Embedding Predications

Haci Halil Kilicoglu

A Thesis

In the Department

of

Computer Science and Software Engineering

Presented in Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy (Computer Science) at

Concordia University

Montréal, Québec, Canada

April 2012

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Haci Halil Kilicoglu**

Entitled: **Embedding Predications**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Computer Science)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

____________________ Hashem Akbari

____________________ Owen Rambow

____________________ Justin Powlowski

____________________ Gregory Butler

____________________ Leila Kosseim

____________________ Thomas C. Rindflesch

____________________ Sabine Bergler

Approved by ____________________

Chair of Department or Graduate Program Director

________ 2012 ____________________

Dr. Robin A. L. Drew, Dean

Faculty of Engineering and Computer Science

ABSTRACT

Embedding Predications

Haci Halil Kilicoglu, Ph.D.

Concordia University, 2012

Written communication is rarely a sequence of simple assertions. More often, in addition to simple assertions, authors express subjectivity, such as beliefs, speculations, opinions, intentions, and desires. Furthermore, they link statements of various kinds to form a coherent discourse that reflects their pragmatic intent. In computational semantics, extraction of simple assertions (propositional meaning) has attracted the greatest attention, while research that focuses on extra-propositional aspects of meaning has remained sparse overall and has been largely limited to narrowly defined categories, such as hedging or sentiment analysis, treated in isolation.

In this thesis, we contribute to the understanding of extra-propositional meaning in natural language understanding, by providing a comprehensive account of the semantic phenomena that occur beyond simple assertions and examining how a coherent discourse is formed from lower level semantic elements. Our approach is linguistically based, and we propose a general, unified treatment of the semantic phenomena involved, within a computationally viable framework. We identify semantic embedding as the core notion involved in expressing extra-propositional meaning. The embedding framework is based on the structural distinction between embedding and atomic predications, the former corresponding to extra-propositional aspects of meaning. It incorporates the notions of predication source, modality scale, and scope. We develop an embedding categorization scheme and a dictionary based on it, which provide the necessary means to interpret extra-propositional meaning with a compositional semantic interpretation methodology. Our syntax-driven methodology exploits syntactic dependencies to construct a semantic embedding graph of a document. Traversing the graph in a bottom-up manner guided by compositional operations, we construct predications corresponding to extra-propositional semantic content, which form the basis for addressing practical tasks. We focus on text from two distinct domains: news articles from the Wall Street Journal, and scientific articles focusing on molecular biology. Adopting a task-based evaluation strategy, we consider the easy adaptability of the core framework to practical tasks that involve some extra-propositional aspect as a measure of its success. The computational tasks we consider include

hedge/uncertainty detection, scope resolution, negation detection, biological event extraction, and attribution resolution. Our competitive results in these tasks demonstrate the viability of our proposal.

Acknowledgments

This thesis would not have been possible without the guidance and support of my advisor, Dr. Sabine Bergler. I wish to thank her for her confidence in me and encouraging style of supervision, which allowed me to explore ideas beyond what I had anticipated when I decided to pursue a PhD in computational linguistics. I will fondly remember our long and casual discussions about language and computation, among other things, over lunch or coffee, satisfying in both intellectual and culinary senses. I hope that we can continue our collaboration in the future, as well.

I feel indebted to my co-supervisor, Dr. Thomas C. Rindflesch, for introducing me to the field of computational linguistics and teaching me how to conduct linguistics research. His support and mentorship as well as his friendship over the years have helped me greatly, both personally and professionally. In the later stages of my research, in particular, his encouraging words about my ideas and writing helped me overcome my anxiety about my dissertation.

I am grateful to the members of my committee, Drs. Leila Kosseim, Gregory Butler, Justin Powlowski, and Owen Rambow, who, despite their busy schedules, took the time to read my dissertation and provided constructive comments and suggestions. In particular, I would like to thank Dr. Kosseim, whose early comments on the relation of my research to discourse coherence provided the impetus to introduce a whole new component to my research.

I would like to express my gratitude to my partner, Zeynep Arsel, who tolerated my grumpiness and gave me a sense of normalcy over the years. I thank my friends, Marcelo Fiszman and Sarah Scotch, for encouraging me to pursue a PhD degree and providing moral support throughout my studies. I also wish to thank my family for their support over the years despite the physical distance; I am particularly grateful to my brother, Ömer Faruk, who gave me the first taste of all things intellectual in my early years. Finally, I would like to thank my fellow graduate students at the CLaC lab, who provided moral support throughout my studies as well as valuable feedback to the earlier versions of the work presented in this manuscript.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

"It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."[1]

(Watson and Crick, 1953)

1.1 Motivation

Categorical assertions are rare. The very fact that the previous sentence, while true, grates a little on the sensibilities of the reader, illustrates this point: it is usually inappropriate to make an assertion without supporting it or presenting it as a tentative statement, particularly in academic writing. Most of human communication is imbued with subjectivity, vagueness, and uncertainty, core aspects of natural language that have long been the subject of philosophical and linguistic inquiry. From a computational perspective, however, most automatic text analysis systems have focused on extracting propositional meaning of categorical assertions, ignoring the kind of meta-information contributed by linguistic expressions of such extra-factual phenomena. Furthermore, such phenomena (beliefs, speculations, opinions, desires, intentions, etc.) as well as categorical assertions do not occur in isolation but are linked to one another in a coherent way to form the larger *discourse*[2]. Discourse interpretation (or text understanding) has long been recognized as one of the

[1] I thank Christopher M. Miller, MD for bringing this sentence to my attention.

[2] While discourse may refer to written as well as spoken communication (dialogue), in the current work, we focus specifically on written text.

major goals in computational linguistics/natural language processing (CL/NLP) research (Hobbs, 1985b); however, attaining it in a general sense has remained elusive.

With the current work, we aim to contribute to our understanding of *extra-propositional meaning*, by investigating the semantic phenomena that occur beyond simple, categorical assertions and how discourse is formed from these lower level semantic elements. Our approach is linguistically based, and we propose a general, unified treatment of the phenomena involved in a computationally viable framework. In recent years, the need for text analysis systems to identify and model extra-factual information has become increasingly clear. This interest is mostly driven by practical concerns. For example, abundance of online opinions and product reviews and the need to mine them has provided a rich context for *sentiment analysis* and *subjectivity* research (see Pang and Lee (2008) for a relatively recent and comprehensive survey of the field). Similarly, the need to distinguish facts from speculative, tentative information in scientific research articles has provided the impetus for research in speculation/hedging detection (Light *et al.*, 2004). The scientific method requires making hypotheses, experimenting, and reasoning to reach tentative and provisional conclusions; therefore, it is not surprising that scientific articles are very rich in non-assertive, speculative statements. Despite these recent research trends, however, most of the CL/NLP research has remained focused on more foundational aspects of text analysis, such as syntactic parsing, named entity recognition and shallow semantic parsing. With the advances made in these important, foundational tasks, we believe it is timely to ask how close we are to the goal of text understanding.

In the contemporary CL/NLP research, two predominant orientations towards text understanding can be distinguished. In the lower level orientation, the research focuses on *propositional meaning* (essentially, "who did what to who?") and the goal is to construct semantic representations that encode such information. For example, the task of *semantic role labeling* (or shallow semantic parsing) (Gildea and Jurafsky, 2002) is largely concerned with this level and aims to identify *predicate-argument structures*, semantic representations in which a predicate is associated with its semantic arguments and the semantic roles of the arguments, such as AGENT and THEME, are specified. Consider the sentence in Example (1). The relevant predicate-argument structures as annotated in PropBank (Palmer *et al.*, 2005), the standard corpus of verbal predicate-argument structures, are given in Table (1.1).

(1) Although Mr. Azoff won't <u>produce</u> films at first, it is possible that he could <u>do</u> so later, the sources <u>said</u>. (wsj_0408)

Such abstract semantic representations can serve as the basis for practical tasks, such as automatic

Predicate	produce.01
Arg0	Mr. Azoff
Arg1	films
ArgM-MOD	will
ArgM-NEG	not
ArgM-TMP	at first

Predicate	do.02
Arg0	he
Arg1	so
ArgM-MOD	could
ArgM-TMP	later

Predicate	say.01
Arg0	the sources
Arg1	Although ... later

Table 1.1: PropBank predicate-argument structures for the sentence in Example (1).

question answering, machine translation, and text mining. There are several things of note in these predicate-argument structures. In the PropBank terminology, Arg0 roughly corresponds to AGENT semantic role and Arg1 to THEME role. On the other hand, arguments prefixed with ArgM correspond to non-core (circumstantial) semantic arguments, ArgM-MOD corresponding to modal verb, ArgM-NEG to negation marker and ArgM-TMP to time argument. While lexical cues for some of the extra-factual phenomena (modal adjective *possible*, modal verb *could* and *won't*, reporting predicate *said*) are recorded within the frames to some extent, their semantic consequences are not made explicit. For example, the notion that *Mr. Azoff will produce films later* is a possibility that is not explicit in the semantic representation. An automatic question answering system based on such frames needs to capture this information to answer a question on whether Mr. Azoff will produce the films later.

In contrast, at a higher level orientation, the focus of research is to explain how textual units (sentences, clauses) relate to each other in creating a coherent textual meaning beyond the sum of the meaning of these units (Hobbs, 1985a; Mann and Thompson, 1988; Asher and Lascarides, 2003). This level of analysis is referred to as *discourse analysis* and the research in this area has focused on modeling text as a set of *coherence relations*[3] (such as Background, Elaboration, Contrast) and automatic identification of such relations based on a particular discourse model (Marcu, 1999; Soricut and Marcu, 2003; Wellner, 2009). One resource for discourse relations is the Penn Discourse TreeBank (PDTB) (Miltsakaki *et al.*, 2004; Prasad *et al.*, 2008), in which discourse relations between textual spans are annotated as a semantic layer on top of the Penn TreeBank (similarly to PropBank). Consider the earlier sentence again. PDTB provides the annotations in Table (1.2) for this sentence.

[3]Coherence relations may also be referred to as *rhetorical relations* or, simply, as *discourse relations*. We will generally refer to them as *discourse relations* in the current work.

Relation	Although:CONTRAST (Ot.Comm)
Arg1	it is possible . . . later (Inh.Null)
Arg2	Mr. Azoff . . . at first (Inh.Null)

Relation	later:PRECEDENCE (Ot.Comm)
Arg1	Although . . . at first (Inh.Null)
Arg2	that he could do so (Inh.Null)

Table 1.2: PDTB discourse relations for the sentence in Example (1).
.

The first discourse relation indicates that there is a CONTRAST relation between the two segments in the sentence, which correspond to Arg1 and Arg2 arguments. In addition to discourse relations and their arguments, PDTB provides attribution information. With respect to the CONTRAST relation, the *source* is someone other than the author, denoted as Ot (Other) and indicated by the cue *the sources* in the sentence. The attribution *type* of the same relation is *assertion*, denoted as Comm and indicated by *said*. Additionally, both arguments of the CONTRAST relation inherit their attribution from the discourse relation itself. The inherited source is denoted as Inh and the inherited type as Null in the representation. The second discourse relation is temporal (PRECEDENCE) and is, again, an assertion and the assertion is attributed to someone other than the author. While these relations and the earlier predicate-argument structures cover different semantic aspects of the same sentence, one is not clearly related to the other. Moreover, even though attribution information is provided to some extent, we still do not have the necessary means to answer the question of whether Mr. Azoff will produce the films later.

This sentence illustrates the apparent gap between these lower-level oriented and higher-level oriented viewpoints of text understanding in CL/NLP research. In general, discourse analysis approaches ignore the propositional meaning encoded within individual discourse units and treat these units as essentially black boxes. On the other side of the coin, approaches focusing on propositional meaning largely ignore the context in which such propositional meaning components appear and how these components relate to one another. Furthermore, extra-factual phenomena are largely ignored in both perspectives. Instead, they are studied separately, often with a focus on a narrowly defined, pragmatic category of interest (such as *hedging* or *uncertainty*) and its detection in text from a particular domain. Such phenomena are not related to higher discourse level and the semantic consequences of the intricate interactions between these phenomena are not accounted for. In summary, a systematic, computational treatment of extra-propositional meaning remains largely lacking.

There are isolated efforts in this regard; however, it seems that they have failed to make a lasting impact. Segmented Discourse Representation Theory (SDRT) (Asher and Lascarides, 2003), for example, aims to bring together various semantic levels from lexical semantics to discourse coherence relations into a formally precise semantic framework. However, its treatment has focused on several important but limited phenomena, such as anaphora and rhetorical relations and its computational applications have been rare. The ontological semantics framework (Nirenburg and Raskin, 2004) has similar goals in that it aims to provide interpretation from propositional to discourse/pragmatic and even stylistic levels. However, it presupposes a large, manually-encoded ontological knowledge base and its implementations have been largely limited to narrow domains. Furthermore, its treatment of some relevant phenomena, for example discourse coherence, has been sparser than its treatment of more basic propositional meaning.

1.2 Approach

In the current work, we attempt to address the gap to some extent. We begin by posing the following questions:

- What lies beyond categorical assertions in discourse?
- How is higher level discourse formed from these categorical assertions and what are the intermediate semantic phenomena?
- Is it possible to model the extra-propositional meaning in a general, domain-independent manner?

We investigate the answers to these questions in two distinct genres and domains of written communication, namely, news articles and the biomedical literature. Our focus on distinct genres and domains is due to both theoretical and practical concerns. We wish to propose a general, domain-independent, unified approach to semantic phenomena occurring beyond categorical assertions and this objective would benefit from taking distinct domains into account. Furthermore, we would like to be able to address practical tasks using this general framework as the core aspect, and such tasks and relevant evaluation resources (corpora) generally have targeted specific domains.

News corpora, particularly the Penn TreeBank (PTB) (Marcus *et al.*, 1993), have formed the basis of much research in natural language processing and have been the standard resources for training probabilistic NLP tools in the general domain. For example, PTB has been used to train most of the state-of-the-art probabilistic syntactic parsers (Klein and Manning, 2003; Charniak and Johnson, 2005). Over the years, various semantic layers have been annotated on top of the syntactic layer of PTB to serve as training and evaluation platforms for various semantic tasks. Verbal and nominal predicate-argument structures have been annotated in PropBank (Palmer *et al.*, 2005) and NomBank (Meyers *et al.*, 2004b), respectively, for semantic role labeling task, while discourse relations have been annotated in Penn Discourse TreeBank (Miltsakaki *et al.*, 2004; Prasad *et al.*, 2008) to serve discourse analysis tasks.

The biomedical literature has emerged as a key resource for NLP research in the last decade. The molecular biology literature, in particular, is growing at an exponential rate. With the overwhelming amount of data, efficient access to existing knowledge is crucial. NLP and text mining techniques are viewed favorably as tools that can assist biomedical and life sciences researchers in their tasks by facilitating biological analyses as well as biological database curation. Most of the research in the field of biomedical natural language processing (bioNLP) has focused on foundational tasks, such as accurate gene/chemical name identification, abbreviation and acronym resolution, and corpus annotation. More recently, tasks that can be subsumed under semantic interpretation have been garnering attention. Extracting explicitly stated assertions in the form of semantic relations from the biomedical literature has been the focus of applications, such as SemRep (Rindflesch and Fiszman, 2003). Other relation extraction applications are more focused with respect to the relation types they are concerned with: protein-protein interactions (see Kabiljo *et al.* (2009) for a relatively recent assessment), biological events (see recent event extraction systems in Kim *et al.* (2011a)), among others. An even more recent trend in bioNLP is the interest in extra-factual phenomena, such as *speculation*, *uncertainty* and *negation*, which are crucial, as they have an effect on the nature and reliability of the underlying scientific claim. Light *et al.* (2004) estimate that 11% of sentences in Medline abstracts contain speculative fragments and argue that speculations, more than established facts, are important for researchers interested in current trends and future directions.

Studying text from these two distinct domains, then, allows us to cast a wide net over extra-propositional content: we consider both high level discourse, generally considered in the context of general English and news articles, and the lower level extra-factual phenomena, more closely associated with the biomedical literature.

In the current work, our main hypothesis is that it is possible to model extra-propositional semantic content in a general, unified manner with a linguistically-motivated perspective and that such a treatment would benefit computational tasks at various levels. We also argue that a linguistically-based approach not only allows a finer-grained text understanding than generally assumed but also can enable moving towards discourse interpretation.

To model semantic content, some representational means is required. We base our discussion on the notion of *predication*, an abstract semantic construct corresponding to a piece of relational meaning, essentially meaning of a simple, declarative statement. In different disciplines, various names have been used to refer to the more or less the same notion, including terms such as *proposition*, *fact*, *assertion*, *semantic relation*, *event*, *eventuality*, or *predicate-argument structure*[4]. We find that the term predication is more neutral with respect to various semantic theories[5] and will be using it as an encompassing term for all kinds and levels of relational meaning.

Why is an abstract construct such as a predication necessary to study semantic content? Such representations are largely taken for granted as theoretical units in many cognitive models of language comprehension (Miller and Johnson-Laird, 1976; Kintsch and van Dijk, 1978; van Dijk and Kintsch, 1983) and their psychological basis has been shown empirically to some extent. For example, Kintsch and Keenan (1973) have shown a correlation between sentence complexity as measured by number of propositions in those sentences and difficulty in processing. Ratcliff and McKoon (1978) used priming effects on recognition latency to study discourse structure and found that it took subjects significantly longer to recognize a word from a sentence different than that previously seen and attributed this priming effect to propositional structure.

Taking predications as our main construct, we make the structural distinction between *atomic* and *embedding* predications, the latter corresponding to the predications that can take as arguments other predications. *Embedding* is generally used in a syntactic sense in linguistics, referring to "the occurrence of one unit as a constituent of another unit at the same rank of the grammatical hierarchy" (Quirk *et al.*, 1985). We use the notion here in a semantic sense; that is, embedding of a predication within another predication as an argument. Throughout this thesis, when we use embedding in a syntactic sense, we will clearly indicate it as such.

The distinction between *atomic* and *embedding* predications is novel within the CL/NLP field. For instance, recall the earlier example, in which the extent of the Arg1 argument of the frame *say.01* corresponds to the entire extent of the discourse relation indicated by *Although*. Therefore, in our view, the predicate-argument structure corresponding to *say* is an embedding one; however, PropBank does not distinguish it from what we would consider atomic predications (*produce.01* and *do.02* frames). We believe that casting the discourse interpretation problem in this structural manner is computationally attractive as it allows us to formulate a compositional approach to text meaning

[4]For subtle but formally significant differences between some of these different notions, we refer the reader to Asher (1993).

[5]For example, the term *proposition* is closely aligned with truth-conditional semantics, and implies a truth value (true or false) for the semantic relation.

and consider different semantic layers in a unified way. We call these different layers collectively as *the embedding layer*. Research in cognitively-based language comprehension lends some support to our characterization. For example, van Dijk and Kintsch (1983) draw parallels between clausal complexity and propositional complexity, using terms *atomic*, *complex*, *composite*, and *compound* to refer to different types of clauses and propositions. In their characterization, a *complex* proposition involves a proposition at a different rank (subordinate), while a *compound* proposition involves two propositions of the same rank (coordinate). *Composite* propositions is a general term for both types. They note that Ratcliff and McKoon (1978) obtained greater priming effects within atomic propositions than between propositions.

1.3 Contributions

We argue that embedding predications largely correspond to a domain-independent layer of extra-propositional semantic content (the embedding layer) and our goal in the current work is to model this layer. By accounting for this layer, our embedding framework aims to bridge the aforementioned gap between lower-level and higher-level oriented viewpoints in a general manner, bringing us closer to the goal of discourse interpretation. Our main contributions are as follows:

1. A linguistically-motivated, structural perspective to textual meaning based on the distinction of atomic and embedding predications
2. A domain-independent categorization scheme of embedding predicate types
3. A dictionary of embedding predicates
4. A bottom-up, compositional semantic interpretation methodology whose goal is to extract predications
   (a) based on semantic dependencies, the embedding categorization scheme and the dictionary of embedding predicates
   (b) from which specific practical tasks can be addressed by defining and tailoring inferential processes

One of our main concerns is to avoid task- or domain-specific optimizations within the embedding framework. Such optimizations often prove beneficial in practical tasks; however, they teach us little about the broader goal of text understanding and also conflict with our desire for generality. Nevertheless, we recognize the need to make our general framework useful for specific practical tasks that have some overlap with the notions explored in this work, as well. In this respect, our approach to addressing such specific tasks is to tailor the output of the core embedding framework to the requirements of these tasks and take its easy adaptability as an indication of the success of the embedding framework.

1.4 Thesis Outline

The thesis is organized as follows: Chapter 2 provides the linguistic foundations of the current work, introducing relevant notions that we will be using throughout this work, such as *scope* and *scalarity*. Furthermore, it provides a survey of related work in theoretical linguistics and CL/NLP as well as in biomedical NLP. The survey section can roughly be divided into two. The first part largely has a lower-level semantic orientation and focuses on the linguistic notions of *modality* and *negation* and related phenomena. The second part is, on the other hand, concerned with higher-level, discourse-based approaches. Chapter 3 presents a theoretical description of the embedding framework. The basic constructs, such as *atomic* and *embedding predications*, are defined and the embedding categorization scheme is introduced, with illustrative examples from Penn TreeBank (Marcus *et al.*, 1993). Furthermore, scopal interactions between embedding predications are described, by appealing to the embedding categorization. Finally, we characterize how a discourse level interpretation can be arrived at based on embedding predications. Chapter 4 provides the lexical and syntactic foundations that underpin the compositional semantic interpretation approach, which is outlined in Chapter 5. Chapter 6 describes the practical tasks that formed the basis for evaluation and experiments that we carried out on biomedical literature as well as on news articles. Chapter 7 presents the lessons learned and suggests several future directions for further research based on the embedding framework.

Chapter 2

Background

We introduced *embedding* in the previous chapter as the structural semantic notion that forms the backbone of the current work. In this chapter, we discuss the linguistic phenomena that are relevant in the context of embedding as well as computational approaches that focus on some semantic aspect expressed via embedding. With respect to linguistics, we confine ourselves mainly to comprehensive studies on core categories, describing basic characteristics and classifications. With respect to computational approaches, we are also concerned with semantic/pragmatic categories with practical implications, such as *hedging*, *factuality*, and *subjectivity*, and emphasize their relation to core linguistic concepts.

We consider both structural and functional aspects of embedding in discussing linguistic phenomena. In Section 2.1, we focus on structural aspects and describe several fundamental notions, such as *clause linking*, *scope*, and *scalarity*. We also introduce the Principle of Compositionality, our basic, structurally-based semantic assumption. In Section 2.2, in terms of linguistic description, we take a largely functional/typological perspective. The discussion in this section is divided into two parts; the first part focuses on the traditional linguistic categories of *modality* and *negation*, while the second part is concerned with discourse structure and coherence, reflecting the lower level/higher level orientation to text understanding, introduced briefly in the first chapter. This section also provides a survey of computational approaches focusing on these categories in natural language processing research (including biomedical NLP), again divided along the distinction of lower level/higher level text understanding.

2.1 Embedding and Linguistic Structure

Embedding, in essence, implies structure, and in this section, we review at a high level several fundamental linguistic notions that have a structural element and a bearing on our approach. First, we discuss clause linking devices of *subordination* and *coordination* from a syntactic as well as discourse semantic point of view. Secondly, we describe *scope*, a structurally-determined unit of semantic influence. Next, we briefly introduce *scalarity*, a semantic notion with structural properties. Finally, we discuss the Principle of Compositionality, which assumes structure and underpins our computational approach.

2.1.1 Clause Linking

In linguistics, the term *clause linking* is used to refer to the ways clauses can be organized in discourse to form more complex units. Two clause linking devices are distinguished traditionally: *subordination* and *coordination*. *Subordination* denotes an asymmetric (hierarchical) relation between the parts of a complex syntactic unit, whereas *coordination* denotes a symmetric (non-hierarchical) relation. The difference between subordination and coordination is often accounted for in terms of two dimensions (Cristofaro, 2003):

1. *Dependency* refers to the fact that a subordinate clause cannot occur in isolation.

2. *Embedding*, in a syntactic sense, refers to the intrinsic property of a subordinate clause to function as a constituent of the matrix clause.

With respect to these two features, subordination involves a (subordinate) clause that is both dependent and embedded, while coordination involves clauses that are both independent and non-embedded. While this distinction holds in some idealized cases, the reality is murkier: it is generally accepted that there is no sharp distinction between the two types of clause linking and that they should be seen as prototypical poles on a continuum (Quirk *et al.*, 1985; Givón, 2001; Cristofaro, 2003). This continuum is illustrated in Figure (2.1).

The terms subordination and coordination have also been used in a discourse semantic sense. Several discourse theories assume the notion that discourse elements can be organized hierarchically or non-hierarchically, referring to discourse relations between such elements as *subordinating relations* and *coordinating relations*, respectively (Mann and Thompson, 1988; Asher and Vieu, 2005). The asymmetry between hierarchical discourse elements have been explained using concepts, such as

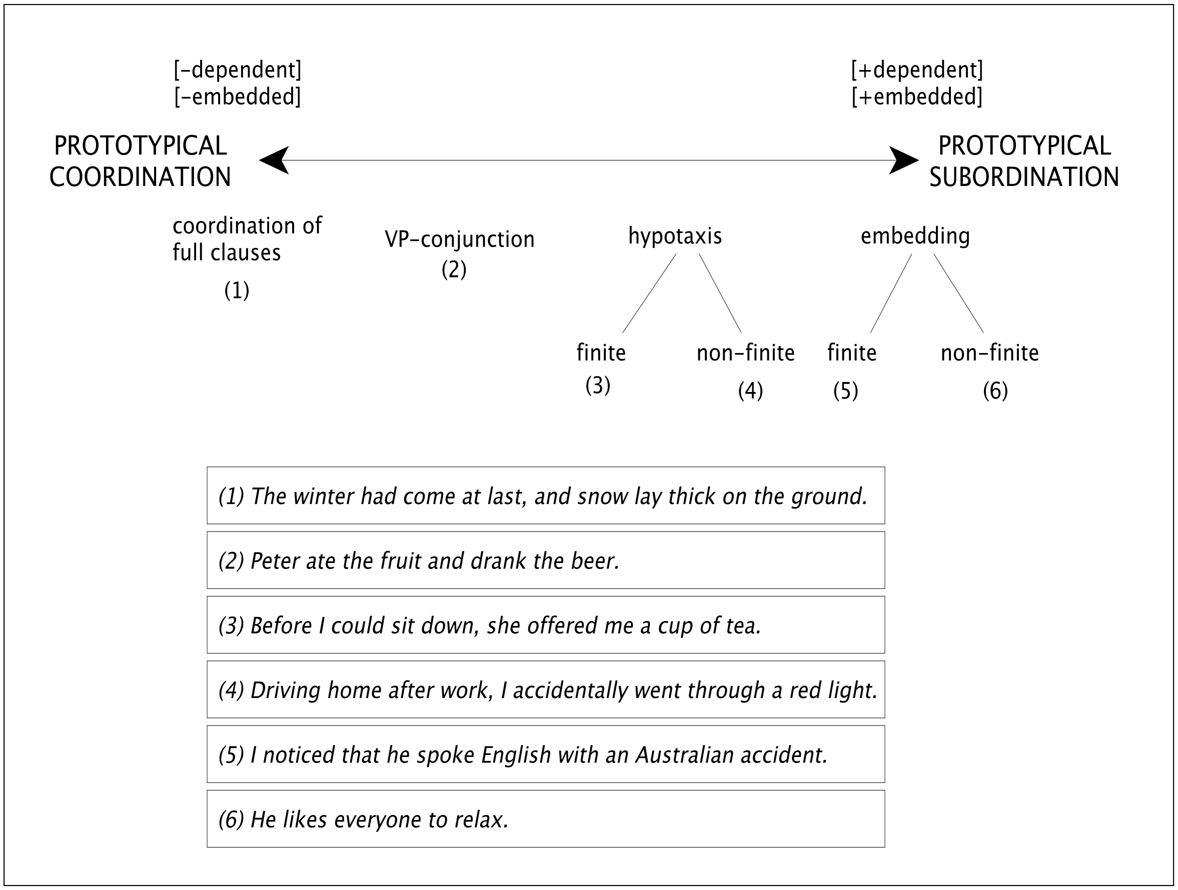

Figure 2.1: Clause linking as a continuum (Cosme, 2008)

foreground/background (Quirk *et al.*, 1985), *nucleus/satellite* (Mann and Thompson, 1988), and *ground/figure* (Langacker, 1987).

The correspondence between the subordination/coordination distinction at the syntactic and discourse semantic levels has also been a matter of debate. Subordinate clauses have often been assumed to carry background information and, thus, reflect hierarchical organization of discourse elements (Matthiessen and Thompson, 1988). However, recent research suggests a more nuanced perspective. Stede (2008) argues that syntactic subordination is only one of the factors that affect saliency in discourse. Blühdorn (2008), based on evidence from discourse connectives (e.g., *and, while*), rules out a parallelism between discourse and syntactic subordination. He argues that the syntactic distinction between coordination and subordination is neutralized at higher levels of discourse and that hierarchical/non-hierarchical relations at the discourse semantic level can be encoded by both types of clause linking devices. Consider the sentences in Example (2), taken from Blühdorn (2008). Example (2a) illustrates a non-hierarchical relation encoded with a coordinating conjunction (underlined) and (2b) a hierarchical relation also indicated with a coordinating conjunction[1]. On the other hand, in Example (2c), the the non-hierarchical relation is indicated by a subordinating conjunction.

(2) (a) The penguins were yellow-brown, <u>and</u> the giraffes were black and white.

(b) Mary went to the library, <u>and</u> she began to feel hungry.

(c) The penguins were yellow-brown, <u>while</u> the giraffes were black and white.

Intuitively, the notion of semantic embedding is linked to syntactic and discourse semantic subordination. However, the gradient nature of clause linking and the flexibility between clause linking at the syntactic level and discourse structure also illustrate the necessity of considering both types of clause linking devices for computational adequacy.

2.1.2 Scope

Scope, as a linguistic notion, is described as the "semantic influence which [some] words have on neighboring parts of a sentence" (Quirk *et al.*, 1985, p.85). Quirk *et al.* (1985) identify such special words as negative forms, pro-forms (including *wh*-words), assertive and nonassertive forms (e.g., quantifiers *some* vs. *any*), and other operators with logical function.

In formal approaches to semantics, scope is widely discussed in relation to quantifiers (e.g., *every*, *some*) and *quantifier scope ambiguity* is a major concern (Chierchia and McConnell-Ginet, 1990). This refers to the meaning ambiguity that arises in a sentence like "Everyone loves someone.", which can be paraphrased either as "everyone has somebody who s/he loves" or as "there is a specific person who everyone loves" depending on whether the universal quantifier *every* has scope over the existential quantifier *some* (the former paraphrase) or whether the scope relation is in the opposite direction (the latter).

Scope is often closely connected with ordering of lexical units (Quirk *et al.*, 1985). For example, a negative form generally has scope over whatever follows and defines it as non-assertive. In Example (3a), *any*, because it follows *never*, is within the scope of negation, while *some* is not (square brackets indicate the scope). However, particularly in the case of negation, ordering of lexical units may not predict the scope correctly. In Example (3b), negation has a *wide scope* reading, corresponding to the paraphrase "It is not the case that everything that glitters is gold", while in Example (3c), the scope of negation is narrowed to the complement clause of *think* (corresponding to the paraphrase "I think that he is not coming").

[1]The relation is hierarchical in the sense that the clauses are temporally ordered.

(3) (a) Some people [never send any Christmas cards].

(b) [All that glitters is not gold].

(c) I do not think [he is coming].

A related notion to scope is *focus*, described as the part of the scope that is more prominently or explicitly influenced by the scoping element (Huddleston and Pullum, 2002). In the following sentence, the scope is within square brackets and the focus is also in bold. The sentence can be paraphrased as "John kicked the ball but not with enough force", implying that John's kicking of the ball in fact took place (Givón, 2001).

(4) John did [not kick the ball **with enough force**].

Identifying semantic influence of lexical items (therefore, their scope) is an important part of our compositional semantic interpretation approach and is aided by syntactic dependency relations (Mel'čuk, 1988). In fact, one of the major components of our compositional approach is to obtain semantic dependencies from syntactic ones, described in detail in Section 5.2. While the notion of scope plays a large role in our approach, we note that quantifier scope ambiguity and semantic focus are outside the scope of our work.

2.1.3 Scalarity

In linguistics, the notion of *scalarity* has often been considered in the context of gradability, comparison, and intensification. Scalarity implies a value scale corresponding to a particular dimension (such as height, temperature, certainty) to which certain textual expressions can be mapped. Scalar nature of gradable adjectives have attracted the most attention (Kennedy and McNally, 2005; Bale, 2011). Sometimes, the reference to a scale is part of the lexical semantics of an expression (such as, *tall*, *short* on *height* scale). In other cases, explicit comparison or intensification (or weakening) situates a linguistic element to a value on a scale or moves it to a new value. In Example (5a), an explicit comparison is used to situate John on *height* scale with respect to Mary, while the adverb *very* strengthens the likelihood that Jorge will win the race, in other words, moves the proposition to a higher value on the likelihood (certainty) scale in Example (5b), due to the fact that the adverb has scope over the adjective *likely*.

(5) (a) John is taller than Mary.

(b) It is very likely that Jorge will win the race.

Horn (1972) proposed the notion of *scalar predication*, collections of predicates P_n ($< P_j, P_{j-1}, \ldots P_2, P_1 >$), where P_n is stronger than P_{n-1} on the relevant scale and he argued that the weaker predicate implies that the speaker believes the negative of the stronger predicate (*scalar implicature*). For example, on the scale of *excellence*, the possible values may be considered <*excellent, very good, good, acceptable*>. When the speaker utters the sentence *The weather is acceptable*, the implicature that can be drawn is that *The weather is not good, very good, or excellent.* Horn also discussed the scalarity of modals, which we take into account in our framework and discuss in the context of modality in Section 2.2.1.

2.1.4 The Principle of Compositionality

The principle of *semantic compositionality*, attributed to Frege, states that the meaning of a complex expression is a function of the meanings of its constituents and of the syntactic rules by which they are combined (Dowty *et al.*, 1981). The main argument for semantic compositionality comes from the notion of *productivity*, which entails that since it is impossible for anyone to have learned the meaning of every composite expression in the way one learns the meaning of lexical units, there must be some function which determines the meaning of composite expressions on the basis of the meaning of lexical units. The implication of the principle for the current work is that it allows constructing meaning representations of clauses from the meaning of lexical units it contains and the syntactic structures that they are involved in.

Although the principle of semantic compositionality is not committed to a specific semantic theory, it is perhaps most closely associated with truth-conditional semantics. However, we must note that the principle is not universally accepted or used. In fact, in natural language processing research, most of the semantics-oriented work disregards it completely, instead focusing on shallow techniques, in which compositionality plays little role. On the other hand, our core notion, *embedding*, has structural emphasis, and we aim to model the meaning shifts engendered at different semantic levels of embedding structure. Therefore, a compositional approach seems most natural for our purposes: the principle of semantic compositionality is the underlying semantic assumption of the current work.

2.2 Functional Aspects of Embedding

In this section, we discuss the research that focuses on semantic/pragmatic functions associated with embedding. We approach the discussion from two opposing perspectives. One is a bottom-up perspective, in which such functions are viewed as extensions of propositional meaning. We have referred to this perspective as the bottom-up viewpoint of the embedding layer earlier and our discussion centers around the linguistic categories of *modality* and *negation* and related concepts. The other is a top-down perspective, in which the embedding layer is considered within the context of discourse coherence and high level pragmatic functions. In addition to linguistic discussion of these functions, we also present a survey of the research on these topics in computational linguistics and biomedical NLP in this section.

2.2.1 Modality and Negation

Adopting a bottom-up viewpoint to the embedding layer, we isolate modality and negation as the two main micro-level linguistic categories that play a role and we discuss these categories here in detail.

Modality

Modality is defined as "the grammaticalized expression of subjective attitudes and opinions of the speaker, including possibility, necessity, obligation, permissibility, ability, desire, and contingency" (Bybee *et al.*, 1994). Within traditional model-theoretic, truth-conditional approaches to semantics, modality is studied with the principles of *modal logic* and the focus is often on modal auxiliaries, most notably *may* and *must*. In fact, modal logic is specifically defined as the "logic of *necessity* and *possibility*, of *must be* and *may be*" (Hughes and Cresswell, 1996). Whereas classical logic is concerned with truth and falsity in the actual world, modal logic extends these notions to *possible worlds*. In this framework, a proposition is identified with the set of worlds in which it is true and modals are conceived as generalized quantifiers over possible worlds: *possibility* corresponds to existential quantification over possible worlds, and *necessity* to universal quantification (Kratzer, 1981). More specifically, a proposition P is *necessary* in a world w if it is true in *all* worlds which are possible relative to w (corresponding to *accessible worlds*) and this situation is denoted as $\Box P$. On the other hand, P is *possible* in a world w if it is true in *at least one* world possible relative to w. This situation is denoted as $\Diamond P$. Due to its focus on modalities of necessity and possibility (subsumed under *alethic modality*), modal logic is generally considered inadequate in addressing the often epistemic nature of natural language modality (Lyons, 1977).

From a functional-typological perspective, modality is considered to be concerned with the *status* of or the *speaker's attitude* towards a proposition, rather than its truth or falsity. There are a number of typological investigations of linguistic modality (Coates, 1983; Palmer, 1986; Auwera and Plungian, 1998; Palmer, 2001; Hengeveld, 2004); however, there is no agreement on precise modal semantic subtypes[2]. Here, we mainly focus on the cross-lingual work of Palmer (1986, 2001), perhaps the most comprehensive account of modality from the functional-typological perspective. He uses the notions of *realis* versus *irrealis* to distinguish between "situations as actualized versus situations as purely within the realm of thought, knowable only through imagination" to explain modality, similar to the actual world/possible worlds distinction made in truth-conditional semantics. As in logic-based approaches, his work also focuses mainly on modal auxiliaries. His coarse-grained modality categorization includes two classes: *propositional modality* and *event modality*, the former referring to the speaker's judgement of the proposition (6a), the latter to the speaker's attitude towards a potential event (6b). The modal expressions are paraphrased in parentheses.

(6) (a) Kate <u>may</u> be at home now. (paraphrased as "It is possible that Kate is at home now.")

(b) Kate <u>may</u> come in now. (paraphrased as "It is possible for Kate to come in now.")

Propositional modality is categorized into *epistemic* and *evidential* types: with *epistemic modality*, speakers express their judgment about the factual status of the proposition, whereas with *evidential modality*, they indicate the nature of evidence they have for its status. Palmer subclassifies epistemic modality further into *Speculative*, *Deductive* and *Assumptive* categories, while the evidential type is subcategorized into *Reported* and *Sensory*. However, he concedes that these distinctions are sometimes blurred, as in the case of *Deductive*, which usually suggests that the judgment was based on evidence, even though the evidence may not be explicit. He, therefore, analyzes this type as being both *epistemic* and *evidential*. Examples for these propositional modality categories, taken from (Palmer, 2001), are given below:

(7) (a) **Speculative:** John <u>may</u> be in his office.

(b) **Deductive:** John <u>must</u> be in his office.

(c) **Assumptive:** John <u>will</u> be in his office.

(d) **Reported:** He is <u>said</u> to be extremely rich.

(e) **Sensory:** I just <u>saw</u> him pack.

Event modality is classified into *deontic* and *dynamic* types. *Deontic modality* is concerned with obligation, permission and promise, and is subcategorized into *Obligative*, *Permissive*, and *Commissive* types, respectively. On the other hand, *dynamic modality* relates to ability or volition,

[2] For a comprehensive comparison of some of these studies, we refer the reader to Nauze (2008).

and its subtypes are, accordingly, *Potential* and *Volitive*. Palmer (2001) gives the following examples, the first concerning permission (*Permissive* subtype of *deontic modality*) and the second ability (*Potential* subtype of *dynamic modality*).

(8) (a) **Deontic:** You may go now.

(b) **Dynamic:** He can run a mile in under four minutes.

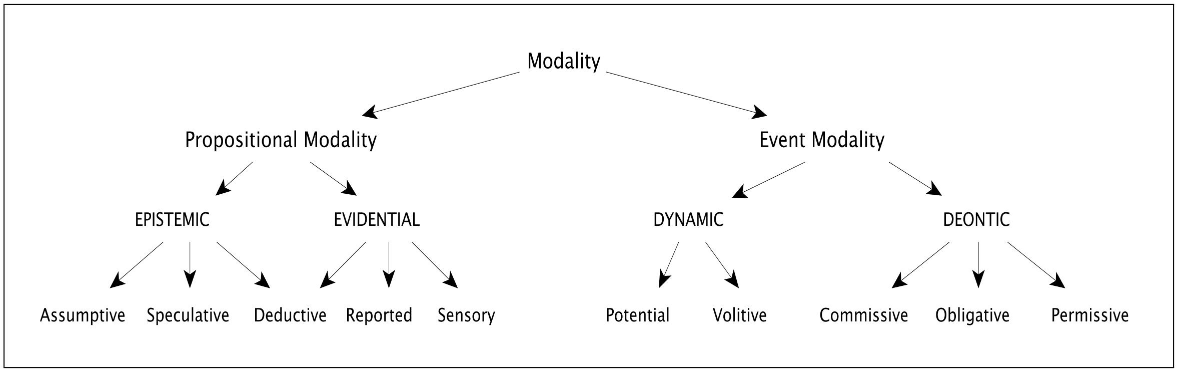

Figure 2.2: Modality categorization of Palmer (2001)

Deontic modality is sometimes referred to as *participant-external modality*, and dynamic modality as *participant-internal modality* (Auwera and Plungian, 1998), reflecting the role of the participant with respect to the proposition. Palmer's categorization of modality types is presented in Figure (2.2).

It is important to note that while Palmer's categorization and most others emphasize the role of the speaker in modal contexts, it is not necessarily the speaker that expresses his/her attitude, as can be seen in reported speech or quotation contexts, an example of which is given in Example (9a). Furthermore, these typological studies, including Palmer's, largely ignore the combination of modals, which occurs often in natural language. Example (9b) illustrates such a sentence from Nauze (2008), in which a combination of *Speculative* and *Obligative* modalities occur.

(9) (a) She said that he might be there.

(b) John may have to pay more taxes.

As stated earlier, natural language modality is mostly epistemic in nature; thus, it is not surprising that this type of modality is also the most widely studied. Epistemic modality is concerned with *factuality* of a proposition. One point of debate regarding epistemic modality concerns the nature of the *epistemic scale* associated with factuality degrees: whether it can be defined in terms of discrete factual values, or whether it represents a continuum. Often, a three-valued discrete system is adopted by linguists: *certainly*, *probably*, and *possibly* (Lyons, 1977; Palmer, 2001). On the other hand, de Haan (1997) analyzes modality as a continuum to present an adequate, cross-lingual account.

