech and the grammatical function only takes into account nouns or pronouns, the **centering** theory is not always applicable. When the filler is not a noun or a pronoun we have created a fake **centering** category called **other**. Thus, according to the training data, it seems that the preference order of the transitions matches the original theory being **continuing** the most common transition.

5.3 Experiments

In the previous section we have proposed the adaptation to the implicit argument filling task of some theories traditionally applied to capture evidence for anaphora and coreference resolution. Since the implicit role reference is a special case of coreference, we expect a similar behaviour also for this case. In fact, our analysis using the training data of SemEval seems to confirm our initial hypothesis. In order to evaluate the potential utility of these sources of evidence we have performed a set of experiments using the SemEval-2010

Task 10 testing-data. In this section, we describe our strategy for solving the implicit arguments and the results of our evaluation.

5.3.1 Processing steps

Any system facing the implicit argument resolution task has to follow the following steps:

1. Select the *frame-elements* that are Null Instantiations and decide if they are are Definite.
2. In case of definite *null instantiation*, locate the corresponding filler.

For the first step, we have followed the strategy explained in Chapter 4.

For the last step, we have modelled the sources of evidence presented previously as features to train a Naive-Bayes algorithm. We applied a maximum-likelihood method without any smoothing function. Thus, having a set of features *f*, for each DNI we select as filler the candidate c that satisfies:

$$\arg\max P(c) \prod_i P(f_i|c)$$

Non-singleton, focus and centering features require a coreference annotation of the document to be analysed. As we explain in Section 3.4.1, the training data of the SemEval task contains manually annotated coreference chains that can be used to exploit these features. However, as the testing data does not contain this type of annotations, we applied an automatic coreference resolution system. We used the software provided by Stanford NLP pipeline[3]. In the following experiments, we present the results obtained using manual and predicted coreference.

5.3.2 Results on the SemEval-2010 test

To evaluate the results of the experiments above we have used the scorer provided for the SemEval-2010 subtask. Table 5.5 shows available precision,

[3] `http://nlp.stanford.edu/software/dcoref.shtml`

recall, F-score and overlapping figures of the different systems using predicted and gold-standard coreference chains[4]. Our simple strategy clearly outperforms Tonelli and Delmonte (2010) (T'10) in terms of both precision and recall. Chen et al. (2010) (C'10) seems to solve more accurately but a more limited number of cases. Our approach also performs better than Tonelli and Delmonte (2011) (T'11). We include the results from Silberer and Frank (2012) (S'12) obtained when using for training a larger corpus extended heuristically (best) and the results obtained with no additional training data (no-extra). Our approach obtains better results in all the cases except when they use extended training data with the gold-standard coreference chains. In this case, our model seems to achieve a similar performance but without exploiting extra training data. Apparently, the system we present in Chapter 4 presents better results but, as we explained previously, a low overlapping score means vague answers. Although our approach outperforms previous works, such a low figures clearly reflect the inherent difficulty of the task.

	Auto Coref				**GS Coref**			
System	**P**	**R**	**F1**	**Over.**	**P**	**R**	**F1**	**Over.**
T'10	-	-	0.01	-				
C'10	**0.25**	0.01	0.02	-				
T'11	0.13	0.06	0.08	-				
S'12 no-extra	0.06	0.09	0.07	-	-	-	0.13	-
S'12 best	0.09	0.11	0.10	-	-	-	**0.18**	-
Chapter 4	0.15	**0.25**	**0.19**	0.54				
This work	0.14	0.18	0.16	**0.89**	0.16	0.20	**0.18**	**0.90**

Table 5.5: Results using SemEval-2010 dataset.

DNI linking experiment: In order to check the sources of evidence independently of the rest of processes, we have performed a second experiment where we assume perfect results for the first step. In other words, we apply our DNI filling strategy just to the correct DNIs in the document. Table 5.6 shows the relevance of a correct DNI identification (the first step of the process). Once again, without extra training data our strategy outperforms the model by Silberer and Frank (2012)[5]. Again, when using extended training

[4]Surprisingly, previous research in the literature do not report results of overlapping.

[5]The rest of the systems do not perform any experiments with gold-standard DNI identification.

data their model seems to perform similar to ours.

	Auto Coref				GS Coref			
System	**P**	**R**	**F1**	**Over.**	**P**	**R**	**F1**	**Over.**
S'12 no-extra					0.26	**0.25**	0.25	-
S'12 best					0.31	**0.25**	**0.28**	-
This work	0.30	0.22	0.26	0.89	**0.33**	0.24	**0.28**	0.89

Table 5.6: Results using SemEval-2010 dataset on the correct DNIs.

Ablation tests: Table 5.7 presents the results using the gold-standard coreference, when leaving out a type of feature one at a time. The table empirically demonstrates that all feature types contribute positively to solve this task. The morpho-syntactic and semantic agreement seem to be the most relevant evidence in terms of precision and recall. That is, identifying the head of the correct filler. On the other hand, syntactic features are the most relevant to detect the correct span of the fillers (with a drop to 0.75 on overlapping).

Source Set	**P**	**R**	**F1**	**Over.**
all	0.33	0.24	0.28	0.89
no-coreference	0.30	0.22	0.25	0.86
no-semantic agreement	0.22	0.22	0.22	0.90
no-discursive	0.29	0.22	0.25	0.82
no-syntactic	0.28	0.21	0.24	0.75

Table 5.7: Ablation tests using the gold-standard coreference.

5.4 Discussion

In order to analyse the limits of the different types of evidence, we used as a reference the results obtained using the gold-standard DNIs and coreference chains (see Table 5.6). As an overall remark, all previous works facing this task agree on the sparsity of the training data. We also observed that this problem affects all sources of evidence we have studied, especially the agreement of semantic types.

Data sparsity for semantic types: The semantic types do not cover the full set of frame-elements. The testing data contains a total of 209 different Frame#frame-elements. 73 of them (around 35%) do not appear on the training data. Another problem appears when the frame-elements have too many different semantic types with very similar probabilities. Without enough information to discriminate correctly the filler, this source of evidence becomes damaging (see Table 5.8).

P	**R**	**F1**	**Over.**
0.21	0.09	0.13	0.61

Table 5.8: Performance of FE having more than 5 semantic types

Outside the same sentence: Recency strongly prioritises the window formed by the same sentence of the lexical-unit of the target DNI and the two previous sentences. However, in 19% of the cases the filler belongs to a sentence outside that window. Furthermore, syntactic based evidences rely on relations between entities in the same sentence. Obviously, adding an artificial node covering the whole document analysis is quite arbitrary. Table 5.9 shows how the performance of our approach decreases strongly when the filler and the lexical-unit are in different sentences.

same sentence				**another sentence**			
P	**R**	**F1**	**Over.**	**P**	**R**	**F1**	**Over.**
0.50	0.34	0.40	0.87	0.20	0.16	0.18	0.96

Table 5.9: Performance when the filler and the lexical-unit are in the same sentence or in another one

Discursive structure: The particular structure of the documents can also affect seriously the performance of the sources of evidence. Table 5.10 presents the results on contexts with at least 10% of entities on a monologue or a dialogue. According to the recency feature, each context is formed by the sentence of the lexical-unit of the target DNI and the two previous sentences. We can observe that the results on mixed contexts are better than in general. Obviously, dialogue features are totally useless in contexts with only monologues or only dialogues.

Singleton fillers: Most of the fillers are entities that belong to a coreference chain. Therefore, these cases heavily depends on a correct coreference

P	R	F1	Over.
0.38	0.29	0.33	0.93

Table 5.10: Performance in mixed contexts with at least 10% of entities of each level

annotation. This is why worse results are obtained when using predicted coreferent chains. Table 5.11 shows the results when the filler belongs or not to a coreference chain. It is important to remind that in this work we have adapted a set of sources of evidence and theories traditionally used for anaphora and coreference resolution. Originally these theories focused just on noun and pronoun entities.

coref-chain				no-coref-chain			
P	R	F1	Over.	P	R	F1	Over.
0.45	0.35	0.39	0.94	0.06	0.04	0.05	0.31

Table 5.11: Performance when the filler belongs to a coreference-chain or not

5.5 Conclusions

We have presented a first attempt to study the behaviour of traditional coreference and anaphora models for the implicit argument resolution task, a special case of coreference. Our analysis shows that these theories and models can be successfully applied for Implicit Semantic Role Labelling as sources of evidence in an existing dataset. In fact, their joint combination improves state of the art results.

However, the sources of evidence proposed are adaptations that focus on nominal entities and pronouns, and on relations within entities and referents belonging to the same sentence. It seems that for these cases it is possible to capture useful evidence. But, as has been showed, the referents of the implicit arguments can be expressed in many other ways and places. Although the approach based on entity coreference has proved to be useful in different works, it also seems to be insufficient.

In the following chapter, a novel perspective is presented for ISRL. We propose that event coreference matching can help to discover explicit occurrences of implicit arguments. In other words, the next chapter describes a first try to introduce event coreference into ISRL.

CHAPTER 6

Completing role labelling via coreferent predicates

In this chapter we describe a novel deterministic algorithm for ISRL that introduces a basic approach for event coreference. The algorithm, called **ImpAr**, combines this strategy with an adapted version of a method for anaphora resolution and it can be applied for any predicate, even without training data available. We motivate this novel approach in Section 6.1 and explain the full algorithm in Section 6.2. Then, Section 6.3 contains the evaluation of the system and Section 6.4 proposes some configuration sets in order to check the different components of the algorithm. Finally, Section 6.5 presents some final remarks.

6.1 Introduction

As shown up to this point, the labelling of implicit semantic roles has been mainly focused just on the search of the proper referent entity on the basis of the features that characterize the possible candidates to be the filler of each semantic role. However, a vaguely studied aspect in this topic is the fact that the event to which the implicit role belongs can be expressed several times and in several forms along the discourse. Thus, the same role can be found explicitly expressed for any other mention of the same event.

[arg_0 Wheat] [np **prices**] remain stubbornly high and **they**ŕe likely to stay that way for months to come.

Table 6.1: A pronoun (*they*) referring to a nominal predicate (**prices**).

[arg_0 The network] had been expected to have [np **losses**] [arg_1 of as much as $20 million] [$arg_3$ on baseball this year]. It isn't clear how much those [np **losses**] may widen because of the short Series.

Table 6.2: Both mention of the predicate **losses** refor to the same event.

In a very related work, Dahl et al. (1987) proposed a strategy to include role annotations for pronouns that refer to nominal predicates. Their method fills the arguments of anaphoric mentions of nominal predicates using previous mentions of the same predicate. The example in Table 6.1 shows that the pronoun *they* could be annotated with the same arg_0 of its referent **prices**, in this case *wheat*.

This approach can be easily extended to solve missing arguments if we make the assumption that in a coherent document the different occurrences of a predicate, including both verbal and nominal forms, tend to be mentions of the same event. Thus, if they are references to the same event, they will share the same argument fillers. Although the simplicity of this hypothesis, recent works on event coreference have shown that it is a baseline really hard to beat (Bejan et al., 2009; Bejan and Harabagiu, 2010; Lee et al., 2012). The potential of this approach for ISRL can be shown in the example of Table 6.2. This analysis includes annotations for the nominal predicate **loss** based on the NomBank structure (Meyers et al., 2004). In this case the annotator identifies, in the first sentence, the arguments arg_0, the entity losing something, arg_1, the thing lost, and arg_3, the source of that loss. In the second sentence there is another instance of the same predicate, **loss**, but in this case no argument has been associated with it. However, the second mention of the predicate could inherit the argument structure of the first one.

In this chapter, a novel deterministic algorithm based on the previous idea is presented. The system, called **ImpAr**, obtains competitive results with respect to supervised methods and furthermore, it can be applied to any predicate without training data.

6.2 ImpAr algorithm

6.2.1 Discoursive coherence of predicates

Exploring the training PropBank/NomBank-based dataset developed by Gerber and Chai (2010, 2012), we observe a very strong discourse effect on the implicit and explicit argument fillers of the predicates. That is, if several instances of the same predicate appear in a well-written discourse, it is very likely that they maintain the same argument fillers. This property holds when joining the different parts-of-speech of the predicates (nominal or verbal) and the explicit or implicit realizations of the argument fillers. For instance, we observed in this corpus that 46% of all implicit arguments share the same filler with the previous instance of the same predicate while only 14% of them have a different filler. The remaining 40% of all implicit arguments correspond to first occurrences of their predicates. That is, these fillers can not be recovered from previous instances of their predicates.

The rationale behind this phenomena seems to be simple. When referring to different aspects of the same event, the writer of a coherent document does not repeat redundant information. They refer to previous predicate instances assuming that the reader already recalls the involved participants. That is, the filler of the different instances of a predicate argument maintain a certain discourse coherence. For instance, in the example of Table 6.2, all the argument positions of the second occurrence of the predicate **loss** are missing, but they can be easily inferred from the previous instance of the same predicate.

Therefore, we propose to exploit this property in order to capture correctly how the fillers of all predicate arguments evolve through a document.

Our algorithm, **ImpAr**, processes the documents sentence by sentence, assuming that sequences of the same predicate (in its nominal or verbal form) share the same argument fillers (explicit or implicit)[1]. Thus, for every *core* argument arg_n of a predicate, **ImpAr** stores its previous known filler as a *default* value. If the arguments of a predicate are explicit, they always replace

[1]Note that the algorithm could also consider sequences of closely related predicates.

default fillers previously captured. When there is no antecedent for a particular implicit argument $iarg_n$, the algorithm tries to find in the surrounding context which participant is the most likely to be the filler according to some salience factors (see Section 6.2.2). For the following instances, without an explicit filler for a particular argument position, the algorithm repeats the same selection process and compares the new implicit candidate with the default one. That is, the default implicit argument of a predicate with no antecedent can change every time the algorithm finds a filler with a greater salience. A damping factor is applied to reduce the salience of distant predicates.

6.2.2 Filling arguments without explicit antecedents

As commented previously, filling the implicit arguments of a predicate has been identified as a particular case of coreference, very close to pronoun resolution (Silberer and Frank, 2012). Consequently, for those implicit arguments that have not explicit antecedents, we propose an adaptation of a classic algorithm for deterministic pronoun resolution. This component of our algorithm follows the RAP approach (Lappin and Leass, 1994). When our algorithm needs to fill an implicit predicate argument without an explicit antecedent it considers a set of candidates within a window formed by the sentence of the predicate and the two previous sentences. Then, the algorithm performs the following steps:

1. Apply two constraints to the candidate list:

   1. 1. All candidates that are already explicit arguments of the predicate are ruled out.

   1. 2. All candidates commanded by the predicate in the dependency tree are ruled out.

2. Select those candidates that are semantically consistent with the semantic category of the implicit argument.

3. Assign a salience score to each candidate.

4. Sort the candidates by their proximity to the predicate of the implicit argument.

5. Select the candidate with the highest salience value.

Quest Medical Inc said it adopted [arg1 a shareholders' rights] [np **plan**] in which rights to purchase shares of common stock will be distributed as a dividend to shareholders of record as of Oct 23.

Table 6.3: Argument arg_0 of predicate **plan** is missing.

As a result, the candidate with the highest salience value is selected as the filler of the implicit argument. Thus, this filler with its corresponding salience weight will be also considered in subsequent instances of the same predicate.

Now, we explain each step in more detail using the example in Table 6.3 where arg_0 is missing for the predicate **plan**.

Filtering. In the first step, the algorithm filters out the candidates that are actual explicit arguments of the predicate or have a syntactic dependency with the predicate, and therefore, they are in the search space of a traditional SRL system.

In our example, the filtering process would remove *[a shareholders' rights]* because it is already the explicit argument arg_1, and *[in which rights to purchase shares of common stock will be distributed as a dividend to shareholders of record as of Oct 23]* because it is syntactically commanded by the predicate **plan**.

Semantic consistency. To determine the semantic coherence between the potential candidates and a predicate argument arg_n, we have exploited the selectional preferences in the same way as in previous works of traditional SRL and implicit argument resolution. First, we have designed a list of very general semantic categories. Second, we have semi-automatically assigned one of them to every predicate argument arg_n in PropBank and NomBank. For this, we have used the semantic annotation provided by the training documents of the CoNLL-2008 dataset. This annotation was performed automatically using the *SuperSenseTagger* (Ciaramita and Altun, 2006) and includes named-entities and WordNet Super-Senses[2]. We have also defined a mapping between the semantic classes provided by the *SuperSenseTagger* and our seven semantic categories (see Table 6.4 for more details). Then, we have acquired the most common categories of each explicit predicate ar-

[2]Lexicographic files according to WordNet terminology.

gument arg_n. **ImpAr** algorithm also uses the *SuperSenseTagger* over the documents to be processed from BNB to check if the candidate belongs to the expected semantic category of the implicit argument to be filled.

Following the example above, *[Quest Medical Inc]* is tagged as an *ORGANIZATION* by the *SuperSenseTagger*. Therefore, it belongs to our semantic category *COGNITIVE*. As the semantic category for the implicit argument $iarg_0$ for the predicate **plan** has been recognized to be also *COGNITIVE*, *[Quest Medical Inc]* remains in the list of candidates as a possible filler.

Semantic category	**Name-entities**	**Super-Senses**
COGNITIVE	*PERSON* *ORGANIZATION* *ANIMAL* ...	noun.person noun.group noun.animal ...
TANGIBLE	*PRODUCT* *SUBSTANCE* ...	noun.artifact noun.object ...
EVENTIVE	*GAME* *DISEASE* ...	noun.act noun.communication ...
RELATIVE		noun.shape noun.attribute ...
LOCATIVE	*LOCATION*	noun.location
TIME	*DATE*	noun.time
MEASURABLE	*QUANTITY* *PERCENT* ...	noun.quantity

Table 6.4: Links between the semantic categories and some name-entities and super-senses.

Salience weighting. In this step, the algorithm assigns to each candidate a set of salience factors that scores its prominence. The *sentence recency* factor prioritizes the candidates that occur close to the same sentence of the predicate. The *subject*, *direct object*, *indirect object* and *non-adverbial* factors weight the salience of the candidate depending on the syntactic role they belong to. Additionally, the head of these syntactic roles are prioritized by the *head* factor. We have used the same weights, listed in Table 6.5, proposed by Lappin and Leass (1994).

Factor type	**weight**
Sentence recency	100
Subject	80
Direct object	50
Indirect object	40
Head	80
Non-adverbial	50

Table 6.5: Weights assigned to each salience factor.

In the example, candidate *[Quest Medical Inc]* is in the same sentence as the predicate **plan**, it belongs to a subject, and, indeed, it is the head of that subject. Hence, the salience score for this candidate is: $100 + 80 + 80 = 260$.