There is also disagreement regarding the status of evidentiality in the modal system. While Palmer (1986, 2001), among others, adopts the view that *evidentiality* is subsumed by modality (corresponding to *evidential* category in Palmer's classification), there is also recent research that considers evidentiality as a separate linguistic system. In her cross-linguistic study of evidentials, Aikhenvald (2004) argues that, although evidentials may acquire secondary epistemic meanings, the association between evidentiality and modality is not cross-linguistically universal. A similar cross-lingual position is taken by de Haan (1999). From a pragmatics-oriented perspective, Chafe (1986) argues that evidentiality subsumes modality and discusses empirical findings for eight classes of evidentiality: *degree of reliability*, *belief*, *induction*, *deduction*, *hedge*, *hearsay evidence*, *sensory evidence* and *expectation*. Some of these classes (e.g., *deduction*, *sensory evidence*) clearly overlap with Palmer's categories, while the status of others (e.g., *expectation*, *hedge*) is not so clear.

In addition to positioning evidentiality outside (but related to) the modal system, de Haan (1999) also proposes an evidential scale (shown below), where the Visual category represents the higher end of the evidential spectrum and the Reportative category the lower end[3]. He argues that the use of a lower evidential implies that the speaker does not have the kind of source of information that would allow him/her to use a higher evidential, making the utterance less reliable or less believable[4].

(10) Visual > Auditory > Nonvisual > Inference > Reportative

Beliefs, desires, expectations, opinions, intentions are functions realized by linguistic modality and are sometimes collectively referred to as *propositional attitudes*. They describe a cognizer's attitude towards a proposition and are often introduced by verbs of cognition, such as *believe*, *think*, *wish*. Such verbs are said to introduce an intensional context for their embedded complement. The main distinguishing feature of propositional attitudes is the explicit expression of the cognizer (*John*), as shown in Example (11) (Dowty *et al.*, 1981).

(11) John <u>believes</u> that Miss America is bald.

Closely related to epistemic modality and evidentiality, *hedging* is a term often used in scientific discourse analysis from a pragmatic viewpoint. The term was introduced by Lakoff (1972), who was mainly concerned with how words and phrases, such as *mainly* and *rather*, make sentences fuzzier or

[3] In his categorization, Visual and Auditory types roughly correspond to Palmer's Sensory category, while Inference and Reportative correspond to Deductive and Reported categories, respectively. It is worth noting that he calls Visual, Auditory , and Nonvisual categories as *direct evidence*, while the other two categories are called *indirect evidence*.

[4] It can thus be said that evidentials indicate *scalar predications* on the scale of reliability.

less fuzzy. In his comprehensive account of hedging in scientific literature, Hyland (1998) described hedges as "weakening the force of a statement, expressing deference to the reader or signaling uncertainty". He proposed a fuzzy model in which he categorizes scientific hedges by their pragmatic purpose: *content-oriented*, *writer-oriented* and *reader-oriented* hedges. Hyland's categorization is duplicated in Figure (2.3), and examples concerning these categories, taken from Hyland (1998), are given below.

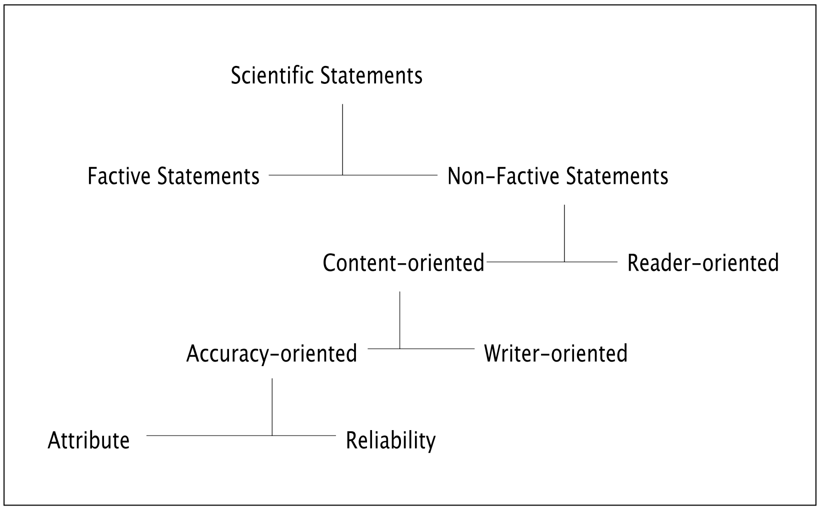

Figure 2.3: Categorization of scientific hedges (Hyland, 1998)

(12) (a) **Attribute:** *This shift could be* <u>*partially*</u> *caused by solvent ...*

(b) **Reliability:** *This modification could* <u>*possibly*</u> *play a role in substrate binding ...*

(c) **Writer-oriented:** <u>*These data indicate that*</u> *phytochrome A possesses the intrinsic ...*

(d) **Reader-oriented** <u>*I believe*</u> *that the major organisational principle of thylakoids is that of continuous unstacking ...*

In summary, *attribute hedges* "specify the extent to which a term accurately describes its reported phenomena", whereas *reliability hedges* are concerned with "conveying the writer's assessment of the certainty of the truth of a proposition". On the other hand, *writer-oriented hedges* allow the writer to avoid personal responsibility for propositional truth, whereas *reader-oriented hedges* help to soften claims. It is important to note that the hedging devices he describes (*epistemic verbs* such as *speculate, believe, suggest*, adverbs such as *possibly, presumably*) may be considered epistemic modality or evidentiality markers by others[5].

[5] In Chafe's view, for example, *presumably* would mark Deduction, not Hedging, even though he identifies a distinct Hedging category.

Overlapping with *hedging* to some extent, *vagueness* is a term used to describe the communicative task of conveying indefinite and imprecise information (Channell, 1994). While it is often assumed that clarity and precision are indicators of good language usage, vagueness is a core aspect of human language. Linguistic use of the term generally covers a wider range of genres than hedging does, and colloquial vagueness markers, such as *like*, *stuff*, are studied more commonly.

Negation

Negation (*negative polarity*) is a core feature of human communication (Horn, 1989). In propositional logic, negation is simply a one-place connective (denoted as $\neg$) that reverses the truth value of a proposition ($\neg P$ is *true* when P is *false*). However, it bears a much wider range of forms, functions and meanings in natural language, as hinted at earlier in Section 2.1.2.

In his seminal book, Horn (1989) investigated negation from a variety of perspectives, incorporating philosophical, psychological, logical, and semantic insights, and proposed a pragmatic framework for negation incorporating the notions of *conversational implicature* and *presupposition*. He made the bipartite distinction of *descriptive negation* versus *metalinguistic negation*. Descriptive negation refers to the negation of the conceptual content of a proposition, in other words, its truth-conditions. On the other hand, metalinguistic negation is defined as "a device for objecting to a previous utterance on any ground whatever" (Horn, 1989) and refers to the effect that negation has on proposition's pragmatic properties, presupposition and implicature. Metalinguistic negation is said to have presupposition- and implicature-canceling properties. In Example (13a), the underlined *not* is not truth-conditional, but is presupposition-canceling, instead (the presupposition being that "There is a king of France."). In Example (13b), *not* is said to be implicature-canceling (the implicature being that "Only some men are chauvinists.").

(13) (a) The king of France is <u>not</u> bald. There is no king of France.

(b) Some men are <u>not</u> chauvinists. All are chauvinists.

We do not describe Horn's extensive framework in more detail here, but we content ourselves with summarizing several interesting features of natural language negation relevant to our work, discussed in more detail in Quirk *et al.* (1985), Huddleston and Pullum (2002), Horn (1989), among others.

With respect to form, negation can be expressed with a variety of lexical categories. Certain words (e.g., *not*, *no*, *nothing*, *neither*, etc.) are *explicitly negative*. On the other hand, the cases of *inherent negation* (14a) and *affixal negation* (14b) are generally considered *implicit negatives*.

(14) (a) She <u>failed</u> to follow the rules.

(b) John is unhappy.

We exemplified the complexity associated with negation scope earlier in Section 2.1.2. Certain predicates (e.g., *believe, think, suppose*, etc.), when negated, allow narrowing of scope to their complement clause. This phenomenon, exemplified in (15a), is called *transferred negation* (Quirk *et al.*, 1985) or *negative raising* (Horn, 1989). These verbs are characterized as having medium strength on a scale of subjective certainty (Horn, 1989). Interestingly, a verb that is stronger (*claim*), as in Example (15b)) and one that is weaker (*hope*), as in Example (15c)) on that scale do not transfer negation.

(15) (a) I do not think [he came].

(b) I do [not claim that he came].

(c) I do [not hope that he came].

Negation can express "less than" or "in between" meaning when used with scalar predicates. The sentence in Example (16a), which can be read as a case of metalinguistic negation, implies that John probably has one or two children (*scalar implicature*). A similar "in between" meaning is achieved via *multiple negation*, where more than one negative appears together in a clause. In logic, two negatives cancel out each other and are equivalent to an affirmative. However, in natural language, this is often not the case. The sentence in Example (16b) is not simply equivalent to the affirmative "You are welcome" as a "not unwelcome" person can be "welcome" or s/he can be situated in the nonexcluded middle between the two contrarily opposed terms (Horn, 1989).

(16) (a) John does not have three children.

(b) You are not unwelcome.

(c) You are not welcome.

Interaction of Epistemic Modality and Negation

Earlier, we mentioned the two approaches to representing the epistemic scale: discrete values vs. the continuum. Horn (1989) presented epistemic modality as an instantiation of scalar predication, and proposed the epistemic scale of *certain, probable/likely*, and *possible.* An interesting outcome of his focus on negation is negative epistemic modality scale: *impossible*, *unlikely*, *uncertain*, which emerges from his use of the Aristotelian Square of Opposition to account for the interaction of negation with epistemic modals, illustrated in Figure (2.4).

The horizontal axis represents positive and negative polarity, while the vertical axis represents the three-valued epistemic scale (*certain, likely, possible*). The logical relations between the nodes is explained in terms of two basic logical laws:

1. **Law of Contradiction (LC):** a statement cannot be both true and false at the same time.

2. **Law of Excluded Middle (LEM):** a statement must be either true or false.

Contradictory propositions satisfy both LC and LEM (for example, *certain(P)* and *not_certain(P)* are contradictions). On the other hand, *contrary* propositions only satisfy LC, and *subcontrary* propositions only LEM.

Computational Approaches to Modality and Negation

In computational linguistics, interest in modality, negation and related phenomena is relatively recent. A recent workshop (Negation and Speculation in NLP (NeSp'2010) (Morante and Sporleder, 2010) and the forthcoming Special Issue on Modality and Negation in the Computational Linguistics journal can be seen as an outcome of this trend. The interest in these phenomena in the natural language processing community is generally focused on more pragmatically defined categories, such as *factuality* or *subjectivity*. While linguistic analysis may form the basis of description and corpus annotation, semantic interpretation approaches are largely based on statistical learning methods, rather than being linguistically-motivated. In this section, we first describe approaches with a clear linguistic bent and then focus on approaches whose objectives are more pragmatic.

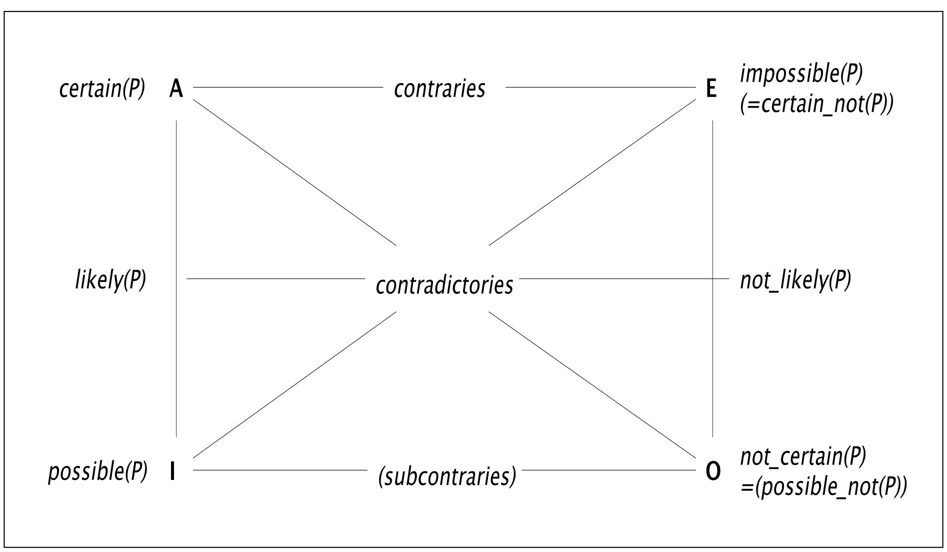

Figure 2.4: Aristotelian Square of Opposition adapted to epistemic scale (Horn, 1989; Saurí, 2008)

Modality and Negation in Ontological Semantics In their ontological semantics framework, Nirenburg and Raskin (2004) limit the category of modality to attitudinal meanings, and define modality as having the following components: *type*, *attributed to*, *scope*, *value*, and *time*. In their linguistically-motivated view, modality can scope over entire propositions, proposition heads, as well as concept instances and instances of properties. Their modality categorization is similar to that of Palmer (2001), but also differs in certain respects. They distinguish *epistemic*, *epiteuctic*, *deontic*, *volitive*, *potential*, *evaluative*, and *saliency* modality types. While *epistemic*, *deontic*, *volitive*, and *potential* clearly correspond to Palmer's categories, the types of *epiteuctic*, *evaluative*, and *saliency* do not. The *epiteuctic* type refers to the "degree of success in attaining the results of the event in its scope" and its values range from "complete failure with no effort expended as in "They never bothered to register to vote." to near success in "He almost broke the world record in pole vaulting." to complete success in "They reached the North pole." The *evaluative* type expresses attitudes, going from the worst (value of 0) to best (value of 1). The markers for this modality include verbs such as *like*, *criticize*, and *hate*. Finally, the *saliency* modality type refers to the "importance that the speaker attaches to a component of text meaning". This type of modality rarely has scope over entire propositions, and is often marked by adjectives, such as *unimportant*. In the ontological semantics framework, this type of modality is said to mark given/new information distinction. An interesting feature of this framework is that negation is subsumed in the category of *epistemic* modality and as being its extreme case (value of 0.0). While their linguistically based treatment is comprehensive, it is not clear how well it performs in practice, as they do not present an evaluation.

Event Factuality/Certainty Event factuality has been considered within the context of TimeML (Pustejovsky *et al.*, 2005b), a specification language designed for representing temporal and event information in text. An early study in this context (Saurí *et al.*, 2006a) was concerned with *event modality*. In this work, they associated modality values with events at the lexical and syntactic level by means of subordination links (SLINKs). SLINK relations can be one of the following types: FACTIVE, COUNTERFACTIVE, EVIDENTIAL, NEGATIVE_EVIDENTIAL, MODAL and CONDITIONAL. For example, a SLINK of type MODAL is created between *investigating* and the event indicated by the verb *participated* in the following sentence:

(17) Officials are <u>investigating</u> whether Rudolph **participated** in all three attacks.

They reported an F1-measure of 70% in identification of subordination links. In later work, they considered the interaction of modality and negation in the context of *event factuality* (Saurí, 2008), which can be seen as a pragmatic category. It is defined as "the level of information expressing the commitment of relevant sources towards the factual nature of events in text", that is, whether events are presented as *facts*, as *counterfacts*, or as *possibilities* according to a particular source. The examples below, from news articles, are taken from Saurí (2008). The events in focus are in bold, and the markers that affect the factuality value are underlined.

(18) (a) **Fact:** Jubilant Red Sox fans **cheered** for players at Fenway Park yesterday.

(b) **Counterfact:** The size of the contingent was <u>not</u> **disclosed**.

(c) **Possibility:** United States <u>may</u> **extend** its naval quarantine to Jordan's Red Sea port of Aqaba.

Saurí proposed a *factuality profiler* grounded on lexical and syntactic expressions of factuality. Drawing from Horn (1989), she modeled *factuality* as an interaction between the parameters of *epistemic modality* and *polarity*. Epistemic modality in this framework expresses the degree of certainty with regard to an event and has one of four possible values: CT (certain), PR (probable), PS (possible), and U (underspecified). On the other hand, polarity conveys whether the event is taking place in the world according to the information source and may have one of three values: + (positive), - (negative), and u (underspecified). From the interaction of these categories, the factuality values shown in Table (2.1) are obtained. A *factuality profile* is modeled as consisting

	Positive (+)	**Negative (-)**	**Underspecified (u)**
Certain (CT)	Fact (CT+)	Counterfact (CT-)	Certain but unknown output (CTU)
Probable (PR)	Probable (PR+)	Not probable (PR-)	*NA*
Possible (PS)	Possible (PS+)	Not certain (PS-)	*NA*
Underspecified (U)	*NA*	*NA*	Unknown or uncommitted (UU)

Table 2.1: Factuality values as interaction of epistemic modality and polarity (Saurí, 2008)

of four components: (a) the *event* in focus, (b) the *factuality value* assigned to the event, as given Table (2.1), (c) the *source* assigning the factuality value, and (d) the *time* of the assignment. She used a top-down algorithm based on a dependency tree and an extensive lexicon to compute factuality profiles for events mentioned in text. The system was evaluated on a corpus annotated for event factuality, called FactBank (Saurí and Pustejovsky, 2009), best result obtained for the CT+ class and the worst for the more challenging PR+ and PS+ classes.

Modality Tagging Baker *et al.* (2010) described the construction of a modality annotation scheme, a modality lexicon, and modality taggers. They defined modality as consisting of three components: a *trigger*, a *target*, and a *holder*. They defined eight types of modalities: *belief*, *requirement*, *permissive*, *intention*, *effort*, *ability*, *success*, and *want*. The resemblance of some of these types to categories described earlier is clear. An interesting aspect of their work was that they considered the interaction of modality with negation. For annotation, these interactions were categorized into fifteen discrete categories, noting the transferred negation of *believe*, for example, and the entailment relations between certain classes (*not require P to be true* entails *permit P to be false*). The resulting discrete categories are as follows: Require, Permit, NotPermit, Succeed, NotSucceed, Effort, NotEffort, Intend, Able, NotAble, Want, NotWant, FirmBelief, and Belief. They reported a modality lexicon and several modality tagging experiments using string-based and structure-based taggers. They applied these taggers to the task of improving machine translation quality.

Modality and Negation in Semantic Role Labeling We briefly described semantic role labeling and annotation of semantic roles in the first chapter and mentioned that two semantic roles ArgM-MOD and ArgM-NEG (modality and negation arguments, respectively) were annotated in the PropBank corpus (Palmer *et al.*, 2005). NomBank (Meyers *et al.*, 2004b) adopts PropBank-style frames for argument-taking nouns. In their annotation, only ArgM-NEG is annotated explicitly. PropBank-style annotations with respect to modality and negation can be considered shallow: while these semantic roles are annotated on syntactic structures, their scope and semantic/pragmatic consequences (factuality, certainty, etc.) are ignored. Semantic role labeling systems trained and evaluated on PropBank generally perform well in identifying ArgM-MOD and ArgM-NEG arguments, since these arguments often simply precede the verbal predicate in the sentence. In the CoNLL 2005 Semantic Role Labeling task (Carreras and Màrquez, 2005), the simple baseline of tagging the modal auxiliary in verbal chunks with ArgM-MOD role yields an already high F_1-score of 88.71 and tagging *n't* and *not* in the target verbal chunk with ArgM-NEG role yields an F_1-score of 91.84. The best systems in the task obtain F_1-scores of 98.47 (Haghighi *et al.*, 2005) and of 98.91 (Màrquez *et al.*, 2005) in tagging these semantic roles, respectively.

Belief and Reported Speech Tagging Prabhakaran *et al.* (2010) focused on *belief tagging* (propositional attitudes) and aimed to determine the propositions that the author believes. They annotated a small corpus, where each head of each proposition was tagged with one of three values expressing the belief level associated with the proposition: (a) *Committed Belief*: the author indicates that s/he believes the proposition, (b) *Non-committed Belief*: the author identifies the proposition as something s/he could believe, but s/he happens not to have a strong belief in, and (c) *Non-applicable*: the proposition is not of a type in which the author is expressing or could express a belief. Using various supervised learning models, they identify committed belief, reporting the best results with an SVM joint model using lexical and syntactic features.

Similar to belief tagging in some respect, *reported speech tagging* has also attracted some attention. One of the earlier, theoretical studies in this area (Bergler, 1992) presented an evidential analysis approach to reported speech in news articles, acknowledging that the embedded clause in reported speech contains the primary information, whereas the matrix clause (*source* and *reporting verb*) provides the evaluative environment (i.e., evidence) for the primary information, affecting its reliability. A lexical semantics for reporting verbs was also developed. This work also formed the basis of an *artificial believer* system (Krestel *et al.*, 2007), which extracted beliefs from reported speech in newspaper articles. Their system marks and enriches reported speech structures in text and then provides these enriched structures to a fuzzy believer, which computes beliefs using fuzzy operations. The system was later applied to the textual entailment task (Krestel *et al.*, 2008, 2009).

Modality and Negation in Sentiment Analysis

The core notion in sentiment analysis is that of *subjectivity* (or *private states*) (Wiebe *et al.*, 2004), defined as the "aspects of language used to express opinions, beliefs, evaluations, and speculations", notions central to discussions of modality and propositional attitudes.

Wiebe *et al.* (2005) developed an annotation scheme and an opinion corpus (MPQA), based on the notion of private states. Two types of private state frames were distinguished: *expressive subjective element frames* and *direct subjective frames*. For factual, non-opinion text, *objective speech event frames* were used. Each frame type has a number of features, including *source* (the person or entity expressing the speech event), *target* (topic of the speech event), and other properties, such as *intensity* (of the sentiment) or *polarity* that only appear in private state frames. Some of these features clearly overlap with those in models of factuality, certainty, and propositional attitudes, discussed earlier. The MPQA corpus has spurred the interest in sentiment analysis research, which we will not discuss here at any length[6].

We will confine ourselves to a few studies that focus on identifying sentiment at the phrase

[6] For a relatively recent survey of the field, see the survey book Pang and Lee (2008).

level that have some modal or negative polarity component. Bethard *et al.* (2006) focused on finding *propositional opinions*, sentential complement clauses of verbs such as *believe* and *claim* that express opinions, and the holders of these opinions, which bears similarities to belief and reported speech tagging. Their approach was based on semantic parsing and learning opinion words by bootstrapping. They concentrated on verbs and extracted verb-specific information from semantic frames such as those that are defined in FrameNet (Baker *et al.*, 1998) and PropBank (Palmer *et al.*, 2005). Kim and Hovy (2006) defined a judgement opinion as consisting of a *valence* (judgement value: positive, negative or neutral), a *holder* and a *topic*. They used structural features from a syntactic parse tree as the basis for a machine learning approach to model the long-distance, structural relation between an opinion holder and an expression.

Negative polarity items (their lexicons, recognition, and effect on sentiment) have also attracted great attention in the field of sentiment analysis. Wilson *et al.* (2005) reported a subjectivity lexicon, where lexical items are assigned polarity values. This lexicon has been widely used in sentiment analysis research. Polanyi and Zaenen (2006) accounted for negative polarity items in terms of *contextual valence shifting*. Negative and positive polarity items were assigned negative and positive scores, respectively, and if a polar expression was negated, its polarity score was inverted (e.g., *successful (+2)* $\rightarrow$ *not successful (-2)*. *Diminishers* (similar to hedges) and *intensifiers* are also taken into account; however, the score is reduced or increased rather than inverted (e.g., *suspicious (-2)* $\rightarrow$ *deeply suspicious (-3)*). Similar approaches have been explored by Kennedy and Inkpen (2006) and Andreevskaia *et al.* (2007), as well. Within the same framework, Taboada *et al.* (2011) assign words with positive and negative polarity values in [-5,+5] range to reflect their sentiment strength. Additionally, intensifiers and diminishers are modelled as percentages over polarity values. For example, *sleazy* has the prior polarity of -3 and *somewhat* is a diminisher with the value -30%. Accordingly, the phrase *somewhat sleazy* has the polarity value of of: -2.1 (-3 *(100% -30%)). On the other hand, negation scoping over a polarity item is modelled as a shift of the polarity value towards the opposite polarity by a fixed amount (taken to be 4). For example, *not sleazy* has the value 1 (-3+4). Another recent thread of research in sentiment analysis, *sentiment composition*, takes the basic idea of valence shifters further. In this approach, the overall sentiment of a phrase is derived compositionally from the sentiment classification of lexical and grammatical constituents and polarity shifts due to negative polarity items play an important role (Moilanen and Pulman, 2007; Choi and Cardie, 2008). For a comprehensive survey on the role of negation in sentiment analysis research, we refer the reader to Wiegand *et al.* (2010).

Vagueness While computational approaches to *hedging* have almost exclusively focused on scientific articles (as we will discuss in the next section), there have been recent studies focusing on *vagueness* expressed in Wikipedia articles. The language in Wikipedia articles is expected to be clear and precise. Vagueness in this context is indicated by hedging and *weasel words*, an undesirable feature according to Wikipedia policy (Ganter and Strube, 2009; Ganter, 2010). Ganter (2010) distinguished between hedges and weasel words, the latter of which she considers as being an evasive form of hedging, and proposed n-gram and POS-based features for weasel word detection. It is important to note that, while hedging is undesired in the descriptive domain of Wikipedia articles, it is "not a strategy to obfuscate or confuse, any more than it is simply a convention of academic style" (Hyland, 1998) in the context of scientific articles.

Modality and Negation in Biomedical Text Processing Recent years have seen much activity regarding modality and negation in biomedical text processing. In fact, the majority of articles that appeared in the NeSp'2010 proceedings (Morante and Sporleder, 2010) concerned the biomedical domain. This activity is due to two factors: (a) automatic processing of clinical reports is a major concern in healthcare informatics, where it is important to recognize whether findings in these reports are asserted, negated or uncertain (Uzuner, 2009) (b) distinguishing facts from speculative or uncertain statements and determining the evidence the author has for a scientific claim is important for biomedical text mining (Light *et al.*, 2004; Wilbur *et al.*, 2006).

In clinical natural language processing, the first work concerning uncertainty was that of Friedman *et al.* (1994), who discussed uncertainty in radiology reports. Their natural language processing system assigns one of five levels of certainty to extracted findings. In the clinical domain, research on negation often focuses on identifying the concepts that are negated, such as diseases or symptoms. Rule-based systems relying on lexical and syntactic information (Mutalik *et al.*, 2001; Chapman *et al.*, 2001; Harkema *et al.*, 2009) as well as supervised learning techniques (Goldin and Chapman, 2003; Averbuch *et al.*, 2004) have been explored in negation detection. The recent i2b2 (Informatics for Integrating Biology to the Bedside) challenge competitions (Uzuner, 2009) have also provided a platform for evaluating systems that extract negated and uncertain statements.

In the context of biomedical articles, the problem of uncertainty and hedging was first posed as a text classification task, where detection of uncertainty amounted to identifying speculative sentences (Light *et al.*, 2004). Some studies (Light *et al.*, 2004; Medlock and Briscoe, 2007; Szarvas, 2008) used machine learning techniques with variants of the "bag-of-words" approach (Light *et al.*, 2004) or weakly supervised techniques (Medlock and Briscoe, 2007; Szarvas, 2008) for this task.

Szarvas (2008) also extended his approach to the clinical domain (radiology reports) and reported relatively poor results on a corpus of biomedical articles from a different source concluding that the portability of hedge classifiers is limited. In a previous study, we approached the classification task from a more linguistically-oriented perspective. We demonstrated a less domain-dependent approach using a hedging dictionary and lexico-syntactic patterns (Kilicoglu and Bergler, 2008).

More recent research that considers modality- and negation-related phenomena in biomedical text processing can be divided into two categories: corpus annotation efforts and scope resolution tasks. With respect to corpora, one of the earliest was the corpus annotation scheme presented by Wilbur *et al.* (2006), in which they propose five qualitative dimensions to characterize scientific sentence fragments: three of these dimensions, *certainty*, *evidence*, and *polarity*, are clearly related to modality and negation. The *certainty* and *evidence* dimensions are 4-valued: in the certainty dimension, the value of 0 indicates complete uncertainty, while the value of 3 represents complete certainty; in the evidence dimension, the value of E0 indicates "no explicit evidence" or "explicit expression of lack of evidence" and E3 explicit evidence (e.g., *Our data demonstrates ...*). Polarity may have positive (P) or negative (N) value. Another dimension, *trend/direction*, indicates whether the text fragment reports an increase/decrease, high/low level in a specific phenomenon, and can be seen as analogous to some valence shifters (diminishers, intensifiers), discussed earlier. The other dimension, *focus*, indicates whether the fragment has scientific content (S), is generic (G) or methodology-based (M). An example annotation from the corpus is given below. For the first fragment, the annotation (1SP3E3+) indicates that the fragment has scientific content, with positive polarity, highest certainty and evidence levels and that it indicates a positive trend.

(19) [We show that treatment with ICG-001 induces apoptosis in colon carcinoma cells,]**1SP3E3+**
[but not in normal colonic epithelial cells.] **2SN3E0**

Despite being intended as a training and evaluation platform for machine learning approaches, this corpus has attracted little attention for these purposes. Shatkay *et al.* (2008) trained and evaluated various SVM-based and Maximum Entropy-based classifiers along these dimensions, reporting best results along the *polarity* dimension and the poorest results on the *trend* dimension.

The main focus of the GENIA event corpus (Kim *et al.*, 2008) is annotating biological events[7]. In their annotation, events are associated with *uncertainty* and *assertion* attributes, linked to modality and negation, respectively. The *uncertainty* feature has one of three values (*certain*, *probable*, and *doubtful*), while the *assertion* feature has two (*exist* or *non-exist*, the latter corresponding roughly

[7] The closest counterpart in the news article domain would be the PropBank corpus.

to negation).

A corpus annotation effort that specifically focused on modality and negation resulted in the BioScope corpus (Vincze *et al.*, 2008), which consists of medical and biological texts specifically annotated for negation and uncertainty together with their linguistic scope. Two annotation instances from the corpus are given below, where the negation or uncertainty cues are underlined and the textual span indicating the scope is in square brackets.

(20) (a) Stable appearance the right kidney [<u>without</u> hydronephrosis].

(b) This result [<u>suggests</u> that the valency of Bi in the material is smaller than +3].

While the annotations are qualitatively similar to those in the GENIA event corpus, the scope is modeled as a textual span that includes the cue in the BioScope corpus. In contrast, in the GENIA event corpus, uncertainty and assertion are features associated with biological events, which are essentially abstract semantic objects.

Several corpus-relaed studies from University of Manchester (Thompson *et al.*, 2008; Nawaz *et al.*, 2010; Thompson *et al.*, 2011) explored ideas similar to those of Wilbur *et al.* (2006), culminating in the *meta-knowledge* model, which subsumes negation and modality, as well as pragmatic intent. Meta-knowledge annotations are applied on top of the event annotations in the GENIA event corpus (Kim *et al.*, 2008). A meta-knowledge annotation consists of 5 basic elements:

1. **Knowledge Type** (KT) captures the general information content (Investigation, Observation, Analysis, Method, Fact, Other).
2. **Certainty Level** (CL) identifies the level of certainty associated with the event (three values: L1, L2, and L3, L1 indicating considerable speculation, L3 indicating certainty).
3. **Polarity** indicates negative or positive polarity.
4. **Manner** indicates rate, level, intensity, or strength of the event (High, Low, Neutral).
5. **Source** indicates the source of the knowledge asserted by the event (Current, Other).

These dimensions are almost identical to those proposed by Wilbur *et al.* (2006), described earlier. However, the meta-knowledge dimensions are applied to events, rather than sentence segments, making them more semantically precise. Another defining feature of the meta-knowledge model is that it incorporates -*dimensions* on top of these basic dimensions. Hyper-dimensions essentially correspond to drawing inferences from combinations of the basic dimensions and can be viewing as corresponding to the pragmatic level. Two hyper-dimensions are defined:

1. **New Knowledge** indicates whether the event in focus describes new knowledge (Yes, No). They infer this hyper-dimension from the values associated with KT, CL, and Source dimensions.

2. **Hypothesis** is determined from KT and CL values (Yes, No).

Source	**Knowledge Type**	**Certainty Level**	**Inference**
Current	Observation	L3	New Knowledge
Current	Analysis	L3	New Knowledge
X	Analysis	L2	Hypothesis
X	Analysis	L1	Hypothesis
X	Investigation	X	Hypothesis

Table 2.2: Inferences for meta-knowledge hyper-dimensions (Thompson *et al.*, 2011)

The inferences presented in Table (2.2) indicate the cases where the event is considered new knowledge or a hypothesis. The meta-knowledge annotated corpus is quite recent, and at the time of this writing, we are not aware of any research that focuses on recognizing these dimensions.

With the availability of corpora annotated with modality- and negation-related phenomena, there has also been research that focuses on interpreting such phenomena. The major tasks in biomedical NLP have been detection of speculation and negation and resolution of their scope. The relevant GENIA annotations formed the basis for the speculation and negation detection subtasks in the two shared task competitions on biological event extraction (BioNLP'09 (Kim *et al.*, 2009) and BioNLP-ST'11 (Kim *et al.*, 2011a)). In these competitions, speculation and negation detection is defined as correctly identifying the speculation and negation instances and the events that they scope over. In the BioNLP'09 competition, we obtained the best results among six participating systems by extending our earlier sentence-level speculation detection system (Kilicoglu and Bergler, 2008) to recognize precise speculation scope as well and applying the same principles to recognition of negation and its scope (Kilicoglu and Bergler, 2009, 2011b).

There have been a number of studies on finding negation and speculation scope based on the BioScope corpus using a variety of statistical learning methods (Morante *et al.*, 2008; Morante and Daelemans, 2009; Özgür and Radev, 2009; Agarwal and Yu, 2010). The CoNLL'10 Shared Task on Hedge Detection (Farkas *et al.*, 2010), based on the BioScope corpus, included two subtasks focusing on hedging in biomedical text: (a) detecting uncertain sentences in biological text (b) detecting the scope of hedging cues in these uncertain sentences. In the first subtask, Tang *et al.* (2010) achieved a F1-score of 86.36% using a cascade of conditional random field and large margin-based models. In the hedging scope resolution task, Morante *et al.* (2010) used a memory-based learning system that relies on syntactic dependencies to obtain the best results in this task (F1-measure of 57.32%). We obtained competitive results in both tasks, extending our earlier work in BioNLP'09 shared task competition, demonstrating the portability of our approach to different scope definitions (Kilicoglu and Bergler, 2010).

2.2.2 Discourse Structure and Coherence

Another way of describing the embedding layer is to take a top-down approach, as done by most research in discourse analysis, where the focus is on explaining how individual segments of text cohere to create textual meaning beyond the sum of the meaning of these units. The nature of the individual units and the relations that hold between them and overall discourse structure are the main topics of inquiry. One level of coherence between textual units is often explained using the notion of *reference* (Halliday and Hasan, 1976), while another level of coherence concerns *discourse relations* (Hobbs, 1985a). In this section, we first briefly describe and exemplify reference relations, and then turn our attention to discourse relation-based theories, discussing them along several criteria. Next, we discuss several core categories of discourse relations, which are also examined from a semantic viewpoint outside discourse theories. We end this section by taking a closer look at discourse and argumentation models of scientific literature.

Reference Relations

Reference relations are concerned with the connectivity of specific items (entities, propositions, etc.) that cannot be interpreted on their own in discourse. In one of the earliest works on discourse structure, Halliday and Hasan (1976) focus on the notion of *cohesion* and one of the main cohesive devices is *reference*. *Anaphora* and *cataphora* are the two main referential devices that create cohesion. Anaphora occurs when the author of a text refers back to an entity or an abstract object previously mentioned in the text, while cataphoric reference is a reference forward. For example, in Example (21a), it is important to capture the fact that the third person pronoun *him* refers to *John*, which appears in the previous sentence. This particular example illustrates an instance of anaphoric reference. The referent is called an *antecedent*. Example (21b) illustrates an instance where the antecedent is an abstract object (an event), rather than an individual, and Example (21c) illustrates a cataphoric reference. While not strictly considered referential relations, *ellipsis* and *substitution* are two other cohesive devices that act in a capacity similar to referential relations. *Ellipsis* refers to omitting words to avoid repetition (as in Example (21d)), and *substitution* refers to substitution of one word for another, more general word in discourse (as in Example (21e)).

(21) (a) [*John*]$_i$ arrived in Edinburgh by train. Max met [*him*]$_i$ at the station.(Individual Anaphora) (Asher and Lascarides, 2003)

(b) John believed [*that Mary was sick*]$_i$. The teacher believed [*it*]$_i$ too. (Event Anaphora) (Asher, 1993)

(c) When [*he*]$_i$ arrived home, [*John*]$_i$ went to sleep. (Cataphora)

(d) Fred [*hit a home run*]$_i$, and then Sally [*did*]$_i$ too. (VP-ellipsis) (Asher, 1993)

(e) I have heard strange [*stories*]$_i$ in my time. But this [*one*]$_i$ was perhaps the strangest one of all. (Substitution)

Reference relations can be viewed as lower level elements of discourse structure, since they simply establish correspondence (or equivalence) between informational units. On the other hand, discourse relations (discussed next) can be considered to operate at the higher level of discourse contributing additional meaning with respect to how different informational units are related.

Discourse Relation Models

One of the cohesive devices described by Halliday and Hasan (1976) is *conjunction*. Conjunction can be defined as a relationship between two linguistic elements (*discourse unit*) that occur in succession. This can be seen as the basic definition of a discourse relation. Halliday and Hasan provide a basic categorization of conjunctive relations: *additive* (*parallel* and *elaboration*), *contrastive*, *causal*, and *temporal*. Much of the research in discourse analysis expands upon these basic notions, first introduced by Halliday and Hasan and formalized and extended further by Hobbs (1985a). Two types of discourse relations are often distinguished: *informational* (Hobbs, 1985a) and *presentational* (Grosz and Sidner, 1986). Informational relations are essentially semantic relations between abstract objects. Causal and temporal relations are often considered informational, for example. On the other hand, presentational relations are concerned with the author's pragmatic intent of affecting the hearer's mental state. To illustrate the difference, consider the sentences below (Moore and Pollack, 1992).

(22) (a) George Bush supports big business.

(b) He's sure to veto House bill 1711.

At the informational level, a causal relation holds between the sentences: the sentence in Example (22a) indicates the cause for the action described in the sentence in Example (22b). At the presentational level, an *evidence* relation holds: the first sentence provides evidence for the second with the goal of convincing the reader of the claim in the latter.

Discourse relation-based theories vary with respect to several criteria: (a) the inventory and granularity of relations, (b) definition of a discourse unit, (c) assumed discourse structure, (d) formal semantic basis for relations, and (e) the degree of correspondence between discourse relations and their markers in text (Wellner, 2009). Below, we discuss several well-known discourse models with regard to these criteria.

Rhetorical Structure Theory Rhetorical Structural Theory (RST), introduced by Mann and Thompson (1988), aims to describe how texts are structured, rather than provide a rigorous formal discourse semantics. In their view, *nuclearity* is the central organizing principle of text structure and a distinction is made between *nucleus* and *satellite*, the former corresponding to the textual unit that expresses the main information and the latter to that which expresses the secondary, peripheral information. In this framework, deletion of a nucleus will make the discourse less coherent, while the deletion of a satellite will make it only less explicit. RST assumes a strict tree representation of discourse structure. Twenty-four primary discourse relations are defined, some corresponding to nucleus-satellite relations (such as Elaboration, Evidence) and some to multinuclear relations (such as Contrast, List). Another way of viewing these two distinct types of discourse relations would be to consider them subordinating and coordinating discourse relations, respectively. Minimal discourse units, called EDUs (elementary discourse units), roughly correspond to clauses but the notion of clause is not made very precise. In addition, no correspondence between discourse relations and their linguistic markers and syntactic constructions is assumed in RST.

Linguistic Discourse Model In Linguistic Discourse Model (LDM) (Polanyi and Scha, 1984; Scha and Polanyi, 1988), syntax plays a major role. Discourse structure is conceived as a constituent tree (similar to a syntactic tree) and the correspondence between discourse units and syntactic constructions is more explicitly fleshed out. LDM is proposed as a computational model and *discourse parsing* (building discourse representations from text) is a major component (Polanyi *et al.*, 2004). Discourse parsing is treated as an extension of syntactic parsing, and is based on three basic context-free rewrite rules: coordination, subordination, and n-ary constructions, which also correspond to their simple inventory of discourse relations.

Segmented Discourse Representation Theory Segmented Discourse Representation Theory (SDRT) (Asher and Lascarides, 2003) is perhaps the most formal and extensive discourse theory. It is an extension of Discourse Representation Theory (DRT) (Kamp, 1981; Kamp and Reyle, 1993), a dynamic, model-theoretic semantic model. DRT aims to dynamically interpret discourse, one sentence at a time, updating a representation of the discourse, called Discourse Representation Structure (DRS). In other words, the meaning of a sentence is a product of the grammar of the sentence and the prior discourse. SDRT extends DRT to the realm of discourse relations, while at the same time formalizing how discourse, clausal and lexical semantics interact, adopting the lexical semantics view of Generative Lexicon (Pustejovsky, 1995). They demonstrate the effect of discourse relations on anaphoric reference (*anaphoric accessibility*), temporal structure, presuppositions, and other important problems in semantics and pragmatics. A hierarchical view of discourse structure is assumed. The inventory of discourse relations is relatively small (Narration, Elaboration, Result, etc.), but they are provided a precise semantics largely based on the work of Hobbs (1985a). In addition, a mechanism is provided for deriving discourse relations using defeasible reasoning. For example, it is assumed that a Narration relation holds between two adjacent sentences (as in Example (23a)), while such interpretation can be overridden based on discourse cues, as in the case of Example (23b), where *because* changes the interpretation of the discourse relation between two sentences to an Explanation relation.

(23) (a) John felt dizzy. He fell on the ground.

(b) John felt dizzy because he fell on the ground.

GraphBank GraphBank (Wolf and Gibson, 2005) is not a discourse theory per se, but a corpus of 135 news articles annotated with discourse relations. It diverges from other discourse approaches in that it is based on an empirical methodology, rather than intuitions and preconceived constraints on discourse structure. Their annotation project revealed a graph representation for discourse with both directed and undirected arcs. The main evidence in favor of this discourse structure representation seems to be crossing dependencies and shared arguments which occurred quite frequently in the corpus. However, Wellner (2009) demonstrates that most of these problems are linked to entity level coherence (a kind of reference) as well as attribution (related to reported speech, evidentiality), which are qualitatively different from discourse relations we have been discussing here. Their inventory of discourse relations consists of 17 relations, mostly overlapping with the inventory described in Hobbs (1985a). The main contribution of GraphBank has been to stimulate discussion of structural constraints on discourse (see, for example, Webber (2006)).

Penn Discourse TreeBank Penn Discourse TreeBank (PDTB) (Miltsakaki *et al.*, 2004; Prasad *et al.*, 2008) is an annotated corpus of discourse relations whose goal is to enable training and formal evaluation of statistical machine learning methods developed for discourse analysis. It provides a layer of discourse annotation over the entire Wall Street Journal portion of the Penn TreeBank (Marcus *et al.*, 1993). Its approach is less theoretical than RST or SDRT and is more data-driven, similar to GraphBank. PDTB does not make any assumptions with respect to discourse structure, making it less amenable to traditional parsing approaches. Discourse segments are arbitrary text spans (phrases, clauses, sentences and their sequences) corresponding to abstract semantic objects (Asher, 1993), essentially predications. In other words, the constraint on what can constitute a discourse segment is semantic. Discourse connectives can be defined lexically: they can be coordinating conjunctions (*and*), subordinating conjunctions (*since*), or adverbials (*however*). Each discourse connective links two arguments: Arg2 is the argument syntactically connected to the discourse connective in the same sentence, while Arg1 is the argument which may lie in the same sentence or anywhere prior in the discourse. The arguments of subordinating and coordinating connectives are determined structurally, based on syntax or adjacency. There is little restriction on the position of Arg1 arguments of discourse adverbials, however. Five types of discourse relations are defined:

1. ExplicitRelation: discourse relations indicated with explicit discourse connectives.
2. ImplicitRelation: discourse relations between adjacent sentences that can be inferred by inserting a discourse connective phrase at the beginning of the second sentence.
3. AltLex: discourse relations between two sentences indicated by some phrase other than a discourse connective and where the insertion of a connective phrase would be redundant.
4. EntRel: two adjacent sentences are related entirely due to entities in those sentences and not over abstract semantic objects.
5. NoRel: no coherence between adjacent sentences in the same paragraph.

Examples for these discourse relation types are given below, where Arg1 argument spans are in italics, Arg2 spans are in bold (Wellner, 2009):

(24) (a) [*Drug makers shouldn't be able to duck liability*]$_{Arg1}$ <u>because</u> [**people couldn't identify precisely which identical drug was used.**]$_{Arg2}$ (ExplicitRelation)

(b) [*John pushed Max.*]$_{Arg1}$ [**Max fell.**]$_{Arg2}$ (ImplicitRelation)

(c) [*In September, she pleaded guilty and paid a \$500 fine.*]$_{Arg1}$ Her alternative [**was 90 days in jail.**]$_{Arg2}$ (AltLex)

(d) [*Howard is 89 years old.*]$_{Arg1}$ [**He turns 90 in a couple of weeks.**]$_{Arg2}$ (EntRel)

(e) [*The new Explorer sport-utility vehicle, set for introduction next spring, will also have rear-seat belts.*]$_{Arg1}$ [**Mr. Leinonen said he expects Ford to meet the deadline easily.**]$_{Arg2}$ (NoRel)

All Explicit, Implicit, and AltLex relations are also annotated with a *sense*, essentially the semantic type of the relation. These semantic types are mostly informational, rather than presentational, perhaps due to the fact that the WSJ corpus consists of news articles[8]. PDTB senses form a hierarchy, and the most specific, distinguishable sense is annotated for each discourse relation instance. Four top level senses are Contingency, Temporal, Comparison, and Expansion. The semantics of these categories are further refined into 16 mid-level categories (types) and 23 low-level categories (subtypes). An example type is Cause, which has two subtypes, *reason* and *result*. For example, the ExplicitRelation in Example (24a) is signaled by the subordinating conjunction *because* and is annotated with the low-level sense of *reason*, whereas the ImplicitRelation in Example (24b) with the *result* sense. On the other hand, the AltLex relation in Example (24c) is annotated with the *contrast* sense.

PDTB also annotates some information secondary to discourse relations, most importantly, attribution (ascription of beliefs and assertions to agents), which apply to entire discourse relations as well as individual arguments. Attribution is interesting, because it is one area where the discourse level and the lower level notions of modality and negation meet. An attribution annotation in PDTB consists of four features, two of which (Source and Type) were briefly discussed in Section 1.1. The other two features are Scopal Polarity and Determinacy, which we will discuss later in the context of attribution resolution in Section 6.4.

With its sizable corpus and the level of annotation that goes deeper than what is generally assumed in discourse frameworks (into modality and negation), PDTB presents a good platform to develop and evaluate practical discourse level applications.

[8]The discourse relation types that can be considered presentational have the Pragmatic prefix (Pragmatic Cause, for example) and were added in the second release of PDTB, PDTB 2.0 (Prasad *et al.*, 2008), even though they are rarely used in annotation.

Discourse Relations Outside Discourse Frameworks

Some basic, informational discourse relations have also been the focus of research that is only tangentially related to discourse analysis. Temporal and causal relations, in particular, have been considered from many perspectives. One interesting feature of such research seems to be that they often provide a finer-grained characterization of the phenomenon in focus than that assumed in discourse theories. Here, we make a few, brief remarks regarding temporal and causal relations.

Temporal Relations Temporal relations are generally viewed as a subclass of informational discourse relations. For example, one of the main sense classes in PDTB is the Temporal class, which is further divided into Asynchronous (with *precedence* and *succession* subtypes) and Synchronous types. In Example (25a), a *precedence* relation holds between the discourse segments.

(25) (a) *Output will be gradually increased* until **it reaches about 11,000 barrels a day.**

(b) *Market technicians were encouraged by the price patterns*, which in the past have preceded **sharp rallies.**

While this level of abstraction is generally found sufficient at the discourse level, research focusing on semantics of temporal relations considers finer-grained temporal relations. For example, in the TimeML (Pustejovsky *et al.*, 2005b) specification language, the case in which an event occurs at *some interval* before another event (BEFORE) is distinguished from the case where it occurs *immediately* before the other event (IMMEDIATELY-BEFORE). While we are informed by TimeML-related research, we do not aim to model or recognize this level of temporal granularity[9].

With regard to temporal relations, we limit ourselves to essentially modeling them at the discourse level granularity, while considering a wider range of expressions, generally ignored within discourse-based approaches. For example, there is a similar precedence relation between highlighted segments in Example (25b). However, no discourse relation is annotated in the PDTB corpus for this instance, since the cue (*preceded*) does not belong in one of the discourse connective classes[10]. Our approach aims to take such instances into account, as well.

Causation Similar to temporal relations, causal relations are also addressed outside discourse frameworks, as a wide range of linguistic expressions are associated with causation. In addition to explicit discourse connectives, such as *therefore, so, because*, causative verbs (*lead, cause*), conditionals (*If ... then ...*) can be used to explicitly indicate causation (Khoo *et al.*, 2002).

[9]We refer the reader interested in the notion of time in language at this fine granularity to the fundamental articles collected in Mani *et al.* (2005).

[10]It is unclear to us why this instance (and similar ones) is not considered an AltLex type of discourse relation.

In discourse approaches, the range of causal relations considered is generally limited. For example, PDTB considers the Cause type with *reason* and *result* subtypes. Similarly, RST considers (Non-)Volitional Cause, (Non-)Volitional Result. From a cognitive semantics perspective, Wolff (2003) proposes the *force dynamics model* of causation, which describes causation as the interaction of two entities, an *affector* and a *patient*[11]. Three types of causal relations are distinguished: CAUSE, ENABLE, and PREVENT. Differences between these relations are accounted for in terms of three dimensions, as shown in Table (2.3). Examples for each relation type are provided in Example (26), with the causal cue underlined.

	Patient tendency for result	Affector-patient concordance	Occurrence of result
CAUSE	N	N	Y
ENABLE	Y	Y	Y
PREVENT	Y	N	N

Table 2.3: Causal relations along three binary dimensions (Wolff, 2003)

(26) (a) The blast caused the boat to heel.

(b) Vitamin B enables the body to digest food.

(c) Corn oil prevents butter from burning.

Again, similar to temporal expressions, discourse frameworks often consider a narrow range of causal expressions. The sentence in Example (27a) is annotated with a *result* relation in PDTB, while a similar relation indicated by means other than a discourse connective is ignored (27b). The fact that discourse relations can be signalled lexically by such "discourse verbs" has been noted in the literature (Danlos, 2006).

(27) (a) *The governor couldn't make it* so **the lieutenant governor welcomed the special guests.**

(b) *The filing on the details of the spinoff* caused **Cray Research stock to jump $2.875 yesterday to close at $38 in New York Stock Exchange composite trading.**

Computational Approaches to Discourse

Co-reference Resolution Coreference resolution has been an active research area in CL/NLP for some time. However, the task has generally been considered independent of discourse analysis

[11] Two entities clearly correspond to arguments with semantic roles of AGENT and THEME, respectively.

and the focus has mostly been on coreference of individuals (entities) rather than propositional objects. State-of-the-art systems in coreference resolution often incorporate a combination of lexical, syntactic, shallow semantic (e.g., WordNet information (Fellbaum, 1998)) and discourse information (e.g., identifying speakers in text) within deterministic or probabilistic frameworks. For more on the current state-of-the-art on unrestricted coreference resolution, we refer the reader to the proceedings of the CoNLL-2011 Shared Task (Modeling Unrestricted Coreference in OntoNotes) (Pradhan, 2011).

One interesting thread of research in this area is concerned with specifically how anaphoric references and discourse structure interact (*anaphoric accessibility*). Some research discusses the ways that discourse structure constrains the antecedents for an anaphora (Asher and Lascarides, 2003; Cristea *et al.*, 2000). In discourse frameworks adopting a hierarchical view of discourse structure, this is sometimes expressed as the *right frontier constraint* (Polanyi *et al.*, 1988): antecedents to the anaphors are introduced in the propositions that reside at the right frontier of the discourse structure that has been formed so far. Constraints in the opposite direction (anaphoric reference constraining discourse relations) have also been empirically demonstrated (Seretan and Cristea, 2002).

Discourse Chunking and Parsing Most computational models of discourse interpretation have focused on building discourse representations automatically from text within one of the discourse relation-based frameworks. Almost all of the computational approaches to this task heavily favor statistical or probabilistic learning methods. Several types of tasks have been attempted. For example, *discourse chunking* focuses on identifying discourse segments. Within the RST framework, Sporleder and Lapata (2005) present a statistical framework for identifying EDUs and labeling them as nucleus/satellite, while Dinesh *et al.* (2005) tackle the problem within the PDTB framework. In *discourse parsing* research, on the other hand, the goal is to build full discourse representations assuming a particular discourse structure. This task has been attempted within RST (Marcu, 1999; Soricut and Marcu, 2003), SDRT (Baldridge and Lascarides, 2005; Baldridge *et al.*, 2007), and PDTB (Wellner, 2009) frameworks. Another task is classsifying discourse relation types, attempted in RST (Marcu and Echihabi, 2002), SDRT (Sporleder and Lascarides, 2005, 2008), GraphBank (Wellner *et al.*, 2006), and PDTB (Wellner, 2009) frameworks.

One main difficulty in discourse parsing and discourse relation classification tasks has been identifying discourse segments accurately (Soricut and Marcu, 2003). Some recent approaches adopt a dependency-based approach to discourse parsing, where discourse relations are assumed to hold between lexical heads of the discourse segments, rather than discourse segments themselves (Baldridge *et al.*, 2007; Wellner, 2009). This characterization simplifies the problem to some extent, avoiding discourse segmentation, allowing shared arguments as well as crossing dependencies.

Models of Scientific Literature Discourse Scientific discourse modeling is considered important for tasks such as automatic summarization of research literature, automated biocuration, and knowledge discovery approaches, such as detecting paradigm shifts (Lisacek *et al.*, 2005), particularly in the biomedical domain. So far, this research area has adopted a macro-level rhetorical, argumentative perspective.

In their *argumentative zoning* approach, Teufel et al. (1999; 2009) assigned sentences in research articles to different zones based on the rhetorical moves of global argumentation and the connections between the current work and the cited research, using domain-independent categories such as AIM (statement of specific research goal, or hypothesis of the paper), NOV_ADV (novelty or advantage of own approach), OWN_MTHD (new knowledge claim, own work: methods), CO_DI (comparison, contrast, difference to other solution), and ANTISUPP (clash with somebody else's results or theory). A corpus of research articles from chemistry and computational linguistics domains were annotated with these zones, and moderate levels of interannotator agreement were reported. Their approach has also been adapted and complemented for different domains. For example, Mizuta and Collier (2004) extended the notion of argumentative zones to biology research articles. Guo *et al.* (2010) annotated 1000 abstracts on cancer risk assessment using Mizuta and Collier's zones. More recently, they also employed weakly supervised models to identify these zones automatically (Guo *et al.*, 2011).

In a similar vein to argumentative zoning, Liakata *et al.* (2010) proposed the CoreSC annotation scheme, in which sentences are annotated on the basis of their role in scientific investigation. Each annotation consists of three layers. The first layer corresponds to the scientific investigation component and is one of the 11 categories: Motivation, Goal, Object, Method, Experiment, Observation, Result, Conclusion, as well as Hypothesis, Model, and Background. The second layer corresponds to whether the presented information is New or Old. The third layer corresponds to a concept identifier that links the instances of the same concept (a kind of reference relation). The CoreSC model was proposed as being complementary to the argumentative zoning approach of Teufel *et al.* (2009) and this complementarity was demonstrated with a corpus annotation study and mapping between CoreSC categories and the categories of Teufel *et al.* (2009).

The well-known rhetorical IMRAD categories (Introduction, Methods, Results and Discussion) (Swales, 1990) formed the basis on which Agarwal and Yu (2009) classify sentences in full-text biomedical articles using rule-based and supervised machine learning methods. They also present a corpus of biomedical sentences categorized along these categories and report substantial interannotator agreement. A similar and simpler approach was taken by de Waard *et al.* (2009), who focus on author's pragmatic intent and use five categories (Hypothesis, Implication, Method, Goal, Result). They show that using very simple regular expressions, they are able to distinguish between these categories on a small set of sentences from research articles.

White (2010) discussed extracting argumentation from the scientific literature using regular expression patterns. She argued that argumentation could be recovered by combining claims recursively into a rhetorical structure. However, she acknowledges the complex, computationally expensive nature of this task and recasts argumentation extraction as the task of finding support and conflict statements. She proposed a fine-grained categorization of such statements, some of which are given below:

1. Evidence is consistent with a hypothesis (Support)
2. Data from two or more experiments agree (Support)
3. Evidence supports a different hypothesis (Conflict)
4. Hypothesis conflicts with another (Conflict)

Categorizing argumentative words into classes such as Causality (e.g., *play a role*), Probability (*putative*), Connection (*however*), she developed a set of patterns to recognize support and conflict statements. However, she concluded that these features are not sufficient to distinguish between support/conflict statements, since the authors use similar concepts whether they are agreeing or disagreeing. Inability to adequately process negation was given as a major source of errors.

Very recently, the BioDRB corpus (Prasad *et al.*, 2011) has been presented. In this corpus, the PDTB methodology and tools have been applied to annotate 24 full-text biomedical articles from the GENIA corpus with PDTB-style discourse relations. The annotation task was somewhat simplified for feasibility. For example, a simpler sense hierarchy than the one used in PDTB was adopted: top level categories and pragmatic (presentational) senses were eliminated, while several senses found to be important in scientific discourse were added (Similarity, Purpose, etc.). Furthermore, attribution features and EntRel type relations were not annotated at all and AltLex and ImplicitRelation types are only considered intra-sententially. In their work, they also experimented with sense detection of explicit connectives, concluding that the connective itself is a reliable predictor of the sense. They found that the sense detection classifier performs poorly when trained on PDTB and tested on BioDRB, attributing this difference to biomedical sublanguage model (Friedman *et al.*, 2002). An interesting correlation was found between discourse relations and IMRAD categories. For example, Temporal relations are frequent in the Methods section, while Contrast and Concession type relations occur more in Results and Discussion sections.

It is worth noting that some of the more bottom-up approaches presented in Section 2.2.1 (including Wilbur *et al.* (2006), Thompson *et al.* (2011)) share some commonalities with the macro-level discourse modeling approaches outlined in this section. For example, Wilbur et al.'s *focus* dimension (Scientific, Methodology, etc.) can be mapped to categories in the IMRAD or CoreSC schemes, highlighting the overlap between bottom-up and macro-level approaches and the need for a more unified approach to embedding layer.

Chapter 3

Theoretical Framework

This chapter is concerned with description of the core embedding framework. We first present the basic assumptions and definitions with simple, constructed examples as well as sentences from the Penn TreeBank corpus (Marcus *et al.*, 1993). We then describe the embedding categorization which underpins our compositional semantic interpretation approach. Next, we discuss the semantic consequences of combination and interaction of embedding predications. The final section of this chapter illustrates how the embedding framework extends beyond sentence level towards discourse level interpretation, based on discourse examples from news articles, as well as from biomedical literature.

3.1 Basic Concepts

The notion of a *predication* underlies our framework, which essentially corresponds to a unit of relational meaning. More precisely:

> **Definition 1.** A **predication** Pr is an n-ary abstract semantic object that consists of a predicate P and n logical arguments.

$$Pr := [P, Arg_{1..n}]$$

Consider the simple, declarative sentence in (28a), which describes an event that can be characterized as a 'building' event involving two participants: a builder and an object that is built. The verb *build* characterizes this event and its syntactic subject and object refer to the participants of the building event. The semantic content of the sentence can be formally represented using the logical form shown in (28b), which essentially states that "There is a boy x and a boat y and x builds y."

Variables x and y stand for the existentially quantified arguments and *build* is the predicate. x is the *logical subject* and y is the *logical object*[1].

(28) (a) A boy builds a boat.

(b) *(∃x,y)build(x,y)* ∧ *boy(x)* ∧ *boat(y)*

In the current work, we assume that all arguments are existentially quantified and do not represent quantification explicitly. Drawing from Hobbs (1985b), we adopt a slightly simplified representation to represent the same content as a *predication*, as shown in Example (29).

(29) $boy(t_1) \wedge boat(t_2) \wedge build(e_1,t_1,t_2)$

The predication ($build(e_1,t_1,t_2)$), as well as the arguments, is given an explicit identifier (e_1). In predications with a single argument, that argument often refers to the *logical object*. With multiple arguments, the first argument of a predication will refer to its *logical subject*, the second argument to its *logical object* and rest of the arguments to non-core *adjuncts*.

3.1.1 Atomic vs. Embedding Predications

Declarative statements asserting simple facts (such as the one in (28a)) are rare in written text. More often, such statements are embedded within additional clauses, serving to alter its semantic content or placing it in context. The following example shows the sentence in (28a) embedded in a clause that expresses a wish, which can be characterized as *dynamic modality* of VOLITIVE type (Palmer, 2001).

(30) (a) A boy wanted to build a boat quickly.

(b) $(\exists e_1, e_2, e_3, x, y) Past(e_1) \wedge want'(e_1, x, e_2) \wedge quick'(e_2, e_3) \wedge build'(e_3, x, y) \wedge boy(x) \wedge boat(y)$

(c) $want{:}\textsc{volitive}(em_1, t_1, em_2) \wedge quickly(em_2, e_3) \wedge build(e_3, t_1, t_2) \wedge boy(t_1) \wedge boat(t_2)$

Both the sentence and its logical form (Examples 30a and 30b, respectively) are taken from Hobbs (1985b). The logical form essentially means that "e_1 occurred in the past, where e_1 is x's wanting e_2, which is the quickness of e_3, which is x's building of y, where x is a boy and y is a boat[2]." In this expression, e_3 (corresponding to the building event) represents a base predication, where all arguments are simple entities, while e_1 and e_2 embed other predications as arguments. Uttering this

[1]We will simply refer to them as subject and object in this document. Syntactic subject and object will be referred to explicitly as such.

[2]The prime (') is essentially a nominalization operator.

sentence, the speaker does not assert the building event as a *fact* anymore. Rather, s/he establishes the boy's wish of building with the infinitival complement, which leaves the factual status of the building event unspecified. Our representation in Example (30c) is almost identical, except the fact that we ignore the tense information (that the wishing occurred in the past)[3] and the embedding predicate *want* is semantically typed (as VOLITIVE).

In our framework, along these lines, we distinguish these two types of predications as *atomic* and *embedding* predications. The building event (e_3) is atomic, while the wanting of building event (em_1) is embedding. The relevant definitions can be given as follows:

Definition 2. A semantic object T is **ontologically simple** if it takes no arguments or if it refers to an entity (individual, physical object, etc.). A predication takes arguments and is, therefore, an **ontologically complex** object. Ontologically simple semantic objects will be referred to simply as **terms** or **semantic terms**[4]. Conjunction of terms is also ontologically simple.

Definition 3. A predication is **atomic**, if all of its arguments are semantic terms.

$$Pr_{atomic} := [P, T_{1..n}]$$

Definition 4. A predication is **embedding**, if it has at least one ontologically complex argument.

$Pr_{embedding} := [P, Arg_{1..n}]$, where $(\exists Arg_i : Arg_i \in PR)$ and PR is the set of all predications.

Definition 5. A **surface element** SU is a single token or a contiguous multi-token unit, which may be associated with an abstract semantic object SEM.

- A surface item that is associated with a semantic object is said to be **semantically bound** ($[\![SU]\!]$ = SEM).

- Otherwise, it is said to be **semantically free** ($[\![SU]\!] = \emptyset$). Being semantically free does not imply that the surface element has no semantic content, but that we do not know what that content is.

[3]We do not deal with tense in the current work, in addition to quantification, as mentioned earlier.

[4]Of course, defining what exactly counts as an entity is a complex issue, which has been the subject of the *ontology* discipline in philosophy. By using the more neutral notion of *semantic term*, we aim to avoid such ontological considerations and also be accommodating to different definitions of an entity as much as possible. For example, while the notion that a physical object (*continuant*) is an entity is generally taken for granted, the status of a process (*occurrent*) is less clear. Since we propose that our approach can integrate with existing systems, we do not make an overt commitment regarding whether a semantic object constitutes a semantic term. In the current framework, when an overt decision has to made regarding the ontological status of a semantic object, we base the decision on whether the object in question takes arguments.

As shown in Example (30c), the identifiers for atomic predications are denoted with the letter e (e_3), while those for embedding predications are denoted with em (em_1 and em_2). The predicational arguments of the embedding predications (that is, arguments that are predications themselves) are denoted simply using reification.

In the current work, we focus on characterization and interpretation of embedding predications. With respect to atomic predications, the framework can adopt one of the following positions:

1. Accept as atomic predications semantic relations generated by an external semantic role labeling or semantic interpretation system.

2. Generate atomic predications using its basic compositional mechanism, provided that terms and predicates relevant to atomic predications are known.

3. Assume that atomic predications can only be signalled by verbs, nominalizations, and adjectival predicates[5] and take such predicates as indicating atomic predications, without being concerned with the inner structure (i.e., their arguments) of these predications.

In this chapter, we adopt the third position above for expository purposes.

In order to clarify embedding predications further and delineate our focus, let us consider the sentence in Example (31a) taken from the Wall Street Journal portion of the Penn TreeBank corpus. Semantic terms are illustrated in Example (31b), atomic predications in Example (31c) and embedding predications in Example (31d).

(31) (a) President Bush <u>will</u> **veto** a bill **funding** the Departments of Labor, Education and Health and Human Services <u>because</u> it <u>would</u> <u>allow</u> federal <u>funding</u> of **abortions** for victims of rape and incest, the White House <u>said</u>. (wsj_2075)

(b) *President_Bush*:PERSON(t_1) $\wedge$ *bill*(t_2) $\wedge$
Departments_of_Labor_Education_and_Health_and_Human_Services:ORGANIZATION(t_3) $\wedge$
federal:ORGANIZATION(t_4) $\wedge$ *the_White_House*:ORGANIZATION(t_5) $\wedge$
victims_of_rape_and_incest:PERSON(t_6)

(c) *veto*:VETO.01$(e_1,t_5,1.0_{epistemic},t_1,t_2)$ $\wedge$ *fund*:FUND.01$(e_2,t_5,1.0_{epistemic},t_2,t_3)$ $\wedge$
abortion:ABORTION.01$(e_3,t_5,1.0_{epistemic},t_6)$

(d) *will*:ASSUMPTIVE$(em_4,t_5,1.0_{epistemic},e_1)$ $\wedge$
funding:PROPOSITIONAL$(em_5,t_5,1.0_{epistemic},t_4,e_3,t_6)$ $\wedge$

[5] This more or less corresponds to the TimeML (Pustejovsky *et al.*, 2005a) notion of event predicates.

allow:ENABLE$(em_6, t_5, 0.8_{epistemic}, t_2, em_5)$ $\wedge$ *would*:ASSUMPTIVE$(em_7, t_5, 1.0_{epistemic}, em_6)$ $\wedge$ *because*:CAUSE$(em_8, t_5, 1.0_{epistemic}, em_7, em_4)$ $\wedge$ *say*:REPORTING$(em_9, WR, 1.0_{epistemic}, t_5, em_8)$

In this example, the surface element *President Bush* corresponds to a semantic term with type PERSON (t_1) and, thus, is semantically bound. On the other hand, the surface item *bill* is semantically free. As illustrated in (31b), we denote semantic terms as *m:*SEM*(id)*, where m corresponds to the textual mention of the entity, SEM to its semantic type, and *id* to its unique identifier. If the entity is semantically free, it is simply denoted as *m(id)*. t_3 (*Departments of Labor, Education and Health and Human Services*) is a semantic term, since it is conjunction of three entities: *Department of Labor*, *Department of Education*, and *Department of Health and Human Services* (Definition 2). On the other hand, t_6 (*victims of rape and incest*) is a semantic term, even though its head, *victim*, takes *rape and incest* as argument. This is due to the fact that t_6 refers to an entity rather than a situation or an event (Definition 2). Predicates such as *victim* are sometimes referred to as *quasi-predicates* (Mel'čuk, 2004) and we consider quasi-predicates as ontologically simple terms.

Atomic predications in the same sentence are shown in Example (31c) and the embedding predications in Example (31d). With this example, we revise the definition and representation of predication, as shown below.

$$Pr := [P, S, MV_{Sc}, Arg_{1..n}]$$

A predication has now two additional features:

- S indicates the *source* of the predication. By default, the source of a predication is the writer (*WR*). But, it may also indicate the semantic object that refers to the source. For example, in the example above, the source of the predication em_8 is given as t_5, which indicates *the White House.*

- MV_{Sc} indicates the *scalar modality value* of the predication in the [0,1] range on a relevant scale. The relevant scales will be introduced in Section 3.2; however, for now, it suffices to say that, by default, an unmarked statement has the scalar modality value of 1 on the *epistemic* scale (essentially, corresponding to a fact).

In the examples throughout, we will denote a predication as *m:*SEM$(id, S, MV_{Sc}, Arg_{1..n})$, where m is the predicate mention (that is, surface element for the predicate) and SEM is the semantic type (or category) of the predicate, and by extension, of the predication.

Underlined predicates in the sentence (*will*, *allow*, *would*, *because*, and *said*) signal embedding predications ($em_{4..9}$). What kind of meaning do these predications encode?

- The main predicate of the sentence (*said*) introduces a reporting context where the source of information, *the White House* (t_5), is introduced by its logical subject. It also makes the embedded predication (em_8 which indicates the causal link cued by *because*) factual according to the source of reporting (*the White House*), but uncommitted by the writer of the text, since s/he merely reports it.

- The predicates *because* and *allow* indicate *causal* connections between their arguments. Take em_6, indicated by *allow*, for example. It establishes a causal link between the bill (t_2) and the federal funding of abortions, encoded by the predication em_5.

- The modal auxiliaries *will* and *would* create intensional contexts for the relations indicated by their main verbs (*veto* and *allow*), affecting their factuality status. The fact that *would* changes the meaning of *allow* with respect to its factuality is captured by embedding the predication indicated by *allow* (em_6) as the logical object argument of the predication indicated by *would* (em_7). Furthermore, the fact that the epistemic modality value of 0.8 is associated with em_6 indicates that the act of allowing is not a fact, but highly probable, according to the source of the predication (t_5).

- The gerund *funding* indicates simple propositional content where one of the arguments happens to be an event[6].

For atomic predications in Example (31c), the available semantic types correspond to PropBank (VETO.01 and FUND.01) or NomBank (ABORTION.01) senses. On the other hand, the semantic categories of embedding predicates (e.g., ENABLE, ASSUMPTIVE) are taken from an embedding categorization scheme, described in more detail in Section 3.2. With respect to arguments, note that all arguments of atomic predications are ontologically simple semantic terms ($t_{1..3}$, t_6). Embedding predications, on the other hand, take both ontologically simple and complex arguments. Some predications (e.g., e_1, e_2, and em_6) have both logical subject and object arguments, while others (e.g., e_3 and em_4) take a single argument, which is their logical object.

[6]Note also that the embedding predicate in this instance, *funding*, is derived from the verb *fund*, which indicates the atomic predication e_2. This further illustrates the fact that the certain predicates can signal both atomic and embedding predications in different contexts, as they can take arguments denoting entities or processes.

Semantic Scope and Embedding Predications

The level of embedding in a sentence can be arbitrarily deep. For example, em_8 in (31d) takes as arguments two other embedding predications, em_4 and em_7, the former taking an atomic predication em_1 as argument, and the latter taking yet another embedding predication em_6 as well as the atomic e_3 as arguments. The predicational structure of the sentence is illustrated in Figure (3.1): atomic predications are represented with green circles and the embedding predications with orange circles.

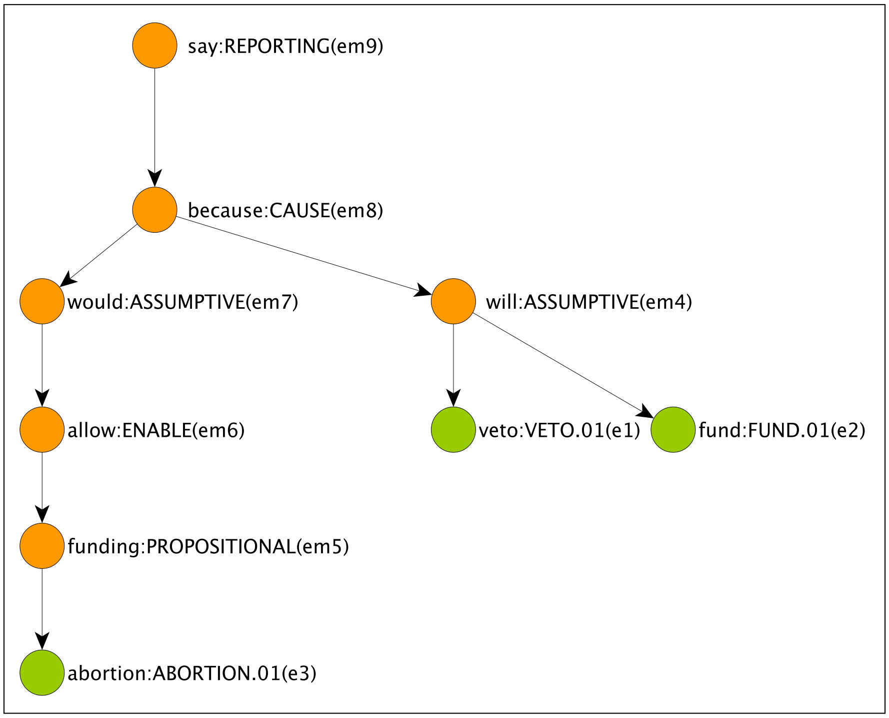

Figure 3.1: The predicational structure corresponding to the sentence in Example (31a)

We use the notion of *semantic scope* to describe the structural relationships between predications. We provide the basic definition below:

Definition 6. A predication Pr_1 **embeds** a predication Pr_2 if Pr_2 is an argument of Pr_1.

$$Pr_1 := [P_1, ..Pr_2, ..]$$

Definition 7. A predication Pr_2 is within the **semantic scope** of a predication Pr_1 (written as $\mathrm{Pr}_1 > \mathrm{Pr}_2$), if one of the following conditions is met:

- Pr_1 embeds Pr_2.
- There is a predication Pr_3, such that Pr_1 embeds Pr_3 and Pr_2 is within the semantic scope of or shares an argument with Pr_3.

$(Pr_1 = [P_1, ..Pr_2, ..]) \vee$

$(Pr_1 = [P_1, ..Pr_3, ..] \wedge (Pr_3 > Pr_2)) \vee$

$(Pr_3 = [P_3, ..X, ..] \wedge Pr_2 = [P_2, ..X, ..] \wedge (Pr_1 > Pr_3))$

$\Rightarrow Pr_1 > Pr_2$

where X is any argument.

The scope relations illustrated in Figure (3.1) can also be represented as follows.

(32) $em_9 > em_8$

$em_8 > em_4 > \{e_1, e_2\}$

$em_8 > em_7 > em_6 > em_5 > e_3$

Scope relations play an important role in propagation of source information as well as scalar modality values, as Example (31) indicates. The source for predications in the scope of em_9 are inherited from the source introduced by em_9 and the modality value of em_6 (0.8) is due to the modality value introduced by its parent predication em_7. These compositional operations will be illustrated in Chapter (5).

3.2 Embedding Categorization

In this section, we present a semantic categorization of embedding predicates. In doing so, our main goal is to pinpoint the kind of semantic information carried by such predicates and and to explore their interactions in a more systematic manner using scope relations. We also aim to synthesize various linguistic typologies and classifications into a unified and computationally viable framework by clarifying and refining some of the nebulous terminology. We distinguish four basic classes of embedding predicates: MODAL, RELATIONAL, VALENCE_SHIFTER and PROPOSITIONAL, each class further divided into subcategories. In a nutshell, MODAL and VALENCE_SHIFTER predicates are concerned with lower level extra-factual phenomena, introducing modal scales or providing meaning shifts with respect to these modal scales as well as with respect to polarity, respectively. On the other hand, RELATIONAL predicates largely operate at the higher discourse coherence level, whereas the PROPOSITIONAL predicates function at the basic propositional level. We discuss these categories in more detail below with examples and illustrate the categories in Figure (3.2). The examples given illustrate the results of the compositional semantic interpretation, which will be discussed in Chapter 5.

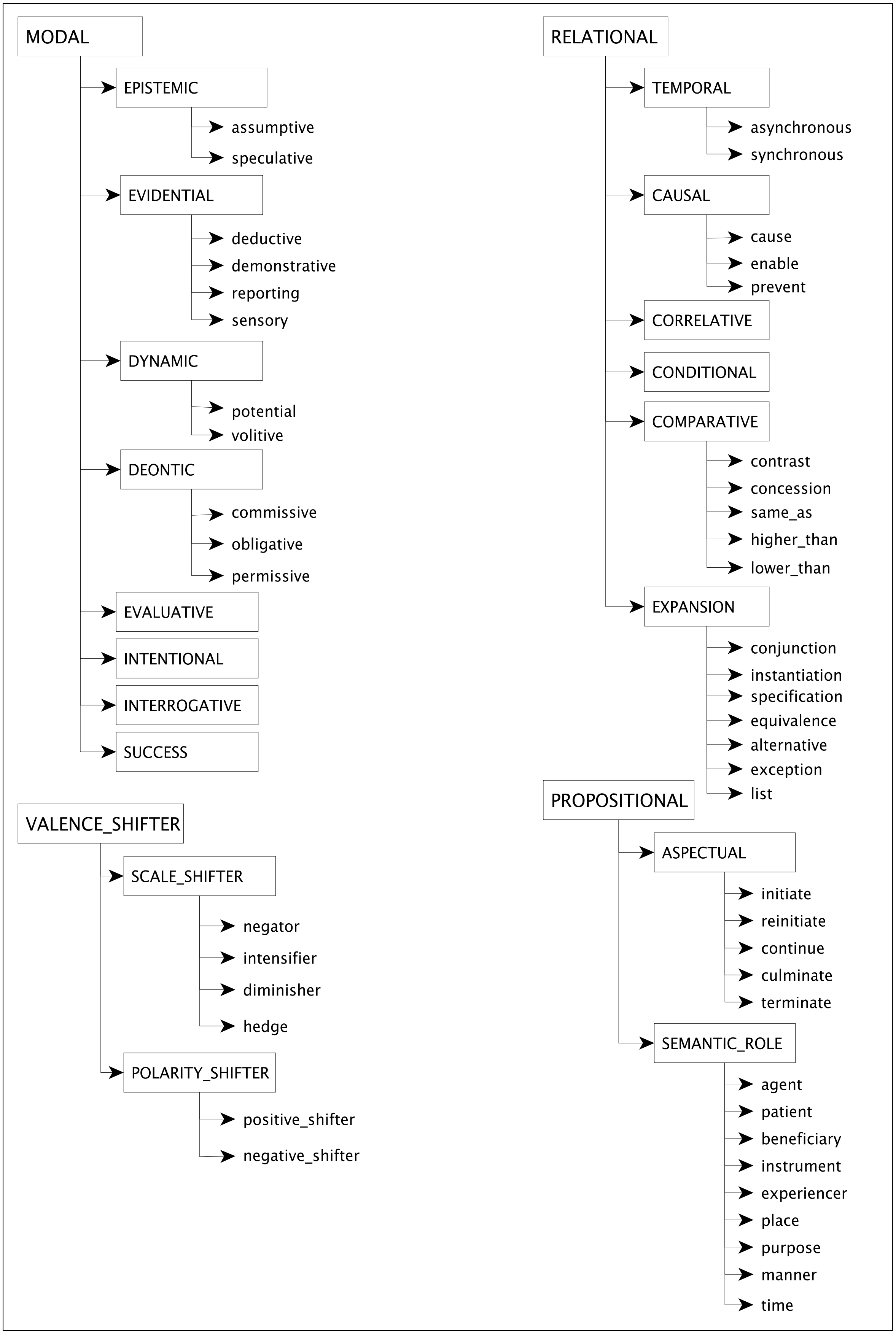

Figure 3.2: Embedding predicate types.

3.2.1 MODAL Category

Definition 8. A modal predicate, $\mathrm{P}_{\text{MODAL}}$, associates the embedded predication, Pr_e, with a *modality value* on a context-dependent *scale*. The scale (Sc) is determined by the semantic category of the modal predicate. The scalar modality value ($\mathrm{MV}_{\mathrm{Sc}}$) is a numerical value between 0 and 1 and indicates how strongly the embedded predication Pr_e is associated with the scale Sc, 1 indicating strongest positive association and 0 negative association.

The modal predicate subcategorization is largely based on and extended from Palmer (2001), presented in Section 2.2.1. To recap, four main categories in his classification are EPISTEMIC, EVIDENTIAL, DYNAMIC, and DEONTIC modalities, each further broken down into subcategories. We extended this basic classification with four other categories described elsewhere. We define and illustrate these categories in the following subsections with examples. As usual, the embedded predicate is in bold, and the embedding (modal) predicate is underlined.

The scalar modality value is partially modeled after Nirenburg and Raskin (2004). In this view, a modality value of zero on the EPISTEMIC scale, for example, corresponds to "The predication Pr_e is not true (a counter-fact)", while a value of 0.5 roughly indicates that "There is a possibility that Pr_e is true." More often, modality values are represented discretely (Lyons, 1977; Palmer, 2001; Saurí, 2008) (Section 2.2.1). In our framework, we favor a contextual, real-valued scale rather than a fixed one since it is more general and flexible, and it allows us to model interactions between different categories more easily.

EPISTEMIC Type

Definition 9. An EPISTEMIC predicate indicates a judgement about the *factual status* of the embedded predication. It associates the embedded predication with a scalar epistemic value on the *epistemic scale* which encodes the degree of factuality (or certainty) of the embedded predication. A predication that corresponds to an unmarked affirmative statement is assumed to have a scalar epistemic value of 1.

Let Pr be an embedding predication signalled by an EPISTEMIC predicate P and Pr_e the embedded predication (that is, $Pr > Pr_e$). Let EPS represent the epistemic scale. The scalar epistemic value of Pr_e ($MV_{EPS}(Pr_e)'$) is a function (f) of the current epistemic value of the embedded predication, denoted as $MV_{EPS}(Pr_e)$, and the scalar modality value introduced by the embedding predicate ($MV(P)$). We will describe these compositional functions in more detail in Section 5.3.3.

$$Sem(P) = \text{EPISTEMIC} \wedge (Pr > Pr_e) \Rightarrow MV_{EPS}(Pr_e)' = f(MV_{EPS}(Pr_e), MV(P))$$

There are various characterizations of the epistemic scale, as discussed earlier in Section 2.2.1. We largely adopt the epistemic scale conceptualized by Horn (1989) (and adapted by Saurí (2008)) which incorporates the interaction of epistemic modality and negation. There are two main differences: (a) We favor a real-valued scale, rather than one with discrete values, and (b) within the epistemic scale, our characterization roughly corresponds to a 5-value system, rather than the 6-value system they propose. The difference is that Saurí (2008) explicitly encodes PS+ (*possible*) and PS- (*not certain*), whereas we do not, since it seems reasonable to assume that if a predication is possibly true, that it is also possibly false. The scale we use can be represented as in Figure (3.3), where *not certain (uncertain)* subsumes intermediate values of factuality (*probable*, *possible*, and *doubtful*) and excludes the endpoints of the scale only (*certain* and *certain not*). Note that an unmarked proposition (an affirmative statement) is considered a fact, and a simply negated affirmative statement is considered a counter-fact.

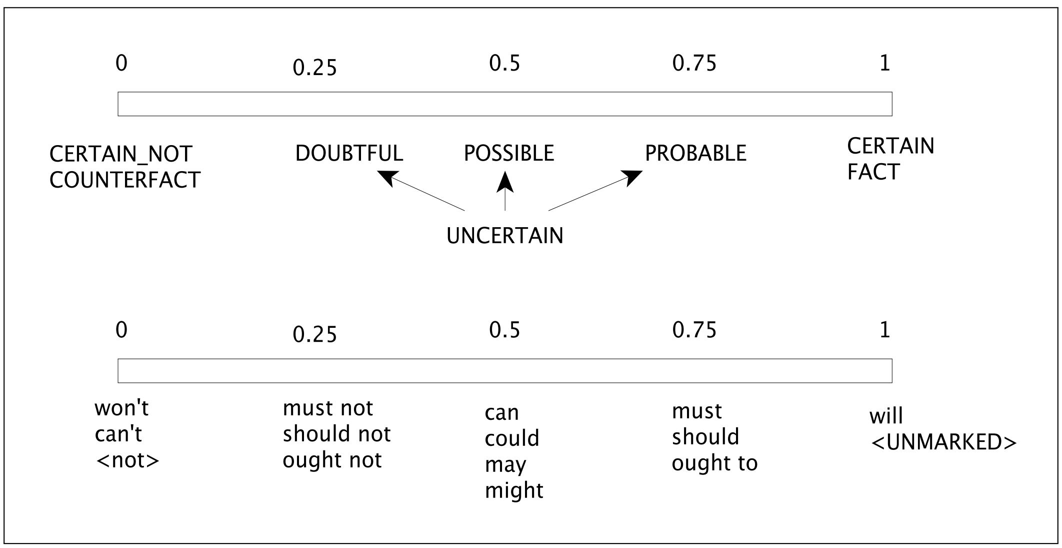

Figure 3.3: The epistemic scale with characteristic values and corresponding modal auxiliaries.