6.2.3 Damping the salience of the default candidate

As the algorithm maintains the default candidate until an explicit filler appears, potential errors produced in the automatic selection process explained above can spread to distant implicit instances, specially when the salience score of the default candidate is high. In order to reduce the impact of these mistakes we include a damping factor that is applied sentence by sentence to the salience value of the default candidate. **ImpAr** applies that damping factor, r, as follows. It assumes that, independently of the initial salience assigned, 100 points of the salience score came from the *sentence recency* factor. Then, the algorithm changes this value multiplying it by r. So, given a salience score s, the value of the score in a following sentence, s', is:

$$s' = s - 100 + 100 \cdot r$$

Obviously, the value of r must be defined without harming excessively those cases where the default candidate has been correctly identified. For this, we studied in the training dataset the cases of implicit arguments filled with the default candidate. Figure 6.1 shows that the influence of the default filler is much higher in near sentences that in more distance ones.

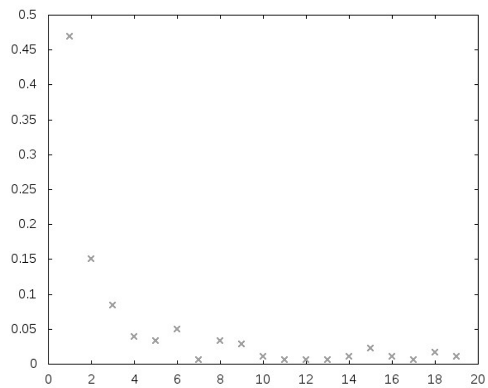

Figure 6.1: Distances between the implicit argument and the default candidate. The y axis indicate the percentage of cases occurring in each sentence distance, expressed in x

We tried to mimic a damping factor following this distribution. That is, to maintain high score salience for the near sentences while strongly decreasing them in the subsequent ones. In this way, if the filler of the implicit argument is wrongly identified, the error only spreads to the nearest instances. If the identification is correct, a lower score for more distance sentences is not too harmful. The distribution shown in Figure 6.1 follows an exponential decay, therefore we have described the damping factor as a curve like the following, where α must be a value within 0 and 1:

$$r = \alpha^d$$

In this function, d stands for the sentence distance and r for the damping factor to apply in that sentence. In this paper, we have decided to set the value of α to 0.5.

$$r = 0.5^d$$

This value maintains the influence of the default fillers with high salience in near sentences. But it decreases that influence strongly in the following sentences.

In order to illustrate the whole process we will use the example in Table 6.3. In that case, *[Quest Medical Inc]* is selected as the arg_0 of **plan** with a salience score of 260. Therefore *[Quest Medical Inc]* becomes the default arg_0 of **plan**. In the following sentence the damping factor is:

$$0.5 = 0.5^1$$

Therefore, its salience score changes to $260 - 100 + 100 \cdot 0.5 = 210$. Then, the algorithm changes the default filler for arg_0 only if it finds a candidate that scores higher in their current context. At two sentence distance, the resulting score for the default filler is $260 - 100 + 100 \cdot 0.25 = 185$. In this way, at more distance sentences, the influence of the default filler of arg_0 becomes smaller.

6.3 Evaluation

			base.	**Gerber & Chai**			**ImpAr**		
	Inst.	**Imp.**	**F_1**	**P**	**R**	**F_1**	**P**	**R**	**F_1**
sale	64	65	36.2	**47.2**	**41.7**	**44.2**	45.7	33.4	38.6
price	121	53	15.4	36.0	32.6	34.2	**45.2**	**53.3**	**49.0**
investor	78	35	9.8	**36.8**	**40.0**	**38.4**	36.7	37.4	37.0
bid	19	26	32.3	23.8	19.2	21.3	**55.1**	**49.2**	**52.0**
plan	25	20	38.5	**78.6**	**55.0**	**64.7**	42.8	40.7	41.7
cost	25	17	34.8	**61.1**	**64.7**	**62.9**	52.9	47.4	50.0
loss	30	12	52.6	**83.3**	**83.3**	**83.3**	52.3	63.5	57.3
loan	11	9	18.2	**42.9**	**33.3**	**37.5**	28.6	20.0	23.5
investment	21	8	0.0	40.0	**25.0**	30.8	**92.9**	23.2	**37.1**
fund	43	6	0.0	14.3	16.7	15.4	**40.0**	**33.3**	**36.4**
Overall	437	246	26.5	44.5	40.4	42.3	**45.2**	**41.5**	**43.3**

Table 6.6: Evaluation with the test. The results from (Gerber and Chai, 2010) are included.

In order to evaluate the performance of the **ImpAr** algorithm, we have followed the evaluation method presented by Gerber and Chai (2010, 2012).

Traditionally, there have been two approaches to develop SRL systems, one based on constituent trees and the other one based on syntactic dependencies. Additionally, the evaluation of both types of systems has been

			base.	Gerber & Chai			ImpAr		
	Inst.	**Imp.**	$\mathbf{F_1}$	**P**	**R**	$\mathbf{F_1}$	**P**	**R**	$\mathbf{F_1}$
sale	184	181	37.3	**59.2**	**44.8**	**51.0**	44.0	37.7	40.6
price	216	138	34.6	**56.0**	48.7	**52.1**	48.0	**52.7**	50.3
investor	160	108	5.1	**46.7**	39.8	**43.0**	24.7	26.0	25.3
bid	88	124	23.8	60.0	36.3	45.2	53.2	**42.2**	**47.0**
plan	100	77	32.3	**59.6**	**44.1**	**50.7**	52.7	**44.1**	48.0
cost	101	86	17.8	**62.5**	**50.9**	**56.1**	46.2	43.0	44.5
loss	104	62	54.7	**72.5**	**59.7**	**65.5**	56.4	54.2	55.2
loan	84	82	31.2	**67.2**	**50.0**	**57.3**	48.0	42.9	45.3
investment	102	52	15.5	32.9	**34.2**	**33.6**	**49.2**	20.8	29.2
fund	108	56	15.5	**80.0**	35.7	**49.4**	53.3	**42.9**	47.5
Overall	1,247	966	28.9	**57.9**	**44.5**	**50.3**	46.0	40.3	43.0

Table 6.7: Evaluation with the full dataset. The results from (Gerber and Chai, 2012) are included.

performed differently. For constituent based SRL systems the scorers evaluate the correct span of the filler, while for dependency based systems the scorer just check if the systems are able to capture the head token of the filler. As shown in Chapter 3, previous works in implicit argument resolution proposed a metric that involves the correct identification of the whole span of the filler. **ImpAr** algorithm works with syntactic dependencies and therefore it only returns the head token of the filler. In order to compare our results with previous works, we had to apply some simple heuristics to guess the correct span of the filler. Obviously, this process inserts some noise in the final evaluation.

We have performed a first evaluation over the test set used in (Gerber and Chai, 2010). This dataset contains 437 predicate instances but just 246 argument positions are implicitly filled. Table 6.6 includes the results obtained by **ImpAr**, the results of the system presented by Gerber and Chai (2010) and the baseline (base.) proposed for the task. Best results are marked in bold[3]. For all predicates, **ImpAr** improves over the baseline (16.8 points higher in the overall F_1). Our system also outperforms the one presented by Gerber and Chai (2010). Interestingly, both systems present very different performances predicate by predicate. For instance, our system obtains much

[3]No proper significance test can be carried out without the the full predictions of all systems involved.

higher results for the predicates **bid** and **fund**, while much lower for **loss** and **loan**. In general, **ImpAr** seems to be more robust since it obtains similar performances for all predicates. In fact, the standard deviation, σ , of F_1 measure is 9.89 for **ImpAr** while this value for the (Gerber and Chai, 2010) system is 19.85.

In a more recent work, Gerber and Chai (2012) presented some improvements of their previous results. In this work, they extended the evaluation of their model using the whole dataset and not just the testing documents. Applying a cross-validated approach they tried to solve some problems that they found in the previous evaluation, like the small size of the testing set. For this work, they also studied a wider set of features, specially, they experimented with some statistics learnt from parts of GigaWord automatically annotated. Table 6.7 shows that the improvement over their previous system was remarkable. The system also seems to be more stable across predicates. For comparison purposes, we also included the performance of **ImpAr** applied over the whole dataset.

The results in Table 6.7 show that, although **ImpAr** still achieves better results in some cases, this time, it cannot beat the overall results obtained by the supervised model. In fact, both systems obtain a similar recall, but the system from (Gerber and Chai, 2012) obtains much higher precision. In both cases, the σ value of F_1 is reduced, 8.82 for **ImpAr** and 8.23 for (Gerber and Chai, 2012). However, **ImpAr** obtains very similar performance independently of the testing dataset what proves the robustness of the algorithm. This suggests that our algorithm can obtain strong results also for other corpus and predicates. Instead, the supervised approach would need a large amount of manual annotations for every predicate to be processed.

6.4 Discussion

6.4.1 Component analysis

In order to assess the contribution of each system component, we also tested the performance of **ImpAr** algorithm when disabling only one of its components. With this evaluations we pretend to highlight the particular contribution of each component. In Table 6.8 we present the results obtained in the following experiments for the two testing sets explained in Section 6.3:

- Exp1: The damping factor is disabled. All selected fillers maintain the same salience over all sentences.
- Exp2: Only explicit fillers are considered as candidates[4].
- Exp3: No default fillers are considered as candidates.

As expected, we observe very similar performances in both datasets. Additionally, the highest loss appears when the default fillers are ruled out (Exp3). In particular, it also seems that the explicit information from previous predicates provides the most correct evidence (Exp2). Also note that for Exp2, the system obtains the highest precision. This means that the most accurate cases are obtained by previous explicit antecedents. Finally, the results in Exp1 shows that the contribution of the damping factor is less relevant. However, disabling this factor has a slight negative effect on the performance.

	test			full		
	P	**R**	**F_1**	**P**	**R**	**F_1**
full	45.2	**41.5**	**43.3**	46.0	**40.3**	**43.0**
Exp1	44.8	41.1	42.9	45.9	40.2	42.8
Exp2	**47.5**	25.4	33.1	**49.0**	25.6	33.6
Exp3	*40.5*	*25.1*	*31.0*	*40.1*	*22.3*	*29.0*
Exp4	52.6	41.5	46.4	54.1	40.3	46.2
Exp5	50.1	46.0	48.0	50.3	44.2	47.1

Table 6.8: Exp1, Exp2 and Exp3 correspond to ablations of the components. Exp4 shows the evaluation only over the arguments in the gold-standard. Exp5 evaluates the system capturing just the head tokens of the constituents.

6.4.2 Missing roles detection

In Chapter 4, we show that before filling an implicit role we need to decide which missing arguments of a predicate mention should be processed. For NomBank/PropBank schemas, we follow a very direct approach and perform the filling task for all core arguments of a predicate that do not appear explicitly in the text. In Table 6.8 we include the performance of our algorithm

[4] That is, implicit arguments without explicit antecedents are not filled.

Ports of Call Inc. reached agreements to **sell** its remaining seven aircraft [$iarg_1$ to buyers] that weren't disclosed].

Table 6.9: The span of the predicted $iarg_0$ does not fit properly the golds-standard [arg_1 *buyers that weren't disclosed*.

when processing only the missing arguments included in the gold-standard (Exp4). As can be seen, the precision in this evaluation increases. For example, in the test data F1 increases from 43.3 to 46.4. These results show that **ImpAr** tries to solve some missing arguments that are not implicit and, therefore, not recoverable from context.

6.4.3 Correct span of the fillers

As explained in Section 6.3, our algorithm works with syntactic dependencies and its predictions only return the head token of the filler. Obtaining the correct constituents from syntactic dependencies is not trivial. In this work we have applied a simple heuristic that returns all the descendant tokens of the predicted head token. This naive process inserts some noise to the evaluation of the system. For example, for the sentence in Table 6.9 our system predicts that the filler of the implicit $iarg_1$ of an instance of the predicate **sale** is *[to buyers]* .

But the actual gold-standard annotation is: [$iarg_1$ *buyers that weren't disclosed*]. Although the head of the constituent, *buyers*, is correctly captured by **ImpAr**, the final prediction is heavily penalized by the scoring method. Table 6.8 presents the results of **ImpAr** when evaluating the head tokens of the constituents only (Exp5). These results show that the current performance of our system can be easily improved applying a more accurate process for capturing the correct span.

6.5 Conclusions

In this chapter we have presented a robust deterministic approach for Implicit Semantic Role Labelling. The method exploits a very simple but relevant discursive coherence property that holds over explicit and implicit arguments of closely related nominal and verbal predicates. This property states that if several instances of the same predicate appear in a well-written discourse, it is very likely that they refer to the same event instance and, in consequence, maintain the same argument fillers. We have shown the importance of this phenomenon for recovering the implicit information about semantic roles. To our knowledge, this is the first empirical study that proves this phenomenon.

Based on these observations, we have developed a new deterministic algorithm, **ImpAr**, that obtains very competitive and robust performances with respect to supervised approaches. That is, it can be applied where there is no available manual annotations to train. The code of this algorithm is publicly available[5] and can be applied to any document. As input it only needs the document with explicit semantic role labeling and Super-Sense annotations. These annotations can be easily obtained from plain text using available tools[6], what makes this algorithm the first effective tool available for implicit SRL.

As it can be easily seen, **ImpAr** has a large margin for improvement. For instance, providing more accurate spans for the fillers. Furthermore, the strategy followed to match coreferent event mentions is quite naive and can be enhanced through a deeper discourse analysis. For example, the system can profit from additional annotations like entity coreference, that has proved its utility in previous works and can help to distinguish if the explicit arguments of different predicates agree or not. Moreover the same events can be expressed in a wide variety of ways, not only by predicates with the strictly same meaning. Indeed, antonyms like **sell** and **buy** could refer to the same event. But there also exists the possibility of studying other kinds of relations between predicates beyond coreference, like implication, causation or precedence. Exploiting this type of relations is the goal of the following chapter.

[5] `http://adimen.si.ehu.es/web/ImpAr`

[6] We recommend mate-tools (Björkelund et al., 2009) and SuperSenseTagger (Ciaramita and Altun, 2006).

CHAPTER 7

Extending event relations for a full role annotation

The current chapter presents an approach that exploits semantic relations between predicates and semantic roles for Implicit Semantic Role Labelling extending the notion of event coreference. We introduce this approach in Section 7.1. In Section 7.2 we explain how we obtain the semantic relations from FrameNet. Later, in Section 7.3, we propose a mapping strategy between the predicates and roles of FrameNet and PropBank/NomBank in order to transfer the semantic relations to the latter resource. Section 7.4 describes the application of this approach for ISRL and includes the corresponding evaluation. Finally, we point some conclusions in Section 7.5

7.1 Introduction

The previous chapter shows the relevance of event coreference for completing predicative annotations when some of the semantic roles are not expressed explicitly. It has been proved that even a simplistic strategy for discovering mentions of the same event instances results on a positive contribution for implicit semantic role labelling. However, as discussed, there is a large room for improvement by extending the relations between the predicates and the roles.

For example, consider a **buying** event. This event can be expressed by different predicates such as **buy** or **sell** depending the point of view of the participants of the event. In fact, the same simple **purchasing** event is composed by two **transfer** sub-events, regarding the *Goods* and the *Money*. This means that for every **buying** event there is always implied a **selling** and a **paying** event. Similarly, for every **selling** event there is always implied a **buying** and a **paying** event, and for every **paying** event there is always implied a **buying** and **selling** events. In other words, the participants of these events are also shared. However, a major obstacle for exploiting this type of semantic relations if the lack of resources encoding this knowledge. Thus, some research have been focused on automatically deriving this type of knowledge from corpora (Chambers and Jurafsky, 2009) or available semantic resources such as WordNet or FrameNet (Aharon et al., 2010).

As pointed out by previous works (Coyne and Rambow, 2009; Aharon et al., 2010), FrameNet, as a manually constructed semantic resource, contains highly accurate representations of this type of knowledge. Thus, in this chapter, we present a novel proposal to derive a large set of frame-element relations from FrameNet and their application to Implicit Semantic Role Labelling (ISRL) based on PropBank/NomBank. For this, we automatically build a wide coverage resource containing semantic relations between argument from PropBank/NomBank from those defined for the FrameNet frame-elements. In summary, the development of the resource consists in the following two steps:

- Defining a set of rules to extend the semantic relations between frame-elements of FrameNet.
- Obtaining an automatic mapping between FrameNet frame-elements and PropBank/NomBank arguments and use that mapping to transfer the semantic relations.

First, each semantic relation between FrameNet (Baker et al., 1998) frames has been defined strictly from a parent frame to their direct children and, consequently, between their frame-elements. But the number of direct relations can be extended. For example the ***SubFrame*** and ***Perspective_on*** relations establish that the child frames are, indeed, particular or more specific points of view of the parent frame. Although all the sibling frames that share this kind of relations with the same parent are obviously interrelated, FrameNet does not not usually connect them directly. For instance, the frames **Commerce_sell** and **Commerce_buy** are connected to **Commerce_goods-transaction** through ***Perspective_on*** but they do no share any relation directly. The rules proposed in these chapter would set a new direct link between both frames.

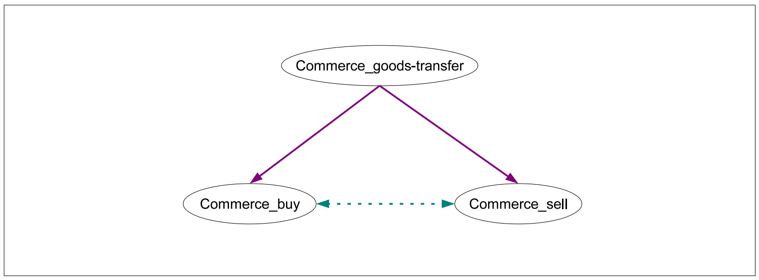

Figure 7.1: Example of extending a ***Perspective_on*** relation. The dotted line represents a relations that does not exist in FrameNet originally.

The second step in order to exploit the relations of FrameNet for PropBank/NomBank based ISRL consists in to transfer FrameNet relations to predicates and arguments of PropBank/NomBank, as show Figure 7.2. For instance, as the predicates **purchase.v** and **sell.v** belongs to **Commerce_buy** and **Commerce_sell** respectively, the new relation obtained in Figure 7.1 could be replicated in NomBank with a proper mapping between the predicate and roles. However, the currently available mappings have some drawbacks. For instance, both SemLink (Palmer, 2009) and its extension, the Predicate Matrix (López de Lacalle et al., 2014b, a) provides mappings between VerbNet, WordNet, FrameNet and PropBank. However, their coverage is still insufficient, specially with respect FrameNet frame-elements and PropBank/NomBank arguments. Thus, we propose a novel method to obtain automatically reliable sets of mappings between these two resources.

Once we have the resource that relates arguments of PropBank/NomBank, the exploitation of this new semantic relations for ISRL can be carried out in very straightforward manner by including for each SRL annotation in the text all the related information available for the arguments and applying an existing ISRL system for processing, like **ImpAr**. Furthermore, using the same system allows to easily evaluate the contribution of the semantic relations and to find out which types of these relations are more adequate for the

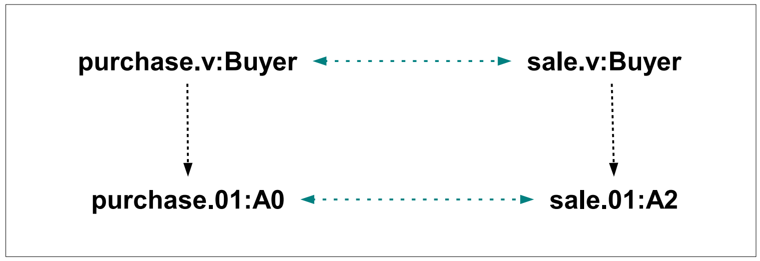

Figure 7.2: Mappings between the resources allow to transfer FN relations to PB/NB.

particular task of ISRL. For instance, those that are more general, like ***Inheritance***, or those, like ***SubFrame***, that are more specific to the scenario. Additionally, it also allows to easily compare against other possible knowledge sources relating argument predicates (Kipper, 2005; Chambers and Jurafsky, 2009; Coyne and Rambow, 2009; Aharon et al., 2010) and mappings between FrameNet and PropBank/NomBank (Palmer, 2009; López de Lacalle et al., 2014b, a).

The example in Figure 7.3 illustrates how the new relations among PropBank/NomBank arguments can help to solve elided roles. This example contains the explicit NomBank annotation of the nominal predicates **purchase.01** and **sale.01**. As can be seen, while the arg_0 of **purchase.01** has been filled, all the roles of **sale.01** are missing. However, according to the new semantic relation showed in Figure 7.2 the arg_2 of **sale.01** could be filled implicitly.

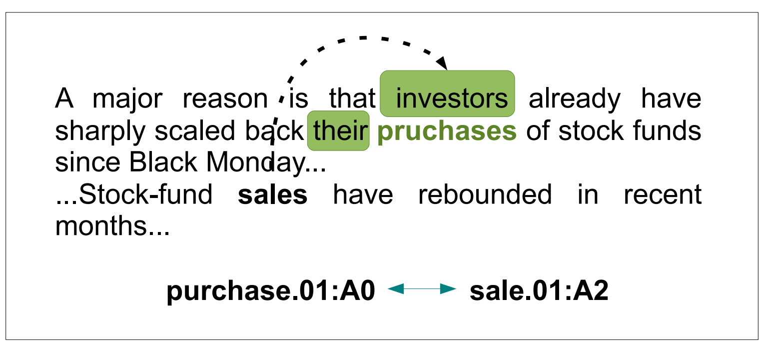

Figure 7.3: Example of a case of implicit arguments solved by the new relations.

Four years ago [$_{Buyer}$ I] BOUGHT^{Commerce_buy} [$_{Goods}$ an old Harmony Sovereign acoustic guitar] [$_{Money}$for £20] [$_{Seller}$ from an absolute prat] .

Table 7.1: An example of FrameNet-based annotation for the frame **Commerce_buy**.

7.2 Extending the relations between FrameNet frames

One of the most interesting characteristics of FrameNet is that its frames are not created as independent entities but being part of a large semantic net formed by a rich set of relations that connect the frames and also their frame-elements. FrameNet frames are related to each other by a fixed set of frame relations. In addition, frame relations are used to define the mapping between corresponding frame-elements in the related frames. Some of the relevant frame relations are ***Inheritance***, ***Perspective_on***, ***Inchoative_of***, ***Causative_of***, ***SubFrame*** and ***Using***. The purpose of defining such kind of information is to facilitate inferencing tasks such as textual entailment.

Regarding to the goals of this dissertation, the semantic and ontological structures described between the different frames of FrameNet can provide richer and more complex information about the events described in a document than those obtained by regular SRL. For example, the sentence in Table 7.1 includes FrameNet based annotations of semantic roles for the verb **buy.n** of the frame **Commerce_buy**.

The annotation above includes the participants of the event that appears explicitly on the text but leaves out many information that is inherent and that a human reader assumes implicitly. To see how FrameNet can help to get a deeper understanding of the events described we can take a look at how the frames interrelate. For instance **Commerce_buy** and the other frames involved with it, as can be seen in Figure 7.4. For that figure and following in the present section the meaning of the colours of the arrows is described in the legend in Figure 7.5. Notice that ***Sibling*** is not an original FrameNet relation that will be explained later.

According to the graph in Figure 7.4, the **Commerce_buy** frame has just a direct link with the **Commerce_goods-transfer** frame through the

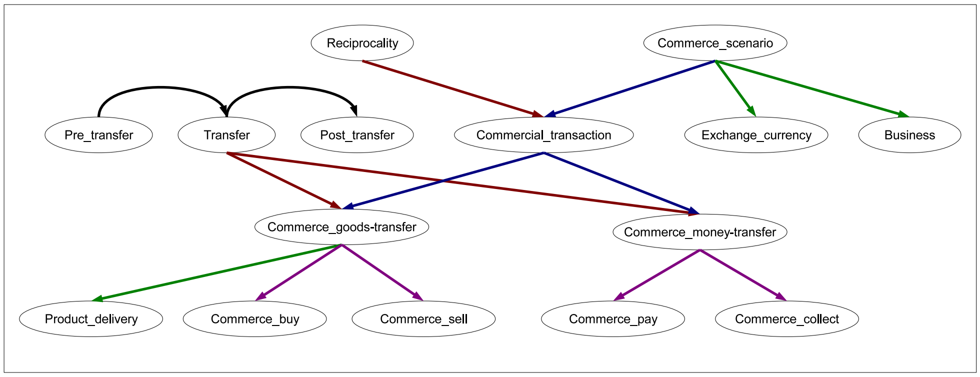

Figure 7.4: Different relations for the Commercial_Transaction frame.

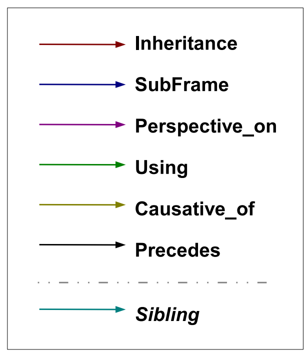

Figure 7.5: Relation type by colour.

Perspective_on relation. A very important question is that the same relations existing between two frames also relates their frame-elements, as show in Figure 7.6. This means that the frame-elements of **Commerce_goods-transfer** and **Commerce_buy** are related as in Table 7.2.

However, the different kind of existing relations allow to establish a huge number of indirect connections. For instance, two frames connected to the same frame by the ***Perspective_on*** relation means that they are referring to the same kind of event but from different points of view. Consequently, for the previous case the links in Table 7.3 between the **Commerce_buy** and **Commerce_sell** frames could be inferred.

Furthermore, this kind of inference can be extended through different relations. Following the previous example, **Commerce_buy** frame could be also connected to the **Commerce_pay** frame using the ***Perspective_on***

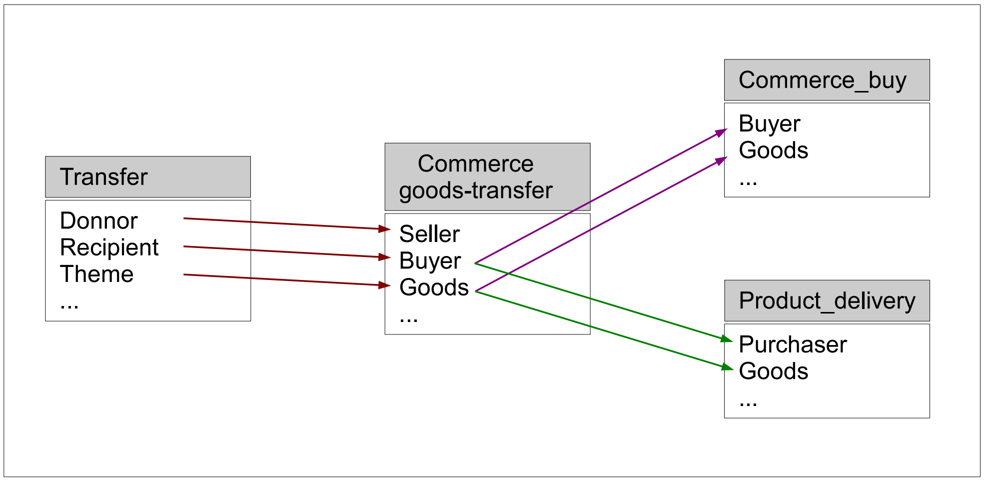

Figure 7.6: Different relations for the **Commercial_Transaction** frame.

Commerce_goods-transfer *Buyer*	=	**Commerce_buy** *Buyer*
Commerce_goods-transfer *Goods*	=	**Commerce_buy** *Goods*
Commerce_goods-transfer *Money*	=	**Commerce_buy** *Money*
Commerce_goods-transfer *Seller*	=	**Commerce_buy** *Seller*

Table 7.2: Relations between the FEs of the frames **Commerce_goods-transfer** and **Commerce_buy**.

relations, like shown in Table 7.4, and the ***SubFrame*** relation, that means that a frame is indeed a part of a parent frame.

The relations also connects frames that do not belong to the same scenarios. For example, both **Commerce_buy** and **Commerce_pay** frames are indirectly connected to the **Transfer** frame by the ***Inheritance*** relation (see Table 7.5 and Table 7.6).

The relations showed in these examples mean that when a **buy.v** event occurs it implies many other events that take place simultaneously or are consequence of each other. Taking into account the relations observed previously, the annotation of sentence presented in Figure 7.4 can be extended as shown in Table 7.7.

Commerce_sell *Buyer*	=	**Commerce_buy** *Buyer*
Commerce_sell *Goods*	=	**Commerce_buy** *Goods*
Commerce_sell *Money*	=	**Commerce_buy** *Money*
Commerce_sell *Seller*	=	**Commerce_buy** *Seller*

Table 7.3: Relations between the FEs of the frames **Commerce_sell** and **Commerce_buy**.

Commerce_pay *Buyer*	=	**Commerce_buy** *Buyer*
Commerce_pay *Goods*	=	**Commerce_buy** *Goods*
Commerce_pay *Money*	=	**Commerce_buy** *Money*
Commerce_pay *Seller*	=	**Commerce_buy** *Seller*

Table 7.4: Relations between the FEs of the frames **Commerce_pay** and **Commerce_buy**.

Transfer *Recipient*	=	**Commerce_buy** *Buyer*
Transfer *Theme*	=	**Commerce_buy** *Goods*
Transfer *Donor*	=	**Commerce_buy** *Seller*

Table 7.5: Relations between the FEs of the frames **Transfer** and **Commerce_buy**.

Transfer *Recipient*	=	**Commerce_pay** *Seller*
Transfer *Theme*	=	**Commerce_pay** *Money*
Transfer *Donor*	=	**Commerce_pay** *Buyer*

Table 7.6: Relations between the FEs of the frames **Transfer** and **Commerce_pay**.

Four years ago I **bought** an old Harmony Sovereighn acoustic guitar for £20 from an absolute prat.

buy$^{Commerce\_buy}$
[$_{Buyer}$ I]
[$_{Goods}$ an old Harmony Sovereign acoustic guitar]
[$_{Money}$ for £20]
[$_{Seller}$ from an absolute prat]

sell$^{Commerce\_sell}$
[$_{Buyer}$ I]
[$_{Goods}$ an old Harmony Sovereign acoustic guitar]
[$_{Money}$ for £20]
[$_{Seller}$ from an absolute prat]

pay$^{Commerce\_pay}$
[$_{Buyer}$ I]
[$_{Goods}$ an old Harmony Sovereign acoustic guitar]
[$_{Money}$ for £20]
[$_{Seller}$ from an absolute prat]

transfer^{Transfer}
[$_{Recipient}$ I]
[$_{Theme}$ an old Harmony Sovereign acoustic guitar]
[$_{Donor}$ from an absolute prat]

transfer^{Transfer}
[$_{Donor}$ I]
[$_{Theme}$ for £20]
[$_{Recipient}$ from an absolute prat]

Table 7.7: Extension of the FrameNet-based annotation for the frame **Commerce_buy**.

7.2.1 Declarative rules for extending the relations

As explained previously, the relations of FrameNet form a huge graph where the participants of different events are interconnected. These links express the implications among the roles. In other words, when the participants that take part of different events are actually the same, as the case of the *Buyer* of the **Commerce_buy** and **Commerce_pay** frames, or when they are one a subtype of the other, like the *Donor* of the **Transfer** frame and the *Seller* of the **Commerce_sell** frame. Unfortunately, FrameNet only encodes direct relations and does not provide any formal representation for them. Thus, we propose a set of rules for inferring new FrameNet relations among frames and frame-elements based on the descriptions included in its technical documentation. Note that currently we do not set any rule for ***See_also***, ***Causative_of*** and ***Inchoative_of*** relations.

7.2.1.1 Inheritance

The ***Inheritance*** relation is defined as the usual *subclass* between two frames. For instance, **Escaping** inherits from **Departing**. This relation can be expressed as *Inheritance(Departing,Escaping)*. Thus, *Inheritance(parent,child)* relation establishes that all the properties of the *parent* frame must be inherited by the *child* frame. This means that all the relations involving the *parent* frame also involve the *child* one. Furthermore, this is a transitive relation. For example, the *Travel* frame is a *subclass* of **Self_motion** which is a *subclass* of **Motion** and **Intentionally_act**. In consequence, **Travel** also inherits frame-elements from **Motion** and **Intentionally_act** frames. Moreover, the **Motion** frame has a ***Causative_of*** relation with **Cause_motion** frame that is inherited by **Self_motion** and **Travel** through the ***Inheritance*** relations. The graph in Figure 7.7 describes the previous relations.

We define these properties by the rules in Table 7.8 where N_n stands for a particular frame or frame-element. As a result of applying these rules to the set of existing relations in FrameNet, as shown in the Figure 7.7, the instantiations and relations in Table 7.9 are produced.

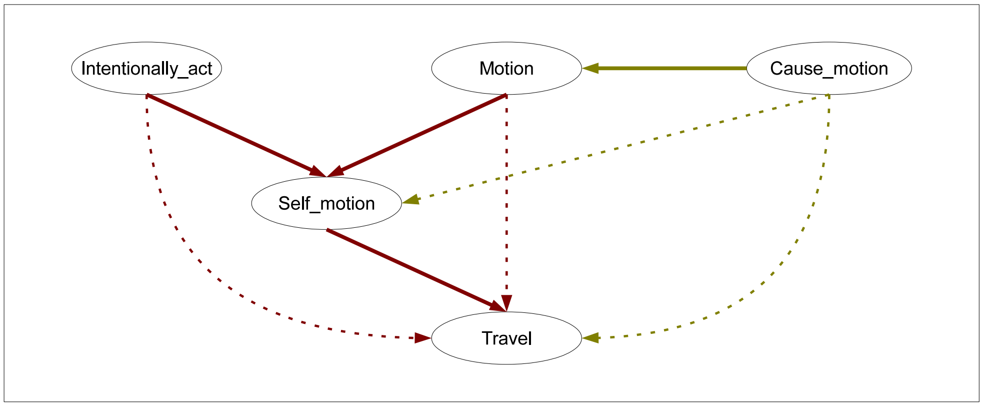

Figure 7.7: Extension through ***Inheritance***.

$$Inheritance(N_1, N_2) \wedge Inheritance(N_2, N_3) \Rightarrow Inheritance(N_1, N_3)$$
$$Inheritance(N_1, N_2) \wedge Relation(N_3, N_1) \Rightarrow Relation(N_3, N_2)$$

Table 7.8: Declarative rules for the ***Inheritance*** relations. ***Relation*** stands for any othe relation type different to ***Inheritance***.

frames:

$Inheritance(Self\_motion, Travel) \wedge Inheritance(Motion, Self\_motion) \Rightarrow$
$\Rightarrow Inheritance(Motion, Travel)$

$Inheritance(Self\_motion, Travel) \wedge Inheritance(Intentionally\_act, Self\_motion) \Rightarrow$
$\Rightarrow Inheritance(Intentionally\_act, Travel)$

$Inheritance(Motion, Self\_motion) \wedge Causative\_of(Cause\_motion, Motion) \Rightarrow$
$\Rightarrow Causative\_of(Cause\_motion, Self\_motion)$

frame-elements:

$Inheritance(Self\_mover, Traveler) \wedge Inheritance(Theme, Self\_mover) \Rightarrow$
$\Rightarrow Inheritance(Theme, Traveler)$

$Inheritance(Self\_mover, Traveler) \wedge Inheritance(Agent, Self\_mover) \Rightarrow$
$\Rightarrow Inheritance(Intentionally\_act\#Agent, Travel\#Traveler)$

$Inheritance(Theme, Self\_mover) \wedge Causative\_of(Theme, Theme) \Rightarrow$
$\Rightarrow Causative\_of(Theme, Self\_mover)$

Table 7.9: Examples of relations between frames and FEs extended through ***Inheritance***.

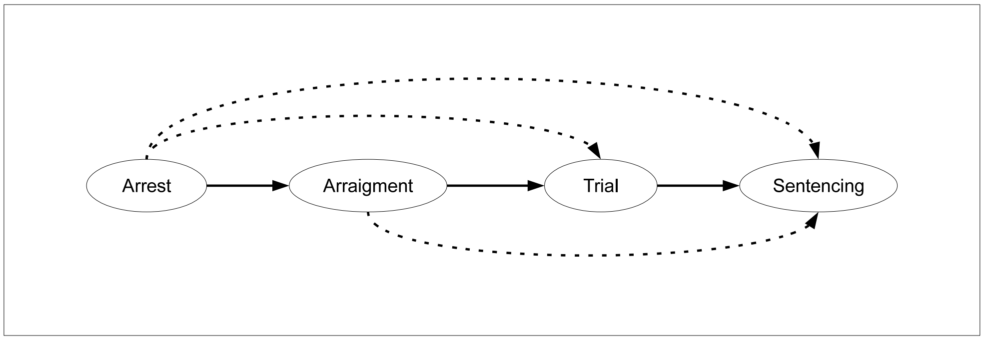

Figure 7.8: Extension through ***Precedes***.

$$Precedes(N_1, N_2) \wedge Precedes(N_2, N_3) \Rightarrow Precedes(N_1, N_3)$$

Table 7.10: Declarative rule for the ***Precedes*** relations.

frames:

$Precedes(Arrest, Arrangement) \wedge Precedes(Arrangement, Trial) \Rightarrow$
$\Rightarrow Precedes(Arrest, Trial)$

$Precedes(Arrest, Trial) \wedge Precedes(Trial, Sentencing) \Rightarrow$
$\Rightarrow Precedes(Arrest, Sentencing)$

frame-elements:

$Precedes(Suspect, Defendant) \wedge Precedes(Defendant, Defendant) \Rightarrow$
$\Rightarrow Precedes(Suspect, Defendant)$

$Precedes(Suspect, Defendant) \wedge Precedes(Defendant, Convict) \Rightarrow$
$\Rightarrow Precedes(Suspect, Convict)$

Table 7.11: Examples of extended ***Precedes*** relations between frames and FEs.