One question that concerns the epistemic modality category is whether the finer-grained epistemic distinctions can be made reliably. Palmer (2001) distinguishes ASSUMPTIVE, SPECULATIVE, and DEDUCTIVE subtypes and illustrates them in the context of modal auxiliaries (*will*, *may*, and *must*, respectively)[7]. However, in non-auxiliary contexts, it seems non-trivial to establish the difference for certain modal items. Consider the verb *believe*. It clearly has epistemic meaning, but is it SPECULATIVE or is it ASSUMPTIVE? Some other non-auxiliary epistemic items are less difficult to

[7]He also concedes that the difference between these types may be difficult to establish and that the DEDUCTIVE type has evidential meaning, as well.

evaluate. For example, the epistemic adjective *possible* seems essentially SPECULATIVE. Other items seem to invoke both speculative and assumptive meanings simultaneously. For example, in the sentence in Example (33a), the meaning contribution of *presumably* seems both speculative and assumptive, and it associates the predication that it embeds (em_2) with an epistemic value of 0.8.

(33) (a) "The most frequent use is home improvement, which presumably **improves** the value of the property," Mr. Durkin says. (wsj_1389)

(b) $value\_of\_the\_property(t_1) \wedge home(t_2) \wedge Mr.\_Durkin(t_3) \wedge improvement(e_1,\ldots) \wedge$
$improve$:PROPOSITIONAL$(em_2,t_3,0.8_{epistemic},e_1,t_1) \wedge$
$presumably$:EPISTEMIC$(em_3,t_3,1.0_{epistemic},em_2)$

Nauze (2008) considers these subtypes too fine-grained and excludes them from his analysis. Based on our observations, our strategy was to adopt the finer-grained categories of SPECULATIVE and ASSUMPTIVE in only clear cases, while using the EPISTEMIC category more generally. The DEDUCTIVE subtype seems more easily distinguishable, and we used it more commonly than the other two subtypes, even though we model it as an EVIDENTIAL subtype, as described in the next section.

EVIDENTIAL Type

Definition 10. An EVIDENTIAL predicate indicates the type of the *evidence* surrounding the embedded predication, whether it is based on observation, hearsay, inference, etc. The type and source of the evidence provides an evaluative context for the embedded predication, in which its *factual status* and *reliability* can be assessed by the reader.

Consider the sentence in Example (34a). The author makes an inference from the data (*The numbers*) regarding the predication indicated by the complement clause. The use of the verb *suggest* and referring to concrete, quantitative data indicate that the author has relatively high confidence (or certainty) for the statement and her/his goal is to persuade the reader of her/his viewpoint, as well. The evidential predicate *suggest* is assigned the evidential subtype DEDUCTIVE.

(34) (a) The numbers suggest that the housing industry is still **suffering** the effects of the Federal Reserve's battle against inflation. (wsj_1782)

(b) $The\_numbers(t_1) \wedge Federal\_Reserve(t_2) \wedge housing\_industry(t_3) \wedge$
$battle(e_1,t_1\ldots) \wedge effect$:CAUSAL$(em_2,t_1,1.0_{epistemic},e_1,t_3) \wedge$
$suffer$:PATIENT$(em_3,t_1,0.8_{epistemic},t_3,em_2) \wedge$
$suggest$:DEDUCTIVE$(em_4,WR,1.0_{epistemic},t_1,em_3)$

We associate predicates of EVIDENTIAL type with the epistemic scale, as well, since they are closely linked to epistemic modality and are often considered to have epistemic meanings, as well (Aikhenvald, 2004; Saurí, 2008).This is a similar position to that taken in Saurí (2008). However, she simply considers evidential markers as a subset of factuality markers, whereas we maintain the distinction between these two types, since we aim to provide fine-grained semantic interpretation, rather than only determine factuality status.

Note that in Example (34b), the source of the embedding predication em_3 as well as the predications in its scope is given as *The numbers* (t_1), whereas at the top level (em_4), the source is simply the writer (WR). EVIDENTIAL predicates, as well as some EPISTEMIC predicates, make the source explicit. The source may refer to the reporter in the context of REPORTING or SENSORY predicates, to the holder of the judgement in the context of EPISTEMIC predicates, or to the evidence in the context of DEDUCTIVE predicates. The default source is assumed to be the writer (WR) and the source is recomputed and percolated down in the context of EVIDENTIAL and EPISTEMIC predicates, similar to the approach taken in Saurí (2008).

Let Pr be an embedding predication signalled by an EVIDENTIAL or EPISTEMIC predicate P and Pr_e the embedded predication (that is, $Pr > Pr_e$). The source of Pr_e, denoted as $S(Pr_e)$), is the logical subject of Pr, if any. Otherwise, it is inherited from the source attribute of Pr.

$$Sem(P) \in \{\text{EVIDENTIAL}, \text{EPISTEMIC}\} \wedge (Pr > Pr_e) \Rightarrow S(Pr_e) = Subject(Pr) \vee S(Pr_e) = S(Pr)$$

Another dimension often discussed with regard to evidentiality is *reliability*. In fact, de Haan (1999) proposes the evidential subcategories of VISUAL, AUDITORY, NONVISUAL, INFERENCE, and REPORTATIVE, and places them along a scale of reliability, VISUAL being the most reliable type of evidence and REPORTATIVE type being the least reliable. For example, a verb of visual perception, such as *see* (*I saw that ...*), can be said to indicate more reliability in the embedded predication than a reporting verb such as *tell* (*I was told that ...*). One problem with this scale is that it does not consider the subtle semantic differences between evidential predicates of the same subtype. For example, consider the REPORTATIVE type and the reporting verbs *claim*, *say*, and *announce*. By using the verb *claim*, the author generally implies more doubt (and less reliability) for the predication corresponding to the underlined fragment, whereas the verb *say* seems neutral, and *announce* indicates the least doubt (and highest reliability) (Bergler, 1992).

(35) (a) The agency said that because MCI's offer had expired AT&T couldn't continue to offer its discount plan. (wsj_767)

(b) The agency claimed that ...

(c) The agency announced that ...

Another issue with the notion of reliability is that it is essentially subjective. In fact, Bergler (1992), in the context of reported speech, states that "depending on knowledge, beliefs, points of view, and interest different readers will evaluate the same instance of reported speech differently" and considers reliability as essentially parallel to certainty. In the same spirit, we do not model reliability explicitly, but consider it as a pragmatic inference on the part of the reader. However, we make the components necessary to assess reliability (source, epistemic strength, etc.) as explicit as possible.

We assume a slightly simpler evidential categorization than that of de Haan (1999) and conflate VISUAL and AUDITORY categories (both of which are considered direct evidence) into SENSORY. The final categorization consists of SENSORY, DEMONSTRATIVE, DEDUCTIVE, and REPORTING subtypes. Examples for SENSORY and DEMONSTRATIVE subcategories are given in Example (36a-b). DEDUCTIVE and REPORTING categories were illustrated earlier in Examples (34) and (31), respectively.

(36) (a) Under a microscope he could actually <u>see</u> that a bit of chromosome 13 was **missing** (wsj_0465)

$he(t_1) \wedge a\_bit\_of\_chromosome\_13(t_2) \wedge missing(e_1,t_1,1.0_{epistemic},t_2) \wedge$

see:SENSORY$(em_2,WR,0.7_{potential},t_1,e_1)$

(b) Government statistics in fact <u>show</u> that the profit rate – net pretax profits divided by capital stock – **peaked** in 1965 at 17.2%. (wsj_1849)

$Government\_statistics(t_1) \wedge the\_profit\_rate(t_2) \wedge$

$peak(e_1,t_1,1.0_{epistemic},t_2\ldots) \wedge show$:DEMONSTRATIVE$(em_2,WR,1.0_{epistemic},t_1,e_1)$

DYNAMIC Type

Definition 11. A DYNAMIC predicate indicates *ability* or *willingness* of an agent towards an event, corresponding to POTENTIAL and VOLITIVE categories, respectively. The subcategories are associated with their own scale of *potential* or *volition*, respectively.

The relevant scales in the context of dynamic predicates are not commonly discussed in the literature. One exception is Nirenburg and Raskin (2004), who associate predicates with *potential* meaning to degrees of an agent's capability (from *able* at the positive end to *unable* at the negative end), and those with *volitive* meaning to degrees of intensity of desire (from *strongly want* to *be interested in* to *do not want*). We roughly adopt this characterization and use separate scales for the subcategories of DYNAMIC type. The scalar values are calculated in the same way as described for epistemic predicates. Examples for each subcategory are presented below. In Example (37a), note that the predication embedded by the predicate with POTENTIAL type (em_2) is positioned on the relevant potential scale and is assigned the value of 1.0 on that scale.

(37) (a) Laboratory tests showed that non-toxic versions of the poisons are capable of **inducing** an immunity to whooping cough, the researchers reported .... (wsj_0739)
non-toxic_versions_of_the_poisons$(t_1) \wedge$ *the_researchers*$(t_2) \wedge$ *Laboratory_tests*$(t_3) \wedge$
immunity$(e_1,t_3,1.0_{epistemic},\ldots) \wedge$ *induce*:CAUSE$(em_2,t_3,1.0_{potential},t_1,e_1) \wedge$
capable:POTENTIAL$(em_3,t_3,1.0_{epistemic},t_1,em_2)$

(b) And we hope to **take advantage** of panics and buy stocks when they plunge. (wsj_2415)
we$(t_1) \wedge$ *take_advantage*$(e_1,WR,1.0_{volitive}\ldots) \wedge$
hope:VOLITIVE$(em_2,WR,1.0_{epistemic},t_1,e_1)$

Similar to EVIDENTIAL predicates, DYNAMIC predicates can also be viewed as having secondary epistemic extensions. However, epistemic status of the embedded predication in the scope of a POTENTIAL or VOLITIVE predications seems to be simply left underspecified (Saurí, 2008). That is, in Example (37b) above, we do not know whether the situation described by e_1 (the act of *taking advantage*) really occurred or will occur.

DEONTIC **Type**

Definition 12. A DEONTIC predicate indicates *obligation*, *permission*, or *command* from an external authority for an event, corresponding to OBLIGATIVE, PERMISSIVE, and COMMISSIVE categories, respectively. Deontic predicates are associated with the *deontic scale*.

The deontic scale, illustrated in Figure (3.4), corresponds to the different degrees of free will in the actions of an agent (Nirenburg and Raskin, 2004); the positive end of the scale corresponds to rigid obligation (or command) from an external authority, while the negative end corresponds to prohibition. In between lie different degrees of free will or optionality[8]. Scalar value calculation is done in the same way as other types of modal predicates. Examples for each DEONTIC subcategory are given below and the scalar effect of DEONTIC predicates on the embedded predications (e_1 in al cases) is illustrated.

(38) (a) Such legislation must be **enacted** by the end of the month. (wsj_2372)

[8]There are also views that obligation and permission scales are separate entities (Frawley, 1992); however, we adopt the position of Nirenburg and Raskin (2004), in that commanding someone to do something entails absence of permission, and vice versa.

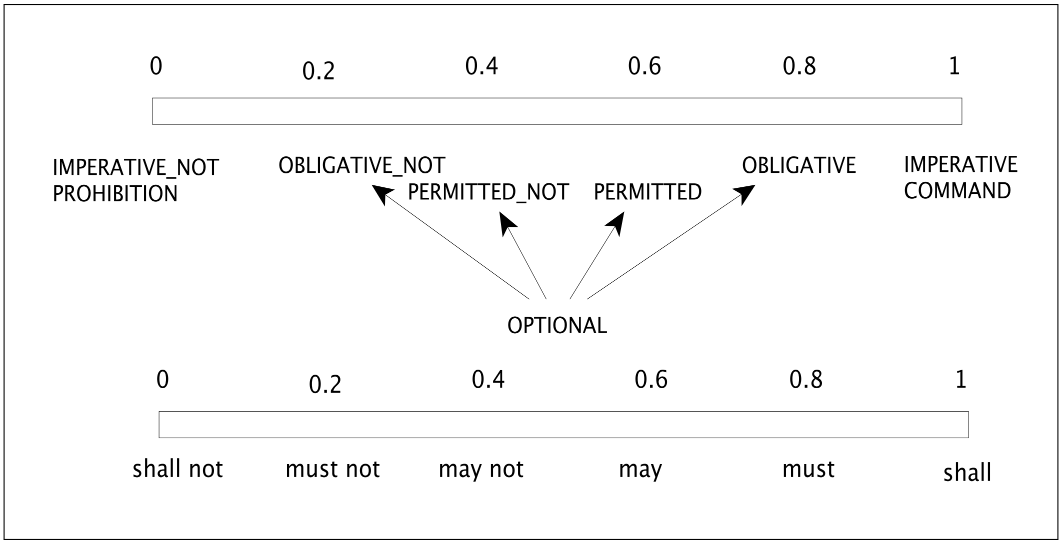

Figure 3.4: The deontic scale with characteristic values and corresponding modal auxiliaries.

$Such\_legislation(t_1) \wedge enact(e_1, WR, 0.8_{deontic}, t_1 \ldots) \wedge$

$must$:OBLIGATIVE$(em_2, WR, 1.0_{epistemic}, e_1)$

(b) But under current rules, they are <u>allowed</u> to **change** just 200 rubles into dollars and other currencies for each trip. (wsj_0934)

$they(t_1) \wedge change(e_1, WR, 0.6_{deontic}, t_1, \ldots) \wedge$

$allow$:PERMISSIVE$(em_2, WR, 1.0_{epistemic}, e_1)$

(c) Thousands of residents of low-lying areas were <u>ordered</u> to **evacuate** as the storm headed north in the Gulf of Mexico with 80 mph winds. (wsj_2356)

$Thousands\_of\_residents(t_1) \wedge evacuate(e_1, WR, 1.0_{deontic}, t_1) \wedge$

$order$:COMMISSIVE$(em_2, WR, 1.0_{epistemic}, e_1)$

Other MODAL Types

We consider four additional modal types: INTENTIONAL, INTERROGATIVE, SUCCESS, and EVALUATIVE. These types are mentioned in discussions of modality and are sometimes adopted as separate categories; however, there appears to be less of a consensus on their modal status. We choose to include them in our categorization, since corpus analysis provides clear evidence that they affect the status of predications they embed and that they occur in considerable amounts.

INTENTIONAL type An INTENTIONAL predicate indicates *effort* of an agent to perform an event (cf. Saurí (2008); Baker *et al.* (2010)). It can be viewed as a stronger type of VOLITIVE modality in which the agent not only wishes but also takes action toward the relevant event[9]. However, in most work on propositional attitudes, intentions and desires are considered separate semantic categories (Bratman, 1987; Prabhakaran *et al.*, 2010), and we adopt this view in the current work. The relevant scale is the *intentional scale*, indicating the level of effort. This type of predicate also indicates a secondary epistemic meaning, which characterizes the predication in scope as *unrealized* (therefore, a future possibility) (Saurí, 2008). An example for this type is given below.

(39) (a) Santa Fe aims to **drill** about 30 wells in this area in 1989 .... (wsj_0725)

(b) $Santa\_Fe(t_1) \wedge drill(e_1, WR, 1.0_{intentional}, t_1, \ldots) \wedge$
aim:INTENTIONAL$(em_2, WR, 1.0_{epistemic}, t_1, e_1)$

Here, the act of drilling is characterized as unrealized, while the intentional strength seems high, due to selection of the strongly intentional verb, *aim*. While often only epistemic effect of this type of predicate is considered (Saurí, 2008), our strategy is to associate these predicates primarily with the intentional scale, and consider its epistemic effect as secondary.

INTERROGATIVE **type** An INTERROGATIVE predicate indicates *questioning* of the predication and inherently expresses some degree of uncertainty or doubt. This type of modality is often only considered within the context of question sentences (Palmer, 2001), however, a subclass of cognition verbs, such as *inquire*, *ask*, *investigate*, have also been examined as interrogative (or rogative) verbs (Carlson and Nirenburg, 1990; Saurí, 2008; Kilicoglu and Bergler, 2011b). Similar to INTENTIONAL type, we model the predicates of this type primarily on a scale of interrogative strength, while their epistemic effect is modeled as a secondary inference. Questions are also considered to be of INTERROGATIVE modality type. Note that in the following example, the source of interrogation (*skeptics*) is also introduced as the logical subject of the INTERROGATIVE predication, and the uncertain status of the embedded predication (e_1) is attributed to this source.

(40) (a) But skeptics question whether asset-backed bonds **offer** sufficient rewards to compensate for the extra risks. (wsj_1635)

(b) $skeptics(t_1) \wedge asset\text{-}backed\_bonds(t_2) \wedge reward(t_3) \wedge$
$offer(e_1, t_1, 1.0_{interrogative}, t_2, t_3) \wedge question$:INTERROGATIVE$(em_2, WR, 1.0_{epistemic}, t_1, e_1)$

SUCCESS **type** A modal predicate of SUCCESS type indicates the *degree of success* associated with the embedded predication (Nirenburg and Raskin, 2004; Baker *et al.*, 2010). Relevant predicates

[9] This seems to be the position taken in Nirenburg and Raskin (2004), for example.

include *implicative verbs* (Karttunen, 1971), such as *manage* and *fail.* The scale of success involves values from full failure (0) to complete accomplishment (1). Similar to INTENTIONAL and INTERROGATIVE types, SUCCESS type of predicates also involve secondary epistemic meanings; a scalar modality value of 0 implies a counter-fact (epistemic value of 0), whereas the value of 1 implies a fact (epistemic value of 1). The main characteristic that distinguishes SUCCESS predicates from typical EPISTEMIC predicates seems to be encoding of a level of expectation and effort (Karttunen, 1971).

(41) (a) It claims the Coast Guard failed to **chart** the rock and refuses to pay damages. (wsj_1470)

(b) $Coast\_Guard(t_1) \wedge rock(t_2) \wedge It(t_3) \wedge chart(e_1,t_3,0.0_{success},t_1,t_2) \wedge$ $fail$:SUCCESS$(em_2,t_3,0.5_{epistemic},t_1,e_1)$

EVALUATIVE **type** An EVALUATIVE predicate indicates a subjective evaluation or attitude toward the embedded predication (Nirenburg and Raskin, 2004). The evaluative scale ranges from positive evaluation to negative evaluation. Evaluative predicates include a wide range of subjective verbs, including *criticize*, *hate*, and *appreciate*, and adjectives, such as *exciting*, *surprised*, which are often considered sentiment-bearing lexical items (positive, negative, or neutral) in sentiment analysis research. Predicates belonging to this type also have secondary epistemic meaning, in that they often signal that the embedded predication is presupposed (therefore, has epistemic value of 1).

(42) (a) Private-sector leaders praised the Conasupo **restructuring**. (wsj_1254)

(b) $Private\text{-}sector\_leaders(t_1) \wedge Conasupo(t_2) \wedge restructuring(e_1,WR,1.0_{evaluative},t_2) \wedge$ $praise$:EVALUATIVE$(em_2,WR,1.0_{epistemic},t_1,e_1)$

In Example (42b), the event of restructuring (e_1) is presupposed, and has epistemic value of 1.

3.2.2 RELATIONAL Category

Definition 13. A RELATIONAL predicate semantically links two semantic objects, at least one of which is a predication, providing a discourse coherence function between them.

A RELATIONAL predication may involve two predications or one semantic term and a predication. Let $Pr_{\text{RELATIONAL}}$ be a RELATIONAL predication indicated by the predicate P and Arg_1 and Arg_2 its arguments.

$$Sem(P) = \text{RELATIONAL} \wedge Pr := [\ldots Arg_1, Arg_2] \Rightarrow Arg_1 \in PR \vee Arg_2 \in PR$$

where PR is the set of all predications.

Consider the earlier example, partially duplicated below, where the subordinating conjunction *because* signals a RELATIONAL predication of CAUSE type between the predications it embeds (em_7 and em_4). The first predication (em_7) corresponds to the main clause (in italics) and the second (em_4) to the subordinate clause (in square brackets). The relational predication (em_8) corresponds to a discourse coherence relation, indicated by the discourse connective *because* and annotated in PDTB (Prasad *et al.*, 2008).

(43) (a) *President Bush <u>will</u> **veto** a bill **funding** the Departments of Labor, Education and Health and Human Services* <u>because</u> [it <u>would</u> <u>allow</u> federal <u>funding</u> of **abortions** for victims of rape and incest], the White House <u>said</u>. (wsj_2075)

(b) $President\_Bush(t_1) \wedge bill(t_2) \wedge the\_White\_House(t_5) \wedge$
$veto(e_1,\ldots) \wedge abortion(e_3,\ldots) \wedge funding$:PROPOSITIONAL$(em_5,t_5,1.0_{epistemic},\ldots e_3\ldots)$
$\wedge$ $allow$:ENABLE$(em_6,t_5,0.8_{epistemic},t_2,em_5) \wedge will$:ASSUMPTIVE$(em_4,t_5,1.0_{epistemic},e_1) \wedge$
$would$:ASSUMPTIVE$(em_7,t_5,1.0_{epistemic},t_2,em_6) \wedge$
$because$:CAUSE$(em_8,t_5,1.0_{epistemic},em_7,em_4)$

In the current work, RELATIONAL predications are conceived as generalized discourse coherence relations. In this view, RELATIONAL predicates are analogous to discourse connectives and their semantic categories to discourse relation types. We take PDTB discourse relation senses as the basis for categorizing RELATIONAL predicates. However, our characterization of RELATIONAL predicates is more liberal and subsumes that assumed in PDTB. In PDTB, discourse connectives are defined lexically: subordinating and coordinating conjunctions (e.g., *although* and *and*, respectively) as well as discourse adverbials (e.g., *then*). An additional class of discourse connectives, *alternate lexicalizations*, correspond to a limited set of sentence-initial expressions that provide discourse connectivity between two sentences, such as “The idea was” and “What’s more”. However, surface realizations of discourse connectives are, in fact, more varied. For example, discourse relations often permeate to the subclausal level, signalled by “discourse verbs” (Danlos, 2006) (e.g., *cause*, *precede*, *compare*), their nominal forms or other predicative nouns, such as *role*. Discourse relations signalled at this level are absent in PDTB. For example, the embedding predication em_6 in the example above is signalled by the discourse verb *allow* and describes a causal relation between its logical subject and object (t_2 and em_5, respectively). But no corresponding discourse relation is annotated in PDTB. Modeling discourse relations at this granularity is important for fully integrating propositional content with high level discourse. The full PDTB sense hierarchy that forms the basis of RELATIONAL subcategories is presented in Figure (3.5). There are several differences between the

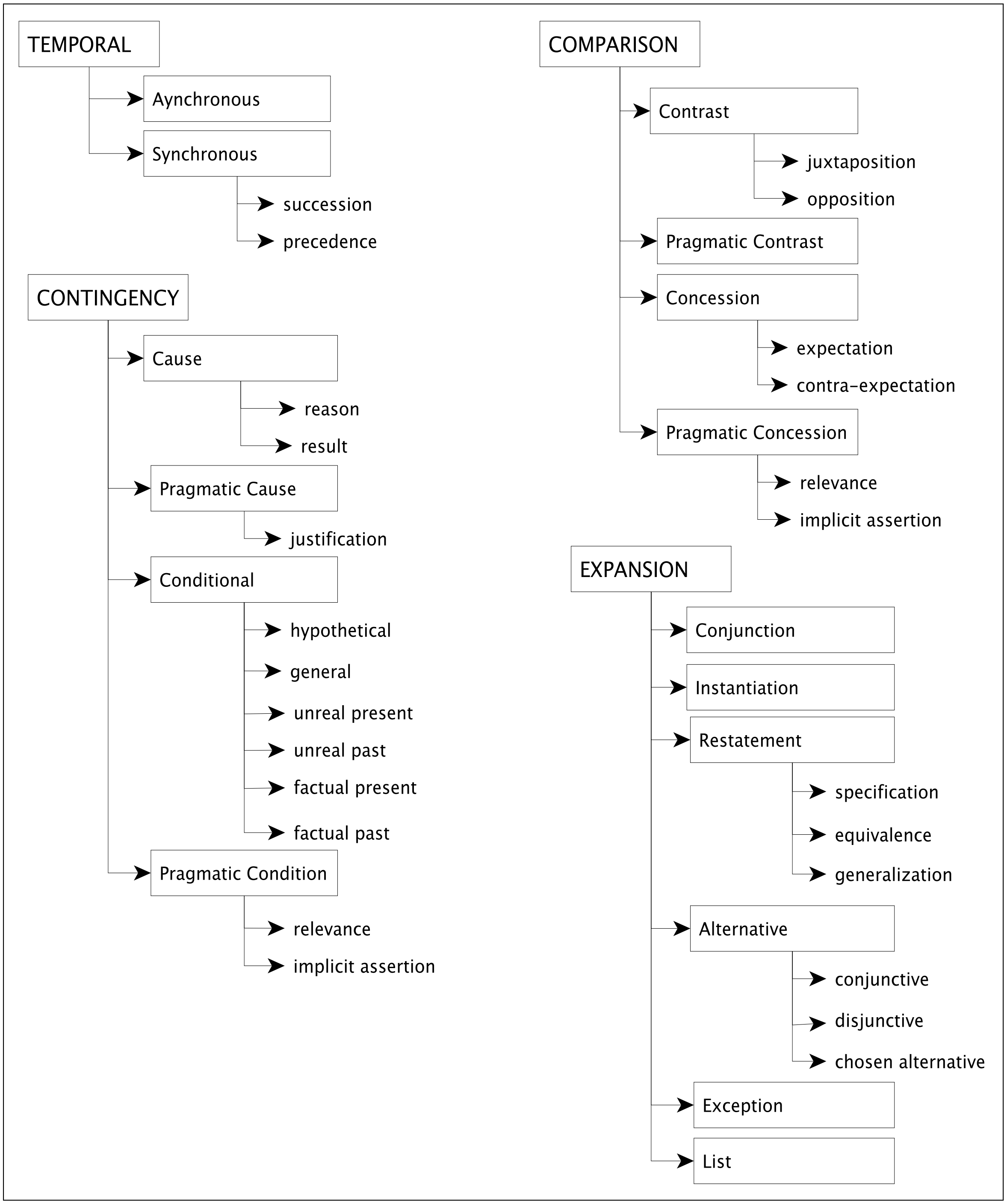

Figure 3.5: The Penn Discourse TreeBank sense hierarchy, taken from PDTB Annotation Manual (The PDTB Research Group, 2008)

PDTB sense inventory and our RELATIONAL subtypes:

1. We adopt a finer-grained classification with respect to cause/effect and comparative relations. PDTB has a single cause/effect relation (Cause). We adopt the tripartite classification of causation often assumed in cognitive semantics literature and consider CAUSE, ENABLE, and PREVENT as subtypes of cause/effect relations (Wolff, 2003). With regard to comparatives, the finer-grained classification of HIGHER_THAN, LOWER_THAN, and SAME_AS are adopted to explicitly represent how the compared items are ranked with respect to each other on a comparative scale (Fiszman *et al.*, 2007). These relations tend to be expressed at the subclausal level more commonly.

2. We assume an additional category of RELATIONAL predications, which we call CORRELATIVE. This class indicates unspecific correlations between two embedding predications, without implying causation. A corresponding PDTB sense does not seem to exist. While this new category is based on work in biomedical literature (Kim *et al.*, 2008; Blake, 2009), we believe it is important to represent such unspecific links explicitly in general.

3. We do not consider Pragmatic type of PDTB senses (Pragmatic_Contrast, Pragmatic_Cause, etc.) These types indicate the cases where the arguments are linked pragmatically and are very rarely used in PDTB annotation.

4. We ignore EntRel type of relations, since they provide entity coherence rather than discourse coherence. We also make the assumption that discourse relations are explicitly indicated and, thus, ignore discourse relations of Implicit type[10].

5. We ignore subtypes in the sense hierarchy (for example, *result* and *reason* for Cause type), since they specify the semantic contribution of the arguments only. In other words, for the predicational arguments Pr_1 and Pr_2, Pr_1-*result*-Pr_2 is equivalent to Pr_2-*reason*-Pr_1.

The resulting RELATIONAL category consists of six subcategories: TEMPORAL, CAUSAL, CORRELATIVE, CONDITIONAL, COMPARATIVE, and EXPANSION. All but CONDITIONAL and CORRELATIVE categories are further subdivided into finer-grained classes. We briefly describe these subcategories below, and exemplify them with realizations at the subclausal level.

[10]This is not to imply that Implicit discourse relations are unimportant. In fact, we illustrate importance of such relations in Section 3.4. However, we leave their treatment largely as future work.

TEMPORAL **type**

TEMPORAL predicates provide temporal coherence between situations described by the embedded predications. We assume two subtypes of TEMPORAL predicates, corresponding to temporally ordered (ASYNCHRONOUS) and temporally overlapping (SYNCHRONOUS) predicates. In an ASYNCHRONOUS predications, the first argument precedes the second in time.

(44) (a) The **crash** of 1929 was followed by a substantial **recovery** before the great Depression and awful bear market of the 1930s began. (wsj_2415)

(b) *crash(*e_1*...)* $\wedge$ *recovery(*e_2*...)* $\wedge$ *follow*:ASYNCHRONOUS*(*em_3*,WR,*$1.0_{epistemic}$*,*e_1*,*e_2*)*

CAUSAL **type**

CAUSAL predicates indicate causal influence between situations described by the embedded predications. CAUSE, ENABLE, and PREVENT are the causal subtypes. CAUSE and ENABLE can be viewed as causal classes with positive polarity, whereas PREVENT is negative in the sense that the expected result does not occur. The difference between CAUSE and ENABLE is more subtle and is explained in terms of force dynamics theory (Wolff, 2003), as described in Section 2.2.2. Typical predicates for these subtypes include *allow* for ENABLE, *stimulate* for CAUSE, and *block* for PREVENT. We duplicate an earlier example below to exemplify the CAUSE subtype of the CAUSAL category.

(45) (a) Laboratory tests showed that non-toxic versions of the poisons are capable of **inducing** an immunity to whooping cough, the researchers reported .... (wsj_0739)

(b) *non-toxic_versions_of_the_poisons(*t_1*)* $\wedge$ *the_researchers(*t_2*)* $\wedge$ *Laboratory_tests(*t_3*)* $\wedge$ *immunity(*e_1*,...)* $\wedge$ *induce*:CAUSE*(*em_2*,*t_3*,*$1.0_{epistemic}$*,*t_1*,*e_1*)*

CORRELATIVE **type**

As mentioned above, CORRELATIVE predicates indicate unspecific relations without implying causation. It does not have any subtypes. The sentence in Example (46a) exemplifies the CORRELATIVE type.

(46) (a) ...Mr. Wilson's **resignation** wasn't related to the sales **shortfall**. (wsj_2342)

(b) *resignation(*e_1*...)* $\wedge$ *shortfall(*e_2*...)* $\wedge$ *relate*:CORRELATIVE*(*em_3*,WR,*$0.0_{epistemic}$*,*e_1*,*e_2*)*

CONDITIONAL **type**

While similar to CAUSAL predicates in the sense that they may also indicate causal influence, CONDITIONAL predicates differ from them with respect to the inferences that can be drawn. A CONDITIONAL predicate makes the factual status of the embedded predication in the logical object position dependent on the factuality of the predication in the subject position. An example is given below, where for the event described as e_1 to occur, e_2 should take place first as a necessary condition. More commonly, CONDITIONAL relations are indicated by conditional clauses introduced by *if*.

(47) (a) The **merger** requires the **approval** of Norwegian authorities. (wsj_1188)

(b) *merger(e_1...) $\wedge$ approval(e_2...) $\wedge$ require*:CONDITIONAL*(em_3,WR,1.0$_{epistemic}$,e_2,e_1)*

COMPARATIVE **type**

COMPARATIVE predicates highlight the similarity or difference between the embedded predications with respect to a property. In addition to scalar comparatives (HIGHER_THAN, LOWER_THAN, and SAME_AS), we assume two other subtypes, based on PDTB: CONTRAST and CONCESSION. The CONTRAST subtype highlights a difference between the embedded predications. The CONCESSION subtype indicates a configuration where one embedded predication causes a situation C, while the other embedded predication implies ¬C. A typical subclausal CONCESSION predicate would be *not suggest* or *not mean*; however, this type of predication is more commonly indicated by discourse connectives, such as *although*.

(48) (a) The **agreement** on Poland contrasts with the major **differences** remaining over the underlying foreign aid bill, ... (wsj_0101)

(b) *agreement(e_1...) $\wedge$ difference(e_2...) $\wedge$ contrast*:CONTRAST*(em_3,WR,1.0$_{epistemic}$,e_1,e_2)*

EXPANSION **type**

The predicates of EXPANSION type serve to expand the discourse by moving the narrative forward. This type and its subtypes are all adapted from PDTB. We briefly describe and exemplify these subtypes below. We also illustrate the semantics of these relations, provided in PDTB (The PDTB Research Group, 2008) and adapted into our framework. In illustrating, we assume that the relation holds between two predicational arguments Pr_1 and Pr_2.

INSTANTIATION The second argument (Pr_2) describes the first (Pr_1) in further detail. Its semantics can be given as: $Pr_1 \wedge Pr_2 \wedge exemplify(Pr_2, Pr_1)$, where *exemplify* holds when its arguments are predications that share arguments and the first predication describes a shared argument in further detail. INSTANTIATION type is most commonly indicated by adverbials such as *for example* or *for instance.*

(49) (a) Microsoft's surprising **strength** is one example of the **difficulty** facing investors looking for reassurances about the financial health of the computer firms. (wsj_2365)

(b) *strength(*$e_1 \ldots$*)* $\wedge$ *difficulty(*$e_2 \ldots$*)* $\wedge$ *example*:INSTANTIATION*(*em_3*, WR, 1.0*$_{epistemic}$*,* e_2*,* e_1*)*

SPECIFICATION **and** EQUIVALENCE These subtypes are in fact subtypes of the RESTATEMENT type in PDTB. Since their semantics are different from each other, we adopt them as separate subtypes of EXPANSION type. Both describe the way the second embedded predication restates the first. Their semantics can be given as:

(50) (a) $Pr_1 \Rightarrow Pr_2$ (SPECIFICATION)

(b) $Pr_1 \Leftrightarrow Pr_2$ (EQUIVALENCE)

The following exemplifies a SPECIFICATION relation. More typically, these relations are expressed using discourse connectives. For example, a more typical EQUIVALENCE predicate is *in other words.*

(51) (a) **Making** computers smaller often means **sacrificing** memory. (wsj_2387)

(b) *make(*$e_1 \ldots$*)* $\wedge$ *sacrifice(*$e_2 \ldots$*)* $\wedge$ *mean*:SPECIFICATION*(*em_3*, WR, 0.8*$_{epistemic}$*,* e_1*,* e_2*)*

ALTERNATIVE Two embedded predications describe alternative situations. Three further subtypes are defined in PDTB: CONJUNCTIVE indicates that both alternatives hold or are possible, whereas DISJUNCTIVE indicates that only one of the alternatives hold and CHOSEN ALTERNATIVE that one alternative is taken.

(52) (a) $Pr_1 \wedge Pr_2$ (CONJUNCTIVE)

(b) $Pr_1 \oplus Pr_2$ (DISJUNCTIVE)

(c) $(Pr_1 \oplus (Pr_2 \wedge \neg Pr_1)) \Rightarrow Pr_2$ (CHOSEN ALTERNATIVE)

The coordinating conjunction *or* is the most common predicate for the CONJUNCTIVE, while the subordinating conjunction *unless* is most commonly associated with the DISJUNCTIVE type and the discourse adverbial *instead* with the CHOSEN ALTERNATIVE type.

EXCEPTION The embedded predication in the object position specifies an exception to the generalization described by the subject. That is, Pr_1 is false because Pr_2 is true and if Pr_2 were false, Pr_1 would be true ($\neg Pr_1 \wedge Pr_2 \wedge (\neg Pr_2 \Rightarrow Pr_1)$).

(53) (a) The companies wouldn't disclose the length of the contract except to say it was a multiyear agreement. (wsj_0372)

(b) *wouldn't(em$_1$...)* $\wedge$ *say(em$_2$...)* $\wedge$ *except*:EXCEPTION*(em$_3$, WR,1.0$_{epistemic}$,em$_1$,em$_2$)*

LIST The embedded predications are members of a list defined in prior discourse. The predications may not be related otherwise. In the following example, *Microsoft* and *Oracle* are members of the *Nasdaq's biggest technology stocks* group defined in previous sentence.

(54) (a) Many of Nasdaq's biggest technology stocks were in the forefront of the rally. Microsoft **added** 2 1/8 to 81 3/4 and Oracle Systems **rose** 1 1/2 to 23 1/4. (wsj_0327)

(b) *add(e$_1$...)* $\wedge$ *rose(e$_2$...)* $\wedge$ *and*:LIST*(em$_3$, WR,1.0$_{epistemic}$,e$_1$,e$_2$)*

CONJUNCTION The second predication provides additional information to the situation described in the first, but is not related to the first predication in any of the ways described for other types of EXPANSION. Its semantics is simply $Pr_1 \wedge Pr_2$.

(55) (a) Roederer Cristal at $90 a bottle **sells out** around the country and Taittinger's Comtes de Champagne Blanc de Blancs is **encroaching upon** that level. (wsj_0071)

(b) *sell_out(e$_1$...)* $\wedge$ *encroach_upon(e$_2$...)* $\wedge$ *and*:CONJUNCTION*(em$_3$, WR,1.0$_{epistemic}$,e$_1$,e$_2$)*

RELATIONAL Predications and Scalarity

Whether RELATIONAL predications open up scales like MODAL predications is a complex issue. Scalarity in the context of COMPARATIVE predications seems obvious, as mentioned in Section 2.1.3. In such predications, sometimes the basis of comparison (or contrast) is not explicit (as in (48) above) and it is merely stated that two predications are being compared or contrasted. In other cases (mostly at the clausal level), the basis of comparison is made explicit, generally by using a gradable adjective, such as *good* or *tall.* Consider the example below. The basis of comparison is *goodness* of understanding. Thus, we can speak of the scale of *goodness* where e_1 is ranked higher than e_2. On the other hand, using another comparative adjective would yield another scale. Therefore, rather than a single relevant scale, comparison statements involving gradable adjectives yield their own scale, dependent on the lexical item used (Fiszman *et al.*, 2007).

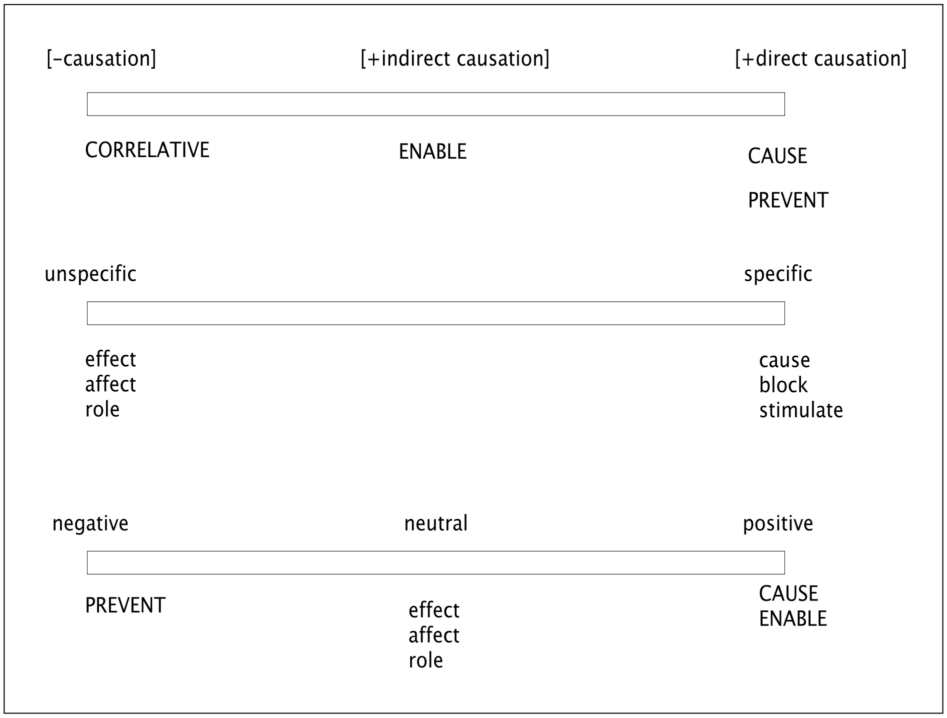

Figure 3.6: Causal dimensions

(56) (a) The truth is, Washington **understands** politics <u>better</u> than economics. (wsj_2229)

(b) $Washington(t_1) \wedge politics(t_2) \wedge economics(t_3) \wedge$

$understand(e_1, WR, 1.0_{epistemic}, t_1, t_2) \wedge understand(e_2, WR, 1.0_{epistemic}, t_1, t_3) \wedge$

$good{:}\textsc{higher\_than}(em_3, WR, 1.0_{epistemic}, e_1, e_2)$

In the context of causation, one can also speak of a *causal scale*, which incorporates the notions of *indirect* and *direct* causation, the difference being concerned with the *mediacy* of the relationship between cause and effect (Comrie, 1985). One can also speak of *specificity* of the causal relation, a generic causal predicate such as *effect* being unspecific with respect to the nature of the causal relation involved, whereas another like *stimulate* being specific. That is, we can speak of *positive effect* or *negative effect*, but not of *negative stimulation*. The polarity of the causal relation is another important aspect of causation, PREVENT can be seen as negative, while ENABLE and CAUSE categories encode positive causation. These distinctions and whether causation can be modeled as being scalar are not much discussed in the literature. We encode these dimensions, illustrated in Figure (3.6), using different categories and granularities for causal predicates. For example, an unspecific predicate like *effect* is encoded as belonging to the coarse-grained category of CAUSAL, which can be considered *neutral* with respect to polarity. The difference between indirect and direct causation as well as polarity (or lack thereof) is also encoded at the level of CAUSE, ENABLE, and

PREVENT. The lack of causation (but existence of some kind of correlation) is encoded with the CORRELATIVE category. Scalarity does not seem to be a feature associated with other types of RELATIONAL predicates.

3.2.3 VALENCE_SHIFTER Category

Definition 14. A VALENCE_SHIFTER predicate engenders a scalar shift of the embedded predication on the relevant scale or changes the polarity of the embedded predication.