7.2.1.2 Precedes

The ***Precedes*** relation expresses sequences of frames that are part of a more general scenario. For example, like shown in Figure 7.8, a **Criminal_process** can be factored into a sequence of frames such as **Arrest**, **Arraignment**, **Trial** and **Sentencing** occurring one after another. Obviously, if the **Arraignment** happens after the **Arrest** and the **Trial** happens after the **Arraignment**, the **Trial** will also befall after the **Arrest**.

In this case, the relation is transitive but the other relations are not transferred. For the sequence of events described in Figure 7.8, the rule in Table 7.10 produces relations like those in Table 7.11.

7.2.1.3 Using

A particular relation in FrameNet is ***Using*** that connects two frames when parts of the scene described by the *child* refers to the *parent* but in a different manner that the ***Inheritance*** relation does. For example, in Figure 7.9 the **Protest** frame describes an event that evokes those covered by **Taking_sides** but it cannot be said that one is a subtype of the other. Although it evokes both frames, **Taking_sides** is neither a *subclass* of **Opinion** nor **Desirable_event**.

This relation is also transitive. The application of the rule in Table 7.12 can generate some new direct relations between frames and frame-elements as the ones in Table 7.13.

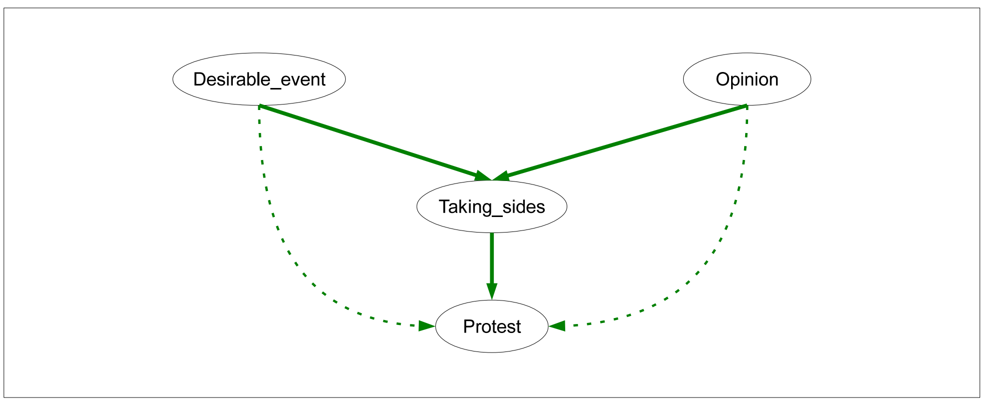

Figure 7.9: Extension through ***Using***.

$$Using(N_1, N_2) \wedge Using(N_2, N_3) \Rightarrow Using(N_1, N_3)$$

Table 7.12: Declarative rule for the ***Using*** relations.

frames:

$Using(Opinion, Taking\_side) \wedge Using(Taking\_sides, Protest) \Rightarrow$
$\Rightarrow Using(Opinion, Protest)$

$Using(Desirable\_event, Taking\_side) \wedge Using(Taking\_sides, Protest) \Rightarrow$
$\Rightarrow Using(Desirable\_event, Protest)$

frame-elements:

$Using(Cognizer, Cognizer) \wedge Using(Cognizer, Protester) \Rightarrow$
$\Rightarrow Using(Cognizer, Protester)$

$Using(State\_of\_affairs, Action) \wedge Using(Action, Action) \Rightarrow$
$\Rightarrow Using(State\_of\_affairs, Action)$

Table 7.13: Examples of extended ***Using*** relations between frames and FEs.

7.2.1.4 SubFrame

The ***SubFrame*** relation is used when a frame is so complex that many other more specific frames can be defined inside it. In these cases, the *children* frames can be seen as sub-parts of the *parent* that share the same participants. For example, the frame **Invasion_scenario** involves the more specific events described by **Invading**, **Conquering** and **Repel** frames.

As seen in the example in Figure 7.10, the particularity of this relation is that the frame-elements of the *children* and the *parent* are indeed participants of the same scene, but *sibling* frames are rarely connected between them. Our proposal, shown in Table 7.14, for this case is to link the *sibling* frames by a new relation we denote ***Sibling***. For the structure of ***SubFrames*** in Figure 7.10, the set of rules in Table 7.14 obtain cases like those in Table 7.15.

$$SubFrame(N_1, N_2) \wedge SubFrame(N_1, N_3) \Rightarrow Sibling(N_2, N_3)$$
$$SubFrame(N_1, N_2) \wedge SubFrame(N_1, N_3) \Rightarrow Sibling(N_3, N_2)$$
$$SubFrame(N_1, N_2) \wedge Relation(N_3, N_1) \Rightarrow Relation(N_3, N_2)$$

Table 7.14: Declarative rules for the ***SubFrame*** relations.

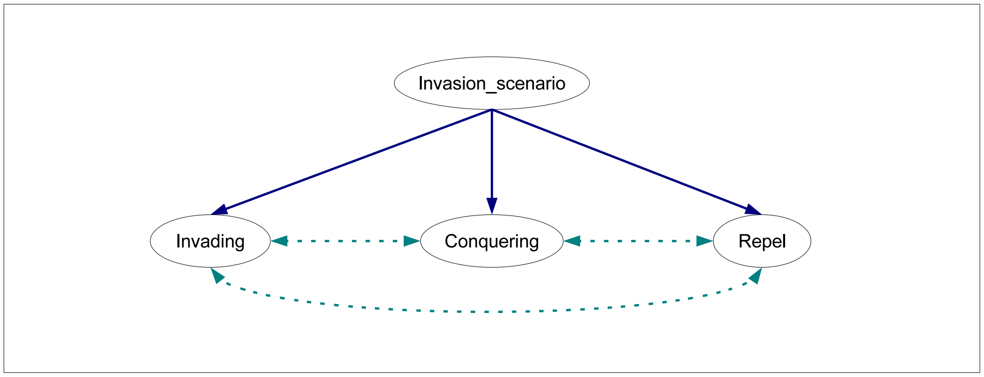

Figure 7.10: Extension through ***SubFrame***.

frames:

$$SubFrame(Invasion\_scenario, Conquering) \wedge SubFrame(Invasion\_scenario, Invasion) \Rightarrow$$
$$\Rightarrow Sibling(Conquering, Invasion)$$

$$SubFrame(Invasion\_scenario, Conquering) \wedge SubFrame(Invasion\_scenario, Invasion) \Rightarrow$$
$$\Rightarrow Sibling(Invasion, Conquering)$$

$$SubFrame(Invasion\_scenario, Conquering) \wedge SubFrame(Invasion\_scenario, Repel) \Rightarrow$$
$$\Rightarrow Sibling(Conquering, Repel)$$

$$SubFrame(Invasion\_scenario, Conquering) \wedge SubFrame(Invasion\_scenario, Repel) \Rightarrow$$
$$\Rightarrow Sibling(Repel, Conquering)$$

frame-elements:

$$SubFrame(Invader, Invader) \wedge SubFrame(Invader, Conqueror) \Rightarrow$$
$$\Rightarrow Sibling(Invading, Conqueror)$$

$$SubFrame(Invader, Invader) \wedge SubFrame(Invader, Conqueror) \Rightarrow$$
$$\Rightarrow Sibling(Conqueror, Invading)$$

$$SubFrame(Invader, Conqueror) \wedge SubFrame(Invader, Enemy) \Rightarrow$$
$$\Rightarrow Sibling(Invader, Enemy)$$

$$SubFrame(Invader, Conqueror) \wedge SubFrame(Invader, Enemy) \Rightarrow$$
$$\Rightarrow Sibling(Enemy, Invader)$$

Table 7.15: Examples of extended relations between frames and FEs through ***SubFrame***.

7.2.1.5 Perspective_on

If different points of view can be taken on a single frame, then those different perspectives can be described by distinct frames and connected to a previous one by the ***Perspective_on*** relation. In fact, the event describing all the frames involved by this relation is exactly the same. Figure 7.11 shows how

Being_born and **Giving_birth** frames express two different points of view of the **Birth_scenario** frame.

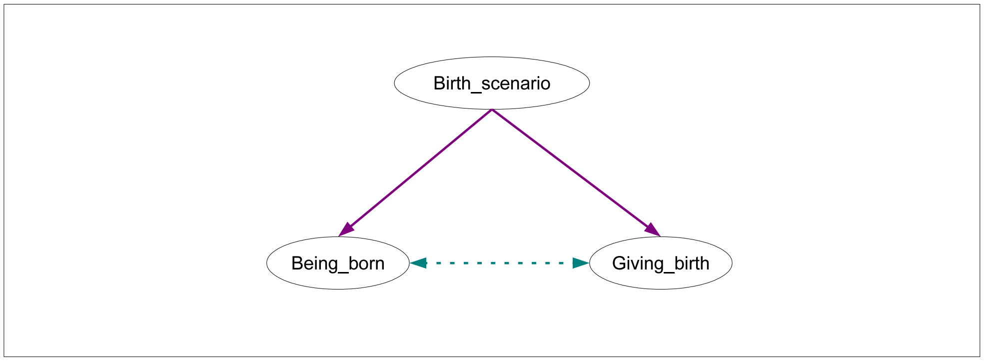

Figure 7.11: Extension through Perspective_on.

$$Perspective\_on(N_1, N_2) \wedge Perspective\_on(N_1, N_3) \Rightarrow Sibling(N_2, N_3)$$
$$Perspective\_on(N_1, N_2) \wedge Perspective\_on(N_1, N_3) \Rightarrow Sibling(N_3, N_2)$$
$$Perspective\_on(N_1, N_2) \wedge Relation(N_3, N_1) \Rightarrow Relation(N_3, N_2)$$

Table 7.16: Declarative rules for the ***Perspective_on*** relations.

frames:

$Perspective\_on(Birth\_scenario, Being\_born) \wedge Perspective\_on(Birth\_scenario, Giving\_birth) \Rightarrow$
$\Rightarrow Sibling(Being\_born, Giving\_birth)$

$Perspective\_on(Birth\_scenario, Being\_born) \wedge Perspective\_on(Birth\_scenario, Giving\_birth) \Rightarrow$
$\Rightarrow Sibling(Giving\_birth, Being\_born)$

frame-elements:

$Perspective\_on(Offspring, Child) \wedge Perspective\_on(Offspring, Child) \Rightarrow$
$\Rightarrow Sibling(Child, Child)$

$Perspective\_on(Offspring, Child) \wedge Perspective\_on(Offspring, Child) \Rightarrow$
$\Rightarrow Sibling(Child, Child)$

Table 7.17: Examples of extended relations between frames and FEs through ***Perspective_On***.

The ***Perspective_on*** relation shares the same particularities with the ***SubFrame*** relation being the frame-elements of the *children* and the *parent* participants of the same scene. In consequence, we also propose the ***Sibling*** relations to connect *sibling* frames, as showed in Table 7.16. In this case, the set of rules described for the ***Perspective_on*** relations can produce the new examples in Table 7.17.

7.2.2 Applying the rules

We have generated the transitive closure of the set of rules described in this section on FramNet 1.3 and we obtain a dramatical increment of the direct relations between frames and frame-elements, as is showed in Table 7.18.

	Direct	Inferred
frames	1,723	4,360
frame-elements	8,486	16,795

Table 7.18: Number of direct relations obtained between frames and between frame-elements.

7.3 Mapping to PropBank

PropBank also describes a large number of predicates with their corresponding role structures. In fact, PropBank covers a much larger number of predicates than FrameNet. This fact, along with a huge annotated corpus, has made PropBank the main paradigm for SRL (Carreras and Màrquez, 2004; Carreras and Màrquez, 2005; Surdeanu et al., 2008; Hajič et al., 2009). However, unlike FrameNet, PropBank defines the arguments, or roles, of each verb individually. In consequence, it becomes very difficult to obtain more abstract generalizations over their verbs and arguments. Moreover, there is no semantic relation defined between the roles of different predicates.

Of course, it is possible to apply the semantic relations of FrameNet to PropBank and NomBank using a proper mapping procedure between these resources. Unfortunately, existing resources such as SemLink (Palmer, 2009) or the Predicate Matrix (López de Lacalle et al., 2014b, a) still do not provide a complete mapping between FrameNet and PropBank/NomBank. For instance, the automatic methods applied to create the Predicate Matrix by extending SemLink have been focused on mapping VerbNet (Kipper, 2005), FrameNet (Baker et al., 1998) and WordNet (Fellbaum, 1998) and they manage to extend the mappings involving PropBank (Palmer et al., 2005) just as a side-effect. Moreover, the current version of the Predicate Matrix does not take into account nominal predicates.

7.3.1 SemLink

SemLink (Palmer, 2009) is a project whose aim is to link together different predicate resources establishing a set of mappings. These mappings make it possible to combine the different information provided by the different lexical resources for tasks such as inferencing, consistency checking, interoperable semantic role labelling, etc.

Having VerbNet (Kipper, 2005) as the central resource, SemLink includes partial mappings to PropBank (Palmer et al., 2005) and FrameNet (Baker et al., 1998), but their coverage presents significant gaps (López de Lacalle et al., 2014a).

The complexity of these mappings relies on the fact that they must be established at two different levels, the lexicon/predicate level and the role level. For example, from the **6,181** different PropBank predicates, just **3,558** have a corresponding VerbNet predicate in SemLink. That is, **2,623** PropBank predicates have no correspondences to VerbNet. Regarding the PropBank arguments and the VerbNet thematic-roles, **7,915** out of **15,871** arguments from PropBank are mapped to a VerbNet thematic-role. That is, only a half of the total PropBank arguments are linked to VerbNet, leaving out the remaining **7,956** arguments.

In the other hand, the alignment between FrameNet and VerbNet proves to be even more incomplete. For example, only **1,730** lexical-units from FrameNet are aligned to, at least, one VerbNet predicate. This number represents only 16% out of the total **10,195** lexical-units of FrameNet.

SemLink also includes the alignment between the roles of both resources. However, unlike PropBank, the roles of FrameNet are defined at a frame-level and not at a predicate level. Therefore, the mapping of the VerbNet thematic-roles and the frame-elements of FrameNet is defined between VerbNet classes and FrameNet frames. Once again, the mapping between VerbNet and FrameNet presents significant gaps and mismatches. For instance, just **825** of the **7,124** frame-elements of FrameNet are linked to a VerbNet thematic-role. That is, 88% of the frame-elements from FrameNet are not aligned to any VerbNet thematic-role.

As result of all these gaps, the mapping that can be obtained from SemLink comprising PropBank and FrameNet is severely sparse. Only **981** predicates of PropBank are linked to a lexical-unit of FrameNet, and just **2,359** arguments of those predicates are mapped to a frame-element.

Moreover, SemLink does not consider nominal predicates, neither those from FrameNet nor those from NomBank.

7.3.2 Predicate Matrix

Predicate Matrix (López de Lacalle et al., 2014b, a) is an extension of SemLink. The Predicate Matrix follows the line of WordFrameNet (Laparra and Rigau, 2009; Laparra et al., 2010; Laparra and Rigau, 2010) for the integration of multiple sources of predicate information including FrameNet, VerbNet, PropBank and WordNet. Applying a variety of automatic methods, the Predicate Matrix extends widely the existing mapping coverage in SemLink.

The resulting mappings included in Predicate Matrix solve some of the existing gaps of SemLink. For example, the number of ProbBank predicates with a least a mapping to a VerbNet class rises to **3,961** and the number of arguments to **9,288**. That means an increase of almost **2,000** of PropBank arguments. On the other hand, the number of frame-elements that result to be mapped in Predicate Matrix, **1,839**, is more than the double of those contained in SemLink, **825**.

Consequently, the connections between PropBank and FrameNet are also extended. In fact, the Predicate Matrix contains **1,982** PropBank predicates and **4,440** arguments mapped to a frame of FrameNet. However, this extension is obtained indirectly as a side-effect of the integration process not because of an on purpose integration between FrameNet and PropBank.

7.3.3 Creating an automatic mapping through cross annotations

Previous works on automatic mapping PropBank/NomBank and FrameNet have focused mainly in the lexicon of both resources. However, as the semantic relations of FrameNet are defined between frame-elements, the only way to define those relations on PropBank/NomBank is by a proper mapping of the roles.

To obtain this mapping automatically we propose to discover the most common correspondences between the annotations of both resources over the same sentences, as in the example in Figure 7.12.

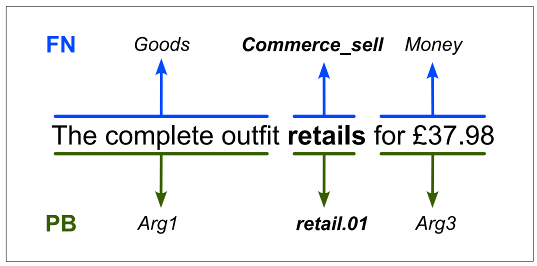

Figure 7.12: Example of matching annotations.

Obviously, we exploit already existing manually annotated FrameNet and PropBank/NomBank corpora that assures a fully reliable annotation. The FrameNet corpora can be divided in two different sets. In the one hand, FrameNet version 1.3 includes 168,519 sample sentences for the 64% of the LUs. In the other hand, the FrameNet corpus contains continuous text annotations for 99 documents from different sources as WikiNews or the American National Corpus. PropBank adds predicate-argument relations to the syntactic trees of the Penn TreeBank Wall Street Journal data, and it has been lately extended with nominal predicates by the NomBank project. For our purposes, we use a subset of PropBank/NomBank distributed by the CoNLL shared-task that comprises 5 sections including 500 different documents.

To automatically obtain the corresponding counterparts of the data presented above we have made use of two available tools that offer state of the art results on semantic role labelling using FrameNet and PropBank/NomBank. For the FrameNet based annotations we use SEMAFOR[1] (Chen et al., 2010) that carries out both frame and frame-element identification with an overall performance of **62.76%** precision and **41.89%** recall. The SEMAFOR package includes a modified version of the MST Parser (McDonald et al., 2005) to obtain the required syntactic dependencies. The PropBank/NomBank based

[1] http://www.ark.cs.cmu.edu/SEMAFOR/

annotation has been done using the mate-tools[2] (Björkelund et al., 2009) a complete multilingual NLP pipeline that includes a high accuracy SRL module that obtains **85.63%** F1 performance.

In this way, we are able to obtain the corpus from FrameNet with manual FrameNet annotations and automatic PropBank/NomBank annotations from the mate-tools. Similarly, we are also able to get the corpus from PropBank/NomBank with manual PropBank/NomBank annotations and automatic FrameNet annotations from SEMAFOR. Thus, by crossing the annotations on both corpora it has been possible to collect the coincidences when the filler of one PropBank/NomBank argument matches a FrameNet frame-element or vice-versa. Then, we have removed those cases we consider too infrequent setting a *threshold* of more than **T** cases per <PropBank/NomBank-argument,FrameNet-frame-element> pair. We have applied different values of **T** obtaining different sets of mappings. Finally we select the most common ones for each predicate. For example, for the predicate **retail.01** we obtain that the arg_1 and the arg_3 match most frequently the frame-elements *Goods* and *Money* of the frame **Commerce_sell** respectively. For this predicate, this method does not provide enough evidence for the rest of arguments. Following this **cross-annotation strategy**, we generate a mapping that connects not only the predicates but also the roles. Furthermore, the mapping includes nominal predicates.

	PropBank		**NomBank**	
	Predicates	**Roles**	**Predicates**	**Roles**
SemLink	1,722	4,394	0	0
PM	2,900	7,493	0	0
T=0	3,446	12,858	3,546	8,722
T=1	2,688	8,232	2,594	5,537
T=4	2,002	4,745	1,718	3,036
T=7	1,651	3,383	1,313	2,133

Table 7.19: Number of mappings obtained with different values of **T** compared to SemLink and PM.

Table 7.19 shows, for different values of **T**, the number of mappings obtained from the **cross-annotation strategy**. The table also shows that the

[2]https://code.google.com/p/mate-tools/

application of this method substantially increase the number of mappings encoded in SemLink and the Predicate Matrix.

7.3.4 Projecting the FrameNet relations to PropBank and NomBank

Once we have the mapping between FrameNet frame-elements and PropBank/NomBank arguments we can project directly relations between frame-elements obtained in Section 7.2. The volume of relations between PropBank/NomBank arguments in the resulting resource depends on the threshold **T** chosen generating the automatic mapping, as shown in Table 7.20. The second column of the table shows that the number of the relations highly increase if we also include the relations between predicates belonging to the same frame of FrameNet.

	arg-Relations	**+SameFrame**
T=0	608,410	847,889
T=1	322,776	457,887
T=4	153,448	217,954
T=7	96,530	136,479

Table 7.20: Number of resulting relations between PropBank/NomBank arguments mappings obtained with different values of **T**.

Note that for this projection we can substitute the mappings we obtain in this section with the mappings from SemLink and Predicate Matrix. In Section 7.4 we present comparison of the results

7.4 Enriching SRL for discovering implicit information

As we propose at the beginning of this chapter, the semantic relations between roles or arguments can be exploited to obtain a more richer representation of the eventive structures in a document, beyond the ones provided by regular SRL. Furthermore, we suggest that the richer this representations are the more sources are available for completing those roles whose fillers have been elided. In order to check these ideas, we have analysed how the inclusion of this information affects an existing system for ISRL.

The algorithm used in this work, **ImpAr**, processes the documents sentence by sentence guessing that every mention of a sequence of the same predicate refers to the same event and, in consequence, all of theme share the same argument fillers. This assumption is useful to capture elided arguments. Thus, for every **core** argument arg_n of a predicate, **ImpAr** stores its previous known filler, explicit or implicit, as a possible filler of the following occurrences of the same argument. When there is enough evidence this "default" filler is changed, for example when an explicit mention of the argument is found.

Originally, this behaviour only affects to mentions of the same predicate in any of its possible forms, including nominalizations. However, semantic relations between roles can be used to extend the SRL annotations allowing **ImpAr** to perform a more complex kind of inference. For instance, in the example of the predicate **retail.01** presented in the Figure 7.12, the relations between roles obtained from FrameNet show that the same event can be expressed in very different manner and, even more, can imply necessarily some other events, as Figure 7.13 shows.

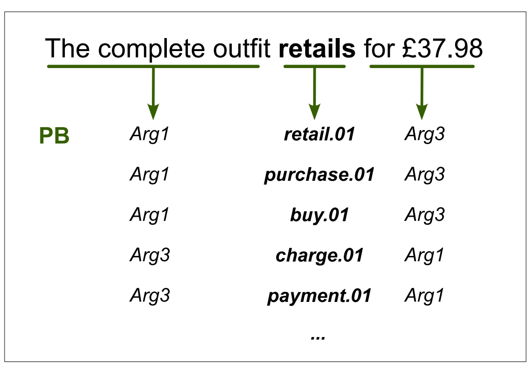

Figure 7.13: Extended annotation for the **retail.01** predicate.

This means that the fillers of the arguments arg_1 and arg_3 can be considered as default fillers for the following mentions of not only the **retail.01** predicate but also **purchase.01**, **buy.01** or **payment.01**.

7.4.1 Evaluation

To empirically measure the contribution on ISRL of the newly acquired relations between roles using **ImpAr**, we use the same evaluation framework as the one described in Chapter 6. Consequently, this involves the same corpus and test sets developed by Gerber and Chai (2010, 2012) and the metric proposed by Ruppenhofer et al. (2010), lately adapted by Gerber and Chai (2010, 2012) which is based on the Dice coefficient. We also use **ImpAr** with the same settings as the ones described in Chapter 6.

			None	**SameFrame**	**All**	**Best**
	Inst.	**Imp.**	$\mathbf{F}_1$	$\mathbf{F}_1$	$\mathbf{F}_1$	$\mathbf{F}_1$
sale	184	181	40.6	40.6	43.0	**45.0**
price	216	138	50.3	50.3	54.7	**55.5**
investor	160	108	25.3	25.3	25.3	25.3
bid	88	124	47.0	47.0	47.0	47.0
plan	100	77	**48.0**	45.0	42.0	45.0
cost	101	86	**44.5**	**44.5**	44.2	44.2
loss	104	62	**55.2**	54.7	54.7	54.7
loan	84	82	45.3	45.3	45.3	45.3
investment	102	52	29.2	29.2	29.2	29.2
fund	108	56	47.5	47.5	47.5	47.5
Overall	1,247	966	43.0	42.7	43.5	**44.5**

Table 7.21: Results of **ImpAr** with different sets of semantic relations.

Table 7.21 shows the results of the performance of **ImpAr** using different sets of relations to enrich original semantic annotations. All the figures presented in the table have been obtained applying the mapping between FrameNet and PropBank/NomBank generated by using a threshold value **T** equal to 4. According to Figure 7.14 the best **ImpAr** performance is given by a value of **T** within 4 and 7. In Table 7.21, **None** column contains the original results of **ImpAr** (see Chapter 6). **SameFrame** are the results using the 64,506 semantic relations we can derive between roles just using the membership relation to the same frame. **All** correspond to the output using all our 217,954 newly acquired semantic relations between roles. **Best** contains the setting that yields the best performance, ruling out all ***Inheritance*** and ***Using*** relations[3]. This setting contains a total of 160,710 relations.

[3]Possibly, not all these relations are counterproductive. However, filtering out only the

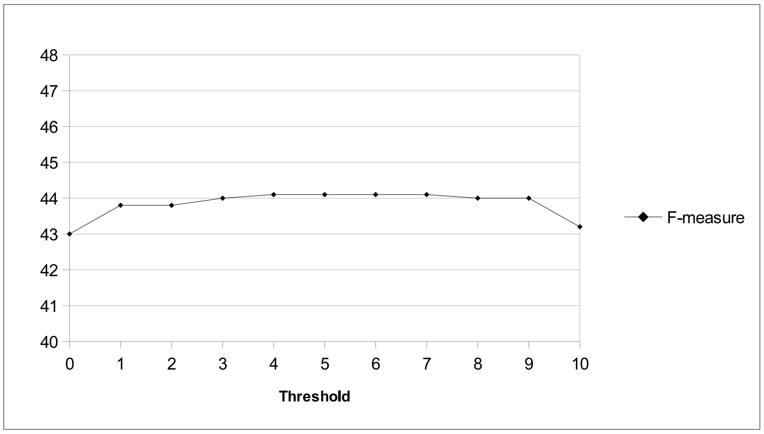

Figure 7.14: Results applying different threshold values for the development of the automatic mapping.

These results show that, on this task and framework, the FrameNet relations affects very little to these predicates. This occurs mainly because the FrameNet lexicon does not cover most of them. For instance, **investor**, **bid** or **fund** does not appear as lexical-units in FrameNet. So, no role relations have been obtained for them and **ImpAr** obtains the same results. Moreover, other lexical-units which appear in FrameNet does not correspond to the appropriate senses of PropBank/NomBank. For instance, this is the case of **investment** which in FrameNet corresponds to the sense of *Besieging*. As the test set does not cover this domain, the application of its role relations has no effect. For predicates having more general senses in FrameNet like **plan**, even the role relations obtained from the same frames are counter-productive. The partial coverage of FrameNet and its intrinsic misalignment with respect to PropBank/NomBank senses seems to be a major drawback. We hope that aligning these resources to WordNet (as the Predicate Matrix does) will provide in the future a more consistent and interoperable resources. However, the new role relations for **sale** and **price** predicates seem to have an important effect on the final results. Possibly due to a proper and complete representation in FrameNet and a correct aligment to PropBank/NomBank.

As expected, according to the **SameFrame** column, adding projected relations for those predicates belonging to the same frame seems to have also

incorrect ones would require to check whether the two related frames belongs to the same domain.

a little effect on the final results.

A very interesting conclusion can be extracted from the results given in **All** and **Best** columns. As said above, the difference within these settings is that the latter does not contain the ***Inheritance*** and ***Using*** relations. Taking into account that those are the relations that connect general frames with more specific ones, it can be inferred that for ISRL the most suitable relations are those that comprises predicates that belong to the same domains. In other words, the knowledge that belongs to the same scenario possibly provides more coherent semantic relations.

Table 7.22 shows the results of **ImpAr** when using no additional relations (**ImpAr**), when using just the ***Inheritance*** and ***Using*** relations (**IU**) and finally, when using the all semantic relations but the ones obtained from ***Inheritance*** and ***Using*** relations (**Best**).

	ImpAr	**IU**	**Best**
sale	40.6	42.0	**45.0**
price	50.3	49.6	**55.5**
cost	**44.5**	44.2	44.2
plan	**48.0**	41.7	45.0
loss	**55.2**	54.7	54.7
Overall	43.0	42.5	**44.5**

Table 7.22: Comparison of the Best results against using just ***Inheritance*** and ***Using*** relations.

Figure 7.15 shows two examples of implicit roles which are negatively affected by the new role relations obtained when applying the *Inheritance* and *Using* rules. In the first one, the arg_0 of plan inherits the filler of the arg_0 of **want.01** (*[Democrats]*) through the **Using** relations that connects both arguments while the correct filler for the arg_0 of **plan.01** is *[Senate leaders]*. In the second example, the **sale.01** predicate wrongly obtains the filler for the arg_1 from the arg_1 of **give.01** through the **Inheritance** relation.

On the other hand, frames that belong to the same scenario describe parts of the same kind of events and thus if the predicates from these frames appear in a text, they are likely to refer to the same scenes or situations. Consequently, the same participants are also shared. This seems to be the reason why relations such as ***SubFrame*** and ***Perspective_on***, like the examples in Figure 7.16, prove to ease the recovering of implicit arguments.

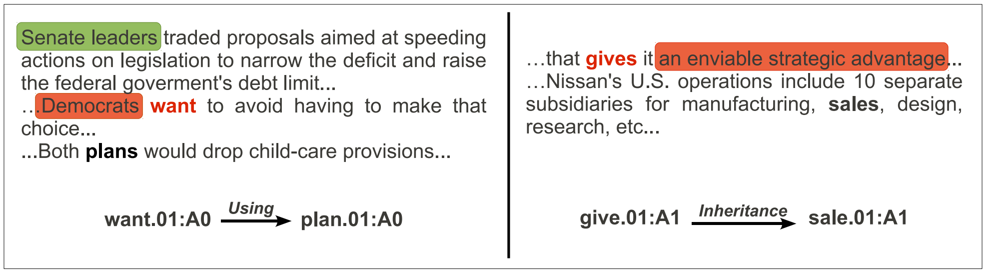

Figure 7.15: Examples of errors committed for ***Inheritance*** and ***Using*** relations. Wrongly identified fillers are highlighted in red. The proper fillers, when exists, are shown in green.

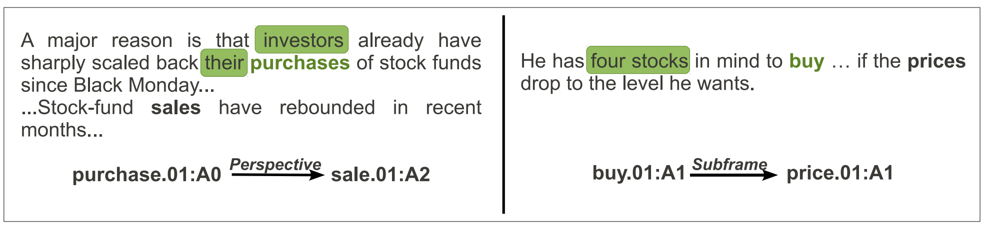

Figure 7.16: Examples of correct matches obtained through ***SubFrame*** and ***Perspective_on*** relations. Correctly identified fillers are highlighted in green.

7.4.1.1 Comparison against other resources

As said previously, there exist alternative resources that provides mappings between FrameNet and PropBank/NomBank. In Section 7.3 we discuss that the most prominent ones are SemLink and the Predicate Matrix, but we also indicate their potential drawback for ISRL. In this section we present a comparison between SemLink, the Predicate Matrix and our role mapping between FrameNet and PropBank/NomBank. We use the three resources to project the frame-element relations acquired from FrameNet (see Section 7.2) to PropBank/NomBank arguments. In all cases we use the same set of relations that obtains the best results. That is, ruling out ***Inheritance*** and ***Using*** relations.

Table 7.23 shows the results obtained when using **ImpAr** (**None**) and the best selection of frame-element relations projected to PropBank/NomBank using our new mappings (**Best**), using SemLink mappings (**SemLink**) and using the Predicate Matrix (**PM**). First, using either SemLink or the Predicate Matrix improves the results of the basic **ImpAr** configuration. However, none of them reaches our new set of projected relations. Interestingly, the Predicate Matrix also offers quite comparable results.

	None	Best	SemLink	PM
sale	40.6	**45.0**	41.7	43.7
price	50.3	**55.5**	50.3	**55.5**
cost	**44.5**	44.2	**44.5**	44.2
plan	**48.0**	45.0	**48.0**	45.0
loss	**55.2**	54.7	**55.2**	54.7
Overall	43.0	**44.5**	43.1	44.0

Table 7.23: Comparative results using the extended FrameNet relations projected through the mappings of SemLink and Predicate Matrix.

The evaluations presented previously show that FrameNet is still far from providing a full and proper coverage of relations between events and participants for ISRL. Aparently, none of the existing potential useful resources seems to contain sufficient knowledge. Now, we present an additional comparison with respect to alternative resources that also relate predicates and roles.

First, we have taken into account two related resources that contain entailment relations derived from FrameNet: LexPar (Coyne and Rambow, 2009) and FRED (Aharon et al., 2010). Both set of relations differ from ours in the type of information from FrameNet that is considered. For example, LexPar generates paraphrase transformations of frame-element patterns associated with verbs using the relations between the frames these verbs belong to. However, the entailment relations between verbs are limited to those cases when they are synonyms or hypernyms/hyponyms in WordNet, or when they are related via the ***Perspective_on*** relation. This makes that the vast majority of the rules are indeed based on WordNet. On the other hand, FRED only derives relations via ***Inheritance***, ***Cause*** and ***Perspective_on***, and limits the entailment according to morphologically derivation or between those lexical-units that are considered to be the *dominants* of the frames.

Table 7.24 shows the evaluation of these resources when we use them into the **ImpAr** algorithm. For these experiments we use the automatic mapping between FrameNet and PropBank/NomBank using $\mathbf{T} = 4$ (see Section 7.3). The results show that the conditions set by Coyne and Rambow (2009) and

Aharon et al. (2010) for the generation of their resources restricts too much the potential contribution of FrameNet relations in ISRL.

	None	**Best**	**LexPar**	**FRED**
P	46.0	**46.3**	45.6	45.8
R	40.3	**42.7**	41.2	41.3
F	43.0	**44.5**	43.3	43.4

Table 7.24: Comparative results with relations from LexPar and FRED.

We can also use the VerbNet classes to group verbs sharing arguments. VerbNet groups semantically related verbs in classes and subclasses. Each of these classes define a set of roles that are supposed to be shared by all the predicates that belong to the same class. For instance, the class called ***get-13.5.1*** (see Figure 7.17) includes verbs like **buy**, **catch**, **choose**, **find** or **get**. The class establishes that these predicates have the same set of semantic roles like *Agent*, *Theme* or *Source*. In other words, the *Agent* of **buy** is semantically related to the *Agent* of **catch**, the *Agent* of **choose** and so on. As SemLink includes very complete mapping between VerbNet and PropBank we can transfer these semantic relations to the arguments of PropBank and use them with **ImpAr** instead of the relations obtained from FrameNet.

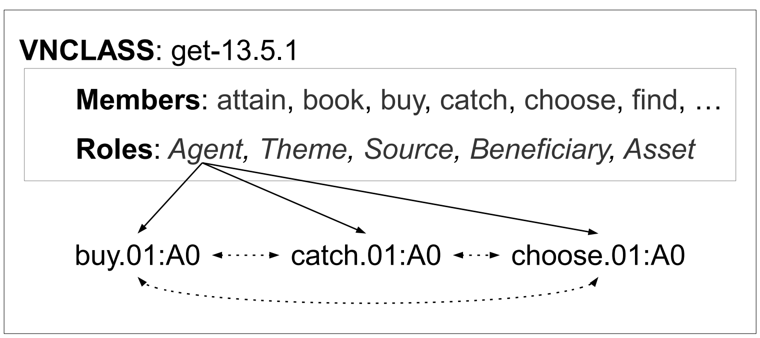

Figure 7.17: Examples of semantic relations obtained from VerbNet.