Contextual valence shifting is a term used to describe the sentiment or polarity shift in a clause brought about by particular lexical items, called *valence shifters* (Polanyi and Zaenen, 2006). Three types of valence shifters are generally distinguished: NEGATOR (e.g., *not*), INTENSIFIER (e.g., *strongly*), and DIMINISHER (e.g., *barely*) (Polanyi and Zaenen, 2006; Kennedy and Inkpen, 2006; Andreevskaia *et al.*, 2007). The overall sentiment of a fragment of text is modeled as the interaction of these valence shifters. We consider these categories in a similar way. In contrast to sentiment analysis approaches, however, we are not only interested in their effect on overall sentiment. Rather, we model them as engendering a scalar shift on the scale associated with the embedded predication, whatever that scale may be, or as changing the polarity of the embedded predication. Accordingly, we distinguish two classes of VALENCE_SHIFTER predicates: SCALE_SHIFTER and POLARITY_SHIFTER. The former consists of NEGATOR, INTENSIFIER, DIMINISHER, and HEDGE subtypes and the latter of POSITIVE_SHIFTER and NEGATIVE_SHIFTER subtypes.

SCALE_SHIFTER type

As the name implies, the SCALE_SHIFTER predicates change the value of the embedded predication (and possibly other predications in its scope) on the relevant scale. Example (57) illustrates the effect of the DIMINISHER *barely*, which lowers the scalar modality values of the predications in its scope (em_2 signalled by *can* and e_1 signalled by *pay*) from 1.0 to 0.8 on the scale they are associated with.

(57) (a) Meanwhile, SCI TV can barely **pay** its cash interest bill, …(wsj_1206)

(b) $SCI\_TV(t_1) \wedge its\_cash\_interest\_bill(t_2) \wedge pay(e_1, WR, 0.8_{potential}, t_1, t_2) \wedge$
$can{:}\text{POTENTIAL}(em_2, WR, 0.8_{epistemic}, e_1) \wedge barely{:}\text{DIMINISHER}(em_3, WR, 1.0_{epistemic}, em_2)$

Example (58) illustrates the effect of the INTENSIFIER *strongly*. It increases the *epistemic* value associated with the predication in its scope (em_2) from 0.7 to 0.9. The *epistemic* value of its immediate embedded predication (*em*$_3$) does not change, since it is already at the right boundary of the scale and cannot be intensified further.

(58) (a) They strongly suggest that Justice's **prosecutions** of Drexel Burnham Lambert, Michael Milken and Princeton/Newport violated notions of fundamental fairness. (wsj_1149)

(b) $They(t_1) \wedge prosecution(e_1, \ldots) \wedge violate$:PROPOSITIONAL$(em_2, t_1, 0.9_{epistemic}, e_1 \ldots)$
$suggest$:SPECULATIVE$(em_3, WR, 1.0_{epistemic}, em_2) \wedge$
$strongly$:INTENSIFIER$(em_4, WR, 1.0_{epistemic}, em_3)$

In Example (59), the effect of the NEGATOR *no* is illustrated. Its effect is to invert the *epistemic* value of the embedded predication (*em*$_2$) with respect to the epistemic scale ($1.0 \rightarrow 0.0$). It further affects the value of the embedded predicationof *em*$_2$ (e_1) in a similar manner.

(59) (a) Thus far there is no indication that they have been **re-supplied** with Stingers or other anti-aircraft weapons. (wsj_2052)

(b) $they(t_1) \wedge Stingers(t_2) \wedge re\text{-}supply(e_1, WR, 0.3_{epistemic}, t_2, t_1) \wedge$
$indication$:DEMONSTRATIVE$(em_2, WR, 0.0_{epistemic}, e_1 \ldots) \wedge$
no:NEGATOR$(em_3, WR, 1.0_{epistemic}, em_2)$

Example (60) illustrates the effect of the HEDGE *essentially*. In contrast to the more inclusive use of the term HEDGE in pragmatics-oriented studies (e.g., Hyland (1998)), we restrict this category largely to the hedge definition of Lakoff (1972) and *attribute hedges* of Hyland (1998). This category of valence shifters involve fuzziness and include predicates such as *essentially*, *in general*, and *mostly*.

(60) (a) Mr. Bush himself essentially acknowledged that he and his aides were trying to **head off** criticism. (wsj_1920)

(b) $Mr.\_Bush(t_1) \wedge head\_off(e_1 \ldots) \wedge try$:INTENTIONAL$(em_2, t_1, 0.8_{epistemic}, \ldots, e_1) \wedge$
$acknowledge$:REPORTING$(em_3, WR, 0.8_{epistemic}, t_1, em_2) \wedge$
$essentially$:HEDGE$(em_4, WR, 1.0_{epistemic}, em_3)$

Rather than shifting or inverting the scalar modality value, the effect of hedges seems to be making the modality value of the embedded predication an interval, rather than a fixed point on the relevant scale, in other words, making it fuzzier. In fact, Lakoff (1972) models such expressions using fuzzy logic operations, such as *dilation*, which increases the fuzziness of a fuzzy set. Since we consider our scalar values to be approximations anyway, we take a simpler approach to hedges. If the modality value of the relevant embedded predication is at the positive or negative end of the scale, we model the fuzziness by decreasing or increasing the modality value (by 0.2). In other cases, the modality value is unchanged. For example, in the example above, the embedded predication indicated by *acknowledge* has a prior epistemic value of 1.0, since the acknowledgement event is factual. The hedge *essentially* acts as a diminisher, lowering the epistemic value to 0.8.

We only illustrated the scalar effect of SCALE_SHIFTER predicates in this section but did not highlight the underlying principles. We will discuss these principles in more detail in Section 3.3.

POLARITY_SHIFTER **type**

Polarity shifting predicates essentially change the polarity of the embedded predication of CAUSAL semantic category. There are only a few such predicates. Consider the sentence in Example (61a).

(61) (a) Maxtor said effects from discontinuing the line may have a positive effect on future earnings and revenue. (wsj_2332)

(b) $Maxtor(t_1) \wedge discontinue(e_1 \ldots) \wedge effect$:CAUSAL$(em_2,t_1,1.0_{epistemic},e_1,_) \wedge$
$effect$:CAUSAL$(em_3,t_1,1.0_{epistemic},em_2,\ldots) \wedge$
$positive$:POSITIVE_SHIFTER$(em_4,t_1,1.0,positive,em_3)$

(c) $positive\_effect$:CAUSE$(em_3,t_1,0.5_{epistemic},em_2,\ldots)$

The POSITIVE_SHIFTER adjective *positive* turns the polarity of the embedded predication indicated by *effect* (em_3 in 61b) from *neutral* to *positive* and allowing us to compose a new relation of positive CAUSE type, shown in (61c).

NEGATIVE_SHIFTER type of predicates act in a similar manner. However, they allow us to compose a relation of PREVENT type, instead of a CAUSE relation.

3.2.4 PROPOSITIONAL Category

Definition 15. An embedding predicate of PROPOSITIONAL type contributes to meaning at the basic propositional level. Predicates belonging to PROPOSITIONAL subtypes, SEMANTIC_ROLE and ASPECTUAL, link embedded predications with one of their semantic attributes. A generic PROPOSITIONAL predicate, on the other hand, essentially behaves like any atomic predicate semantically.

Let us begin with the generic PROPOSITIONAL predicates. These are largely due to our structural approach and to the fact that we assume that atomic predications can be indicated by nominal predicates, which implies that all verbal predicates that can take a nominalized object can be embedding predicates. Recall the earlier illustrative example, partially duplicated below.

(62) (a) President Bush will **veto** a bill **funding** the Departments of Labor, Education and Health and Human Services because it would allow federal funding of **abortions** for victims of rape and incest, the White House said. (wsj_2075)

(b) $President\_Bush(t_1) \wedge bill(t_2) \wedge federal(t_4) \wedge$
$Departments\_of\_Labor,\_Education\_and\_Health\_and\_Human\_Services(t_3) \wedge$
$the\_White\_House(t_5) \wedge victims\_of\_rape\_and\_incest(t_6) \wedge$
$veto(e_1,\ldots,t_1,t_2) \wedge fund(e_2,\ldots,t_2,t_3) \wedge abortion(e_3,\ldots) \wedge$
$funding$:PROPOSITIONAL$(em_5,t_5,1.0_{epistemic},t_4,e_3,t_6)$

The nominal predicate *funding* indicates an embedding predication with PROPOSITIONAL type (em_5). Two of its arguments are semantic terms while the other is an atomic predication indicated by *abortions*. On the other hand, the predication indicated by the related verbal predicate *fund* is represented as an atomic predication (e_2), since both of its arguments are semantic terms. Semantically, both predications refer to similar situations. In fact, the verb *fund* and the nominal *funding* have the same frame semantics in PropBank (Palmer *et al.*, 2005) and in NomBank (Meyers *et al.*, 2004b), respectively, where Arg1 (roughly, logical object) corresponds to the "thing financed". A thing financed can clearly be an entity or may refer to a situation or event (*fund schools* vs. *fund road construction*). In the current work, while these are treated as being structurally distinct predications, the final semantic interpretation must be able to accommodate the fact that they essentially refer to the same semantic situations. We view this as a separate post-processing step where the equivalence of an embedding predication of generic PROPOSITIONAL type with a semantic frame is established. The generic PROPOSITIONAL predicates are not explicitly encoded in the embedding dictionary, but are determined from structural constraints[11]. In that sense, this generic category acts as a catch-all category for embedding predicates that do not fit in one of the embedding categories.

Predicates belonging to two subtypes of PROPOSITIONAL category, SEMANTIC_ROLE and ASPECTUAL, serve at the propositional level, while their contribution is more specific than that made by the generic PROPOSITIONAL predicates.

Let Pr be an embedding predication signalled by a predicate P belonging to a PROPOSITIONAL subcategory and Pr_e the embedded predication ($Pr > Pr_e$). Let ATT be the semantic subcategory associated with the predicate P and Arg a logical argument of Pr different from Pr_e.

$$Pr := [\ldots Arg,\ Pr_e, \ldots]$$

[11] This will be described in detail in Section 5.2.1.

Then, *Arg* is a semantic attribute of Pr_e, indicated by ATT. The effect of this configuration on the embedded predication Pr_e can be given as follows:

$$Sem(P) = \textsc{att} \wedge (Pr > Pr_e) \Rightarrow \textsc{att}(Pr_e) = Arg$$

SEMANTIC_ROLE **type**

Embedding predicates of SEMANTIC ROLE type link the embedded predication with one of its semantic arguments and specifies its semantic role. In other words, ATT above refers to a semantic role. Consider the fragment in Example (63a). The verbal predicate (*undergo*) takes a nominalized predicate (*restructuring*) as its syntactic object. The other syntactic argument of the verbal predicate, *Spain's seventh largest bank*, serves as the semantic argument of the embedded predicate (*restructuring*) and has the semantic role PATIENT. The representation in Example (63b) corresponds to the intermediate interpretation where the predication corresponding to the nominal predicate *restructuring* (e_1) is underspecified in that it is not assigned any arguments. The representation in Example (63c) shows the final interpretation where the argument *Spain's seventh largest bank* is assigned as an argument of the nominal predicate (with PATIENT semantic role), due to the fact that *undergo* is a PROPOSITIONAL predicate with PATIENT subtype[12].

(63) (a) ... Spain's seventh largest bank is <u>undergoing</u> a tough **restructuring** that analysts say may be the first step toward the bank's privatization. (wsj_0616)

(b) $Spain's\_seventh\_largest\_bank(t_1) \wedge restructuring(e_1,\_) \wedge$
$undergo{:}\textsc{patient}(em_2, WR, 1.0_{epistemic}, e_1, t_1)$

(c) $Spain's\_seventh\_largest\_bank(t_1) \wedge restructuring(e_1,t_1) \wedge \textsc{patient}(e_1) = t_1$
$undergo{:}\textsc{patient}(em_2, WR, 1.0_{epistemic}, e_1, t_1)$

Verbs functioning in this way are numerous (e.g., *perform* corresponding to AGENT role, *experience* to EXPERIENCER role) (called "theta verbs" in Power (2007)). Derivational forms of these verbs may also function in the same way (e.g., *Spain's seventh largest bank's undergoing of a tough restructuring*).

In the NomBank project, some of these verbs are subsumed under the category of *support verbs* (mostly consisting of *light verbs*, such as *take*, *give*, *get*) and they are annotated as SUPPORT in the NomBank corpus (Meyers *et al.*, 2004b). They define a *support verb* as "a verb which takes at least two arguments NP_1 and XP_2 such that XP_2 is an argument of the head of NP_1" (Meyers

[12]This final interpretation is the interpretation given in the NomBank corpus (Meyers *et al.*, 2004b).

et al., 2004a). While they discuss the argument sharing aspect of support verbs, their semantic contribution to the propositional content is largely ignored.

The semantic roles we assume in the current work and example verbal predicates are presented in Table (3.1). It is interesting to note that the verb *involve* is highly ambiguous with respect to the semantic roles it can encode, as indicated in Power (2007).

Subcategory	**Examples**	**Subcategory**	**Examples**
AGENT	*perform, involve*	BENEFICIARY	*benefit, involve*
PATIENT	*suffer, undergo*	INSTRUMENT	*employ, involve*
EXPERIENCER	*involve, experience*	PURPOSE	*serve,involve*
MANNER	*involve,characterize*	TIME	*happen, occur*
LOCATION	*house, occur*		

Table 3.1: SEMANTIC_ROLE subcategories and example predicates

ASPECTUAL **type**

An ASPECTUAL type of embedding predicate specifies the *aspect* attribute of the event described by the embedding predication. Therefore, ATT above refers to aspect with this type of predicate. Aspect is generally considered a core category of verbal meaning (in addition to tense and modality). While we do not attempt a full interpretation of aspect within the scope of the current work, we make explicit aspectual information as it relates to embedding predicates. We adopt the aspectual classification in TimeML (Pustejovsky *et al.*, 2005b) as our basis in classification of ASPECTUAL types. The main aspectual classes specified in TimeML are: INITIATE, CULMINATE, TERMINATE, CONTINUE, REINITIATE. We illustrate CULMINATE, CONTINUE, and REINITIATE classes below:

(64) (a) In February 1989, when the Soviets said they had completed their **pullout**, the U.S. cut it further.(wsj_2052)

the_Soviets(t_1) ∧ *pullout*(e_1,t_1,*1.0*$_{epistemic}$,t_1) ∧

complete:CULMINATE(em_2,t_1,*1.0*$_{epistemic}$,t_1,e_1) ∧ ASPECT(e_1) = CULMINATE

(b) Options markets stopped **trading** in many securities. (wsj_2222)

Options_markets(t_1) ∧ *trade*(e_1,*WR*,*1.0*$_{epistemic}$,t_1) ∧

stop:TERMINATE(em_2,*WR*,*1.0*$_{epistemic}$,t_1,e_1) ∧ ASPECT(e_1) = TERMINATE

(c) As a result, he reignited the **inflation** that Mrs. Thatcher, through a long and costly effort, had subdued. (wsj_0571)

$he(t_1) \wedge inflation(e_1, WR, 1.0_{epistemic}, \text{-}) \wedge$

$reignite$:REINITIATE$(em_2, WR, 1.0_{epistemic}, t_1, e_1) \wedge$ ASPECT(e_1) = REINITIATE

In contrast to SEMANTIC ROLE predicates, the factuality effect of ASPECTUAL predicates is evident, and has been discussed in Saurí (2008). For example, in (64b), the trading event (e_1) is portrayed as happening in the past, and not at the current time.

3.3 Scopal Influence of Embedding Predications

In the previous section, we introduced the embedding predicate categorization and showed how an embedding predication affects the meaning of its predicational argument in the logical object position. To summarize, MODAL and VALENCE_SHIFTER types of predications cause meaning shifts in their predicational object argument by placing them on a modal scale or changing their value on the scale they are already associated with, or by affecting their polarity. We also hinted in discussing the SCALE_SHIFTER predicates that the scalar influence of some predicates may extend beyond immediate arguments to other predications in scope. In contrast, RELATIONAL and PROPOSITIONAL type of embedding predications do not have such effects.

In this section, we take the discussion a step further and examine how an embedding predication interacts with a predication further down in its scope, that is, a predication that is in its semantic scope but that is not its argument. In other words, what is the extent of its scopal influence? We will first illustrate such influence on several examples and then discuss the generalizations that we draw based on these examples and others in the Penn TreeBank corpus.

First, take the sentence in Example (65a). The semantic categories and prior scalar values associated with relevant embedding predicates in the sentence are given in (65b). Our representation is given in (65c) and it is also graphically illustrated in Figure (3.7). Note that the predication indicated by the negative particle *don't* is under the scope of that indicated by the epistemic predicate *believe*, rather than vice versa as might be expected. This is due to negative raising (Horn, 1989) associated with the verbal predicate *believe*.

(65) (a) <u>Maybe</u> people <u>don't</u> <u>believe</u> I <u>want</u> to **give** this money **away**. (wsj_1409)

(b) *maybe(*SPECULATIVE*,0.5), believe(*EPISTEMIC*, 0.8), want(*VOLITIVE*, 1.0)*

(c) $people(t_1) \wedge I(t_2) \wedge this\_money(t_3) \wedge give\_away(e_1, t_1, 0.0_{volitive}, t_2, t_3) \wedge$

$want$:VOLITIVE$(em_2, t_1, 0.2_{epistemic}, t_2, e_1) \wedge$

$don't$:NEGATOR$(em_3, t_1, 0.5_{epistemic}, em_2) \wedge$

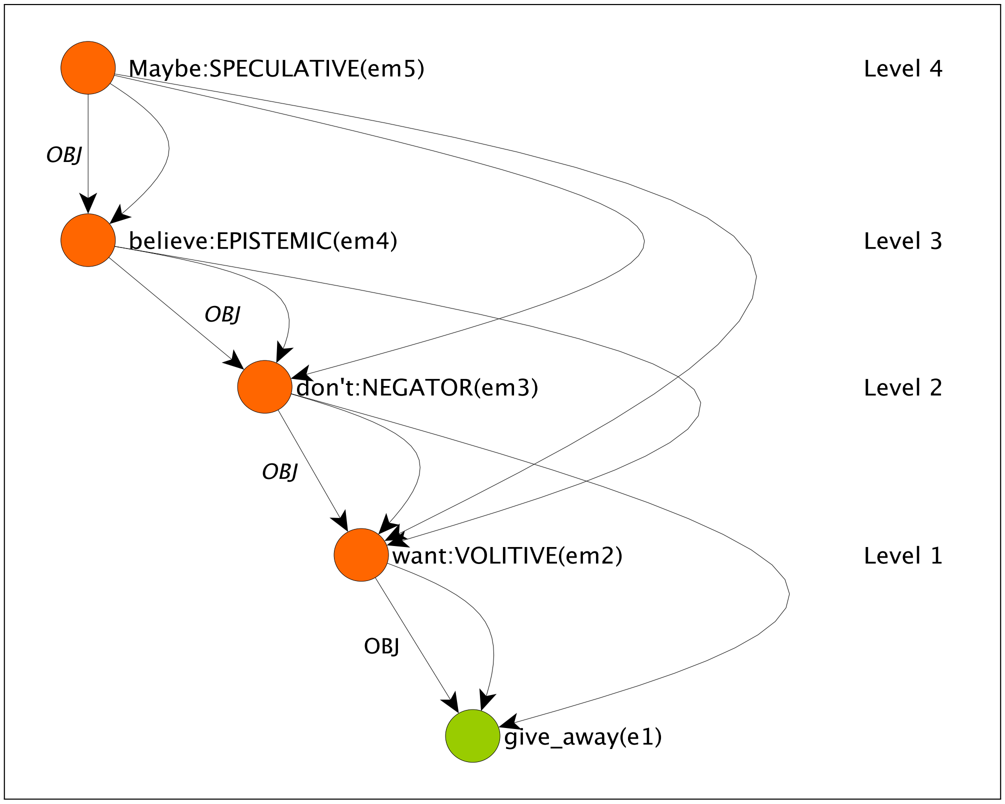

Figure 3.7: Scope relations and interactions relating to Example (65). Curved lines illustrate scopal influence.

$believe$:EPISTEMIC$(em_4, WR, 0.5_{epistemic}, t_1, em_3) \wedge$

$maybe$:EPISTEMIC$(em_5, WR, 1.0_{epistemic}, em_4)$

Going bottom-up, the first predicate of real interest here in terms of embedding is the VOLITIVE predicate *want*. At this level of interpretation (marked as Level 1 in Figure (3.7)), two scalar operations take place. First, its argument (e_1) is placed on the VOLITIVE scale and assigned a volitive modality value of 1. Second, the VOLITIVE predication itself (em_2) is implicitly associated with the EPISTEMIC scale and is assigned a value of 1, due to being unmarked at this level. More informally, we represent that at this level, "the act of the writer's giving away money" is *desired*, and "the act of the writer's wanting to give away money" is a *fact*.

Next up is the negative particle *don't*. As a predicate of NEGATOR type, its function is one of scale shifting of the predication in scope and it can interact with any modal scale. The predications in its scope (em_2 and em_1) are on the EPISTEMIC and VOLITIVE scales, respectively, and both are inverted with respect to their relevant scale ($MV_{epistemic}(em_2) = 0$ and $MV_{volitive}(e_1) = 0$). That is to say, "the act of giving away money" is now *undesired*, while "the act of wanting" is a *counter-fact*. At this level, the embedding predication indicated by the negative particle is also constructed (em_3)

and is implicitly placed on the epistemic scale with the value of 1. That is, "the act of not wanting" becomes a *fact*.

Moving another level up (to Level 3), we encounter the EPISTEMIC predicate *believe*, which has been assigned the prior epistemic value of 0.8. This predicate also introduces the source (or the belief holder) in its logical subject position (*people*). In its scope are the embedding predications em_3 and em_2 as well as the atomic e_1. The effect of introducing the source at this level is to associate the predications in scope with this source. In terms of scales, the effect of *believe* extends down to em_3 and em_2 only, since these predications are associated with the EPISTEMIC scale. The atomic e_1 is not affected since it is associated with the VOLITIVE scale. The epistemic value of em_3 is lowered from 1 to 0.8 (the prior epistemic value of *believe*). On the other hand, the scalar value of em_2, which at this point was 0, is shifted by the same amount in the reverse direction to 0.2. More informally, the effect of these operations can be paraphrased as follows:

- "The case that I do not want to give this money away" has *high probability* according to *people* (0.8). (em_3)
- "The case that I want to give this money away" is *doubtful* according to *people* (0.2). (em_2)

Needless to say, at this level, also, a predication associated with *believe* (em_4) is constructed with the default epistemic value of 1.0, which corresponds to "the act of the people believing" according to the author.

Finally, at the top level (Level 4), the SPECULATIVE predicate *maybe* (with the prior epistemic value of 0.5) has in its scope the embedding predications em_4, em_3, and em_2 and the atomic e_1. Its effect extends down to all embedding predications, since they are all on the epistemic scale already. em_4 gets the epistemic value percolated down from the predicate *maybe* (0.5). The epistemic value of em_3 is lowered from 0.8 to 0.5 while that of em_2 remains unchanged at 0.2. Additionally, as expected, a predication associated with *maybe* (em_5) is also constructed with epistemic value of 1. More informally, at the end, the result of operations can be paraphrased as:

- "That people believe that I don't want to give this money" is *possible* (0.5) (em_4)
- "That I do not want to give this money away", according to *people*, is *possible* (0.5). (em_3)
- "That I want to give this money away", according to *people*, is *doubtful* (0.2). (em_2)

This example illustrated the scopal influence of epistemic predicates, which extends to predications in the scope that are also on the epistemic scale.

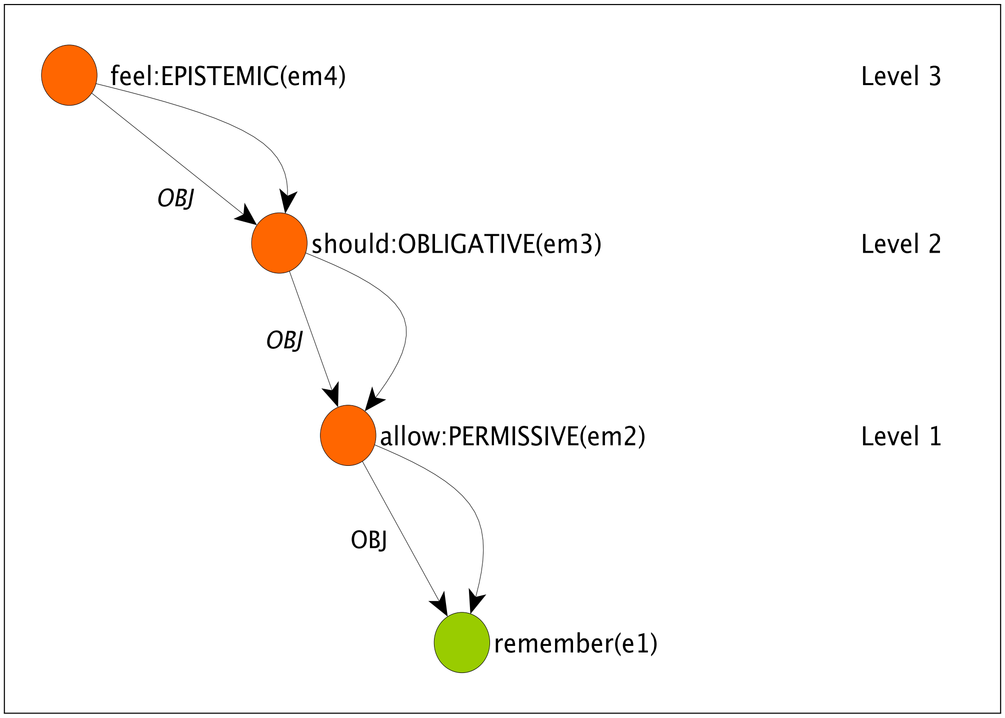

Figure 3.8: Predications and compositional interactions relating to Example (66). Curved lines indicate scopal influence.

Now, let us consider the sentence in Example (66a), whose corresponding predications and relevant compositional processes are illustrated in Figure (3.8).

(66) (a) I <u>feel</u> people <u>should</u> be <u>allowed</u> to **remember** players . . . (wsj_0214)

(b) *feel(*EPISTEMIC*,0.8), should(*OBLIGATIVE*,0.8), allow(*PERMISSIVE*, 0.6)*

(c) $I(t_1) \wedge people(t_2) \wedge players(t_3) \wedge remember(e_1,t_1,0.6_{deontic},\ t_2,t_3) \wedge$

$allow$:PERMISSIVE$(em_2,t_1,0.8_{deontic},e_1) \wedge$

$should$:OBLIGATIVE$(em_3,t_1,0.8_{epistemic},em_2) \wedge$

$feel$:EPISTEMIC$(em_4,WR,1.0_{epistemic},t_1,em_3)$

As expected, at Level 1, the deontic predicate *allow* opens up the DEONTIC scale for its argument (e_1) and assigns it its lexically defined deontic value (0.6). Going one level up, we encounter another deontic predicate (*should*) (prior deontic value of 0.8). While it interacts with its predicational argument (em_2), its influence does not extend to the other predication in its scope (the atomic e_1), even though this atomic predication is on the DEONTIC scale as well. In other words, with the modal auxiliary *should*, the obligation lies not in the act of remembering but in the deontic act of permitting to remember. Therefore, the deontic value of e_1 (remembering) does not change, while the deontic value of em_2 is set to 0.8. Finally, at the top level, the scopal influence of the epistemic predicate *feel* extends only to its predicational argument em_3 ("the act of being obligated to allow ..."), whose epistemic value is lowered from 1 to 0.8. There is no epistemic effect due to the predicate *feel* on the predication indicated by *allow* (em_2) and further down. This example illustrates that the scopal influence of deontic predications do not extend beyond their object arguments.

Finally, consider the sentence in Example (67a). Here, we will ignore the fragment *a lawyer for Burlington* and focus specifically on the interaction between the sc reporting predicate *said* and embedding predication corresponding to the complement clause *the Beebes' symptoms were not related to carpetings* (em_2). Does the scopal influence of the predicate *said* extend to em_2? Since the embedded predication in question (em_2) as well as the only intermediate predication (em_3) are on the epistemic scale, it may be expected that *said* would influence the scalar value of em_2, similar to the earlier sentence in Example (65). However, this is not the case, since the source of the intermediate predication is different than that of both em_2 and that of the current predication (em_4). We say that the introduction of an intermediate source (*the company*) serves to *block* scopal influence.

(67) (a) Anthony J. Iaciofano, a lawyer for Burlington, <u>said</u> the company <u>believes</u> the Beebes' symptoms were <u>not</u> **related** to the carpetings ... (wsj_1946)

(b) $the\_company(t_1) \wedge Anthony\_J.\_Iaciofano(t_2) \wedge relate(e_1, t_1, 0.2_{epistemic}, \ldots) \wedge$
not:NEGATOR$(em_2, t_1, 0.8_{epistemic}, e_1) \wedge$
$believe$:EPISTEMIC$(em_3, t_2, 1.0_{epistemic}, t_1, em_2) \wedge$
say:EPISTEMIC$(em_4, WR, 1.0_{epistemic}, t_2, em_3)$

These examples as well as analysis of other sentences from the corpus reveal several generalizations with regard to scopal influence of embedding predications. We summarize these generalizations as follows:

1. Initially, every predication is assigned to the EPISTEMIC scale with the value of 1.0.

2. A MODAL predicate places its logical object argument on its corresponding scale. The corresponding scale for all EVIDENTIAL and EPISTEMIC types is the epistemic scale, and that for the DEONTIC subtypes is the deontic scale. All other MODAL types open up their own specific scale.

3. A SCALE_SHIFTER predicate does not introduce a new scale but changes the existing scalar value of its predicational object argument.

4. A POLARITY_SHIFTER predicate, if it embeds a CAUSAL predication, combines with this embedded predication to create a new predication of CAUSE or PREVENT type (for POSITIVE_SHIFTER

and NEGATIVE_SHIFTER subtypes, respectively).

5. An embedding predicate has scalar influence on an embedded predication which is in its scope but not its argument if there is no scopal influence blocking due to intermediate source-introducing predicates and one of the following holds:

   (a) the embedding predicate type is one that is primarily associated with the epistemic scale (that is, a subtype of EPISTEMIC or EVIDENTIAL types) and the intermediate predications are either of SCALE_SHIFTER type or are of a type primarily associated with the epistemic scale OR

   (b) the embedding predicate is of SCALE_SHIFTER type and at most a single intermediate predication has a modal type OR

   (c) the embedding predicate is a non-EPISTEMIC or non-EVIDENTIAL type and intermediate predications are only of SCALE_SHIFTER type.

This last generalization can be defined more formally as follows:

Definition 16. Let P be an embedding predicate associated with embedding predication Pr, Pr_e an embedded predication and Pr_i the set of all intermediate predications between Pr and Pr_e. Pr interacts with Pr_e iff

- $Sem(P) \in \{\text{EVIDENTIAL}, \text{EPISTEMIC}\} \wedge (\forall Pr_i : Sem(Pr_i) \in \{\text{EVIDENTIAL}, \text{EPISTEMIC}, \text{SCALE\_SHIFTER}\})$ (5a)

- $Sem(P) \in \{\text{SCALE\_SHIFTER}\} \wedge ((\forall Pr_i : Sem(Pr_i) = \text{SCALE\_SHIFTER}) \vee (\exists! Pr_i : Sem(Pr_i) \in \{\text{MODAL}\}))$ (5b)

- $Sem(P) \in \{\text{MODAL}\} \wedge Sem(P) \notin \{\text{EVIDENTIAL,EPISTEMIC}\} \wedge (\forall Pr_i : Sem(Pr_i) = \text{SCALE\_SHIFTER})$ (5c)

3.3.1 Embedding Predications and Factuality

An important question that comes up in the context of scopal influence of embedding predications is the issue of factuality; that is, what is the epistemic or factual status of a particular event (or any kind of predication) with respect to different sources mentioned in the text? Does the author see the particular event as a fact, a counter-fact, or a possibility? How about the other cognizers (sources) mentioned in the text? We have outlined the work of Saurí (2008) earlier, which can be considered the most comprehensive work on factuality from a computational perspective. Since her work overlaps with the current work to some extent, in this section, I will take a closer look at an example to highlight the factuality inferences she draws from the sentence and whether our predications can predict some of these inferences.

The sentence I will examine is given in Example (68a) and is taken from Saurí (2008). The corresponding predications we construct are given in (68b). In Saurí (2008), factuality values are assigned to event expressions, in this case, to verbal predicates *aware*, *know*, and *is*, which indicate predications em_3, em_2, and e_1 below, respectively.

(68) (a) Mary is <u>not</u> <u>aware</u> that John <u>knows</u> that Paul is the **father**.

(b) $Mary(t_1) \wedge John(t_2) \wedge Paul(t_3) \wedge$
$father(e_1, t_2, 1.0_{epistemic}, t_3) \wedge know{:}\text{EPISTEMIC}(em_2, t_1, 0.5_{epistemic}, t_2, e_1) \wedge$
$aware{:}\text{EPISTEMIC}(em_3, WR, 0.0_{epistemic}, t_1, em_2) \wedge$
$not{:}\text{NEGATOR}(em_4, WR, 1.0_{epistemic}, em_3)$

The factuality profiles for these predications are given as follows in Saurí (2008). We explain the inference and indicate whether our representation predicts the inference.

- $f(em_3, WR) =$ CT- (According to the writer, Mary's being aware of John's knowledge of Paul's paternity, is a *counter-fact*. In our representation, em_3 has the epistemic value of 0, which corresponds to a counter-fact. Therefore, we predict the inference correctly.)

- $f(em_2, WR) =$ CT+ (According to the writer, John's knowing of Paul's paternity, is a fact. Our representation does not make a prediction regarding this factuality.)

- $f(em_2, t_1) =$ Uu (According to Mary, John's knowing of Paul's paternity, is unknown. This corresponds to em_2, which has the epistemic value of 0.5, which corresponds to a possibility as well as uncertainty in our framework. So, we can consider our prediction partially correct.)

- $f(e_1, WR) =$ CT+ (Paul's paternity according to the writer is a fact. Our representation does not make a prediction regarding this.)

- $f(e_1, t_1) =$ Uu (Paul's paternity according to Mary is unknown. Again, our representation does not make a prediction regarding this.)

- $f(e_1, t_2) =$ CT+ (Paul's paternity according to John is a fact. This is precisely the information encoded by e_1).

- *f*(e_1,t_{2_1}) = Uu (Paul's paternity according to Mary's knowledge of John's state of information is unknown. Again, we do not make a prediction.)

We reiterate that our goal in the current work is not to make these factuality inferences fully but to represent what is explicitly stated in text in a minimal manner while also making such factuality inferences possible. In this spirit, each predication is only associated with its immediate source (or cognizer), rather than being evaluated with respect to all nested sources in the context (like Saurí (2008) does). That is to say, the information *Paul is the father* (e_1) is only represented with respect to *John* and not to *Mary* or to the writer. To accommodate evaluation at each level, it would seem necessary to extend the current framework such that the epistemic effect of each embedding predicate is lexically encoded not only with respect to the cognizer but also with respect to the anchor (the source that presents the act of cognition). For example, for the embedding predicate *know*, this would entail encoding two separate epistemic values rather than one. The additional value would be the epistemic status according to the anchor and, in the case of *know*, would also be 1.0. That is to say, the predicate *know* commits both the cognizer (the entity who knows) and the anchor (the entity who reports the knowing) to the factuality of the predication expressed by the complement clause of the predicate. This level of factuality inference is beyond the scope of the current work.

Another related issue is the factuality of predications that are on non-epistemic scales. Recall the earlier sentence, "Maybe people don't believe that I want to give this money away." Our representation encoded that the atomic predication indicated by the predicate *give away* was on the VOLITIVE scale with value 0. In the factuality framework of Saurí (2008), the factuality of this predication would be Uu. The mere fact that we place the predication on the volitive scale, rather than the epistemic scale, allows us to make the correct prediction that the epistemic status of the predication is unspecific.

3.4 Towards Discourse Interpretation

So far, our discussion focused on intra-sentential embedding predications, even though we proposed the embedding framework as the basis for moving towards discourse interpretation earlier. In this section, we discuss the mechanisms that play a role in this move. Before delving into the mechanisms and issues involved, it is necessary to define more precisely what is meant by discourse interpretation. We identify the two major, interrelated components of discourse interpretation as:

- Moving to the inter-sentential level in semantic interpretation.
- Linking semantic content encoded by individual predications to reveal the discourse structure and the rhetorical contribution of individual predications and sentences.

Inter-sentential semantic interpretation necessitates resolving *reference relations*, described in Section 2.2.2. The specific tasks include anaphora resolution for entities and predications (individual and event anaphora), identification of coreferring expressions, resolution of VP ellipsis, as well as identifying implicit predicational arguments that are locally uninstantiated but have referents in the wider discourse context (*null instantiation*). The sentence in Example (69a), taken from Ruppenhofer *et al.* (2010), illustrates an entity anaphora indicated by the pronoun *he*, which should be resolved to *the defendant* in the previous sentence. Furthermore, *murder* in the first sentence fills an argument role for the verb *cleared* in the second. The underspecified semantic representation before resolution of reference relations is given in (69b), where underspecified arguments are represented as underscore (_). The coreference chain between the personal pronoun *he* and *the defendant* is denoted as the COREF relation in (69c) and the final semantic interpretation after the resolution of this coreference chain in (69d). Note that in the end, the logical subjects of predications e_1 and e_2 remain unresolved, since they are not mentioned in the text (the agent for trying and clearing the defendant).

(69) (a) In a lengthy court case *the defendant* was tried for *murder*. In the end, <u>he</u> was <u>cleared</u>.

(b) $the\_defendant(t_1) \wedge murder(t_2) \wedge he(PPRO_3) \wedge$
$try(e_1, WR, 1.0_{epistemic}, \_, t_1, t_2) \wedge clear(e_2, WR, 1.0_{epistemic}, \_, PPRO_3, \_)$

(c) $COREF(PPRO_3, t_1)$

(d) $try(e_1, WR, 1.0_{epistemic}, \_, t_1, t_2) \wedge clear(e_2, WR, 1.0_{epistemic}, \_, t_1, t_2)$

The tasks concerning reference relations have been largely in the purview of coreference resolution and semantic role labeling research in NLP/CL. We mentioned a coreference resolution shared task earlier (Pradhan, 2011). In addition, as part of the SemEval semantic evaluation competition, a semantic evaluation challenge that focuses on null instantiation has been proposed, although challenge participation was limited (SemEval'2010 Task 9: Linking Events to Their Participants in Discourse (Ruppenhofer *et al.*, 2010)).

Linking semantic content to reveal discourse structure and rhetorical roles, on the other hand, has been the focus of top-down discourse analysis research, as discussed earlier in Section 2.2.2, the main tasks being *discourse chunking* (or *segmentation*) and *discourse parsing*, the former concerned with identifying internally coherent discourse segments and the latter with identifying the discourse relations that hold between segments. In general, explicit representation of semantic content of discourse segments has been ignored and more attention has been paid to the nature of discourse relations between discourse segments and how the discourse should be structured. In the current work, our goal is not to present state-of-the-art algorithms for discourse chunking or discourse parsing[13], but rather to investigate whether and how discourse structure and rhetorical contribution of discourse segments and sentences emerge from the lower level predications (in particular embedding ones). It is also important to note that the relationship between discourse relations and lower level phenomena may not be unidirectional. From an opposite perspective, the effect of discourse relations on resolving lower level referential phenomena, in particular, anaphora, VP ellipsis, etc., has also been shown (Asher and Lascarides, 2003). An approach to discourse interpretation would benefit from exploring the interactions between these components and levels.

While discourse analysis in the general sense emphasizes relational aspects of discourse, in discourse models focusing on scientific text, the focus seems to be firmly on identifying the role of a sentence with respect to global argumentation or what function they serve in scientific investigation[14]. Furthermore, the rhetorical roles identified in such scientific discourse theories are quite different than the relation types often discussed in general discourse theories.

In the light of these observations, how does our framework take such discourse phenomena into account? Let me begin with discourse relations. In our framework, discourse relations do not have a special status and simply correspond to predications with RELATIONAL types. Therefore, rather than holding between arbitrary textual segments, they hold between abstract semantic objects (generally predications)[15]. As has been shown in many earlier examples, simpler discourse relations are intra-sentential and are indicated explicitly by embedding predicates of various types (most importantly, subordinating and coordinating conjunctions, discourse adverbials). Our embedding categorization subsumes classes, adapted largely from PDTB, that are discursive in nature, specifically RELATIONAL subtypes. Therefore, identification of intra-sentential discourse relations as well as linking of semantic predications that are intra-sentential is a natural extension of the mechanisms discussed so far, they are treated just as any other embedding predication. The questions more pertinent for this section

[13]For a survey on these aspects of discourse analysis, Wellner (2009) is a good reference.

[14]Very recent BioDRB corpus (Prasad *et al.*, 2011) is an exception, taking its basic notions from PDTB (Prasad *et al.*, 2008).

[15]Resolving the correspondence between a predication and its textual span is seen as an orthogonal problem to discourse interpretation. Our view that discourse relations hold between predications rather than textual segments is similar to and is consistent with head-driven discourse parsing approaches advocated by Baldridge and Lascarides (2005) and Wellner (2009) (among others) which model discourse relations as holding between the lexical heads of discourse segments.

are:

1. How can we identify such links across sentences (inter-sentential discourse relations)?
2. What are the contributions of embedding predications discussed so far on identifying various components of such relations?
3. Can discourse relations assist in resolving lower level semantic phenomena?

The main mechanism to answer the first question is coreference resolution. Semantic typing of embedding predications as well as their scalar value and source features allow us to draw inferences to assist in answering the second question. In answering the third question, we will discuss how knowledge of discourse relations can help in narrowing down the candidates for coreference resolution.

We begin by discussing one area where MODAL predications are relevant in the context of discourse relations, namely, attribution, in the next section. Then, we illustrate the main issues and principles in moving towards discourse interpretation by analyzing two discourses in great detail. The first discourse is a paragraph from a news article from the PTB corpus and the analysis is guided by the PDTB discourse relation annotations for the given paragraph. The second discourse is a biomedical abstract from the GENIA event corpus and the analysis is guided by the meta-knowledge model annotations (Thompson *et al.*, 2011) for the given abstract as well as argumentative zoning scheme of Teufel *et al.* (2009). As a caveat, note that we have not operationalized all the mechanisms discussed in these analyses; however, we believe that it is important to clarify the main issues and principles first and that integrating them into the framework incrementally is a matter of implementation. After the analyses are presented, in Section 3.4.4, we summarize the main issues and findings that emerge from the analyses.

3.4.1 Discourse Relation Attribution and MODAL Predications

Attribution refers to the source or ownership of an abstract semantic object (a predication), often indicated by embedding constructions. The role of attribution in discourse relations and parsing as well as its effect on factuality of discourse segments has been demonstrated (Dinesh *et al.*, 2005; Prasad *et al.*, 2007; Danlos and Rambow, 2011). Attribution is an area where MODAL predications are significant to the interpretation of discourse. For example, consider the sentences in Example (70a-b) (Danlos and Rambow, 2011). Our representation for Example (70b) is presented in (70c).