The results of this experiment are shown in Table 7.25. Interestingly, VerbNet seems to harm the performance of the algorithm. This behaviour is possibly caused by the same reason why the ***Inheritance*** and ***Using*** relations of FrameNet also damage the results. Although the verbs included in a class of VerbNet are indeed semantically related, they do not necessarily belong to the same domain or scenario. A VerbNet class can contain both general verbs, like **get**, and more specific ones, like **buy**.

	None	Best	VNC	VNSC
P	46.0	**46.3**	42.9	44.7
R	40.3	**42.7**	39.3	39.8
F	43.0	**44.5**	41.0	42.1

Table 7.25: Comparison with respect to VerbNet relations. VNC stands for relations between all the predicates of the same class. The relations of VNSC takes into account the subclasses of VerbNet.

Finally, we have performed a last comparison against the Narrative Schemas (NS) database (Chambers and Jurafsky, 2010). This resource contains automatically obtained sequences of events that commonly occur in text, where the *subjects* or the *objects* of the verbs are the same entity. The database contains sets of sequences of 6, 8, 10 and 12 verbs and the reliability of each sequence is weighted by a score value. Figure 7.18 shows an example of a common sequence where the same entity is the *subject* of 6 different verbs. As *subjects* and *objects* of the predicates correspond usually to the arg_0 and arg_1 respectively in PropBank the mapping between these two resources can be established quite straightforwardly. Thus, we define a relation between all the arguments belonging to the same schema.

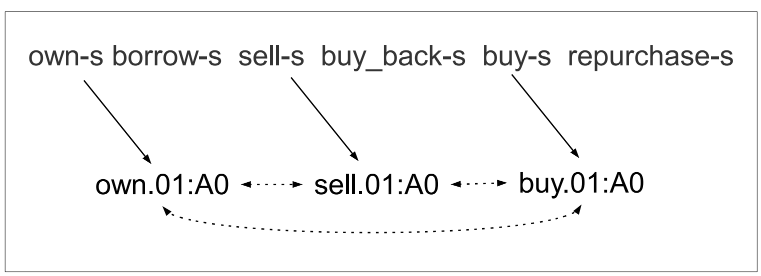

Figure 7.18: Examples of semantic relations obtained from the Narrative Schemas.

Table 7.26 shows the performance of the different sets of relations obtained from the narrative schemas with respect to the basic **ImpAr** (**None**) and the best configuration (**Best**). The code $nsXtY$ corresponds to the relations obtained from the schemas of size X that have at least a score value Y. Notoriously, the performance of the **ImpAr** using the relations obtained from the narrative schemas is worse than not applying any semantic relation at all. These poor results can be explained by the fact that the relations of the Narrative Schemas just encode a common temporal ordering, not implications between the related roles. The example in Figure 7.18 shows how the arg_0 of **buy.01** ends up being connected to the arg_0 of **sell.01**. In other words, the Narrative Schemas relates the *Buyer* with the *Seller*. Obviously, with respect to **ImpAr** this set of relations could be harmful.

	None	**Best**	**ns6t0**	**ns6t12**	**ns6t14**	**ns8t0**	**ns12t0**	**ns12t35**
P	46.0	**46.3**	24.1	28.4	45.0	26.6	32.7	33.2
R	40.3	**42.7**	24.7	28.4	40.2	26.9	32.7	33.1
F	43.0	**44.5**	24.4	28.4	42.4	26.7	32.7	33.1

Table 7.26: Comparison with relations from Narrative Schemas.

7.5 Concluding remarks

In this chapter we have proposed that relations that connect semantically events and roles can contribute positively to ISRL. For the approach described we have made use of the relations manually defined between the frames of FrameNet. Before applying this knowledge in a framework based on PropBank/NomBank we have had to perform two steps. First, we have proposed a set of logical rules to extend all the relations existing in FrameNet. Second, we have presented a method to map the predicates and roles contained in PropBank/NomBank and FrameNet. As a result of these tasks we have obtained a novel resource that connects arguments of different predicates of PropBank and NomBank through semantic relations. We have shown that including these relations into the **ImpAr** improves its performance.

The experiments presented in this chapter show that not all possible relations between events and participants are suitable for ISRL. According to our results, it seems that relations connecting frames that do not belong to the same domains should be avoided. For example, although the predicates **want.01** and **plan.01** or the predicates **get.01** and **buy.01** can be semantically related they do not evoke necessarily the same events. Relating the roles of such kind of events has proved to be harmful. On the other, properly connecting roles of predicates that describe a particular scenario, like **buy.01**, **sell.01** or **pay.01** seems to be helpful. However, acquiring this kind of knowledge is a complex task because manually created sources do not contain enough coverage and automatic approaches do not always generate the kind of relations required. Obtaining such kind of information will be a main focus in our future work.

CONCLUSION AND FURTHER WORK

CHAPTER 8

Conclusion and further work

This last chapter presents a summary (Section 8.1) that reviews the goals we have reached during our research on Implicit Semantic Role Labelling. In Section 8.2 we list the research papers we have published that are related with this work. Finally, Section 8.3 proposes some possible future lines of research.

8.1 Summary

Implicit Semantic Role Labelling is the task of recovering semantic roles beyond the syntactically close context of the predicates. Thus, this task aims to extend the scope of traditional Semantic Role Labelling systems in order to complete the explicit role structures trying to advance one step forward the current state of the art on natural language processing. The relevance of this type of analysis can be seen in the example shown in Figure 8.1. In this case, a traditional Semantic Role Labelling (SRL) system only manages to annotate the roles of the predicate **sell.01** and misses all the participants of the events targeted by the predicates **price.01** and **sale.01**. Indeed, Gerber and Chai (2010) pointed out that solving implicit arguments the coverage of role structures can increase by **71%**.

Armstrong World Industries Inc. agreed in principle to **sell** its carpet operations to Shaw Industries Inc. .

The **price** wasn't disclosed but one analyst estimated that it was $150 million.

Moreover, such a **sale** could help Armstrong reassure its investors.

Figure 8.1: Traditional SRL annotation with missing arguments.

In this research, we have described some approaches that combine techniques of SRL with different types of Coreference Resolution (CR) proving that **Implicit Semantic Roles Labelling** depends on complex relations between the elements of the discourse. We have also tried different strategies to overcome the need of manual annotated data.

Our first approach followed similar previous works adapting a set of traditional anaphora resolution models for the implicit argument resolution task. Our evaluation shows that a model trained with these features can improve state of the art results. The sources of evidence proposed are adaptations that focus on nominal entities and pronouns, most of them originally focusing on looking for referents in the same sentence. For that reason, it seems that they can provide useful information for cases like the example in Figure 8.2, where the filler of the missing arg_2 of **price.01** is recoverable within sentence boundaries. In order to avoid problems derived from the small training set available the set of features is as lexically independent as possible.

Our second approach proposed that elided arguments can be recovered from other mentions of the same events. We have exploited a relevant discursive property that states that several mentions of the same predicate tend to refer to the same event instance maintaining the same role fillers. Figure 8.3 shows an example where some missing arguments of the predicate **sale.01** can be obtained from a previous mention of the same event. In other words, we introduce the first attempt to label implicit roles by solving event coreference. We have developed a new deterministic algorithm, **ImpAr**, that obtains very competitive and robust performances with respect to supervised approaches. Furthermore, it can be applied where there is no available

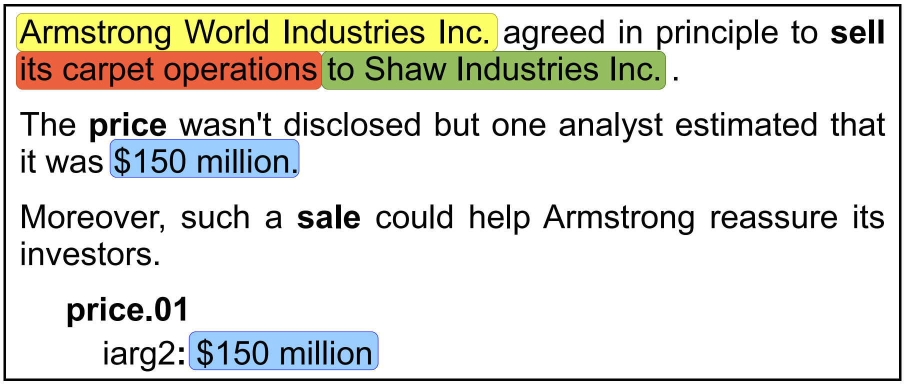

Figure 8.2: ISRL as anaphora resolution.

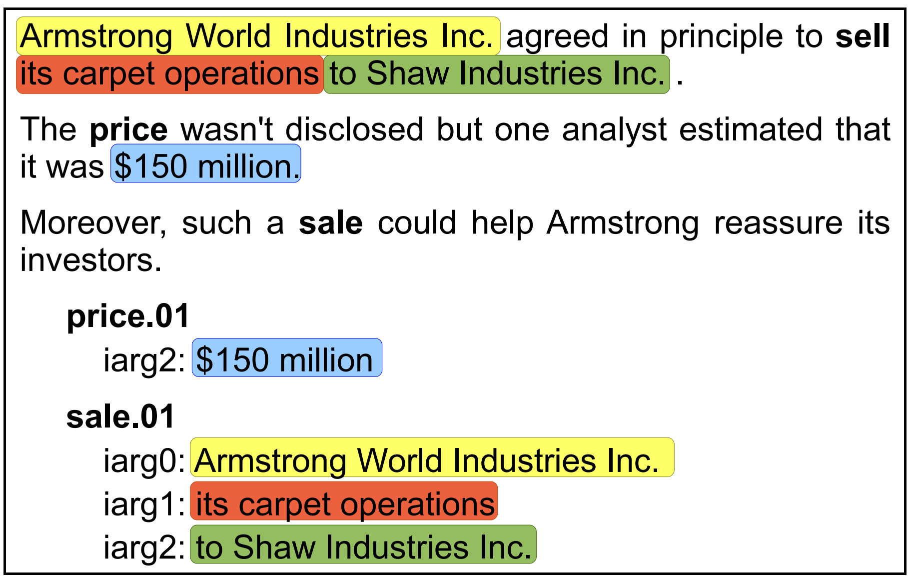

Figure 8.3: ISRL via event coreference.

manual annotations for training. The algorithm is publicly available[1].

The previous approach has a large margin for improvement. For that reason, we have tried to include other types of semantic relations between the predicates and roles. We have seen that relations that describe implication,

[1] `http://adimen.si.ehu.es/web/ImpAr`

causation or precedence can positively affect ISRL because these relations possibly connect predicates involving the same scenario. We have taken advantage of the semantic relations described in FrameNet to extend the behaviour of the **ImpAr** algorithm and we have proved that we can extend the number of cases solved. For example, the arg_3 of the predicates **sell.01** and **sale.01**, and the arg_0 and arg_1 of the predicate **price.01** in Figure 8.4 can be solved if we know that there exists an entailment relation between these predicates. That is, for every **sale** event there is always a **price**.

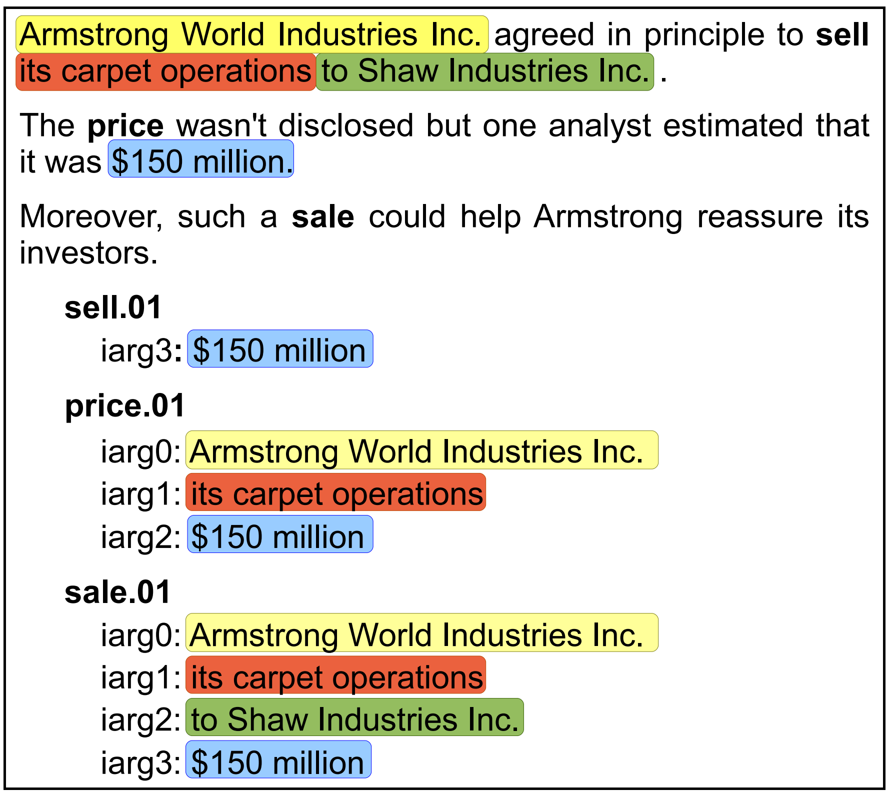

Figure 8.4: ISRL including semantic relations between events.

In summary, we have studied **Implicit Semantic Role Labelling** with respect to three different types of coreferent relations in the discourse, solving implicit roles as anaphora resolution, taking into account event coreference and finally, including some entailment relations between roles and predicates. We have also developed novel methods that overcome the lack of training data: lexically independent features, deterministic algorithms that do not require any training data and wide coverage knowledge bases to obtain rules between the events. Our empirical evaluations have proved that our approaches obtain similar or better performances than supervised systems. Finally, as a result of this research, we have developed a set of robust and open domain tools and resources that are freely available.

8.2 Publications

Below, we present cronologically the list of publications related with the research described in this document:

- Laparra E. and Rigau G. *Exploiting Explicit Annotations and Semantic Types for Implicit Argument Resolution*. 6th IEEE International Conference on Semantic Computing (ICSC'12). Palermo, Italy. 2012.

  The contributions of the previous publication are described in Chapter 4.

- Laparra E. and Rigau G. *Sources of Evidence for Implicit Argument Resolution*. 10th International Conference on Computational Semantics (IWCS'13). Postdam, Germany. 2013.

  This study and its corresponding experiments are presented in Chapter 5.

- Laparra E. and Rigau G. *ImpAr: A Deterministic Algorithm for Implicit Semantic Role Labelling*. The 51st Annual Meeting of the Association for Computational Linguistics (ACL'2013). Sofia, Bulgaria. 2013.

  The algorithm and the evaluation described in this paper are included in Chapter 6.

- Laparra E., López de Lacalle M., Aldabe I. and Rigau G. *Predicate Matrix: Automatically extending the interoperability between predicative resources*. Language Resources and Evaluation. 2014. *(Submitted)*

  This publication contains part of the contributions presented in Chapter 7.

The following references are not covered but are very closely related to this thesis:

- Laparra E. and Rigau G. *Integrating WordNet and FrameNet using a knowledge-based Word Sense Disambiguation algorithm*. Proceedings of Recent Advances in Natural Language Processing (RANLP'09). Borovets, Bulgaria, September, 2009.

- Laparra E., Rigau G. and Cuadros M. *Exploring the integration of WordNet and FrameNet*. Proceedings of the 5th Global WordNet Conference (GWC'10), Mumbai, India. January, 2010.

- Laparra E. and Rigau G. *eXtended WordFrameNet*. 7th international conference on Language Resources and Evaluation (LREC'10). La Valetta, Malta. 2010.

- López de Lacalle M., Laparra E. and Rigau G. *First steps towards a Predicate Matrix*. Proceedings of the 7th Global WordNet Conference (GWC'14). Tartu, Estonia. 2014.

- López de Lacalle M., Laparra E. and Rigau G. *Predicate Matrix: extending SemLink through WordNet mappings*. Proceedings of the 9th Language Resources and Evaluation Conference (LREC'14). Reykjavik, Iceland. 2014.

We also present references to other works produced during the development of the present research:

- Álvez J., Atserias J., Carrera J., Climent S., Laparra E., Oliver A. and Rigau G. *Complete and Consistent Annotation of WordNet using the Top Concept Ontology*. 6th international conference on Language Resources and Evaluation (LREC'08), Marrakesh, Morroco. 2008.

- Alecha M., Álvez J., Hermo M. and Laparra E. *A New Proposal for Using First-Order Theorem Provers to Reason with OWL DL Ontologies*. Proceedings Spanish Conference on Programming and Computer Languages (PROLE'09). San Sebastián, Spain. 2009

- Alonso L., Castellón I, Laparra E., Rigau G. *Evaluación de métodos semi-automáticos para la conexión entre FrameNet y SenSem*. Proceedings of the 25th Annual Meeting of Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN'09). San Sebastián, Spain. 2009.

- Agirre E., Casillas A., Díaz de Ilarraza A., Estarrona A., Fernandez K., Gojenola K., Laparra E., Rigau G., Soroa A. *The KYOTO Project*. Proceedings of the 25th Annual Meeting of Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN'09). San Sebastián, Spain. 2009.

- Vossen P., Bosma W., Cuadros M., Laparra E. and Rigau G. *Integrating a large domain ontology of species into WordNet*. 7th international conference on Language Resources and Evaluation (LREC'10). La Valetta, Malta. 2010.

- Gonzalez-Agirre A., Laparra E. and Rigau G. *Multilingual Central Repository version 3.0: upgrading a very large lexical knowledge base*. Proceedings of the 6th Global WordNet Conference (GWC'12), Matsue, Japan. January, 2012.

- Gonzalez-Agirre A., Laparra E. and Rigau G. *Multilingual Central Repository version 3.0*. 8th international conference on Language Resources and Evaluation (LREC'12). Istambul, Turkey. 2012.

- Laparra E., Rigau G. and Vossen P. *Mapping WordNet to the Kyoto ontology*. 8th international conference on Language Resources and Evaluation (LREC'12). Istambul, Turkey. 2012.

- Agerri R., Agirre E., Aldabe I., Altuna B., Beloki Z., Laparra E., López de Lacalle M., Rigau G., Soroa A., Urizar R. *The NewsReader project*. Proceedings of the 30th Annual Meeting of Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN'14). Girona, Spain. Procesamiento del Lenguaje Natural. Vol. 53 pp. pp 215-218. ISSN: 1135-5948. 2014.

8.3 Future work

Although there are two different datasets available for evaluating **Implicit Semantic Role Labelling** they include very different kinds of additional annotations. For example, the one based on FrameNet (Ruppenhofer et al., 2009) contains gold-standard coreference chains which are not part of the dataset based on PropBank/NomBank (Gerber and Chai, 2010, 2012). Furthermore, while the syntactic analysis of the first is based on constituents, the latter provides syntactic dependencies. For each experiment presented throughout this work we have focused just in the one that fits better for each approach. However, it would be desirable to complete both datasets with the same sets of annotations for a more complete evaluation framework. This non-trivial task is one of our next steps in order to evaluate not only the strategies described in this document but also the future extensions of this research.

First of all, many of the techniques used in this research for **Implicit Semantic Role Labelling** are quite basic and naive. Our goal has been to prove the suitability of the ideas behind them and, although we have empirically proved that our approaches obtain successful results, there are still large avenues for improvement. One of the places that we plan to investigate in a near future is on the strategy we apply for coreference resolution. On the one hand, the anaphora resolution approach of **ImpAr** uses a set of very general semantic classes to filter out the candidates to be fillers of the implicit roles. However, sometimes our classification can be too general. For example, both the arg_1 of the predicate **drink.01** and the arg_1 of the predicate **write.01** are classified as *TANGIBLE*. Obviously, we can apply more fine grained *selectional preferences* for a better filtering process. On the other hand, at this point, **ImpAr** just checks the lemmas of the predicates to determine if their mentions refer to the same event. We are currently studying to include a nominal entity coreference analysis to refine this process because it can help to know if the explicit roles of different predicates are actually the same. If the explicit roles are not the same, the predicates are not coreferent. Using this approach, we expect to improve the current version of the algorithm.

We are also working on the automatic acquisition of entailment relations between predicates and semantic roles. We studied an approach that consists on acquiring new relations from existing knowledge sources. The main advantage of this approach is that the resources have been manually built and,

invest.01
arg_1: thing invested
arg_2: invested in
spend.01
arg_1: thing bought, commodity
arg_3: price paid, money

Table 8.1: Descriptions of some arguments in PropBank

in consequence, the information is very reliable. However, in many cases it is very difficult to exploit this type of resources. Consider the role descriptions from PropBank in Table 8.1. The descriptions of these arguments can provide some evidence to establish an entailment relation among them. However, as they are written in natural language it turns to be quite hard to process them automatically.

Another possibility is to obtain the entailment relations from corpus, following works like Chambers and Jurafsky (2009); Cheung et al. (2013). We are currently working on the development of unsupervised probabilistic models designed for obtaining scenarios similar to those included in FrameNet. This scenarios will group semantically related roles of different predicates belonging to the same cluster. Following with the example in Table 8.1, the algorithm could generate an scenario including the predicates **invest.01** and **spend.01** and two roles, one merging the arg_1 of **invest.01** and the arg_3 of **spend.01**, and another one merging the arg_2 of **invest.01** and the arg_1 of **spend.01**.

Although we have focused our research on English only, a great advantage of the strategies we have developed is that they are easily extended to other languages. For example, we only need few steps to adapt **ImpAr** to process Spanish texts. AnCora provides a knowledge base describing the role structures in a very similar way to PropBank. Thus, the only actual requirement is to learn the *selectional preferences* for the arguments contained in this resource. Furthermore, AnCora includes manual annotations for implicit roles in Spanish (Taulé et al., 2012) allowing the evaluation of the performance of **ImpAr** in this language. In fact, potentially, our algorithm could be applied to any language as long as there is available a traditional SRL system and a semantic resource with predicates and roles for it.

The bomb exploded in a crowded marketplace. Five civilians were killed, including two children. Al Qaeda claimed responsibility.

Table 8.2: A very complex case of implicit roles.

The judge did not hesitate to sentence the murderer. Police had found at his home a written confession.

Table 8.3: ISRL could help to find the referent of *his*.

Finally, we have seen that **Implicit Semantic Role Labelling** depends on divers relations between the elements of the discourse. In this thesis we have studied just some of them, but the manner in how the information is realized in a text can adopt very complex forms. For example, in the sentences in Table 8.2 there are many implicit roles, like *who exploded the bomb*, *what killed the civilians* or *what did Al Qaeda claimed responsibility for*. Notice that solving these implicit roles may require a *global* re-interpretation process each time a new explicit or implicit information emerges.

Moreover, implicit roles can be a key piece of information for the resolution of other types of coreferent elements. The example in Table 8.3 shows a very difficult case for anaphora resolution. However, knowing that the pronoun *his* is the filler of the implicit *writer* of the *confession* can give evidence to infer that it refers to the *murderer* and not to the *judge*. These examples evince the need to take into account **Implicit Semantic Role Labelling** for a full automatic understanding of natural language.

Bibliography

Agirre, E. and Edmonds, P. (2007). *Word Sense Disambiguation: Algorithms and Applications*. Springer Publishing Company, Incorporated, 1st edition.

Agirre, E. and Soroa, A. (2009). Personalizing pagerank for word sense disambiguation. In *Proceedings of the 12th conference of the European chapter of the Association for Computational Linguistics (EACL-2009)*, Athens, Greece. Eurpean Association for Computational Linguistics.

Aharon, R. B., Szpektor, I., and Dagan, I. (2010). Generating entailment rules from framenet. In *Proceedings of the ACL 2010 Conference Short Papers*, ACLShort '10, pages 241–246, Uppsala, Sweden.

Ahn, D. (2006). The stages of event extraction. In *Proceedings of the Workshop on Annotating and Reasoning about Time and Events*, pages 1–8, Sydney, Australia.

Alshawi, H. (1987). *Memory and Context for Language Interpretation*. Cambridge University Press.

Álvez, J., Atserias, J., Carrera, J., Climent, S., Laparra, E., Oliver, A., and Rigau, G. (2008). Complete and consistent annotation of wordnet using the top concept ontology. In *LREC*.

Ando, R. K. and Zhang, T. (2005). A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, pages 1817–1853.

Aone, C. and Bennett, S. W. (1994). Discourse tagging tool and discourse-tagged multilingual corpora. In *In Proceedings of the International workshop on sharable natural language resources*.

Bagga, A. and Baldwin, B. (1998). Algorithms for scoring coreference chains. In *In The First International Conference on Language Resources and Evaluation Workshop on Linguistics Coreference*, pages 563–566.

Bagga, A. and Baldwin, B. (1999). Cross-document event coreference: Annotations, experiments, and observations. In *In Proc. ACL-99 Workshop on Coreference and Its Applications*, pages 1–8.

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The berkeley framenet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, ACL '98, pages 86–90, Montreal, Quebec, Canada.

Bejan, C. A. and Harabagiu, S. (2010). Unsupervised event coreference resolution with rich linguistic features. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL'10, pages 1412–1422, Uppsala, Sweden.

Bejan, C. A., Titsworth, M., Hickl, A., and Harabagiu, S. M. (2009). Non-parametric bayesian models for unsupervised event coreference resolution. In *NIPS*, pages 73–81.

Bell, E. (1934). Exponential numbers. *American Mathematical Monthly*, pages 411–419.

Bengtson, E. and Roth, D. (2008). Understanding the value of features for coreference resolution. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, Honolulu, Hawaii.

Björkelund, A., Hafdell, L., and Nugues, P. (2009). Multilingual semantic role labeling. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task*, CoNLL '09, pages 43–48, Boulder, Colorado, USA.

Blanco, E. and Moldovan, D. (2014). Leveraging verb-argument structures to infer semantic relations. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 145–154, Gothenburg, Sweden.

Brennan, S. E., Friedman, M. W., and Pollard, C. J. (1987). A centering approach to pronouns. In *Meeting of the Association for Computational Linguistics*.

Burger, J. D. and Connolly, D. (1992). Probabilistic resolution of anaphoric reference. In *AAAI Fall Symposium on Probabilistic Approaches to Natural Language*, pages 17–24, Cambridge, Massachusetts, USA.

Byron, D. K. and Stent, A. (1998). A Preliminary Model of Centering in Dialog. In *Meeting of the Association for Computational Linguistics*, pages 1475–1477.

Carreras, X. and Màrquez, L. (2004). Introduction to the CoNLL-2004 Shared Task: Semantic Role Labeling. In *Proceedings of the Eigth Conference on Computational Natural Language Learning (CoNLL-2004)*, pages 89–97, Boston, MA, USA.

Carreras, X. and Màrquez, L. (2005). Introduction to the conll-2005 shared task: Semantic role labeling. In *Proceedings of the 9th Conference on Computational Natural Language Learning*, CoNLL '05, pages 152–164, Ann Arbor, Michigan, USA.

Carter, D. (1987). *Interpreting Anaphors in Natural Language Texts*. Ellis Horwood Ltd.

Chambers, N. and Jurafsky, D. (2009). Unsupervised learning of narrative schemas and their participants. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2*, ACL '09, pages 602–610, Suntec, Singapore.

Chambers, N. and Jurafsky, D. (2010). A database of narrative schemas. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta.

Charniak, E. (1972). Toward a model of children's story comprehension. Technical report, Massachusetts Institute of Technology, Cambridge, MA, USA.

Chen, B., Su, J., Pan, S. J., and Tan, C. L. (2011). A unified event coreference resolution by integrating multiple resolvers. In *In Proceedings of 5th*

International Joint Conference on Natural Language Processing, Chiang Mai.

Chen, D., Schneider, N., Das, D., and Smith, N. A. (2010). Semafor: Frame argument resolution with log-linear models. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, SemEval '10, pages 264–267, Los Angeles, California, USA.

Chen, Z. and Ji, H. (2009). Graph-based event coreference resolution. In *Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing*, TextGraphs-4, pages 54–57, Suntec, Singapore.

Cheung, J. C. K., Poon, H., and Vanderwende, L. (2013). Probabilistic frame induction. In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 9-14, 2013, Westin Peachtree Plaza Hotel, Atlanta, Georgia, USA.*

Chinchor, N. A. (1998). Overview of muc-7/met-2. In *Proceedings of the Seventh Message Understanding Conference (MUC-7).*

Ciaramita, M. and Altun, Y. (2006). Broad-coverage sense disambiguation and information extraction with a supersense sequence tagger. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, EMNLP '06, pages 594–602, Sydney, Australia.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, pages 2493–2537.

Coyne, B. and Rambow, O. (2009). Lexpar: A freely available english paraphrase lexicon automatically extracted from net. In *Proceedings of the 3rd IEEE International Conference on Semantic Computing (ICSC 2009), 14-16 September 2009, Berkeley, CA, USA*, pages 53–58.

Culotta, A., Wick, M., and Mccallum, A. (2007). First-Order Probabilistic Models for Coreference Resolution. In *Proceedings of NAACL HLT 2007*, Rochester, NY.

Cybulska, A. and Vossen, P. (2013). Semantic relations between events and their time, locations and participants for event coreference resolution. In

Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013, Hissar, Bulgaria.

Dahl, D. A. (1992). Pundit - natural language interfaces. In *Logic Programming in Action*, volume 636 of *Lecture Notes in Computer Science*, pages 176–185. Springer Berlin Heidelberg.

Dahl, D. A., Palmer, M. S., and Passonneau, R. J. (1987). Nominalizations in pundit. In *In Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, ACL '87, pages 131–139, Stanford, California, USA.

Das, D., Schneider, N., Chen, D., and Smith, N. A. (2010). Probabilistic frame-semantic parsing. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '10, Los Angeles, California.

Daumé, III, H. and Marcu, D. (2005). A large-scale exploration of effective global features for a joint entity detection and tracking model. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, Vancouver, British Columbia, Canada.

Davidson, D. (1969). The individuation of events. In *Essays in Honor of Carl G. Hempel*, volume 24 of *Synthese Library*, pages 216–234. Springer Netherlands.

Davidson, D. (1985). The individuation of events. In *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, pages 172–176. Oxford: Blackwell.

Díaz de Ilarraza Sanchez, A., Rodríguez Hontoria, H., and Verdejo Maíllo, F. (1990). A mechanism for ellipsis resolution in dialogued systems. In *Proceedings of the 13th Conference on Computational Linguistics - Volume 3*, COLING '90, Helsinki, Finland.

Doddington, G. R., Mitchell, A., Przybocki, M. A., Ramshaw, L. A., Strassel, S., and Weischedel, R. M. (2004). The automatic content extraction (ace) program - tasks, data, and evaluation. In *LREC*. European Language Resources Association.

Erk, K. and Pado, S. (2006). Shalmaneser - a flexible toolbox for semantic role assignment. In *Proceedings of LREC 2006*, Genoa, Italy.

Feizabadi, S. P. and Padó, S. (2014). Crowdsourcing annotation of non-local semantic roles. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, volume 2: Short Papers*, pages 226–230, Gothenburg, Sweden.

Fellbaum, C. (1998). *WordNet: an electronic lexical database*. MIT Press.

Ferrández, A. and Peral, J. (2000). A computational approach to zero-pronouns in spanish. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, pages 166–172, Hong Kong.

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 and Speech*, volume 280, pages 20–32, New York.

Finkel, J. R. and Manning, C. D. (2008). Enforcing transitivity in coreference resolution. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, HLT-Short '08, Columbus, Ohio.

Fisher, D., Soderland, S., Feng, F., and Lehnert, W. (1995). Description of the umass system as used for muc-6. In *Proceedings of the 6th Conference on Message Understanding*, MUC6 '95, pages 127–140, Columbia, Maryland.

Gale, W. A., Church, K. W., and Yarowsky, D. (1992). Estimating upper and lower bounds on the performance of word sense disambiguation. In *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*. ACL.

Ge, N., Hale, J., and Charniak, E. (1998). A statistical approach to anaphora resolution. In *Proceedings of the Sixth Workshop on Very Large Corpora*, pages 161–171.

Gerber, M. and Chai, J. (2012). Semantic role labeling of implicit arguments for nominal predicates. *Computational Linguistics*, 38(4):755–798.

Gerber, M. and Chai, J. Y. (2010). Beyond nombank: a study of implicit arguments for nominal predicates. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, pages 1583–1592, Uppsala, Sweden.

Gildea, D. and Jurafsky, D. (2000). Automatic labeling of semantic roles. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, ACL '00, pages 512–520, Hong Kong.

Gildea, D. and Jurafsky, D. (2002). Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288.

Glavaš, G. and Šnajder, J. (2014). Construction and evaluation of event graphs. *Natural Language Engineering*, FirstView.

Gorinski, P., Ruppenhofer, J., and Sporleder, C. (2013). Towards weakly supervised resolution of null instantiations. In *Proceedings of the 10th International Conference on Computational Semantics*, IWCS '13, pages 119–130, Potsdam, Germany.

Grenager, T. and Manning, C. D. (2006). Unsupervised discovery of a statistical verb lexicon. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, EMNLP '06, pages 1–8, Sydney, Australia.

Grishman, R. and Sundheim, B. (1995). Design of the MUC-6 evaluation. In *Proceedings of the Sixth Message Understanding Conference (MUC-6)*, pages 1–11, San Francisco.

Grosz, B. J., Joshi, A. K., and Weinstein, S. (1995). Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics*, 21:203–225.

Grosz, B. J. and Sidner, C. L. (1986). Attention, Intentions, and the Structure of Discourse. *Computational Linguistics*, 12:175–204.

Hacioglu, K., Pradhan, S., Ward, W., Martin, J. H., and Jurafsky, D. (2004). Semantic role labeling by tagging syntactic chunks. In *In Proceedings of CoNLL 2004 Shared Task*, pages 110–113, Ann Arbor, Michigan.

Haghighi, A. and Klein, D. (2007). Unsupervised Coreference Resolution in a Nonparametric Bayesian Model. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, ACL'07, pages 848–855, Prague, Czech Republic.

Haghighi, A. and Klein, D. (2010). Coreference resolution in a modular, entity-centered model. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '10, Los Angeles, California.

Hajič, J., Ciaramita, M., Johansson, R., Kawahara, D., Martí, M. A., Màrquez, L., Meyers, A., Nivre, J., Padó, S., Štěpánek, J., Straňák, P., Surdeanu, M., Xue, N., and Zhang, Y. (2009). The CoNLL-2009 shared task: Syntactic and semantic dependencies in multiple languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task*, CoNLL '09, pages 1–18, Boulder, Colorado, USA.

Harabagiu, S. and Moldovan, D. (1998). Knowledge processing on an extended wordnet. *WordNet - An Electronic Lexical Database*, pages 379–405.

Hirschman, L., Palmer, M., Dowding, J., Dahl, D., Linebarger, M., Passonneau, R., Land, F.-M., Ball, C., and Weir, C. (1989). The pundit natural-language processing systemt. In *Proceedings of the Annual AI Systems in Government Conference*, pages 234–243, Washington, DC.

Hirst, G. (1987). *Semantic Interpretation and the Resolution of Ambiguity.* Cambridge University Press.

Hobbs, J. R. (1977). Pronoun resolution. *Intelligence/sigart Bulletin*, pages 28–28.

Hobbs, J. R., Stickel, M. E., Appelt, D. E., and Martin, P. (1993). Interpretation as abduction. *Artificial Intelligence*, 63(1-2):69–142.

Hull, R. D. and Gomez, F. (1996). Semantic interpretation of nominalizations. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence - Volume 2*, AAAI'96, pages 1062–1068, Portland, Oregon.

Humphreys, K., Gaizauskas, R., and Azzam, S. (1997). Event coreference for information extraction. In *Proceedings of the Workshop on Operational*

Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts, 35th Meeting of ACL, pages 75–81.

Iida, R., Inui, K., and Matsumoto, Y. (2007). Zero-anaphora resolution by learning rich syntactic pattern features. *ACM Transactions on Asian Language Information Processing*, pages 1:1–1:22.

Iida, R. and Poesio, M. (2011). A cross-lingual ilp solution to zero anaphora resolution. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, pages 804–813, Portland, Oregon.

Izquierdo, R., Suárez, A., and Rigau, G. (2009). An empirical study on class-based word sense disambiguation. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '09, Athens, Greece.

Jiang, Z. P. and Ng, H. T. (2006). Semantic role labeling of nombank: A maximum entropy approach. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, EMNLP '06, pages 138–145, Sydney, Australia.

Johansson, R. and Nugues, P. (2008). The effect of syntactic representation on semantic role labeling. In *Proceedings of the 22Nd International Conference on Computational Linguistics - Volume 1*, COLING '08, pages 393–400, Manchester, United Kingdom.

Jurafsky, D. and Martin, J. H. (2009). *Speech and Language Processing (2nd Edition) (Prentice Hall Series in Artificial Intelligence)*. Prentice Hall.

Kamp, H. (1981). A theory of truth and semantic representation. In *Formal Semantics - the Essential Readings*, pages 189–222. Blackwell.

Kehler, A. (1997). Probabilistic coreference in information extraction. In *In Proceedings of the Second Conference on Empirical Methods in Natural Language Processing*, pages 163–173.

Kennedy, C. and Boguraev, B. (1996). Anaphora for everyone: Pronominal anaphora resoluation without a parser. In *Proceedings of the 16th Conference on Computational Linguistics - Volume 1*, COLING '96, pages 113–118, Copenhagen, Denmark.