(70) (a) (Fred will go to Dax for Christmas)$_\alpha$. Afterwards (he will go to Pau)$_\beta$.

(b) (Fred will go to Dax for Christmas)$_\alpha$. Jane claims that afterwards (he will go to Pau)$_\beta$.

(c) $Pau(t_1) \wedge Jane(t_2) \wedge he(PPRO_3) \wedge$
$go(e_1,t_2,0.5_{epistemic},PPRO_3,t_1) \wedge will$:ASSUMPTIVE$(em_2,t_2,0.5_{epistemic},e_1) \wedge$
$claim$:SPECULATIVE$(em_3,WR,1.0_{epistemic},t_2,em_2)$

The work of Danlos and Rambow (2011) is within the SDRT framework, which proposes the notion of *veridical discourse relations*. A discourse relation is considered *veridical* if the semantic contents of its arguments are also true. Such relations include Narration, Contrast, Elaboration, among others, in the SDRT inventory of discourse relations. In the sentences in Example (70), a Narration relation holds between the segments α and β. As predicted, in Example (70a), the writer commits to the veridicality of the segment β. In contrast, in Example (70b), the segment β is attributed to Jane and due to semantics of the verb *claim*, the factuality inference is that the writer does not commit to the veridicality of this segment, which is not predicted based on the discourse relation type. Based on such examples, Danlos and Rambow (2011) point out that it is necessary to evaluate the veridicality of discourse segments according to different sources (Jane as well as the writer in Example (70)) and that attribution of discourse segments should be annotated. This position is clearly similar to that of Saurí (2008), who was not specifically concerned with discourse relations or segments, but proposed evaluation of event factuality with respect to sources indicated in text.

Whereas attribution is sometimes considered a specific type of discourse relation (for example, in RST (Mann and Thompson, 1988) and GraphBank (Wolf and Gibson, 2005)), in PDTB, it is annotated with respect to discourse relations and their arguments, albeit, at a shallow level. Some nuances with regard to attribution are ignored in PDTB annotation (as discussed in Saurí and Pustejovsky (2009) and Danlos and Rambow (2011)); however, it remains the only attribution-annotated discourse relation corpus at this time.

Within the embedding framework, source and scalar value features of predications allow us to make predictions regarding attribution features. For example, in Example (70c), the predication corresponding to the top level element within the segment β (em_2) has epistemic value of 0.5, indicating that the writer does not commit to it as a fact. We discuss the task of discourse attribution resolution in more detail in Section 6.4.

3.4.2 Discourse Interpretation of a Paragraph of a News Article

In this section, we provide an analysis of the paragraph in Figure (3.9) from a discourse interpretation point of view. We start with the analysis of the first sentence, shown in Example (71). Since this is the first sentence of our discourse and no referential relations are present, its semantic representation

...Mr. Greenspan's decision to keep quiet also prompted a near-mutiny within the Fed's ranks.
A "senior Fed official" spoke on Saturday after the market swoon to both the Washington Post and the New York Times, saying the Fed was prepared to provide as much credit as the markets needed.
The statement angered Chairman Greenspan, but it was greeted with applause by the Bush administration and the financial markets.
And, while the mutinous Fed member hasn't gone public, some Fed governors, most notably Vice Chairman Manuel Johnson, are known to have disagreed with the chairman's decision to remain silent. ...

Figure 3.9: A paragraph from wsj_0598

is not much different than analyses provided in earlier sections. Simply atomic and embedding predications are identified. Some arguments are omitted for readability.

(71) (a) Mr. Greenspan's decision to keep **quiet** also prompted a **near-mutiny** within the Fed's ranks. (S1)

(b) $\textit{Mr. Greenspan}(t_1) \wedge \textit{quiet}(e_1, WR, 1.0_{epistemic}, t_1) \wedge$
$\textit{keep}:\text{CONTINUE}(em_2, WR, 1.0_{intentional}, t_1, e_1) \wedge \textit{near-mutiny}(e_3, WR, 1.0_{epistemic} \ldots) \wedge$
$\textit{decision}:\text{INTENTIONAL}(em_4, WR, 1.0_{epistemic}, t_1, em_2) \wedge$
$\textit{prompt}:\text{CAUSE}(em_5, WR, 1.0_{epistemic}, em_4, e_3)$

We note two discourse-related phenomena here: (a) the discourse connective *also* links this sentence to prior discourse context (indicating a CONJUNCTION type of discourse relation), which we ignored to keep the example manageable, (b) there is an embedding predication (em_5) indicated by *prompt*, which corresponds to an intra-sentential discourse relation of CAUSE type[16]. CAUSE is a *veridical* discourse relation (Asher and Lascarides, 2003), indicating that both its arguments are factual, unless, of course, the connective and its arguments are embedded in a matrix clause which may alter their factuality (a modal attribution context for example). The fact that both arguments of this embedding predication (em_4 and e_3) are on the epistemic scale with value 1.0 seems to allow us to infer this type of discourse relation between them.

The second sentence and its interpretation are given in Example (72). The intra-sentential interpretation given in (72b) is relatively straightforward. The more interesting question that arises is whether and how semantic content of this sentence is related to that of the previous sentence. We notice that there is no explicit discourse connective that indicates a link between these sentences.

[16]Of course, it must be noted again that this particular relation is not considered a *discourse relation* in PDTB, since it is not indicated by an explicit discourse connective or an alternate lexicalization.

However, it is possible to speak of an implicit relation: this sentence *exemplifies* the situation described in the previous sentence. In fact, adding a sentence-initial *For example* would not change the meaning. Therefore, we can characterize the relation between the two sentences as one of INSTANTIATION. In the PDTB framework, the relation is simply between the textual spans of this and the previous sentence, whereas we represent it as an embedding predication (em_{12} holds between the top level predications from these sentences, em_5 and em_{11}). The lack of an explicit indicator for this relation is shown with the keyword IMPLICIT[17].

(72) (a) A "senior Fed official" **spoke** on Saturday after the market **swoon** to both the Washington Post and the New York Times, saying the Fed was prepared to **provide** as much credit as the markets needed. (S2)

(b) *A_"senior_Fed_official"(t₂)* $\wedge$ *the_Fed(t₃)* $\wedge$ $speak(e_6, WR, 1.0_{epistemic}, t_2 \dots) \wedge$
$swoon(e_7, WR, 1.0_{epistemic} \dots) \wedge provide(e_8, t_2, 1.0_{intentional}, t_3, \dots) \wedge$
prepared:INTENTIONAL$(em_9, t_2, 1.0_{epistemic}, em_3, e_8) \wedge$
say:REPORTING$(em_{10}, WR, 1.0_{epistemic}, t_2, em_9) \wedge$
IMPLICIT:CONJUNCTION$(em_{11}, WR, 1.0_{epistemic}, e_6, em_{10})$

(c) *IMPLICIT*:INSTANTIATION$(em_{12}, WR, 1.0_{epistemic}, em_{11}, em_5)$

How can the discourse relation between two sentences be inferred in the absence of an explicit connective? One method could be defeasible reasoning. In the SDRT framework, Asher and Lascarides (2003) assume that a Narration relation holds between two adjacent sentences (analogous to EXPANSION relation in the embedding categorization), while such interpretation can be overridden based on explicit discourse cues. INSTANTIATION is a subtype of EXPANSION, so in this case, this simple principle seems to work to some extent, however, it is not precise enough. Since the top level predications in both sentences (em_5 and em_{11}) are factual, it may be inferred that the discourse relation between them should be a veridical one, even though in this particular case, this principle does not help narrow the relation down, since Narration is veridical, as well. To precisely identify the discourse relation, it seems necessary to uncover the set-instance relation between the phrase *Fed's ranks* in the first sentence and the indefinite *A "senior Fed official"* in the second sentence as well as the fact that situation described in the second sentence is temporally subsumed by the situation described in the first. This type of world and temporal knowledge is beyond the scope of this work. However, similarly to Asher and Lascarides (2003), in the absence of explicit discourse

[17]The lack of a specific trigger for the top level predication in this sentence, em_{11}, is also indicated by the keyword IMPLICIT. This predication is triggered by the infinitival complementation.

connectives, we can assume that an EXPANSION relation holds between top level predications in adjacent sentences.

The third sentence in Example (73a) is interesting in more ways, since it brings coreference into the picture. First, note that an intra-sentential discourse relation is expressed with the coordinating conjunction *but*, most commonly associated with the CONTRAST type of discourse relation. CONTRAST is also a veridical relation, and the fact that both arguments of the corresponding predication are factual (on the epistemic scale with the value of 1.0) seem to allow and predict this discourse relation. In other words, this discourse relation emerges from its predicational arguments. More importantly, to interpret this sentence, it is necessary to link the definite noun phrase *the statement* and the pronominal anaphor *it* to their coreferring expressions. Intra-sententially, it may be possible to link the pronominal anaphor *it* to the definite NP *the statement*. However, to resolve the referent of *the statement*, it is necessary to analyze prior discourse. One thing that helps us predict the referring expression here is that the predicate *statement* is typed as a REPORTING predicate. Therefore, we can restrict the referring expression to be an embedding predication of the same type. Analyzing the previous sentence, it is possible to infer that *the statement* (and, therefore, *it*, as well) refer to embedding predication em_{10}. The changes in representation from pre- to post-coreference resolution is illustrated in (73b) and (73d), respectively. Anaphoric expressions are denoted by their lexical category (DPRO for demonstrative pronoun and DET for determiner). The coreference chain is illustrated in (73c).

(73) (a) The statement angered Chairman Greenspan, but it was **greeted** with applause by the Bush administration and the financial markets. (S3)

(b) $\mathit{Chairman\_Greenspan}(t_4) \wedge \mathit{The\_statement}{:}\textsc{reporting}(DET_5) \wedge it(DPRO_6) \wedge$
$anger{:}\textsc{propositional}(em_{13}, WR, 1.0_{epistemic}, DET_5, t_4) \wedge$
$greet(e_{14}, WR, 1.0_{epistemic}, \ldots, DPRO_6) \wedge but{:}\textsc{contrast}(em_{15}, WR, 1.0_{epistemic}, em_{13}, em_{14})$

(c) $COREF(DPRO_6, DET_5, em_{10})$

(d) $anger{:}\textsc{propositional}(em_{13}, WR, 1.0_{epistemic}, em_{10}, t_4) \wedge$
$greet(e_{14}, WR, 1.0_{epistemic}, \ldots, em_{10}) \wedge but{:}\textsc{contrast}(em_{15}, WR, 1.0_{epistemic}, em_{13}, em_{14})$
$IMPLICIT{:}\textsc{cause}(em_{16}, WR, 1.0_{epistemic}, em_{11}, em_{15})$

It is further necessary to identify the relation between the current sentence and the previous one. As with the previous sentence, there is no explicit discourse connective that links two sentences. It seems possible to characterize the implicit relation between these two sentences as a CAUSE relation, due to the coreference relation between the sentences and lexical semantics of the verbs *anger* and the verbal phrase *greet with applause*, which incorporate causation. On the other hand, in PDTB, an EntRel relation is annotated between the sentences for this instance, which entails that the sentences are related by the sake of being about the same entities. We can reach the same conclusion since we found the antecedent for the anaphoric expressions of the current sentence in the previous sentence.

In the fourth and the last sentence in our short discourse, we ignore the fragment "most notably Vice Chairman Manuel Johnson" for brevity. There are a number of anaphoric expressions in this sentence, "the mutinous Fed member", "the chairman", and "the chairman's decision to remain silent". The coreference chains for these expressions are given in (74c). Pre-coreference resolution representation is presented in (74b), and the predications that are updated due to coreference resolution in (74d).

(74) (a) And, while the mutinous Fed member hasn't <u>gone</u> **public**, some Fed governors, most notably Vice Chairman Manuel Johnson, are <u>known</u> to have <u>disagreed</u> with the chairman's <u>decision</u> to <u>remain</u> **silent**. (S4)

(b) $the\_mutinous\_Fed\_member(DET_7) \wedge some\_Fed\_governor(t_8) \wedge$

$the\_chairman(DET_9) \wedge public(e_{17}, WR, 0.0_{epistemic}, DET_7) \wedge$

$go{:}\textsc{initiate}(em_{18}, WR, 0.0_{epistemic}, DET_7, e_{17}) \wedge$

$n't{:}\textsc{negator}(em_{19}, WR, 1.0_{epistemic}, em_{18} \wedge silent(e_{20}, WR, 1.0_{intentional}, DET_9) \wedge$

$remain{:}\textsc{continue}(em_{21}, WR, 1.0_{intentional}, DET_9, e_{20}) \wedge$

$decision{:}\textsc{intentional}(em_{22}, WR, 1.0_{epistemic}, DET_9, e_{21}) \wedge$

$disagree{:}\textsc{propositional}(em_{23}, WR, 1.0_{epistemic}, t_8, em_{22}) \wedge$

$know{:}\textsc{epistemic}(em_{24}, WR, 1.0_{epistemic}, em_{23}) \wedge$

$while{:}\textsc{contrast}(em_{25}, WR, 1.0_{epistemic}, e_{19}, em_{24}) \wedge$

$and{:}\textsc{conjunction}(em_{26}, WR, 1.0_{epistemic}, em_{11}, em_{25})$

(c) $COREF(DET_7, t_2) \wedge COREF(DET_9, t_4, t_1) \wedge COREF(em_{22}, em_4)$

(d) $public(e_{17}, WR, 0.0_{epistemic}, t_2) \wedge go{:}\textsc{initiate}(em_{18}, WR, 0.0_{epistemic}, t_2, e_{17}) \wedge$

$n't{:}\textsc{negator}(em_{19}, WR, 1.0_{epistemic}, em_{18} \wedge silent(e_{20}, WR, 1.0_{intentional}, t_4) \wedge$

$remain{:}\textsc{continue}(em_{21}, WR, 1.0_{intentional}, t_4, e_{20}) \wedge$

$decision{:}\textsc{intentional}(em_{22}, WR, 1.0_{epistemic}, t_4, e_{21}) \wedge$

Intra-sententially, the subordinating conjunction *while* indicates a discourse relation of CONTRAST relation. The sentence-initial coordinating conjunction *And*, on the other hand, indicates an inter-sentential discourse relation. We know that its argument in the current sentence is the top level embedding predication, em_{25}, the CONTRAST relation. The question is then finding the predication that encodes the other argument in prior discourse. Similarly to the example in the previous sentence (the case of *the statement*), this is where coreference chains may be of assistance. According to Veins Theory (Cristea *et al.*, 1998), which aims to explain the interaction of referential phenomena with the discourse structure within the RST framework, reference chains in text are associated to sets of structurally related units, even though they may be distant in text. A similar approach is explored in Centering Theory (Grosz *et al.*, 1995) within the discourse framework of Grosz and Sidner (1986). Prasad *et al.* (2010) propose the following coreference evaluation rules inspired by Centering Theory to identify the prior sentence in discourse that acts as an argument to the discourse connective in the current sentence:

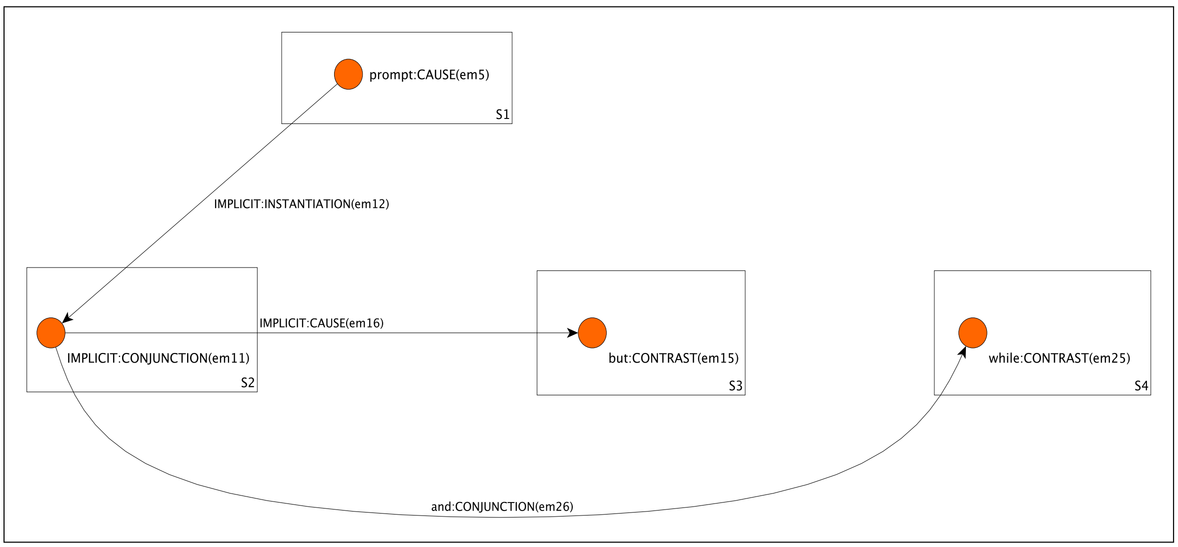

Figure 3.10: Graphical representation of the relationships between sentences in the paragraph in Figure (3.9).

1. If the entity mention in the connective's sentence has a pronominal form, the argument is the first sentence linked via the coreference chain for this entity.

2. If the entity mention in the connective's sentence has a non-pronominal form, then the argument is the first sentence that has a non-pronominal mention of the entity in the coreference chain for that entity.

Using the second rule, it is possible to infer that the argument of the discourse connective *And* is in the second sentence (S2). This is because the first entity mention in the current sentence (*the mutinous Fed member*) has a non-pronominal form, and its coreference chain leads us to its antecedent (*A "senior Fed official*) located in that sentence. We further can assume that the argument is top level predication encoded in that sentence (em_{11}). The resulting discourse structure with only the top level predications from each sentence is shown is presented in Figure (3.10).

PMID - 10089566
TITLE - Involvement of adenylate cyclase and p70(S6)-kinase activation in IL-10 up-regulation in human monocytes by gp41 envelope protein of human immunodeficiency virus type 1
ABSTRACT - Our previous results show that recombinant gp41 (aa565-647), the extracellular domain of HIV-1 transmembrane glycoprotein, stimulates interleukin-10 (IL-10) production in human monocytes. The signal cascade transducing this effect is not yet clear. In this study, we examined whether gp41-induced IL-10 up-regulation is mediated by the previously described synergistic activation of cAMP and NF-kappaB pathways. gp41 induced cAMP accumulation in monocytes in a time- and concentration-dependent manner and the adenylate cyclase inhibitor SQ 22536 suppressed gp41-induced IL-10 production in monocytes. In contrast, gp41 failed to stimulate NF-kappaB binding activity in as much as no NF-kappaB bound to the main NF-kappaB-binding site 2 of the IL-10 promoter after addition of gp41. We also examined the involvement of other signal transduction pathways. Specific inhibitors of p70(S6)-kinase (rapamycin), and Gi protein (pertussis toxin), prevented induction of IL-10 production by gp41 in monocytes, while inhibitors of the phosphatidylinositol 3-kinase (PI 3-kinase) (wortmannin) and mitogen-activated protein kinase (MAPK) pathway (PD 98059) did not. Thus HIV-1 gp41-induced IL-10 up-regulation in monocytes may not involve NF-kappaB, MAPK, or PI 3-kinase activation, but rather may operate through activation of adenylate cyclase and pertussis-toxin-sensitive Gi/Go protein to effect p70(S6)-kinase activation.

Figure 3.11: Medline abstract 10089566.

Our previous results show that em_2. The signal cascade transducing this effect is not yet clear. In this study, we examined whether em_2 is mediated by the previously described e_3 and e_4. t_1 induced e_5 and t_6 suppressed em_2. In contrast, t_1 failed to stimulate e_6 in as much as no e_7 after e_8. We also examined the involvement of other signal transduction pathways. t_9 and t_{10} prevented em_2 while t_{13} and t_{14} did not. Thus em_2 may not involve e_4, e_9, or e_{10}, but rather may operate through e_{11} and e_{12} to effect e_{13}.

Figure 3.12: Abstract of PMID 10089566 with entities (semantic terms) and some events substituted with their identifiers.

3.4.3 Discourse Interpretation of a Medline Abstract

Let us now turn our attention to the abstract of a PubMed article (PMID 10089566) and examine it from a discourse interpretation perspective. This abstract has been annotated in the GENIA event corpus (Kim *et al.*, 2008) for events and in the meta-knowledge annotation corpus (Thompson *et al.*, 2011) for meta-knowledge dimensions, including Certainty Level, Knowledge Type, Source, Polarity.

We will refer to these annotations when appropriate and we will also discuss the argumentative zones of sentences based on the approach of Teufel *et al.* (2009) as well as the discourse relations holding between predications. The full citation (with its title and abstract) is given in Figure (3.11).

We will ignore the title due to its special status in discourse and focus on the abstract only. There are large number of entities and events mentioned in the abstract; however, some of them corefer or are equivalent. Others are irrelevant for our illustration purposes. Therefore, to avoid repetition, we will identify them in advance and substitute their occurrences in text with their identifiers. These equivalences are given in Table (3.2) and the abstracts with substitutions in Figure (3.12).

Representation	**Sentences**	**Textual expressions**
gp41(t_1)	S1, S3, S4, S5, S7, S8	*gp41, aa565-647, the extracellular domain of HIV-1 transmembrane protein*
IL-10(t_2)	S1, S3, S4, S7,S8	*interleukin-10, IL-10*
cAMP(t_3)	S3, S4	*cAMP*
NF-κB(t_4)	S3, S5, S8	*NF-κB*
adenylate_cyclase(t_5)	S4, S8	*adenylate cyclase*
SQ_22536(t_6)	S4	*SQ 22536, adenylate cyclase inhibitor*
main_NF-kappaB-binding_site_2_of_the_IL-10_promoter(t_7)	S5	*the main NF-κ B-binding site 2 of the IL-10 promoter*
p70(S6)-kinase(t_8)	S7, S8	*p70(S6)-kinase*
rapamycin(t_9)	S7	*rapamycin, specific inhibitors of [p70(S6)-kinase*
Gi_protein(t_{10})	S7	*Gi protein*
pertussis_toxin(t_{11})	S7, S8	*pertussis toxin, specific inhibitors of ... Gi protein*
PI_3-kinase(t_{12})	S7, S8	*PI 3-kinase, phosphatidylinositol 3-kinase*
wortmannin(t_{13})	S7	*wortmannin, inhibitors of the phosphatidylinositol 3-kinase*
MAPK(t_{14})	S7, S8	*MAPK, mitogen-activated protein kinase*
PD_98059(t_{15})	S7	*PD 98059, inhibitors of the ... mitogen-activated protein kinase (MAPK) pathway*
pertussis-toxin-sensitive_Gi/Go_protein(t_{16})	S8	*pertussis-toxin-sensitive Gi/Go protein*

Representation	Sentences	Textual expressions
we(t_{17})	S3, S6	*we*
production(e_1,t_2)	S1, S3, S4, S7,S8	*IL-10 production, IL-10 up-regulation*
activation(e_3,t_3)	S3	*activation of cAMP ... pathways*
activation(e_4,t_4)	S3, S8	*activation of ... NF-κB pathways, NF-κB ... activation*
accumulation(e_5,t_3)	S4	*cAMP accumulation*
binding_activity(e_6,t_4)	S5	*NF-κB binding activity*
bind(e_7,t_4,t_7)	S5	*NF-κB bound to the main ... IL-10 promoter*
addition(e_8,t_1)	S5	*addition of gp41*
activation(e_9,t_{14})	S8	*MAPK ... activation*
activation(e_{10},t_{12})	S8	*PI 3-kinase activation*
activation(e_{11},t_5)	S8	*activation of adenylate cyclase*
activation(e_{12},t_{16})	S8	*activation of ... pertussis-toxin-sensitive Gi/Go protein*
activation(e_{13},t_8)	S8	*p70(S6)-kinase activation*
induce:CAUSE(em_2,t_1,e_1)	S1, S3, S4, S7, S8	*gp41 ... stimulates interleukin-10 ... production, gp41-induced IL-10 up-regulation, IL-10 production by gp41, HIV-1 gp41-induced IL-10 upregulation*

Table 3.2: Entities and main predications in PMID 10089566.

We begin with the first sentence. Its semantic interpretation is independent, and we simply determine the atomic and embedding predications. In Example (75b), note that some predications are duplicated to illustrate their source and scalar value information. We note a clausal level CAUSE relation which is in the scope of a modal predication of DEMONSTRATIVE type. The logical subject of this type of predication indicates the *source of evidence*, which, in this case, is previous work of the authors[18]. In the GENIA corpus, e_1 and em_2 are annotated as events. In the meta-knowledge annotation, the former is annotated as being of *Other* Knowledge Type, while the latter as *Analysis*

[18]In a cross-document discourse interpretation approach, it could be important to link this expression to the actual document where these previous results were reported; however, this is clearly a very challenging task that is beyond the scope of the current work. Rather, our focus is on interpretation at the document level.

Knowledge Type and *Other* Source. *Other* Knowledge Type is the default category, so we will not be concerned with it. But, more important is whether it is possible to infer that em_2 is an Analysis statement. The fact that this predication is in the scope of a DEMONSTRATIVE predication seems to allow this inference. Furthermore, we determined the source of this predication as t_{17} (*Our previous results*), a standard way of referring to non-current work and we can easily infer the Source of this predication as *Other*. At a higher, sentence-level granularity, if we consider the argumentative zones of Teufel *et al.* (2009), we can conclude that the zone of this sentence is PREV_OWN, which, according to its definition, corresponds to a neutral description of a knowledge claim (significant for the current work) held by authors in a previous paper. Again, this category for this sentence can be inferred based on the source of the main claim of the sentence (em_2).

(75) (a) Our previous results show that recombinant gp41 (aa565-647), the extracellular domain of HIV-1 transmembrane glycoprotein, stimulates interleukin-10 (IL-10) **production** in human monocytes. (S1)

(b) Our previous results show that em_2.

(c) $Our\_previous\_result(t_{18}) \wedge production(e_1, t_{18}, 1.0_{epistemic} \ldots) \wedge$
$stimulates$:CAUSE$(em_2, t_{18}, 1.0_{epistemic} \ldots) \wedge$
$show$:DEMONSTRATIVE$(em_3, WR, 1.0_{epistemic}, t_{18}, em_2)$

The second sentence, while it is not associated with any event or meta-knowledge annotation, is quite significant for the discourse progression. Note the anaphoric expression, *this effect*, which needs to be resolved from the prior discourse. The head of this demonstrative noun phrase, *effect*, is typed as a CAUSAL predicate. We infer that the referent should be a predication with this type or a subtype. In the previous sentence, em_2 with the CAUSE subtype satisfies this constraint and is therefore assigned as the referent, allowing us to resolve the argument of the predication em_{16}. The top level predication is of CONCESSION type and links the current sentence to the previous sentence. The link to the previous sentence may not be difficult to infer, since there is only one sentence prior to the current one and there is a structural relationship between the two sentences due to coreference. However, it is more difficult to infer that the link is to the predication that is not at the top level of the previous sentence (it is to em_2). Knowing that the top level predication in the previous sentence essentially provides *attribution* for the primary information in the complement clause may help us draw that inference. Finally, what kind of rhetorical move does this sentence indicate? Again, using argumentative zone categories, we can say that this is a GAP_WEAK statement indicating lack of solution in the field. The lack of certainty associated with *the signal cascade* (e_{14} has epistemic value of 0.3) seems to point in the direction of this argumentative zone.

(76) (a) The signal **cascade** transducing this effect is not yet clear. (S2)

(b) $signal(t_{19}) \wedge cascade(e_{14}, WR, 0.3_{epistemic}, t_{19}) \wedge this\_effect$:CAUSAL$(DPRO_{15}) \wedge$
$transduce$:CAUSE$(em_{16}, WR, 1.0_{epistemic}, t_9, DPRO_{15}) \wedge$
$clear$:EPISTEMIC$(em_{17}, WR, 0.0_{epistemic}, e_{14}) \wedge$
not:NEGATOR$(em_{18}, WR, 1.0_{epistemic}, em_{17}) \wedge$
yet:CONCESSION$(em_{19}, WR, 1.0_{epistemic}, _, em_{18})$

(c) $COREF(DPRO_{15}, em_2)$

(d) $transduce$:CAUSE$(em_{16}, WR, 1.0_{epistemic}, t_9, em_2) \wedge$
yet:CONCESSION$(em_{19}, WR, 1.0_{epistemic}, em_2, em_{18})$

The third sentence describes the goal of the current study, and therefore, its argumentative zone is AIM (defined as the statement of specific research goal). Having an INTERROGATIVE predication as the top level predication of the sentence seems to allow this inference, as well. With respect to meta-knowledge annotations, em_{23} (mediation) is characterized with the Knowledge Type Investigation, which, again, can be inferred due to the fact that it is on the INTERROGATIVE scale. On the other hand, the predications e_3 and e_4 (cAMP and NF-κB activation pathways) are assigned as Source the value *Other*. This can be inferred structurally, since both predications are in the scope of em_{22} (indicated by the adverbial *previously*). How is the semantic information in this sentence linked to prior discourse? This sentence clearly expands upon the first sentence (S1) rather than the previous one. However, the link is an implicit one that seems to be licensed at the lexical level, by the contrast of *Our previous results* with *In this study, we ...*.

(77) (a) In this study, we examined whether gp41-induced IL-10 **up-regulation** is mediated by the previously described synergistic activation of cAMP and NF-κB pathways. (S3)

(b) In this study, we examined whether em_2 is mediated by the previously described e_3 and e_4.

(c) and:CONJUNCTION$(em_{20}, t_{17}, 1.0_{epistemic}, e_3, e_4) \wedge$
$describe$:REPORTING$(em_{21}, t_{17}, 1.0_{epistemic}, _, em_{20}) \wedge$
$previously$:ASYNCHRONOUS$(em_{22}, t_{17}, 1.0_{epistemic}, em_{21}, em_{23}) \wedge$
$mediate$:ENABLE$(em_{23}, t_{17}, 1.0_{interrogative}, em_{20}, em_2) \wedge$

$examine$:INTERROGATIVE$(em_{24}, WR, 1.0_{epistemic}, t_{19}, em_{23}) \wedge$

$IMPLICIT$:CONJUNCTION$(em_{25}, WR, 1.0_{epistemic}, em_3, em_{24})$

The next sentence summarizes the results of the experiments conducted for the current work. As such, it can be categorized as a OWN_RES argumentative zone, which indicates a measurable/objective outcome of own work. A lack of modal predications and the prominence of causal predications in this sentence as well as the use of past tense leads us to this inference. All the GENIA events in this sentence have default meta-knowledge annotations, which, again, can be inferred from the lack of modal predications. The rhetorical contribution of the sentence is to elaborate on the previous sentence, achieved without an explicit discourse connective. The relation, therefore, can be considered an INSTANTIATION relation (em_{29}). The fact that em_2 is referred to in both sentence is a discourse clue that leads us to this inference.

(78) (a) gp41 <u>induced</u> cAMP **accumulation** in monocytes in a time- and concentration-dependent manner <u>and</u> the adenylate cyclase inhibitor SQ 22536 <u>suppressed</u> gp41-<u>induced</u> IL-10 **production** in monocytes. (S4)

(b) t_1 induced e_5 and t_6 suppressed em_2.

(c) $induce$:CAUSE$(em_{26}, WR, 1.0_{epistemic}, t_1, e_5) \wedge$

$suppress$:PREVENT$(em_{27}, WR, 1.0_{epistemic}, t_6, em_2) \wedge$

and:CONJUNCTION$(em_{28}, WR, 1.0_{epistemic}, em_{26}, em_{27}) \wedge$

$IMPLICIT$:INSTANTIATION$(em_{29}, WR, 1.0_{epistemic}, em_{28}, em_{24})$

The next sentence is quite similar to the previous one in terms of its content: it describes objectively the results of experiments carried out in the current work. Therefore, its argumentative zone can be characterized as OWN_RES. Again, the lack of modal predications points in this direction. In terms of GENIA events and meta-knowledge annotations, the predications em_{30} (stimulation) and e_7 are considered to have Polarity value *Negative*, which can easily be inferred since the former is on the *success* scale with the value 0 and the latter on the *epistemic* scale with the same value (both corresponding to counter-facts). This sentence is also rich with respect to discourse connectives: *after* indicating a temporal relation, *in as much as* providing an elaboration function, and *In contrast* linking the current sentence to the previous sentence with a CONTRAST relation. While these discourse relations can simply be inferred from the discourse connectives, it is also worth noting that the prominence of negated predications in this sentence in contrast to their absence in the previous sentence also seems to suggest a CONTRAST relation.

(79) (a) <u>In contrast</u>, gp41 <u>failed</u> to <u>stimulate</u> NF-kappaB **binding activity** <u>in as much as</u> <u>no</u> NF-kappaB **bound** to the main NF-kappaB-binding site 2 of the IL-10 promoter <u>after</u> addition of gp41. (S5)

(b) In contrast, t_1 failed to stimulate e_6 in as much as no e_7 after e_8.

(c) *stimulate*:CAUSE$(em_{30}, WR, 0.0_{success}, t_1, e_6) \wedge$
fail:SUCCESS$(em_{31}, WR, 1.0_{epistemic}, t_1, em_{30}) \wedge$
$bind(e_7, WR, 0.0_{epistemic}, t_4, t_7) \wedge$ *no*:NEGATOR$(em_{32}, WR, 1.0_{epistemic}, e_7) \wedge$
after:ASYNCHRONOUS$(em_{33}, WR, 1.0_{epistemic}, e_8, em_{32}) \wedge$
in_as_much_as:EXPANSION$(em_{34}, WR, 1.0_{epistemic}, e_{31}, em_{33}) \wedge$
in_contrast:CONTRAST$(em_{35}, WR, 1.0_{epistemic}, em_{34}, em_{28})$

The next sentence also describes a research statement, and similarly to S3, it can be characterized as being in the AIM zone, which, again, can be inferred from the fact that the top level intra-sentential predication of the sentence is an INTERROGATIVE one. No GENIA event or meta-knowledge annotations are provided for this sentence, since no specific biomolecular entities are present. On the other hand, a discourse relation is indicated by the discourse adverbial *also*; one of the arguments is the top level predication em_{38}, while the other argument lies in prior discourse. Which piece of semantic information forms the other argument? The structural similarity between the current sentence and S3 (both have INTERROGATIVE predications at the top level, for example), its dissimilarity to the two prior sentences in addition to the conjunctive discourse connective (*also*) seem to suggest that S3 is the relevant sentence and its top level predication em_{24} is the other argument. Resolving the argument of this discourse relation has another consequence. Note that in Example (80b) below, an argument of em_{38} (involvement) is left unspecified. The implicit, null-instantiated argument can be recovered from prior discourse. Since this sentence expands S3 most directly, it seems most reasonable to examine that sentence for an argument, In fact, Gerber and Chai (2010) use this heuristic as a feature to recover implicit arguments for nominal predicates. Examining the predications in that sentence, em_{23} seems to be the most appropriate, as its semantic category is consonant with that of em_{38} (ENABLE is a subtype of CAUSAL). Therefore, it is possible to assign as the implicit argument of em_{38} the logical object of em_{23} (em_2). That is, the author is speaking of *involvement of other signal transduction pathways in gp41-induced IL-10 upregulation* in this sentence.

(80) (a) We <u>also</u> <u>examined</u> the <u>involvement</u> of other signal transduction pathways. (S6)

(b) $signal(t_{20}) \wedge transduction\_pathway(e_{36}, t_{17}, 1.0_{epistemic}, t_{20}) \wedge$
involvement:CAUSAL$(em_{37}, t_{17}, 1.0_{interrogative}, e_{33}, \text{-}) \wedge$

examine:INTERROGATIVE$(em_{38}, WR, 1.0_{epistemic},\ t_{17}, em_{37}) \wedge$

also:CONJUNCTION$(em_{39}, WR, 1.0_{epistemic}, e_{38}, em_{24})$

(c) *involvement*:CAUSAL$(em_{37}, t_{17}, 1.0_{interrogative}, e_{33}, em_{2})$

S7, even though quite complicated from a non-expert perspective with the number of entities discussed, is structurally more or less straightforward, as shown in (81b). In fact, it is possible to think of it as parallel to the concatenation of sentences S4 and S5: the first portion of the sentence (up to *while*) exhibiting similarities with S4 and the second portion (beginning with *while*) with S5, with a CONTRAST relation between them. Therefore, the same facts (lack of modal predications and prominence of causal relations) seem to indicate that this sentence is also in the argumentative zone of OWN_RES. Furthermore, this sentence is linked to the previous sentence, in the same way S4 is linked to S3, with an INSTANTIATION relation. The set-instance relation between *transduction pathways* in the previous sentence and *PI 3-kinase and MAPK pathway* in this sentence indicate this relation implicitly. With regard to GENIA event and meta-knowledge annotations, it is worth noting that the Polarity value of two predications are set to Negative (em_{43} and em_{44}), and we capture this as well, since they both have epistemic value of 0.

(81) (a) Specific **inhibitors** of p70(S6)-kinase (rapamycin), and Gi protein (pertussis toxin), prevented induction of IL-10 **production** by gp41 in monocytes, while **inhibitors** of the phosphatidylinositol 3-kinase (PI 3-kinase) (wortmannin) and mitogen-activated protein kinase (MAPK) pathway (PD 98059) did not. (S7)

(b) t_9 and t_{10} prevented em_2 while t_{13} and t_{14} did not.

(c) *prevent*:PREVENT$(em_{40}, WR, 1.0_{epistemic}, t_9, em_2) \wedge$

prevent:PREVENT$(em_{41}, WR, 1.0_{epistemic}, t_{10}, em_2) \wedge$

and:CONJUNCTION$(em_{42}, WR, 1.0_{epistemic}, em_{40}, em_{41}) \wedge$

prevent:PREVENT$(em_{43}, WR, 0.0_{epistemic}, t_{13}, em_2) \wedge$

prevent:PREVENT$(em_{44}, WR, 0.0_{epistemic}, t_{14}, em_2) \wedge$

not:NEGATOR$(em_{45}, WR, 1.0_{epistemic}, e_{43}) \wedge$ *not*:NEGATOR$(em_{46}, WR, 1.0_{epistemic}, e_{44}) \wedge$

and:CONJUNCTION$(em_{47}, WR, 1.0_{epistemic}, em_{45}, em_{46}) \wedge$

while:CONTRAST$(em_{48}, WR, 1.0_{epistemic},\ em_{42}, em_{47}) \wedge$

IMPLICIT:INSTANTIATION$(em_{49}, WR, 1.0_{epistemic}, em_{48}, em_{38})$

The final sentence of the abstract is also the most crucial sentence, as it reports the conclusions of the authors. As such, it can be categorized as being in the argumentative zone of OWN_CONC (defined as findings, non-measurable conclusions of own work). The prominence of explicitly epistemic predications (SPECULATIVE) that occupy middle points on the epistemic scale as well as the fact that this is simply the last sentence of current discourse lead us to make this inference. This sentence is also rich with regard to explicit discourse cues. *but rather* indicates an ALTERNATIVE relation between two intra-sentential predications, which can be recovered solely on the basis of the discourse connective. On the other hand, the sentence-initial connective *Thus* provides a link from the top level predication (em_{58}) to the prior discourse, although it seems difficult to link it to a single sentence or predication. Instead, it seems to link to the entire semantic content of sentences S4, S5, and S7, all in OWN_RES argumentative zone, reporting experimental results. This fact is captured by the implicit discourse relation em_{59}, which allows us to treat these sentences as a single discourse unit. This discourse relation then becomes an argument of the discourse relation indicated by *Thus*.

(82) (a) Thus HIV-1 gp41-induced IL-10 up-regulation in monocytes may not involve NF-kappaB, MAPK, or PI 3-kinase activation, but rather may operate through activation of adenylate cyclase and pertussis-toxin-sensitive Gi/Go protein to effect p70(S6)-kinase activation. (S8)

(b) Thus em_2 may not involve e_4, e_9, or e_{10}, but rather may operate through e_{11} and e_{12} to effect e_{13}.

(c) *or*:CONJUNCTION$(em_{50}, WR, 1.0_{epistemic}, e_4, em_9, em_{10}) \wedge$
involve:CAUSAL$(em_{51}, WR, 0.3_{epistemic}, em_{50}, em_2) \wedge$
may:SPECULATIVE$(em_{52}, WR, 0.0_{epistemic}, em_{51}) \wedge$
not:NEGATOR$(em_{53}, WR, 1.0_{epistemic}, em_{52}) \wedge$
and:CONJUNCTION$(em_{54}, WR, 1.0_{epistemic}, e_{11}, e_{12}) \wedge$
effect:CAUSAL$(em_{55}, WR, 1.0_{epistemic}, em_{54}, em_{13}) \wedge$
operate:PATIENT$(em_{56}, WR, 0.5_{epistemic}, em_2, em_{54}) \wedge$
may:SPECULATIVE$(em_{57}, WR, 1.0_{epistemic}, em_{56}) \wedge$
but_rather:ALTERNATIVE$(em_{58}, WR, 1.0_{epistemic}, em_{57}, em_{53}) \wedge$
IMPLICIT:CONJUNCTION$(em_{59}, WR, 1.0_{epistemic}, em_{35}, em_{48}) \wedge$
thus:CAUSE$(em_{60}, WR, 1.0_{epistemic}, em_{59}, em_{58})$

When it comes to GENIA events and meta-knowledge annotations, in this sentence, the points that stand out are the following:

- em_{51} indicated by *involve* is presented as an Analysis statement with *Negative* Polarity and L1 (Low) Certainty Level. The Knowledge Type (Analysis) is recoverable from the fact that this predication is within the scope of a SPECULATIVE predication, and Polarity (Negative) from the fact that it is also in the scope of a NEGATOR type of predication. On the other hand, the epistemic value of the predication (0.3) points to a low factuality value, corresponding to L1 Certainty Level.

- em_{56} indicated by *operate* is presented as an Analysis statement with L1 Certainty Level. These are also recoverable from the fact that this predication is within the scope of a SPECULATIVE predication and has epistemic value of 0.5.

The discourse structure of this abstract, with only top level predications explicitly shown, in given in Figure (3.13).

3.4.4 Where Are We In Discourse Interpretation?

In the last two sections, we attempted to illustrate how the embedding framework can form the basis for moving towards discourse interpretation by analyzing two discourses in detail. As stated earlier, not all the possible mechanisms that have been discussed are currently operationalized; however, we believe and tried to demonstrate that the embedding framework provides the basis for implementation.

From our discussion in the previous sections, three main challenges emerge with respect to moving from the current embedding framework towards discourse interpretation: (a) accurate coreference resolution, (b) identification of implicit relations between predications and sentences, and (c) lack of a single, universally accepted discourse model that can be specifically targeted. We have begun addressing the issue of coreference resolution. In fact, we have implemented a coreference resolution module within the embedding framework (Section 5.4), currently limited to biomedical text even though the problems in extending it to general domain are smaller, focused issues (such as identifying meronymic relations) and can be handled by extending the framework. The second issue, identification of implicit relations, is more problematic. Our framework has been firmly based on predicates and explicit encoding of their lexical semantic and syntactic features in a dictionary. In this section, we simply suggested some mechanisms, such as using the semantic categories of predications or coreference resolution, to identify these implicit relations. Finally, the lack of a single, perfect discourse model is particularly evident in scientific discourse models. Should we adapt the embedding framework to address discourse interpretation as a task of identifying argumentative zones? Or is it more beneficial to adapt it to the meta-knowledge model? Clearly, while their approaches are different, they attempt to address overlapping, discourse-related problems. Furthermore, different practical tasks may require adoption of different models. In this section, we simply showed how one can move from our application-neutral predications to identify argumentative zones of sentences or meta-knowledge dimensions, such as Certainty, Knowledge Type or Polarity. While the basis of identifying such knowledge is made explicit within the framework to a large extent, actual identification of such knowledge and its evaluation remains future research. We also note that improvements in coreference resolution as well as addressing implicit relations in a general way are likely to contribute to these specific tasks, as well.

3.5 Conclusions

In this chapter, we discussed the core aspects of the embedding framework in detail. We began by providing the basic definitions, such as atomic and embedding predications and semantic scope. We introduced the embedding predicate categorization, drawn from various linguistic classifications and augmented with corpus analysis. This categorization scheme forms the basis for the compositional semantic interpretation approach. We did not get into details of the procedural aspects of the composition, which is left to Chapter 5; however, we clearly illustrated the outcome of the procedure on numerous examples, mostly from the PTB corpus. At a conceptual level, we closely aligned MODAL and VALENCE_SHIFTER predicates with lower level extra-factual phenomena and the RELATIONAL predicates with the higher level discourse. We also explained how embedding predications contribute to basic propositional meaning, via PROPOSITIONAL predicates. Next, we discussed scopal influence of embedding predications, analyzing how an embedding predication influences other predications in its semantic scope. We have drawn generalizations from corpus analysis using the knowledge made explicit by the embedding framework in order to explain the scopal influence behavior. Finally, we illustrated how the framework can serve the goal of discourse interpretation by examining two discourses in detail and highlighting the main issues and principles involved. We concluded by suggesting ways in moving forward in this direction, namely, accurate coreference resolution and identification of implicit relations between predications and sentences.

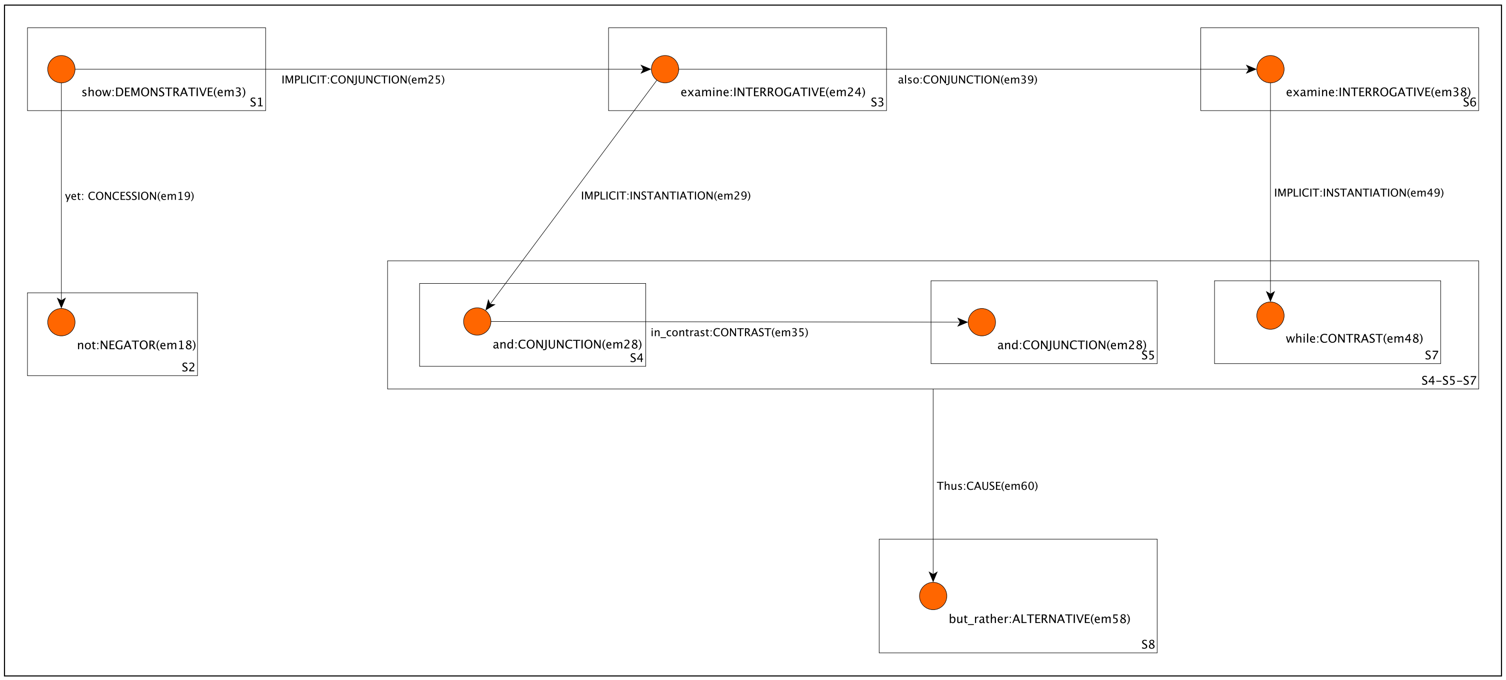

Figure 3.13: Graphical representation of the relationships between sentences in the abstract in Figure (3.11).

Chapter 4

Lexical and Syntactic Considerations

Our semantic interpretation, which will be described in Chapter 5, is based on a compositional, rule-based approach that relies on lexical semantic knowledge and syntactic structure as input. Lexical semantic knowledge is encoded in a dictionary of embedding predicates, in which such predicates are associated with one of more semantic senses (corresponding to embedding categories) and with various other features that play a role in compositional interpretation. Relevant syntactic structures, on the other hand, are captured with typed syntactic dependency relations between individual words. In this section, we discuss the embedding predicate dictionary and the relevant syntactic phenomena in more detail. We conclude with a brief description of dependency parsing, which allows us to identify typed syntactic dependency relations.

4.1 Dictionary of Embedding Predicates

Lexical knowledge regarding individual predicates relevant to embedding are stored in an embedding predicate dictionary. In the dictionary, such predicates are mapped to embedding categories that they indicate, described in Section 3.2. We have been developing the embedding dictionary incrementally, beginning with our early work on speculation detection (Kilicoglu and Bergler, 2008). We later extended and refined this dictionary using various corpora and linguistic classifications. Construction of the dictionary also involved a fair amount of manual work, since most of the relevant predicates harvested from other resources do not neatly fit into one of the embedding categories.

Additionally, we target deeper levels of meaning distinctions than assumed in most other linguistic resources. Being a core aspect of the compositional analysis, it needs to accommodate the requirements of such an analysis.

Currently, the dictionary consists of 910 embedding predicate entries. Each predicate entry in the dictionary has the following features:

Lemma The lemma of the embedding predicate. For generality, we allow multi-word predicates in addition to single-word lexical units. For multi-word lexical items, the lemma of each word is specified.

Part-of-speech The part-of-speech tag of the predicate. For multi-word predicates, again, the part-of-speech tag of each lemma is specified.

Senses Semantic senses associated with the predicate.

Lemma and part-of-speech are lexical features, while each sense describes syntactic and semantic information regarding one particular sense of the predicate. Each sense is associated with the following features:

Category Sense name, corresponding to an embedding predicate category.

Prior scalar value The scalar modality value associated with the predicate, if any. The relevant scale is determined from the sense. By default, this value is 1.0.

Scope Type Whether the predicate allows a *narrow* or *wide* scope reading. If a predicate has narrow scope, transfer of negation to its complement is allowed (see Section 2.2.1). On the other hand, wide scope allows a predicate to have scope over the predicate that syntactically embeds it. This value is set to *default* otherwise, meaning that neither a narrow nor wide scope reading is allowed.

Embedding Types These correspond to the semantic dependency classes that are used to identify the logical object argument of the predicate. They correspond to the edge labels of the semantic embedding graph, which will be discussed in detail in Section 5.2. By default, this feature is an empty list, meaning that default logical object argument identification rules can be applied.

Argument Inversion Whether the predicate requires argument inversion, that is, its syntactic subject should be taken as its logical object and its syntactic object as its logical object. By default, this value is set to *false*.

Probability If this sense is derived from a corpus, its likelihood of indicating the corresponding semantic category. If the sense is not derived from a corpus, it is 1.0. The probability is calculated as:

$$P(p,S) = w(S{:}p)/w(p)$$

where $w(S{:}p)$ is the number of times the predicate p occurs as a trigger for a semantic category S in the relevant corpus and $w(p)$ is the frequency of the predicate.

Sense features, *prior scalar value*, *scope type*, *embedding types* and *argument inversion*, allow us to map syntactic structure to predications in compositional analysis, thus, operating at the syntax-semantics interface. The *probability* feature is not a core aspect of the dictionary; however, if available, it can be used as the basis of a simple word sense disambiguation scheme in interpretation, when a predicate has more than one sense. We illustrate these features on several examples below.

Predicate	*may*	
Lemma [POS]	*may*[MD]	
Sense.01	Category	SPECULATIVE
	Prior scalar value	0.5
	Embedding types	AUX
Sense.02	Category	PERMISSIVE
	Prior scalar value	0.6
	Embedding types	AUX

Table 4.1: Dictionary entry for *may*

The entry in Table (4.1) indicates that the modal auxiliary *may* is associated with two modal senses, one SPECULATIVE and the other PERMISSIVE, with differing prior scalar values. Prior scalar value of 0.5 for the SPECULATIVE type means that the predication embedded by the predicate *may* will be assigned the epistemic value of 0.5 initially. Since there is more than one sense associated with the predicate and the embedding types corresponding to both senses are the same (AUX), we can speak of the sense ambiguity of the modal *may*. While the framework currently does not specifically handle ambiguity, in our experiments described in Chapter 6, we have used several criteria to narrow the senses down in an effort to reduce ambiguity. Another thing to notice about the entry above is that two sense features, *scope type* and *argument inversion*, are not specified explicitly, indicating default values (*default* and *false*, respectively).

The entry for the verbal predicate *result* is given in Table (4.2). Here, first, note that *prior scalar value* and *scope type* features implicitly take on the default values 1.0 and *default*. More importantly,

Predicate	*result*	
Lemma [POS]	*result*[VB]	
Sense.01	Category	CAUSE
	Embedding types	PREP_IN
	Argument inversion	false
	Probability	0.14
Sense.02	Category	CAUSE
	Embedding types	PREP_FROM
	Argument inversion	true
	Probability	0.14

Table 4.2: Dictionary entry for *result*

note that two separate senses are created even though the categories of both senses are the same (CAUSE). This is due to the fact that logical subject/object arguments of the verbal predicate *result* differ, depending on the preposition that cues its syntactic object. If the preposition is *in* (as in *X results in Y*), the relevant embedding class is PREP_IN and, therefore, the first sense is selected and the argument inversion is taken to be *false*. Thus, we conclude that the syntactic object corresponds to logical object as well. That is, Y is the logical object and X is the logical subject. However, if the preposition is *from* (as in *X results from Y*), the selectional restrictions of the second sense is satisfied, which also leads to identifying Y as the logical subject and X as the logical object since the *argument inversion* feature is taken to be *true*. Finally, note that the probability feature is set to 0.14, calculated from the statistics in GENIA corpus, from which the predicate was derived.

Predicate	*believe*	
Lemma [POS]	*believe*[VB]	
Sense.01	Category	ASSUMPTIVE
	Prior scalar value	0.8
	Embedding types	THAT_COMP,XCOMP,PREP_IN
	Scope type	narrow
Sense.02	Category	REPORTING
	Prior scalar value	0.8
	Embedding types	THAT_COMP
	Scope type	narrow

Table 4.3: Dictionary entry for *believe*

The main thing to notice about the entry for the verbal predicate *believe* (in Table 4.3) is that the *scope type* feature is assigned the non-default value *narrow*, indicating that it allows the transfer of negation to its complement, which is taken into account in the composition phase. Furthermore, the two senses have common embedding types as well as differing ones, which may allow disambiguation depending on the context.

Predicate	*in contrast*	
Lemma [POS]	*in*[IN],*contrast*[NN]	
Sense.01	Category	CONTRAST
	Embedding types	PREP_IN
	Argument inversion	false
Sense.02	Category	CONTRAST
	Embedding types	PREP_TO,PREP_WITH
	Argument inversion	true

Table 4.4: Dictionary entry for *in contrast*

Similarly to *result* above, *in contrast* also has two senses belonging to the same category but different selectional restrictions (in Table 4.4). Furthermore, note that it is a multi-word lexical unit and that lexical features of both words, *in* and *contrast*, are specified in the entry.

The main embedding categories and current dictionary sense counts for each category are given in Table (4.5).

Category	**Sense Count**
MODAL	776
PROPOSITIONAL	72
RELATIONAL	393
VALENCE_SHIFTER	71

Table 4.5: Embedding categories and relevant predicate sense counts in the dictionary

4.1.1 Resources for Embedding Predicates

The dictionary of embedding predicates has been compiled from several relevant resources semi-automatically and expanded manually. Linguistic knowledge and intuitions as well as empirical evidence have played a role in manual expansion. In this section, we briefly discuss the external resources we exploited in creating the dictionary. We distinguish two types of resources: (a) word lists provided in various books and articles and (b) corpora. We begin with the word lists and conclude with PDTB and GENIA-related corpora.

Hedging dictionary of Kilicoglu and Bergler (2008)

The current embedding predicate dictionary was initially conceived as a word list of hedging markers. The hedging dictionary was based on the categorization of lexical hedging markers of Hyland (1998) and was extended using synonymy information from WordNet (Fellbaum, 1998) as well as nominalization information from the UMLS Specialist Lexicon (McCray *et al.*, 1994). Some markers were also derived from the Hedge Classification Dataset (Medlock and Briscoe, 2007). Each marker was then semi-automatically assigned a weight based on its hedging strength and classified into categories, including epistemic verbs, numerical hedges, etc. A total of 198 markers was identified in this way, mostly corresponding to EPISTEMIC, EVIDENTIAL, and VALENCE_SHIFTER categories[1]. Integrating this word list into the embedding dictionary involved encoding their sense information explicitly. Hedging weights played a partial role in determining prior scalar values and hedging categories in determining the sense categories and relevant syntactic classes. For example, the verb *know* is marked as an epistemic *unhedger* in the hedging dictionary, indicating that it is relevant to hedging only when it is negated. Based on this information, we assigned it to the category of EPISTEMIC with the prior scalar value of 1.0. All 198 predicates are included in the dictionary.

Factuality markers of Saurí (2008)

Saurí (2008) provides a comprehensive list of factuality markers and classifies them into categories based on their factuality behavior. These mostly correspond to modal predicates, but there are non-modal items, as well (e.g., *invite, entitle*, etc.) In addition to markers and their factuality classes, she also provides the syntactic dependencies corresponding to the syntactic relation between the factuality marker and its embedded event. Her focus is on determining the factuality status of the embedded events, rather than semantically classifying these markers. Therefore, the classes of these markers are not directly applicable to our embedding categories. Some of her classes are quite large and heterogenous, bringing together predicates that cannot be considered to be related, other than through the fact that the events they embed have similar factual status. For example, consider the class named *want*. It can be assumed that the predicates in this class can be mapped to VOLITIVE embedding category, given the name. However, predicates such as *lead* and *depend* in this category are simply not modal items, while other predicates in the same class (e.g., *want*, *wish*) can be considered VOLITIVE. The only similarity between these items is that same inference can be made with respect to the factual status of their embedded event. On the other hand, other classes

[1] A detailed description of development of this hedging dictionary can be found in Kilicoglu and Bergler (2008).

have far fewer items and are more homogenous. For example, the *confirm* class has four predicates, *confirm*, *confirmation*, *proof*, and *recognition*, all of which can be considered to correspond to the DEMONSTRATIVE category with high scalar values.

Using several simple heuristics based on marker categories, we mapped the factuality markers to dictionary entries. For example, the predicates in the *confirm* class were simply assigned the DEMONSTRATIVE category with the scalar value of 1.0. The predicates of the *attempt* class were again simply assigned the INTENTIONAL category with the scalar value of 1.0. Those in the *fail* class were assigned the SUCCESS category with the scalar value of 0.0. The initial value assignment was followed by inspection and manual update of the categories and scalar values, if necessary. We also largely adopted the syntactic types she assigned to individual predicates as the basis for embedding types. The dictionary currently uses 254 predicates (corresponding to 384 senses) derived from the classification provided by Saurí (2008).

Propositional attitude markers of Nirenburg and Raskin (2004)

In their book on ontological semantics, Nirenburg and Raskin (2004) devote a section to modality. While they do not explicitly list all the markers that they use in their various implementations of the framework, their discussion includes a fair number of markers and their modal categories. Since their modal categories largely overlap with and are parallel to our modal categories, we used their markers and the categories they are assigned to. In some cases, we assigned finer-grained categories (such as SPECULATIVE rather than just EPISTEMIC, as they do), while in other cases, the mapping was direct (their EPITEUCTIC category was mapped to the SUCCESS class). No syntactic information regarding these markers is provided in Nirenburg and Raskin (2004). Again, the entries derived from this resource were manually updated. While they discuss scalar values for modal markers, they do not provide these values explicitly. However, we derived some of these values from the examples that they provide and used them.

We also used an earlier and somewhat different version of the modal categorization and markers they describe in Carlson and Nirenburg (1990). Combined, we derived 49 predicates (96 senses) from modality markers in the ontological semantics framework.

Modality lexicon of Baker *et al.* (2010)

The modality categorization of Baker *et al.* (2010) resembles that of Nirenburg and Raskin (2004) and is also similar to ours. They provide a modality lexicon where modal items are semantically typed and the syntactic relation between the modal predicate and its complement is explicitly encoded.

We used 63 predicates and 110 senses from their modality lexicon.

Abstract verbs of Power (2007)

In the context of natural language generation, Power (2007) describes *abstract verbs*, essentially verbs that express abstract relationships, pushing events and other relational semantic objects down into nominalizations. He distinguishes two classes: one expressing discourse relations (such as *induce* for causal relations) and another expressing participant roles (such as *perform* for AGENT role). The first class of predicates is relevant as RELATIONAL predicates, while the second is relevant as SEMANTIC_ROLE predicates. We used 24 predicates and 52 senses from his classification.

Polarity Lexicon of Wilson *et al.* (2005)

Polarity Lexicon is a major resource in the field of sentiment analysis. Each item in this lexicon is classified as having *positive*, *negative*, or *neutral* prior polarity. While sentiment-bearing words are not our focus in the current work, we derived a fair number of implicitly negative items, such as *absence*, from this lexicon (45 predicates, 55 senses).

Causal verbs from Wolff (2003)

A small number of periphrastic causal verbs are listed in Wolff (2003), categorized along CAUSE, PREVENT, and ENABLE types. Since our categorization of causal types was adopted from this article, we added these verbs to the dictionary and directly used his categorization (12 predicates, 23 senses).

Aspectual verbs from TimeML annotation guidelines (Saurí *et al.*, 2006b)

We derived a small list of aspectual verbs from the TimeML annotation guidelines. These predicates were provided as examples of verbs that provide aspectual links (ALINKs) with events. We have 16 such predicates, each with a unique sense.

Explicit discourse connectives of PDTB 2.0

We automatically extracted the explicit discourse connectives and their senses from the PDTB corpus (Prasad *et al.*, 2008) as the basis of our RELATIONAL predicates. Since our RELATIONAL categories are slightly different from the discourse relation senses in PDTB, we defined simple mappings between these categories. For example, both *precedence* and *succession* categories were mapped to the ASYNCHRONOUS category. Probabilities for individual senses were also calculated. A total of 122 RELATIONAL predicates (181 senses) were derived from PDTB.

Triggers from GENIA-related corpora

We also automatically extracted CAUSAL and CORRELATIVE predicates from GENIA-related corpora, which include the GENIA event corpus (Kim *et al.*, 2008) and BioNLP shared task corpora (Kim *et al.*, 2009, 2011a), the latter based primarily on the former.

Most of the extraction from GENIA-related corpora has been done earlier, in the context of BioNLP shared task competitions. For those competitions, we had extracted reliable trigger expressions from training data using several simple heuristics, based on part-of-speech categories and probability as a trigger (Kilicoglu and Bergler, 2011b). Only some of these trigger expressions are relevant as embedding predicates; in particular, those that correspond to GENIA regulation events are relevant as causal predicates. We had not mapped these trigger expressions to embedding categories explicitly. In the context of the current work, we took the predicates that trigger POSITIVE_REGULATION events as predicates for the CAUSE category and those that trigger NEGATIVE_REGULATION as predicates for the PREVENT category. Triggers for REGULATION events were taken to indicate the less-specific CAUSAL category. We then manually recategorized some of the CAUSE predicates into the ENABLE category. Predicates for the CORRELATIVE category were drawn from the GENIA event corpus itself, using the same heuristics. For this category, events with the type CORRELATION were considered. Similar to PDTB, probabilities for each predicate-embedding category combination are included as the basis for simple disambiguation.

While Speculation and Negation are annotated to some extent in the GENIA corpus, their trigger expressions are not. During the BioNLP shared task competitions, we have used their Speculation and Negation annotations to extend our list of MODAL and NEGATOR predicates. We derived a total of 140 predicates (203 senses) from the GENIA-related corpora.

4.2 Relevant Syntactic Phenomena

In this section, we discuss the syntactic constructions that play a role in the embedding framework in more detail. So far, taking the TimeML notion of an event predicate as the basis, we assumed that atomic predications can be signalled by verbal, nominalized, and adjectival predicates. The logical arguments of an atomic predication typically correspond to the syntactic arguments of the predicate. Consider an earlier example, duplicated below for reference, where atomic predicates are in bold and the embedding predicates are underlined. The arguments of the atomic predication indicated by the verbal predicate *veto* are its syntactic subject (*President Bush*) and its direct object (*a bill*). The arguments of the first *funding* predicate (verbal) are its direct object (*Departments of Labor, Education, and Health and Human Services*) and the noun phrase that it modifies (*a bill*). On the other hand, the nominalized predicate *abortions* is not syntactically linked to its logical subject argument (*victims of rape and incest*).

(83) President Bush will **veto** a bill **funding** the Departments of Labor, Education and Health and Human Services because it would allow federal funding of **abortions** for victims of rape and incest, the White House said.

In the current work, we are not concerned with the inner structure of atomic predications per se. However, since we are casting a wide net and not limiting ourselves to a particular lexical category of predicates (verbs, nouns, etc.)[2], it quickly becomes apparent that we need to consider the wide range of lexical categories and syntactic constructions that they appear in. Furthermore, the fact that an atomic predication can be signalled by a nominal predicate essentially means that an embedding predication can be indicated by a verbal predicate (taking the nominal atomic predicate as a syntactic argument), as well as the corresponding nominal and adjectival forms. In other words, any lexical category that can indicate an atomic predication can also indicate an embedding predication. For instance, in Example (83), the verbal predicate *allow* indicates an embedding predication; its syntactic subject and object corresponding to its semantic arguments, as well. Furthermore, modal auxiliaries (e.g., *would*, *will*, *may*), subordinating and coordinating conjunctions (e.g., *because*, *but*), some adverbials (e.g.,*possibly*), certain determiners (e.g., *no*), and even pronouns (*nobody*) need to be taken into account for an adequate characterization. Some of the examples presented in Chapter 3 had embedding predicates of these categories.

In order to adequately account for the lexical items in all these categories with respect to predication, we also need to consider the syntactic constructions they occur in. Due to the variety of lexical categories that are of interest, syntactic constructions that play a role in expressing embedding predications are also quite varied. The major classes that need to be taken into account are subordination and coordination, various syntactic modification types (adverbial, adjectival, modal), as well as basic verbal and nominal predication. We examine these syntactic constructions below, also commenting on the correlation between syntactic forms and semantic embedding. Note that we will not discuss basic verbal, nominal, and adjectival predication, as they are not specific to embedding predications and because we already presented some examples earlier.

[2]This strategy of focusing on a particular lexical category has been employed gainfully in large-scale projects such as PropBank (Palmer *et al.*, 2005) and NomBank (Meyers *et al.*, 2004b) and related research.

4.2.1 Syntactic Modification

Syntactic modification in the form of adverbial, modal, and adjectival modification plays a role in semantic embedding. While elements of these classes (i.e., adverbs, adjectives, modal auxiliaries) are syntactically dependent on the verbal or nominal predicates that they modify, they may semantically embed the predication indicated by the governing predicates, especially with respect to epistemic content. We provide several examples below, where syntactic modification alters the meaning of the embedded predications epistemically.

(84) (a) New Environmentalism probably **started** in 1962 with the publication of Rachel Carson's book "Silent Spring." (wsj_2021)

$New\_Environmentalism(t_1) \wedge publication(e_1 \ldots) \wedge$

$start$:INITIATE$(em_2, WR, 0.7_{epistemic}, e_1, t_1 \ldots)$

$probably$:SPECULATIVE$(em_3, WR, 1.0_{epistemic}, em_2)$

(b) ... Mr. Reagan was pulled into discussing the possible **elimination** of nuclear weapons without consulting American allies. (wsj_0288)

$nuclear\_weapon(t_1) \wedge eliminate(e_1, WR, 0.5_{epistemic}, t_1) \wedge$

$possible$:SPECULATIVE$(em_2, WR, 1.0_{epistemic}, em_1)$

(c) Antitrust laws provide that injured parties may be **reimbursed** for lawyers' fees. (wsj_2433)

$injured\_parties(t_1) \wedge lawyers'\_fee(t_2) \wedge Antitrust\_laws(t_3) \wedge$

$reimburse(e_1, t_3, 0.5_{epistemic}, t_1, t_2) \wedge may$:SPECULATIVE$(em_2, t_3, 1.0_{epistemic}, e_1)$

In Example (84a), the adverbial *probably* is syntactically dependent on the main verbal predicate *started* but the predication it encodes embeds the predication signalled by the verbal predicate (*adverbial modification*). Example (84b) shows an instance of adjectival modification: just as adverbs modify verbal predicates, adjectives may modify nominalized predicates, similarly encoding semantic embedding. Similar to adverbs, modal auxiliaries also modify verbal predicates to alter the meaning contributed by the verbal predicate, as Example (84c) shows (*modal modification*).

As their name implies, all modal auxiliaries contribute modal meaning to the embedded predications. On the other hand, this semantic behavior is not generalizable to all adjectives or adverbs. For example, consider the earlier example "A boy wanted to build a boat quickly." We represented the meaning contribution of the adverb *quickly* as a predication that has scope over the predication indicated by the verb *build*. However, the effect of the adverb is, in fact, restricting the building event with respect to its MANNER attribute, which is fundamentally different from the meaning contribution of adverbs like *possibly* or *likely*, which provide epistemic content. The same situation holds for adjectives, as well (*quick run* vs. *possible run*). With respect to these lexical categories, sometimes the terms *modal adjectives* and *modal adverbs* are used to describe particular words that are of more interest within the scope of the current work. In the context of adjectives, Mendes and Amaro (2009) distinguish *property-ascribing adjectives* and *non-restricting adjectives*, the former referring to those adjectives that restrict the properties of the nouns they modify (e.g., *blue book*) and the latter to those that behave as semantic operators (e.g., *false diamond*, *alleged murderer*). They model the meaning contribution of non-restricting adjectives by proposing that they semantically embed the modified noun into a modal context. As the discussion above suggests, we take the same position here and extend it further to adverbs, arguing that syntactic modification is simply isomorphic to semantic embedding with respect to property-ascribing adjectives and adverbs.

The semantic contribution of syntactic modification is not limited to modal adverbs, adjectives, and auxiliaries. It may also express negation via negative particles, as in Example (85a) as well as other types of embedding categories, such as temporal categories (Example 85b), via non-restricting adjectives and adverbs.

(85) (a) "We have no plans at this time to **pay off** those notes," he said. (wsj_2031)

$We(t_1) \wedge note(t_2) \wedge he(t_3) \wedge pay\_off(e_1,t_3,0.0_{intentional},t_1,t_2) \wedge$

$plan$:INTENTIONAL$(em_2,t_3,0.0_{epistemic},t_1,e_1) \wedge no$:NEGATOR$(em_3,t_3,1.0_{epistemic},em_2)$

(b) Dealers say the firm apparently has wanted to publicize its recent **buying** and subsequent **selling** of 30-year bonds ... (wsj_0671)

$the\_firm(t_1) \wedge 30\text{-}year\_bond(t_2) \wedge Dealers(t_3) \wedge$

$buy(e_1,t_3,\ldots,t_1,t_2\ldots) \wedge sell(e_1,t_3,\ldots,t_1,t_2,\ldots) \wedge$

$subsequent$:ASYNCHRONOUS$(em_3,t_3,1.0_{epistemic},e_1,e_2)$

4.2.2 Syntactic Subordination

As we mentioned earlier in Section 2.1.1, the embedding framework is intuitively linked to syntactic subordination. Syntactic subordination is generally taken to subsume the notions of *complementation*, *adverbial subordination*, and *relativization*. We examine these notions more closely below.

Complementation

Among subordination classes, complementation is the major mechanism that has a bearing on semantic embedding. In fact, Givón (2001) notes that "a systematic isomorphism obtains between the syntactic and semantic dimensions of complementation." This is illustrated in the following sentence:

(86) (a) Conservationists say that drift-net fishing threatens to **wipe out** much of the world's tuna stocks in a few years. (wsj_1250)

(b) $Conservationists(t_1) \wedge say$:REPORTING$(em_1, WR, 1.0_{epistemic}, t_1, em_2) \wedge threaten(em_2 \ldots)$

Here, the main verb of the sentence, *say*, takes a clausal complement, predicated by the verb *threaten*. This syntactic complementation corresponds to an instance of semantic embedding, where the embedding predication indicated by the main verb (em_1) embeds the predication indicated by the complement verb (em_2). In addition to *finite complementation* as in the example above, *infinitival complementation* is also common and performs a similar function, as the complementation between *fail* and *reach* show in the sentence below, as well as complementation between *threaten* and *wipe* in Example (86). This is again an instance where the syntactic subordination is isomorphic to semantic embedding, as illustrated in Example (87b).

(87) (a) The cuts are **necessary** because Congress and the administration have failed to reach **agreement** on a deficit-cutting bill. (wsj_2384)

(b) $Congress\_and\_the\_administration(t_1) \wedge The\_cut(t_2) \wedge agreement(e_1, WR, 0.0_{epistemic} \ldots)$
$\wedge fail$:SUCCESS$(em_2, WR, 1.0_{epistemic}, t_1, em_3) \wedge reach$:SUCCESS$(em_3, WR, 0.0_{epistemic}, t_1, e_1)$
$\wedge\ necessary(e_4, WR, 1.0_{epistemic}, t_2) \wedge because$:CAUSE$(em_5, WR, 1.0_{epistemic}, em_2, e_4)$

Adverbial Subordination

Adverbial subordination is one of the major devices that provide discourse connectivity and coherence. In the previous example (87a), discourse connectivity is signalled with the subordinating conjunction *because*, which provides a causal connection between the subordinate adverbial clause (corresponding to em_2) and the main clause (corresponding to e_4). In this example, it is also important to note that while the syntactic relation between the predicates *necessary* and *failed* is a clear case of syntactic subordination (the former is the main predicate of the matrix clause and the latter is the main predicate subordinate adverbial clause), there is no semantic embedding relation between the corresponding predications ($e_4 \not> em_1$ and $em_1 \not> e_4$), both simply function as arguments of the same predication (em_3). As such, in the case of adverbial subordination, we cannot speak of an isomorphism between syntactic subordination and semantic embedding.

Relativization

Relativization is the type of subordination in which a subordinate clause (called a *relative clause*) post-modifies a noun phrase or a pronoun. Relative clauses may be introduced by relative pronouns (e.g., *who*, *which*) or the complementizer *that*[3]. Relativization is not specifically linked to semantic embedding; atomic predications can also be signaled with relativization. In the case of relativization, the subordination between the head of the matrix clause and the head of the embedded clause in the syntactic domain is reversed in the case of semantic embedding: the predication indicated by the embedded clause has scope over that indicated by the matrix clause head. In the example below, the predication indicated by the embedded predicate (*allow*) takes that indicated by the relative clause head (*agreement*) as an argument and, therefore, embeds it.

(88) (a) The TVA, in fact, decided to proceed with the bond offering following an **agreement** last week with the Financing Bank, which <u>allows</u> TVA to <u>keep</u> **borrowing** short term from the bank for two years . . . (wsj_1943)

(b) $TVA(t_1) \wedge agreement(e_1,\ \ldots) \wedge borrow(e_2, \ldots) \wedge$ $keep{:}\text{CONTINUE}(em_3, WR, 1.0_{epistemic}, t_1, e_2) \wedge allow{:}\text{ENABLE}(em_4, WR, 1.0_{epistemic}, e_1, em_3)$

4.2.3 Syntactic Coordination

Coordination is the other major clause-linking device, as mentioned earlier when it was contrasted with subordination. The discourse coherence function of coordinating conjunctions (such as *and* and *but*) is well-documented. An example is shown below, where the conjunction *and* coordinates the full clauses headed by the verbal predicates *base* and signals a discourse CONJUNCTION relation between the atomic predications signalled by them (e_1 and e_2).

(89) (a) United Illuminating is **based** in New Haven, Conn., <u>and</u> Northeast is **based** in Hartford, Conn. (wsj_0013)

(b) $United\_Illuminating(t_1) \wedge New\_Haven,\_Conn.(t_2) \wedge Northeast(t_3) \wedge Hartford,\_Conn.(t_4)$ $\wedge\ base(e_1, WR, 1.0_{epistemic}, t_1, t_2) \wedge base(e_2, WR, 1.0_{epistemic}, t_3, t_4) \wedge$ $and{:}\text{CONJUNCTION}(em_3, WR, 1.0_{epistemic}, e_1, e_2)$

At the intra-clausal level, coordination gets more complex, since items from practically any lexical category and phrases of any kind can be coordinated. Furthermore, arguments may be eliminated to avoid repetition (*coordination ellipsis*) and such elliptical cases need to be resolved. VP-coordination,

[3]In reduced relative clauses, the relativizing marker may be omitted.

NP-coordination as well as NP head- and modifier-coordination are among the coordination types that need to be considered at the intra-clausal level. Examples for these classes are provided below, details irrelevant to the discussion of coordination have been omitted for clarity.

(90) (a) She [said Wheeler group was profitable but wouldn't **give** figures]. (wsj_1666) (VP-coordination)

She(t$_1$) $\wedge$ *say*:REPORTING*(em$_2$, WR,1.0$_{epistemic}$,t$_1$...)* $\wedge$

wouldn't:ASSUMPTIVE*(em$_3$, WR,0.0$_{epistemic}$,t$_1$...)* $\wedge$

but:CONCESSION*(em$_1$, WR,1.0$_{epistemic}$,em$_2$,em$_3$)*

(b) ...the price surge of 1979-80 precipitated an expansion of [mine **production** and scrap **recovery**] ...(wsj_1554) (NP-coordination)

mine(t$_1$) $\wedge$ *scrap(t$_2$)* $\wedge$ *production(e$_1$,...,t$_1$)* $\wedge$ *recovery(e$_2$,...,t$_2$)* $\wedge$

and:CONJUNCTION*(em$_1$, WR,1.0$_{epistemic}$,e$_1$,e$_2$)* $\wedge$

expansion:PROPOSITIONAL*(em$_2$, WR,1.0$_{epistemic}$,em$_1$)*

(c) Britain and all of Europe need to reconsider the prospects for European integration in light of the possible [**reunification** and **neutralization**] of Germany (wsj_0571) (NP head-coordination)

Germany(t$_1$) $\wedge$ *reunification(e$_1$,...,t$_1$)* $\wedge$ *neutralization(e$_2$,...,t$_1$)* $\wedge$

and:CONJUNCTION*(em$_3$, WR,0.5$_{epistemic}$,e$_1$,e$_2$)* $\wedge$

possible:SPECULATIVE*(em$_4$, WR,1.0$_{epistemic}$,em$_3$)*

(d) The [unnecessary and inappropriate] **use** of the hospital, and not the actual need for a particular procedure, has been the main focus," the panel said. (wsj_1793) (NP modifier-coordination)

hospital(t$_1$) $\wedge$ *panel(t$_2$)* $\wedge$ *use(e$_1$,...,t$_1$)* $\wedge$ *unnecessary*:OBLIGATIVE*(em$_2$,t$_2$,1.0$_{epistemic}$,e$_1$)*

$\wedge$ *inappropriate*:EVALUATIVE*(em$_3$,t$_2$,1.0$_{epistemic}$,e$_1$)* $\wedge$

and:CONJUNCTION*(em$_4$,t$_2$,1.0$_{epistemic}$,em$_2$,em$_3$)*

Examples (90a,c,d) illustrate cases of coordination ellipsis. In order to identify the arguments of predications accurately in these cases, we need to resolve the sentence elements that are shared between the conjuncts. In (90a), this element is the subject (*She*) (*She said ... but* {*She*} *wouldn't give figures*). In (90c), the shared element is *Germany* and, in (90d), it is the nominal predicate *use*. In these examples, similar to adverbial subordination, there is no semantic embedding relation between the conjuncts; both are embedded by the same predication (to *em*$_1$ in (90a), for example).

4.3 Dependency Parsing

Knowing the kinds of syntactic phenomena that are relevant in semantic embedding, how do we capture such syntactic constructions in a sentence automatically? Syntactic parsing is the task of identifying the syntactic structure of a sentence, and there are various formalisms in representing such structure. Two major formalisms are phrase structure and dependency grammars. The phrase structure formalism (introduced by Chomsky) is constituency-based: a sentence is represented as a nesting of multi-word constituents. On the other hand, in dependency grammar (Mel'čuk, 1988), a sentence is represented as a collection of relations between individual words, generally typed grammatical relations, such as nominal subject or indirect object. Dependency grammar has recently grown in popularity, particularly, in relation extraction tasks, because it reveals information about predicate-argument structure more directly than phrase structures and has the ability to capture long-range grammatical relations directly. In addition, it is considered to be more suitable for flexible or free word order languages.

Analysis of syntactic dependency relations is at the core of our framework and is the main mechanism in moving from syntactic structure to semantic predications. In dependency representation, a relation is formalized as a directed grammatical relationship involving two words (*governor* and *dependent*), and a sentence is represented as a graph, where the nodes correspond to words and the edges to grammatical relations between them. Dependency relations can be directly identified from the sentence (native dependency parsing) or they can be identified from phrase structure parses. Among native dependency parsers, MST Parser (McDonald *et al.*, 2006) and MALT Parser (Nivre *et al.*, 2006) performed best in the CoNLL-X Shared Task on Multilingual Dependency Parsing (Buchholz and Marsi, 2006). On the other hand, the Stanford dependency parser (de Marneffe *et al.*, 2006) extracts dependencies from phrase structure parses derived from a Penn Treebank-based parser, using rules defined on phrase structure. In the current work, we adopt the Stanford dependency parser for providing syntactic information.

4.3.1 Stanford Dependency Scheme

In Stanford dependency scheme, grammatical relations are arranged as a hierarchy, rooted with the most generic relation, *dependent* (abbreviated as *dep*)[4]. Each syntactic dependency is written as *abbreviated_dependency_name(governor, dependent)*, where governor and dependent are individual words in the sentence. For example, several dependency relations from the sentence in Example (91a)

[4]The entire grammatical relation hierarchy, which contains a total of 55 grammatical relations, is presented in de Marneffe and Manning (2008).

are given below. The full dependency graph of the sentence is presented in Figure (4.1).

(91) (a) Mr. Bush has said he personally approves of abortions in the cases of rape, incest and danger to the life of the mother.

(b) *nn(Bush,Mr.)* [nn: *noun compound modifier*]
nsubj(approves,Bush) [nsubj: *nominal subject*]
advmod(approves, personally) [advmod: *adverbial modifier*]
ccomp(said, approves) [ccomp: *clausal complement*]

The Stanford dependency parser provides dependency output of one of four types: *basic*, *collapsed*, *collapsed with propagation of conjunct dependencies*, and *collapsed preserving a tree structure*. Basic format preserves the tree structure, while in the collapsed output, the tree structure is broken and a directed graph is output. In this format, dependencies involving prepositions and conjunctions are collapsed to get direct relationships between content words. Collapsed with propagation option, as the name implies, propagates conjunct relations. In collapsed preserving a tree structure, those dependencies that break the tree are omitted.

The Stanford dependency parser has seen considerable use in relation extraction tasks. In the biomedical domain, the Stanford dependency scheme has also been proposed as a common representation for biological information extraction applications (Clegg and Shepherd, 2007). Björne *et al.* (2008) argue that Stanford "collapsed" dependency representation, in particular, closely captures the relevant semantics of biomedical text. The fact that it can extract dependency relations from a phrase structure parse makes it attractive as it can be combined with a highly accurate phrase structure parser for increased accuracy. In fact, in the latest BioNLP Shared Task on Event Extraction (Kim *et al.*, 2011a), best performing systems employed a combination of Charniak parser adapted to the biomedical domain (a highly accurate phrase structure parser) (McClosky and Charniak, 2008) with the Stanford dependency parser for identifying the underlying syntactic structure of sentences. In the current work, we adopt Stanford dependency relations as the source of syntactic information. We will describe how they are used later in Section 5.2. We use the default collapsed dependency with propagation of conjunct dependencies format in the experiments discussed in Chapter 6.

4.4 Conclusions

With this section, we presented the lexical semantic and syntactic information that underpin our semantic interpretation approach. The lexical semantic information regarding embedding predicates is encoded in a dictionary which consists of more than 900 predicates with more than 1300 senses, semi-automatically harvested from various resources and manually updated. The syntactic constructions that play a role in encoding embedding predications were also discussed and exemplified. We concluded this chapter by introducing syntactic dependency formalism and the Stanford dependency scheme, which provides the syntactic information regarding the syntactic constructions discussed earlier. With the material introduced in this chapter at our disposal, we can now turn to describing our compositional semantic interpretation approach.