Kipper, K. (2005). *VerbNet: A broad-coverage, comprehensive verb lexicon.* PhD thesis, University of Pennsylvania.

Kipper, K., Dang, H. T., and Palmer, M. (2000). Class-based construction of a verb lexicon. In *Seventeenth National Conference on Artificial Intelligence*, AAAI-2000, Austin.

Koomen, P., Punyakanok, V., Roth, D., and Yih, W.-t. (2005). Generalized inference with multiple semantic role labeling systems. In *Proceedings of the Ninth Conference on Computational Natural Language Learning*, CONLL'05, pages 181–184, Ann Arbor, Michigan.

Lang, J. and Lapata, M. (2010). Unsupervised induction of semantic roles. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '10, pages 939–947, Los Angeles, California.

Lang, J. and Lapata, M. (2011a). Unsupervised semantic role induction via split-merge clustering. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, pages 1117–1126, Portland, Oregon.

Lang, J. and Lapata, M. (2011b). Unsupervised semantic role induction with graph partitioning. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 1320–1331, Edinburgh, United Kingdom.

Laparra, E. and Rigau, G. (2009). Integrating wordnet and framenet using a knowledge-based word sense disambiguation algorithm. In *Proceedings of RANLP*, Borovets, Bulgaria.

Laparra, E. and Rigau, G. (2010). extended wordframenet. In *LREC*.

Laparra, E., Rigau, G., and Cuadros, M. (2010). Exploring the integration of wordnet and framenet. In *Proceedings of the 5th Global WordNet Conference (GWC 2010), Mumbai, India.*

Lapata, M. (2000). The automatic interpretation of nominalizations. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, pages 716–721. AAAI Press.

Lappin, S. and Leass, H. J. (1994). An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20(4):535–561.

Lee, H., Chang, A., Peirsman, Y., Chambers, N., Surdeanu, M., and Jurafsky, D. (2013). Deterministic coreference resolution based on entity-centric, precision-ranked rules. *Computational Linguistics*, pages 885–916.

Lee, H., Peirsman, Y., Chang, A., Chambers, N., Surdeanu, M., and Jurafsky, D. (2011). Stanford's multi-pass sieve coreference resolution system at the conll-2011 shared task. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, CONLL Shared Task '11, Portland, Oregon.

Lee, H., Recasens, M., Chang, A. X., Surdeanu, M., and Jurafsky, D. (2012). Joint entity and event coreference resolution across documents. In *EMNLP-CoNLL*, pages 489–500, Jeju Island, Korea.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago University Press.

Lin, Z., Ng, H. T., and Kan, M.-Y. (2014). A pdtb-styled end-to-end discourse parser. *Natural Language Engineering*, 20:151–184.

Litkowski, K. (2004). Senseval-3 task: Automatic labeling of semantic roles. In *Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, pages 9–12, Barcelona, Spain.

Liu, C. and Ng, H. T. (2007). Semantic role labeling of nombank: A maximum entropy approach. In *In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, ACL '07, pages 208–215, Prague, Czech Republic.

Lockman, A. and Klappholz, A. D. (1980). Toward a procedural model of contextual reference resolution. *Discourse Processes 3*.

López de Lacalle, M., Laparra, E., and Rigau, G. (2014a). First steps towards a predicate matrix. In *Proceedings of the 7th Global WordNet Conference (GWC2014)*, Tartu, Estonia.

López de Lacalle, M., Laparra, E., and Rigau, G. (2014b). Predicate matrix: extending semlink through wordnet mappings. In *Proceedings of* *the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland.

Luo, X. (2005). On coreference resolution performance metrics. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, pages 25–32, Vancouver, British Columbia, Canada.

Luo, X., Ittycheriah, A., Jing, H., Kambhatla, N., and Roukos, S. (2004). A mention-synchronous coreference resolution algorithm based on the bell tree. In *Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics*, ACL '04, Barcelona, Spain.

Mani, I., Verhagen, M., Wellner, B., Lee, C. M., and Pustejovsky, J. (2006). Machine learning of temporal relations. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, ACL-44, Sydney, Australia.

Mann, W. C. and Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.

Marcus, M., Kim, G., Marcinkiewicz, M. A., Macintyre, R., Bies, A., Ferguson, M., Katz, K., and Schasberger, B. (1994). The penn treebank: Annotating predicate argument structure. In *In ARPA Human Language Technology Workshop*, pages 114–119.

Marcus, M. P., Marcin-kiewicz, M. A., and Santorini, B. (1993). Building a large annotated corpus of English: the penn treebank. *Computational Linguistics*, 19:313–330.

Mccarthy, J. and Lehnert, W. G. (1995). Using decision trees for coreference resolution. In *In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 1050–1055.

McDonald, R., Crammer, K., and Pereira, F. (2005). Online large-margin training of dependency parsers. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 91–98, Ann Arbor, Michigan.

Meyers, A., Macleod, C., Yangarber, R., Grishman, R., Barrett, L., and Reeves, R. (1998). Using nomlex to produce nominalization patterns for information extraction. In *Proceedings: the Computational Treatment of Nominals*.

Meyers, A., Reeves, R., Macleod, C., Szekely, R., Zielinska, V., Young, B., and Grishman, R. (2004). The nombank project: An interim report. In *In Proceedings of the NAACL/HLT Workshop on Frontiers in Corpus Annotation*, HLT-NAACL '04, pages 24–31, Boston, Massachusetts, USA.

Miller, G., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. (1990). Five Papers on WordNet. *Special Issue of International Journal of Lexicography*, 3(4).

Miller, G., Leacock, C., Tengi, R., and Bunker, R. (1993). A Semantic Concordance. In *Proceedings of the ARPA Workshop on Human Language Technology*.

Mitkov, R. (2002). *Anaphora Resolution*. London: Longman.

Modi, A., Titov, I., and Klementiev, A. (2012). Unsupervised induction of frame-semantic representations. In *Proceedings of the NAACL-HLT Workshop on the Induction of Linguistic Structure*, WILS '12, pages 1–7, Montreal, Canada.

Moor, T., Roth, M., and Frank, A. (2013). Predicate-specific annotations for implicit role binding: Corpus annotation, data analysis and evaluation experiments. In *Proceedings of the 10th International Conference on Computational Semantics*, IWCS '13, pages 369–375, Potsdam, Germany.

Morton, T. S. (2000). Coreference for nlp applications. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, pages 173–180, Hong Kong.

Nadeau, D. and Sekine, S. (2007). A survey of named entity recognition and classification. *Lingvisticae Investigationes*, 30(1):3–26.

Ng, V. and Cardie, C. (2002). Improving machine learning approaches to coreference resolution. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, Philadelphia, Pennsylvania.

Ngai, G., Wu, D., Carpuat, M., shing Wang, C., and yung Wang, C. (2004). Semantic role labeling with boosting, svms, maximum entropy, snow, and decision lists. In *In Proceedings of SENSEVAL-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*.

Palmer, M. (2009). Semlink: Linking propbank, verbnet and framenet. In *Proceedings of the Generative Lexicon Conference*, pages 9–15.

Palmer, M., Gildea, D., and Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

Palmer, M. S., Dahl, D. A., Schiffman, R. J., Hirschman, L., Linebarger, M., and Dowding, J. (1986). Recovering implicit information. In *Proceedings of the 24th annual meeting on Association for Computational Linguistics*, ACL '86, pages 10–19, New York, New York, USA.

Peris, A., Taulé, M., Rodríguez, H., and Bertran Ibarz, M. (2013). Liarc: Labeling implicit arguments in spanish deverbal nominalizations. In *Computational Linguistics and Intelligent Text Processing*, volume 7816 of *Lecture Notes in Computer Science*, pages 423–434. Springer Berlin Heidelberg.

Poesio, M., Patel, A., and Eugenio, B. D. (2006). Discourse Structure and Anaphora in Tutorial Dialogues: An Empirical Analysis of Two Theories of the Global Focus. *Research on Language and Computation*, 4:229–257.

Poesio, M., Ponzetto, S., and Versley, Y. (2011). Computational models of anaphora resolution: A survey. *Linguistic Issues in Language Technology*.

Poesio, M., Stevenson, R., Di Eugenio, B., and Hitzeman, J. (2004). Centering: A parametric theory and its instantiations. *Computational Linguistics*, 30(3):309–363.

Poesio, M. and Vieira, R. (1998). A corpus-based investigation of definite description use. *Computational Linguistics*, pages 183–216.

Poesio, M., Vieira, R., and Teufel, S. (1997). Resolving bridging references in unrestricted text. In *Proceedings of the ACL Workshop on Operational Factors in Robust Anaphora Resolution*, pages 1–6, Madrid, Spain.

Poon, H. and Domingos, P. (2008). Joint unsupervised coreference resolution with markov logic. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, Honolulu, Hawaii.

Poon, H. and Domingos, P. (2009). Unsupervised semantic parsing. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1*, EMNLP '09, pages 1–10, Singapore.

Pradhan, S., Ramshaw, L., Marcus, M., Palmer, M., Weischedel, R., and Xue, N. (2011). Conll-2011 shared task: Modeling unrestricted coreference in ontonotes. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, CONLL Shared Task '11, pages 1–27, Portland, Oregon.

Pradhan, S. S., Hovy, E., Marcus, M., Palmer, M., Ramshaw, L., and Weischedel, R. (2007). Ontonotes: A unified relational semantic representation. *2012 IEEE Sixth International Conference on Semantic Computing*, pages 517–526.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., and Webber, B. (2008). The penn discourse treebank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco.

Prince, E. F. (1981). Toward a taxonomy of given-new information. In *Syntax and semantics: Vol. 14. Radical Pragmatics*, pages 223–255. Academic Press, New York.

Pustejovsky, J., Hanks, P., Sauri, R., See, A., Gaizauskas, R., Setzer, A., Radev, D., Sundheim, B., Day, D., Ferro, L., and Lazo, M. (2003a). The TIMEBANK corpus. In *Proceedings of Corpus Linguistics 2003*, Lancaster.

Pustejovsky, J., no, J. C., Ingria, R., SaurÃ, R., Gaizauskas, R., Setzer, A., and Katz, G. (2003b). Timeml: Robust specification of event and temporal expressions in text. In *in Fifth International Workshop on Computational Semantics (IWCS-5)*.

Quine, W. V. (1985). Events and reification. In *Actions and Events: Perspectives on the Philosophy of Davidson*, pages 162–71. Blackwell.

Raghunathan, K., Lee, H., Rangarajan, S., Chambers, N., Surdeanu, M., Jurafsky, D., and Manning, C. (2010). A multi-pass sieve for coreference resolution. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10, Cambridge, Massachusetts.

Rahman, A. and Ng, V. (2009). Supervised models for coreference resolution. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*.

Recasens, M., Màrquez, L., Sapena, E., Martí, M. A., Taulé, M., Hoste, V., Poesio, M., and Versley, Y. (2010). Semeval-2010 task 1: Coreference resolution in multiple languages. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, SemEval '10, pages 1–8, Los Angeles, California.

Recasens, M. and Martí, M. (2010). Ancora-co: Coreferentially annotated corpora for spanish and catalan. *Language Resources and Evaluation*, 44(4):315–345.

Reinhart, T. (1976). *The syntactic domain of anaphora*. MIT Linguistics Dissertations. Massachusetts Institute of Technology.

Rello, L. and Ilisei, I. (2009). A rule-based approach to the identification of spanish zero pronouns. In *Proceedings of the Student Research Workshop at RANLP'09*, pages 60–65, Borovets, Bulgaria.

Richardson, S. D., Dolan, W. B., and Vanderwende, L. (1998). Mindnet: Acquiring and structuring semantic information from text. In *COLING-ACL*, pages 1098–1102.

Ruppenhofer, J., Gorinski, P., and Sporleder, C. (2011). In search of missing arguments: A linguistic approach. In *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011*, RANLP '11, pages 331–338, Hissar, Bulgaria.

Ruppenhofer, J., Sporleder, C., Morante, R., Baker, C., and Palmer, M. (2009). Semeval-2010 task 10: Linking events and their participants in discourse. In *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions*, DEW '09, Stroudsburg, PA, USA.

Ruppenhofer, J., Sporleder, C., Morante, R., Baker, C., and Palmer, M. (2010). Semeval-2010 task 10: Linking events and their participants in discourse. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, SemEval '10, pages 45–50, Los Angeles, California, USA.

Saeboe, K. (1996). Anaphoric presuppositions and zero anaphora. *Linguistics and Philosophy*, 19(2):187–209.

Sapena, E., Padró, L., and Turmo, J. (2010). A global relaxation labeling approach to coreference resolution. In *Proceedings of 23rd International Conference on Computational Linguistics, COLING*, Beijing, China.

Schilder, F., Katz, G., and Pustejovsky, J. (2007). Annotating, extracting and reasoning about time and events. In *Annotating, Extracting and Reasoning about Time and Events*, volume 4795 of *Lecture Notes in Computer Science*, pages 1–6. Springer Berlin Heidelberg.

Shopen, T. (1973). Ellipsis as grammatical indeterminacy. *Foundations of Language*, 10:65.

Sidner, C. (1986). Focusing in the comprehension of definite anaphora. In *Readings in Natural Language Processing*, pages 363–394. Morgan Kaufmann.

Sidner, C. L. (1978). The use of focus as a tool for disambiguation of definite noun phrases. In *Proceedings of the 1978 workshop on Theoretical issues in natural language processing*, TINLAP '78, pages 86–95, Stroudsburg, PA, USA. Association for Computational Linguistics.

Silberer, C. and Frank, A. (2012). Casting implicit role linking as an anaphora resolution task. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics*, *SEM '12, pages 1–10, Montréal, Canada.

Soon, W. M., Ng, H. T., and Lim, C. Y. (1999). Corpus-Based Learning for Noun-Phrase Coreference Resolution. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-99)*, College Park, Maryland.

Soon, W. M., Ng, H. T., and Lim, D. C. Y. (2001). A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, pages 521–544.

Stent, A. J. and Bangalore, S. (2010). Interaction between dialog structure and coreference resolution. In *IEEE Workshop on Spoken Language Technology*.

Stern, A. and Dagan, I. (2014). Recognizing implied predicate-argument relationships in textual inference. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 739–744, Baltimore, Maryland.

Strube, M. and Hahn, U. (1999). Functional centering: Grounding referential coherence in information structure. *Computational Linguistics*, 25(3):309–344.

Surdeanu, M., Johansson, R., Meyers, A., Màrquez, L., and Nivre, J. (2008). The CoNLL-2008 shared task on joint parsing of syntactic and semantic dependencies. In *Proceedings of the Twelfth Conference on Natural Language Learning*, CoNLL '08, pages 159–177, Manchester, United Kingdom.

Surdeanu, M. and Turmo, J. (2005). Semantic role labeling using complete syntactic analysis. In *Proceedings of the Ninth Conference on Computational Natural Language Learning*, CONLL '05, Ann Arbor, Michigan.

Swier, R. and Stevenson, S. (2004). Unsupervised semantic role labelling. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, EMNLP'04, pages 95–102, Barcelona, Spain.

Taulé, M., Martí, M. A., Peris, A., Rodríguez, H., Moreno, L., and Moreda, P. (2012). Iarg-ancora: Anotación de los corpus ancora con argumentos implícitos. *Procesamiento del Lenguaje Natural*, 49:181–184.

Tetreault, J. R. (2002). Implicit role reference. In *International Symposium on Reference Resolution for Natural Language Processing*, pages 109–115, Alicante, Spain.

Titov, I. and Klementiev, A. (2011). A bayesian model for unsupervised semantic parsing. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, pages 1445–1455, Portland, Oregon.

Tjong Kim Sang, E. F. and De Meulder, F. (2003). Introduction to the conll-2003 shared task: Language-independent named entity recognition.

In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003 - Volume 4*, CONLL '03, Edmonton, Canada.

Tonelli, S. and Delmonte, R. (2010). Venses++: Adapting a deep semantic processing system to the identification of null instantiations. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, SemEval '10, pages 296–299, Los Angeles, California, USA.

Tonelli, S. and Delmonte, R. (2011). Desperately seeking implicit arguments in text. In *Proceedings of the ACL 2011 Workshop on Relational Models of Semantics*, RELMS '11, pages 54–62, Portland, Oregon, USA.

Vieira, R. and Poesio, M. (2000). An empirically based system for processing definite descriptions. *Computational Linguistics*, pages 539–593.

Vilain, M., Burger, J., Aberdeen, J., Connolly, D., and Hirschman, L. (1995). A model-theoretic coreference scoring scheme. In *Proceedings of the 6th Conference on Message Understanding*, MUC6 '95, pages 45–52.

Webber, B., Egg, M., and KordoniI, V. (2012). Discourse structure and language technology. *Natural Language Engineering*, 18:437–490.

Weischedel, R. M. and Sondheimer, N. K. (1982). An improved heuristic for ellipsis processing. In *Proceedings of the 20th Annual Meeting on Association for Computational Linguistics*, ACL '82, Toronto, Ontario, Canada.

Whittemore, G., Macpherson, M., and Carlson, G. (1991). Event-building through role-filling and anaphora resolution. In *Proceedings of the 29th annual meeting on Association for Computational Linguistics*, ACL '91, pages 17–24, Berkeley, California, USA.

Wilks, Y. (1975). An intelligent analyzer and understander of english. *Communications of the ACM*, pages 264–274.

Wilks, Y. A. (1986). An intelligent analyzer and understander of english. In *Readings in Natural Language Processing*, pages 193–203. Morgan Kaufmann Publishers Inc.

Winograd, T. (1972). Understanding natural language. *Cognitive Psychology*, 3:1 – 191.

Yang, X., Su, J., and Tan, C. L. (2005). A twin-candidate model of coreference resolution with non-anaphor identification capability. In *Proceedings of the Second International Joint Conference on Natural Language Processing*, IJCNLP'05, Jeju Island, Korea.

Yang, X., Zhou, G., Su, J., and Tan, C. L. (2003). Coreference resolution using competition learning approach. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, ACL '03, Sapporo, Japan.

Yeh, C.-L. and Chen, Y.-C. (2007). Zero anaphora resolution in chinese with shallow parsing. *Journal of Chinese Language and Computing*, 17:41–56.

Zhao, H., Chen, W., Kazama, J., Uchimoto, K., and Torisawa, K. (2009). Multilingual dependency learning: Exploiting rich features for tagging syntactic and semantic dependencies. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task*, CoNLL'09, pages 61–66, Boulder, Colorado.

Zhao, S. and Ng, H. T. (2007). Identification and resolution of chinese zero pronouns: A machine learning approach. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP'07, pages 541–550.