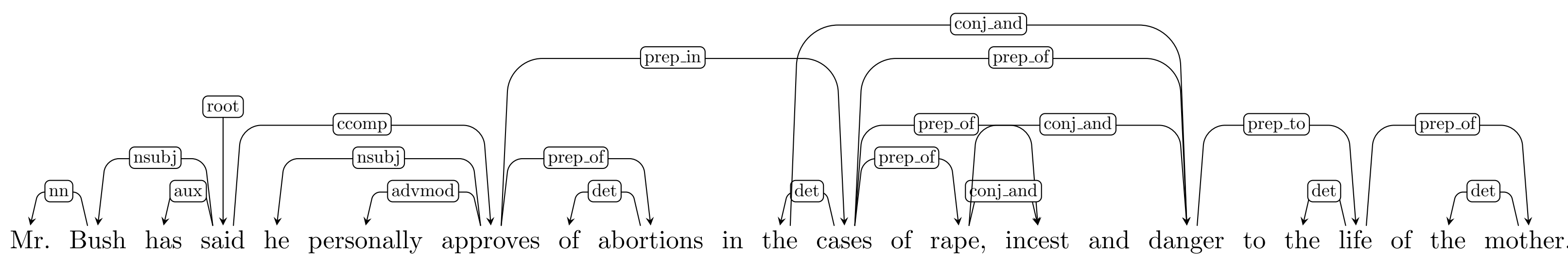

Figure 4.1: The dependency parse representation of a sentence from wsj_2075, extracted by the Stanford dependency parser (de Marneffe *et al.*, 2006), based on the phrase structure of the sentence provided in PTB (Marcus *et al.*, 1993).

Chapter 5

Compositional Construction of Predications

So far, we simply presented the meaning representations derived from sentences in the form of atomic and embedding predications but largely ignored how these representations are constructed automatically from the surface elements. In the previous chapter, we described the main resources that we rely on in constructing these predications: namely, the embedding predicate dictionary and syntactic dependency relations. In this chapter, we will describe the procedure that takes these as input and automatically derives meaning representations.

Predication construction is a bottom-up, compositional process. While the focus of the current work is embedding predications, the compositional approach can extend to construction of atomic predications, as well. The main difference is that embedding predications are associated with semantic classes from the embedding categorization, whereas atomic predications are not.

The composition phase builds mainly on the following components:

- Syntactic dependency parse of each sentence in the document.
- Word information, including lemma, part-of-speech, and positional information.
- The embedding predicate dictionary.
- (Optionally) additional semantic information associated with the document, in the form of semantic terms and atomic predications.

Additional semantic information allows the framework to integrate with an external relation extraction system or use existing term and relation annotations. However, for the sake of discussion here, we will mostly assume that no additional semantic information is provided.

The overall methodology is in the tradition of graph-based semantic representations (Sowa, 1984) and it aims at a level of representation similar to the deep-syntactic level proposed in Meaning-Text Theory (Mel'čuk, 1988). Using the components listed above, we first construct a semantic embedding graph representing the content of the document and make semantic dependencies explicit, guided by transformation rules, described in Section 5.2. Next, we compose predications by traversing the embedding graph in a bottom-up manner, guided by several compositional operations, such as argument identification and source propagation, discussed in Section 5.3. Limited coreference resolution is also performed in predication composition, and we conclude this chapter by describing the coreference resolution module. The pipeline is illustrated in Figure (5.1).

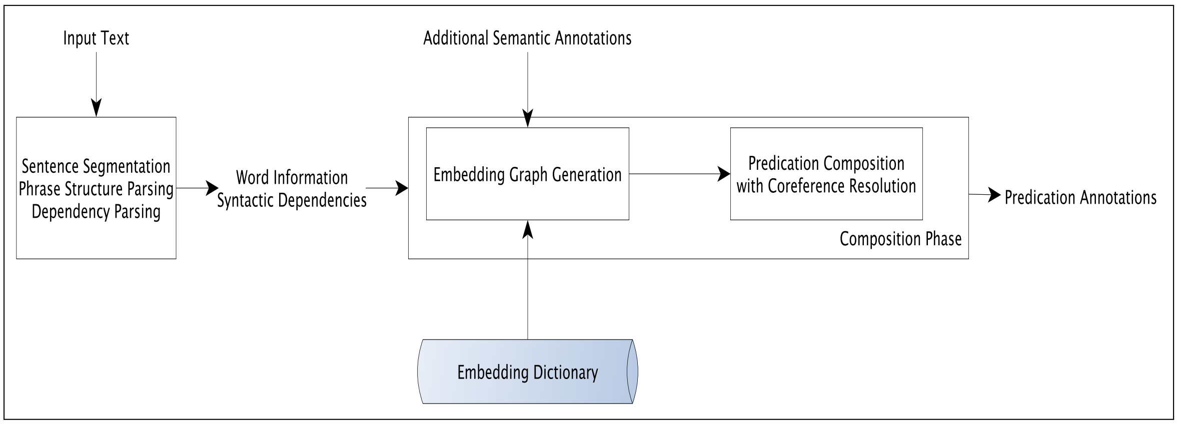

Figure 5.1: The pipeline for predication construction.

5.1 Pre-processing

Two components required for the composition phase, a syntactic dependency parse of each sentence and word information, are obtained in the pre-processing step. In this step, each document is first segmented into sentences using simple regular expressions for detecting sentence boundaries. Next, each sentence is parsed using the re-ranking parser of Charniak and Johnson (2005)[1] (a phrase structure parser) and syntactic dependencies are extracted from the resulting parse trees using the Stanford dependency parser (de Marneffe *et al.*, 2006), which also provides word information, including lemma and positional information.

[1]For processing biomedical text, we use the version of this parser adapted to the biomedical domain (McClosky and Charniak, 2008).

5.2 From Syntactic Dependencies to Semantic Embedding Graph

As the first step in semantic composition, we convert syntactic dependencies from sentences of the document (D) into a semantically-enriched directed, acyclic semantic embedding graph (G_D), whose nodes correspond to *surface elements* of the document, V, and whose labeled arcs correspond to semantic *embedding relations* between surface elements, E (G_D := (V, E)). A semantic embedding graph is aimed at capturing the semantic structure of a document and allowing a bottom-up semantic interpretation.

Definition 17. An **embedding relation** E in the semantic embedding graph G holds between two surface elements A and B and has type T.

$$E := T(A, B)$$

The surface element A is said to **dominate** B in the semantic embedding graph G.

$$A >_d B$$

If the surface elements A and B are semantically bound, the semantic object associated with A embeds (and has scope over) that associated with B.

$$(A >_d B) \wedge [\![A]\!] \neq \emptyset \wedge [\![B]\!] \neq \emptyset \Rightarrow [\![A]\!] > [\![B]\!]$$

An embedding relation is clearly similar to a typed syntactic dependency, in that it involves two elements, a *governor* and a *dependent*, and has a type. What distinguish it from a typed syntactic dependency are the following:

- Elements of an embedding relation can be multi-word units optionally bound to semantic objects, whereas an element of a syntactic dependency is a single, semantically free word.
- Direction of an embedding relation reflects the semantic dependency between its elements, rather than their syntactic dependency. More specifically, $A >_d B$ does not imply that B is syntactically dependent on A.
- An embedding relation can cross sentence boundaries.

Intra-sentential semantic dependencies are derived from the syntactic dependencies of the sentence via transformation rules, described next. Inter-sentential semantic dependencies, on the other hand, are captured with two special embedding relations, PREV and COREF, concerned with adjacency of sentences and coreference, respectively.

5.2.1 Syntactic Dependency Transformation Rules

The conversion of syntactic dependencies to embedding relations is guided by a set of intra-sentential transformation rules that are applied sequentially. The initial input to the transformation rules is simply the syntactic dependency graph of the sentence. Each of the subsequent rules takes a partially transformed semantic embedding graph corresponding to the sentence (G_S) as well as word information as input and returns as output a set of surface elements and embedding relations, a transformed graph of the sentence (G_S'). Transformation rules serve several functions:

1. Enriching the embedding graph with semantic information, including:
   (a) semantic information already provided (term and/or predicate mentions).
   (b) generic semantic information that can be inferred from lexical properties of the surface elements and the graph structure.
2. Capturing semantic dependency behavior from syntactic dependencies, by addressing the relevant syntactic phenomena and the (non-)isomorphism between syntactic structures and semantic dependencies, discussed in Section 4.2.
3. Correcting syntactic dependencies that are systemically misidentified by the dependency parser, while, at the same time, transforming them to embedding relations.

It is important to note that a transformation may not be necessary when the syntactic dependency under consideration is isomorphic to an embedding relation, that is, it reflects the direction of the semantic dependency between the elements of the dependency accurately. We describe the transformation rules below. Note that the labels for embedding relations are sometimes the same as syntactic dependency relations, while sometimes we adopt new labels that distinguish them better. Embedding relation labels are denoted in uppercase to distinguish them from syntactic dependency labels.

Semantic Enrichment of the Embedding Graph

Two transformation rules are specifically concerned with binding semantic objects with surface elements. The first, Imposing Semantic Information, is concerned with semantic information already provided, as embedding predicates or additional term/predicate annotations, whereas the second, Generic Semantic Object Binding, uses structural properties of the graph and lexical properties of the surface elements to bind these surface elements to generic semantic objects. A generic semantic object is an entity or a predicate that is not assigned an explicit semantic type (or category). This is not to imply that object in question does not have a semantic interpretation. Rather, this means that our system does not know about its semantics beyond the knowledge of whether it is an entity or a predicate.

Imposing Semantic Information If there are embedding predicate mentions in the sentence, the relevant senses in the dictionary are associated with the surface elements corresponding to these mentions. If the predicate is a single-word lexical unit, the corresponding node is semantically bound to these senses. If the predicate is a multi-word expression, the graph needs to be updated, as well. The nodes in the input graph that correspond to the words of the mention are collapsed into a single node to allow its treatment as a single unit and the new node is semantically bound. This further necessitates that all edges that the words of the mention were involved in be redirected to the new node. For example, consider the sentence below, whose transformation is illustrated in Example (92b) and in Figure (5.2) as (a) → (b)[2]. The adverbial *on the other hand* is associated with the semantic category CONTRAST in the embedding dictionary. With the transformation, its treatment as a single surface element with the CONTRAST type is allowed. Also note that the dependency between the verb *retained* and *hand* is substituted with one between the verb and *on the other hand*, since the *hand* node was collapsed into the new node corresponding to the multi-word expression *on the other hand*.

(92) (a) <u>On the other hand</u>, Mr. Cheney <u>retained</u> all those new land forces.

(b) *prep_on(retained, hand)*, *det(hand,the)*, *amod(hand,other)* → *PREP_ON(retained,On the other hand:*CONTRAST*)*

(c) *PREP_ON(retained,On the other hand:*CONTRAST*)* → *ADVMOD(On the other hand:*CONTRAST*,retained)*

In addition to embedding predicate mentions, this transformation can also deal with externally supplied semantic annotations in the form of semantic terms (entities) and predicates, if any.

Generic Semantic Object Binding This transformation is concerned with semantically enriching the embedding graph based on graph and lexical properties of the surface elements. For example, pronominal and definite noun phrase anaphoric expressions in a sentence are identified and are bound to a semantic object of ANAPHOR type, allowing coreference resolution in later steps.

[2]The transformation illustrated as (b) → (c) in Figure (5.2) and in Example (92c) is Discourse Connective Transformation, which will be discussed shortly.

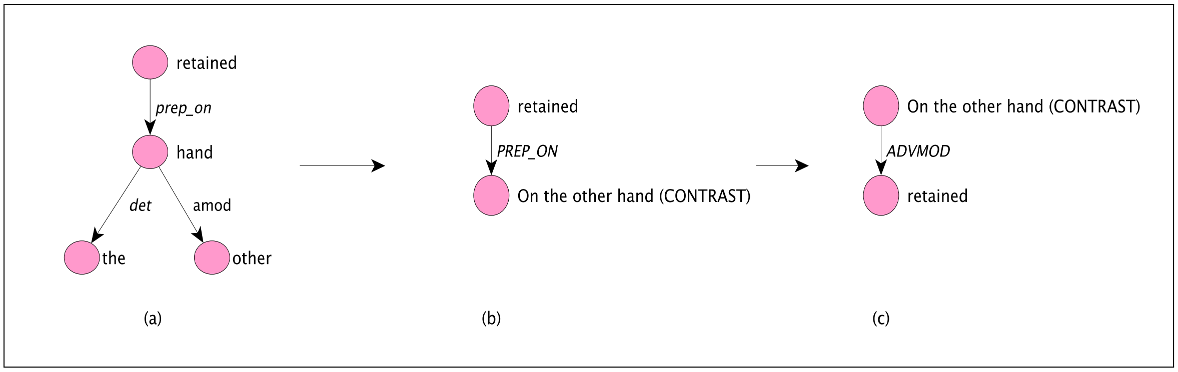

Figure 5.2: Imposing Semantic Information and Discourse Connective Transformation rules applied to the sentence in Example (92a).

For example, in the graph for the fragment in Example (93a), the node for the personal pronoun *he* is bound to an ANAPHOR object, as shown in (93b).

(93) (a) Mr. Bush has said he personally approves of abortions ...

(b) *nsubj(approves,he)* → *NSUBJ(approves:*PROPOSITIONAL*,he:*ANAPHOR*)*

nsubj(said,Mr. Bush) → *NSUBJ(said:*REPORTING*,Mr. Bush:*ENTITY*)*

prep_of(approves,abortions) → *PREP_OF(approves:*PROPOSITIONAL*,abortion:*PREDICATE*)*

Another use of this transformation is to associate semantically free surface elements with generic semantic objects. For example, if a surface element is a leaf node in the graph or is nominal (its head is a noun) but not a nominalization, it is bound to a generic entity object. On the other hand, if it is a nominalization, or a gerund, or a non-leaf verb, adjective or adverb, then it is bound to a generic predicate. In the example above, the surface element *Mr. Bush* is bound to a generic entity, whereas *approves* and *abortions* are bound to generic predicates, as shown in (93b)[3]. This transformation rule, which is applied as the final transformation, allows us to compose semantically underspecified predications, if needed.

Capturing Semantic Dependencies

Most of the transformation rules are concerned with capturing the semantic dependencies explicitly from syntactic dependencies, based on the properties of the surface elements involved, such as their lexical category, or whether they are semantically bound. These transformation rules allow us to take into account the syntactic phenomena relevant to embedding, such as syntactic subordination

[3]Note that the surface element *approves* is bound to a PROPOSITIONAL predicate, which is essentially a generic embedding predicate, whereas *abortions* is bound as an generic atomic predicate, due to their position on the graph.

and coordination, and, at the same time, address the non-isomorphism between syntactic structure and semantic dependencies.

Dependency Direction Inversion For particular types of syntactic dependencies, largely encoding syntactic modification instances, the syntactically dependent element semantically embeds (or dominates) the syntactically dominant element. Therefore, the direction of the dependency between the elements is inverted to reflect the semantic dependency. The syntactic dependency types that function this way are *partmod* (participial modifier), *aux* (auxiliary), *auxpass* (passive auxiliary), *cop* (copula), *mark* (marker), *neg* (negation modifier), *rcmod* (relative clause modifier), and *parataxis* (parataxis). For some dependency types, inversion is allowed only in restricted cases. For example, for the *advmod* (adverbial modifier) dependency, it is allowed only if the adverbial that serves as the dependent is *non-restricting* (see Section 4.2.1). Consider the fragment in Example (94a), where there is a marker dependency between *searched* and *while*.

(94) (a) But he put off a firm decision <u>while</u> his aides and legislators <u>searched</u> for a compromise . . .

(b) *mark(searched,while)* $\rightarrow$ *MARK(while,searched)*

(c) *prt(put,off)* $\rightarrow \emptyset$

(d) *advcl(put off,while)* $\rightarrow$ *ADVCL(while,put off)*

The semantic dependency between them is captured by reversing the direction of the dependency, as illustrated with the embedding relation in (94b). This is also illustrated as the transformation (a) $\rightarrow$ (b) in Figure (5.3).

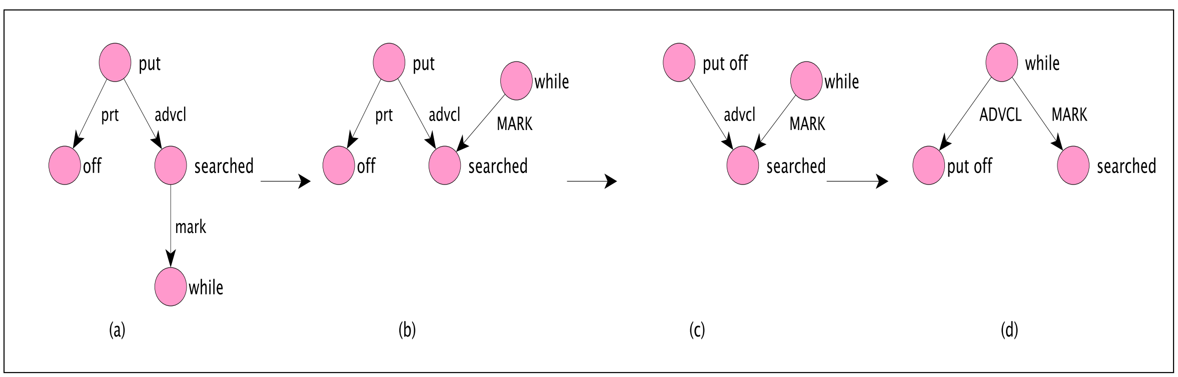

Figure 5.3: Dependency Direction Inversion, Verb Particle Transformation and Adverbial Clause Transformation rules applied to the sentence in Example (94a).

Phrasal Verb Transformation This transformation collapses a phrasal verb particle with the verb itself, allowing us to treat them as a single semantic unit. The syntactic dependency *prt* (particle) is sought between the verb and the particle. The result of the transformation is to remove this dependency and collapse the verb and particle into a single surface element. This transformation occurs with the sentence fragment in Example (94a), in which the nodes corresponding to *put* and *off* are collapsed into a new node for *put off*. This is illustrated in Example (94b) as well as in Figure (5.3) as the transformation (b) → (c).

Adverbial Clause Transformation Adverbial clauses often indicate discourse relations and this transformation is aimed at capturing the semantic dependency between discourse elements. The syntactic dependency configuration for an adverbial clause is indicated by two dependencies: (a) *advcl* (adverbial clause modifier) dependency, and (b) *advmod* (adverbial modifier) or *mark* (marker) dependency. Assume that the subordinating conjunction marking the adverbial clause is the surface element a, the main predicate of the adverbial clause is b, and the verbal predicate that is modified by the adverbial clause is c. In this configuration, the syntactic dependencies are *advcl(c,b)* and *syn_dep(a,b)*, where *syn_dep* is one of *mark* or *advmod* dependencies. This configuration is substituted with *ADVCL(a,c)* and *SYN_DEP(a,b)*, reflecting the direction of the semantic dependency indicated by the discourse connective, a. This transformation occurs in the sentence fragment in Example (94a) and is illustrated in Example (94d) as well as in Figure (5.3) as the transformation (c) → (d).

Discourse Connective Transformation Discourse connectives are often multi-word adverbials (e.g., *in contrast*, *on the other hand*). If such an expression, a, begins with a preposition p, there is usually a surface element b and a dependency of type *prep_p*, where the multi-word surface element is in the dependent position (*prep_p(b,a)*). We convert this dependency to *ADVMOD(a,b)* to reflect the direction and nature of the semantic dependency. This transformation occurs in sentence in Example (92a) and illustrated in Example (92c), as well as in Figure (5.2) as (b) → (c).

This transformation also accommodates a side effect of the collapsed dependency output format of the Stanford dependency parser, which may treat subordinating conjunctions as part of the dependency labels rather than individual words. Consider the fragment in Example (95a) and the syntactic dependency of the *prep_despite* type extracted from this sentence. Since such subordinating conjunctions often act as discourse connectives, we break down the relevant syntactic dependency into two embedding relations, while also making their discourse connective function more explicit.

The transformation is illustrated in Example (95b).

(95) (a) <u>Despite</u> valiant <u>efforts</u> by Finance Minister Mailson Ferreira da Nobrega, inflation <u>came</u> to 36% in September alone . . .

(b) *prep_despite(came,efforts)* $\rightarrow$ *ADVCL(Despite,came), MARK(Despite,efforts)*

Coordination Transformation The coordination between two surface elements *a* and *b* indicated by the conjunction *c* is encoded by the syntactic dependency *conj_c(a,b)* in the Stanford dependency scheme. The coordination transformation primarily serves to reflect the semantic dependency between the conjunction and the conjuncts, by splitting this dependency into two: *CC(c,a)* and *CC(c,b)*. Consider the fragment in Example (96).

(96) (a) United Illuminating is <u>based</u> in New Haven, Conn., <u>and</u> Northeast is <u>based</u> in Hartford, Conn.

(b) *conj_and(based,based)* $\rightarrow$ *CC(and,based)*, *CC(and,based)*

Serial coordination between more than two surface elements indicated by two or more dependencies is also transformed. Generally, in a serial coordination scenario (*a,b, and d*, for example), the syntactic dependencies are given as *conj_and(a,d)* and *conj_and(b,d)*. These are transformed into three semantic dependencies *CC(and,a)*, *CC(and,b)* and *CC(and,d)*.

Verbal Dependency Transformation This transformation is concerned with transforming the relations that a verb is involved in as the dependent. More specifically, it performs two functions: (a) making clausal complement dependencies more specific based on the complementizer and pruning redundant dependencies, (b) reordering embedding relations, in which the verb is the dependent, into an embedding chain ending in the verb so as to reflect the semantic scope relations between the surface elements explicitly.

To illustrate, consider the sentence in Example (97a), where the verb of interest is *suffer*. The clausal complement function is illustrated in Example (97b). The dependencies are collapsed into a single embedding relation with the type *THAT_COMP*, which readily encodes the complementizer, *that*, while making the *complm* dependency redundant. On the other hand, the embedding relations obtained from earlier transformations are given in Example (97c). From these three embedding relations, it is clear that *suffer* is in the semantic scope of *said*, *would*, and *if*, but how should these three surface elements be ordered to capture the semantic dependencies accurately? We use the principle of proximity to determine this. With this principle, we assume that the surface element closer to the verb at the surface level has smaller scope and is therefore lower in the embedding chain[4]. While there are exceptions to this principle, it seems to be a relatively good approximation. With this principle, we determine the embedding relations illustrated in Example (97d), from those in Example (97b-c).

(97) (a) Mr. Nemeth said in parliament that Czechoslovakia and Hungary would suffer environmental damage if the twin dams were built as planned.

(b) *ccomp(said,suffer)*, *complm(said,that)* $\rightarrow$ *THAT_COMP(said,suffer)*

(c) *AUX(would,suffer)*, *ADVCL(if,suffer)*[5]

(d) *THAT_COMP(said,if)*, *ADVCL(if,would)*, *AUX(would,suffer)*

Noun Phrase-Internal Transformation This transformation considers the NP-internal modifier dependencies, which include *nn* (nominal modifier), *amod* (adjectival modifier), *quantmod* (quantifier phrase modifier), *measure* (measure-phrase modifier), *det* (determiner), *num* (numeric modifier) and *poss* (possessive). The assumption here is that if all elements of the noun phrase are semantically free, then we can treat it as a single unit, ignoring all the NP-internal dependencies. Thus, if a semantically free head noun a is linked to a set of semantically free modifier elements through a set of syntactic dependencies with the types listed above, all of these syntactic dependencies are pruned and all the surface elements are collapsed into a single unit. Consider the noun phrase *the White House*. The following two NP-internal syntactic dependencies hold: *nn(House,White)* and *det(House,the)*. This transformation results in pruning these dependencies and substituting individual surface elements corresponding to each word with a single surface element corresponding to *the White House*.

Complex Adjectival Phrase Transformation This transformation applies to complex, hyphenated adjectives, which appear quite commonly in biomedical text. Consider the phrase in Example (98a). Generally, a single dependency of *amod* (adjectival modifier) type holds for such a phrase, as shown in Example (98b). This transformation applies when it is known that the right side of the hyphen in the adjective (*induced*) corresponds to a semantic predicate. In these cases, we split the complex adjective into two, while also substituting the dependency with two embedding relations. Two embedding relations created in this case, labeled *SUBJ* and *OBJ* are shown in Example (98b).

[4] This was inspired by the Proximity Principle of Givón (2001).

[5] The first embedding relation is due to Dependency Direction Inversion and the second to Adverbial Clause Transformation.

(98) (a) ... gp41-<u>induced</u> upregulation ...

(b) *amod(upregulation,gp41-induced)* $\rightarrow$ *SUBJ(induced,gp41)*, *OBJ(induced,upregulation)*

If the adjective is a present participle rather than a past participle (e.g., *inducing* instead of *induced*), then the dependents of the embedding relations are reversed.

Wide/Narrow Scope Transformation This transformation is aimed at capturing the wide or narrow scope associated with some embedding predicates. Consider the fragment in Example (99a). The effect of narrow scope of the predicate *think* is illustrated in Example (99b). The rest of the graph may also be updated, if necessary, so that the relations incoming to the negative particle are instead redirected to the embedding predicate with narrow scope.

(99) (a) We do<u>n't</u> <u>think</u> this will affect that.

(b) *NEG(n't,think), CCOMP(think,will)*[6] $\rightarrow$ *CCOMP(think,n't), NEG(n't,will)*

The effect of a predicate with wide scope is illustrated in Example (100). The wide scope predicate is the negative determiner *No*. The resulting embedding relation type is *NEG*, since currently all the wide scope predicates are negative items. The rest of the graph is also updated, if necessary. In this example, the embedding relation of *ADVCL* type is redirected to the negative determiner *No* as a result of the update.

(100) (a) <u>No</u> meeting is scheduled because the expansion ...

(b) *det(meeting,No)*, *nsubjpass(scheduled,meeting)*, *ADVCL(because,scheduled)*[7] $\rightarrow$
ADVCL(because,No), *NEG(No,scheduled), NSUBJPASS(scheduled,meeting)*

Corrective Transformations

Two transformation rules are designed to accommodate syntactic dependencies that are systemically misidentified and capture the semantic dependencies between the surface elements involved in these dependencies. These transformation rules are Prepositional Phrase Attachment Transformation and Modifier Coordination Transformation. They both rely on heuristics based on the involvement of semantically bound surface elements.

Prepositional Phrase Attachment Transformation A problem with prepositional phrase attachment problem arises in sentences like *I saw the man with the telescope*, where there is a structural

[6] These embedding relations are due to the Dependency Direction Inversion rule.
[7] This embedding relation is due to Adverbial Clause Transformation.

ambiguity, since the resulting meaning depends on whether the prepositional phrase *with the telescope* attaches to the noun phrase *the man* or to the verb *saw*. Incorrect attachment of prepositional phrases is a major problem of constituent parsers, leading to syntactic dependency errors by Stanford dependency parser. Schuman and Bergler (2006) outline some heuristics for accurately identifying the prepositional phrase attachment points. This transformation takes these heuristics as the basis for correcting some of the potentially erroneous syntactic dependencies while transforming them into semantic dependencies. It applies to a limited set of nominal embedding predicates, such as *role* and *influence*, whose syntactic behavior is explicitly specified in the dictionary. Consider the sentence in Example (101a). Note that the PP *in Ferranti* should attach to the NP headed by *role*. However, the relevant syntactic dependencies indicate, incorrectly, that it attaches to the verb *diminish*. The fact that *role* takes a syntactic object indicated by the preposition *in* is encoded in the dictionary since the predicate has PREP_IN as an embedding type. This, in addition to the fact that *role* is the direct object of the verb allow us to infer that the correct attachment point is the NP headed by *role*. The transformation is illustrated in Example (101b).

(101) (a) A consortium bid, however, would diminish GEC's direct <u>role</u> in Ferranti ...

(b) *prep_in(diminish,Ferranti)*, *dobj(diminish,role)* →
PREP_IN(role,Ferranti), DOBJ(diminish,role)

Modifier Coordination Transformation Another corrective transformation involves modifier coordination, again often misidentified by constituent parsers. We correct these dependencies in limited cases where the modifiers correspond to same type of semantic objects; that is, both either correspond to predicates or entities. Consider the fragment in Example (102a), taken from a molecular biology text, where the modifiers are underlined, and the head noun is *activation*. The modifiers correspond to PROTEIN terms. The *conj_and* syntactic dependency, as shown in Example (102b) indicates, incorrectly, that the coordination is between one of the modifiers *adenylate cyclase* and the head noun, *activation*. Since the modifier of this NP, *p70(S6)-kinase* has the same semantic type as the conjunct *adenylate cyclase*, the transformation illustrated in Example (102b) is applied.

(102) (a) ... <u>adenylate cyclase</u> and <u>p70(S6)-kinase</u> activation ...

(b) *conj_and(adenylate cyclase:*PROTEIN*,activation)*,
*amod(activation,p70(S6)-kinase:*PROTEIN*)* →
AMOD(activation,and),
*CC(and,adenylate cyclase:*PROTEIN*), CC(and,p70(S6)-kinase:*PROTEIN*)*

5.2.2 PREV and COREF Embedding Relations

Once the intra-sentential transformations are complete, we finalize the semantic embedding graph of the document by considering two types of special embedding relations:

PREV A semantic dependency that holds between the topmost nodes associated with adjacent sentences in order to reflect the sequence of sentences. If the governor element of a PREV embedding relation is the topmost node of the current sentence, its dependent is the previous sentence.

COREF A coreference relation that holds between an anaphoric element and its antecedent. The antecedent may be in the same sentence as the anaphor or in a prior sentence. COREF relations are the result of coreference resolution, described in Section 5.4.

These special relations allow us to move to the inter-sentential level in semantic interpretation. A portion of an example document embedding graph, corresponding to the sentence given earlier in Figure (4.1), is given in Figure (5.4) below. Note that two senses are associated with the surface item *and* (CONJUNCTION and LIST), indicating an ambiguity.

5.3 Composing Predications

After constructing the document embedding graph, we traverse it in a bottom-up manner and compose predications. At this stage, it is important to remember that we refer to the revised definition of predication here, represented as follows.

$$Pr =: [P, S, MV_{Sc}, ARG_{1..n}]$$

Recall that S refers to the *source*, and MV_{Sc} to the *scalar modality value* on the *Sc* scale in the [0,1] range. Atomic predications initially take an epistemic scalar value of 1.0. The source is assigned as *WR*, referring to the author of the text.

Predication composition involves five operations, each of which addresses a component of the predication. With *argument identification* rules, we determine the logical arguments, $ARG_{1..n}$, of a predication. *Source propagation* allows us to identify the source feature (S), whereas *scalar modality value composition* is concerned with calculation and update of the scalar modality value feature (MV_{Sc}). The two other operations, *polarity composition* and *argument propagation*, apply in more limited cases, the former to change the semantic category of the predication and the latter to extend argument identification based on graph structure and constraints.

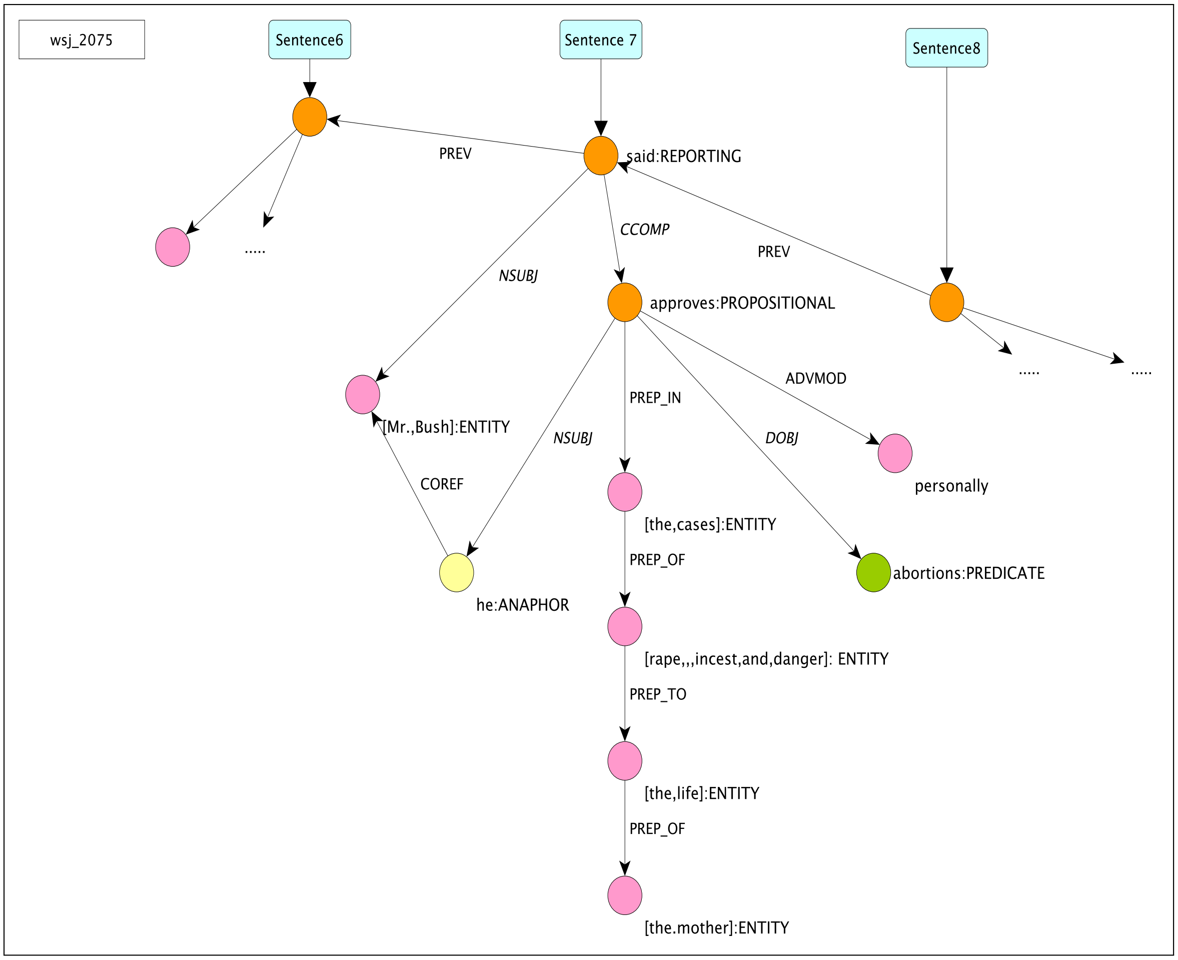

Figure 5.4: Partial semantic graph for the document wsj_2075. In the focus is the part corresponding to the sentence *Mr. Bush has said he personally approves of abortions in the cases of rape, incest and danger to the life of the mother.*

5.3.1 Argument Identification

Argument identification is concerned with determining the logical arguments of a predication, based on the embedding graph. It is guided by *argument identification rules*, each of which defines a mapping from a lexical category and an embedding relation type to a logical argument type (logical subject, object, or adjunct) and applies to a predicate belonging to the lexical category and serving as the governing element of an embedding relation of the specified type. An argument identification rule may be limited to some predicates, or predicates may be excluded from it. A more formal definition is below.

Definition 18. An **argument identification rule** $R{:}Q{\rightarrow}A$ is a typing function. Q is a 4-tuple $\langle T{,}POS{,}IN{,}EX\rangle$, where

- T is an embedding relation type

- *POS* is a part-of-speech category, such as verb (VB) or noun (NN).
- *IN* and *EX* are sets denoting inclusion and exclusion constraints, respectively

and A is the set of logical argument types (A = {Object, Subject, Adjunct}). A predicate sense S_P *satisfies* a constraint C if the lemma of the predicate or the sense category is included in C.

$$Lemma(P) \in C \vee Sem(S_P) \in C \Rightarrow satisfies(S_P, C)$$

Let V be a surface element corresponding to a non-leaf node in the embedding graph and E an embedding relation, such that E = $T(V, V_e)$. An argument identification rule R applies to the pair *(V,E)* and assigns the surface element V_e as the logical argument of type A for V, if

$$Part\_of\_speech(V) = POS \wedge (satisfies(V, IN) \vee (\neg satisfies(V, IN) \wedge \neg satisfies(V, EX))) \Rightarrow$$
$$applies\_to(R, V, E) \wedge A(V) = V_e$$

Some argument identification rules are exemplified in Table (5.1). The embedding relation type T may be specified lexically in the dictionary for each sense as an *embedding type* feature for logical arguments of Object type. The embedding relation types for Subject and Adjunct arguments, on the other hand, are determined based on lexical category and inclusion and exclusion criteria. We currently use about 80 such rules. All the children nodes of a non-leaf node are recursively processed for logical arguments with these rules.

Embedding Relation Type (T)	**POS**	**Inclusions (IN)**	**Exclusions (EX)**	**Argument Type (A)**
PREP_ON	NN	*influence,impact,effect*	-	Object
AGENT	VB	-	-	Subject
NSUBJPASS	VB	-	-	Object
WHETHER_COMP	VB	INTERROGATIVE	-	Object
PREP_IN	NN	-	*effect,influence,importance*	Adjunct

Table 5.1: Illustration of several argument identification rules. Each rule R is given as R:Q→A, where Q=⟨*T,POS,IN,EX*⟩. Note that inclusion and exclusion constraints may apply to embedding categories, as well as to specific predicates.

5.3.2 Source Propagation

Source propagation is the process of identifying the *source* (S) feature of relevant predications in the context of a source-introducing predicate. Source-introducing predicates are modal predicates that explicitly introduce the EPISTEMIC scale, namely, predicates with EPISTEMIC or EVIDENTIAL senses.

The source that is introduced by such a predicate (S_{Pr}) is its logical subject, if one exists. If the predicate is verbal and is passivized (as in, *It is believed that*), the source element is assigned the value of GEN (generic). If neither is the case, no source propagation is performed.

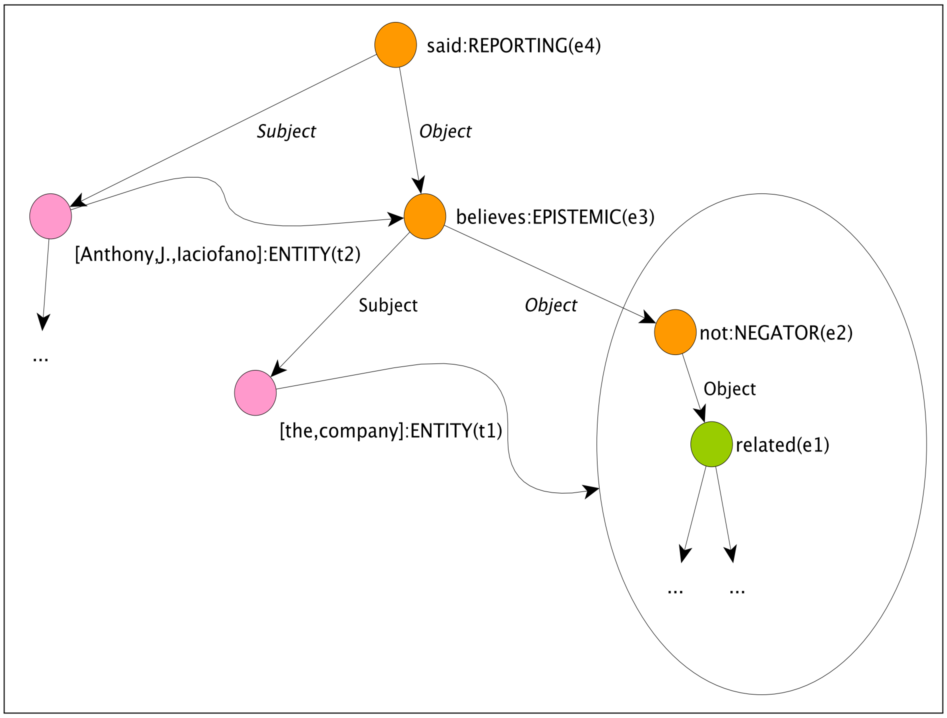

Figure 5.5: Illustration of source propagation for the sentence in Example (103a). The curved lines link the source with the predications that are affected.

Once the source value is identified, how do we determine the predications whose source value should be updated? The relevant predications are those that are in the scope of the current source-introducing predicate but not in the scope of another source-introducing predicate. Once these relevant predications are determined, their *source* value is set to S_{Pr}. The propagation occurs recursively until another source-introducing predicate or an atomic predicate is encountered. For example, consider the sentence in Example (103a), which was discussed earlier and is duplicated here. The corresponding predications are also illustrated in Figure (5.5). where the semantic terms functioning as the source are linked to predications they affect with curved lines. Note that *Anthony J. Iaciofano* (t_2) is only the source of the predication em_3 and not of em_2 and e_1, since both of these are in the scope of em_3, another source-introducing predicate, as well.

(103) (a) Anthony J. Iaciofano, a lawyer for Burlington, said the company believes the Beebes' symptoms were not related to the carpetings . . . (wsj_1946)

(b) $the\_company(t_1) \wedge Anthony\_J.\_Iaciofano(t_2) \wedge relate(e_1,t_1,0.2_{epistemic}, \ldots) \wedge$

not:NEGATOR$(em_2,t_1,0.8_{epistemic},e_1)$ $\wedge$

believe:EPISTEMIC$(em_3,t_2,1.0_{epistemic},t_1,em_2)$ $\wedge$

say:EPISTEMIC$(em_4,WR,1.0_{epistemic},t_2,em_3)$

5.3.3 Scalar Modality Value Composition

In Section 3.3, we discussed in detail how the influence of embedding predications extend to the predications in their scope. Scalar modality value composition is essentially the procedure of determining the relevant scale for a predication and its modality value on this scale, and essentially the principles outlined in that section are followed. To recap:

1. At the time of composition, every predication is assigned to EPISTEMIC scale with the value of 1.0.

2. A MODAL predicate places its logical object argument on the relevant MODAL scale and assigns to it its *prior scalar value*, specified in the embedding dictionary.

3. A VALENCE_SHIFTER predicate does not introduce a new scale but changes the existing scalar value or polarity of its logical object argument.

4. The scalar influence of an embedding predicate extends beyond its immediate predicational argument to other predications in its scope, if certain constraints, discussed in detail in Section 3.3 are met.

In this section, we assume that we have a predicate P which indicates an embedding predication Pr and a predication in its scope (Pr_e). We also assume that Pr has scalar influence over Pr_e, due to one of the principles above. The question that we try to answer in this section is how the scalar modality value of Pr_e is updated due to this scalar influence. The update procedure differs, based on whether the predicate P is a MODAL or a SCALE_SHIFTER predicate.

The update procedure, in the context of MODAL predicates, is affected by the following values:

- The *prior scalar value* of the relevant sense of the embedding predicate (*Scalar_Value*($\mathrm{P}_{\mathrm{MODAL}}$)), specified in the embedding dictionary.

- The current *scalar modality value* associated with the embedded predication ($\mathrm{MV}_{Sc}(\mathrm{Pr}_e)$).

The scalar value updates in the context of modal predicates based on these factors are illustrated in Table (5.2), where the first two columns correspond to the values above and the last column represents the new scalar modality value of the embedded predication after the update. X and Y represent any value in the range of [0,1]. For example, the computation in row 3 applies when the current scalar modality value of the embedded predication is 0.6 (Y) and the prior scalar value of the embedding predicate is 0.8 and results in increasing the value of the embedded predication to 0.8 (min(0.9,0.6+0.2)).

	***Scalar_Value*(P$_{\text{MODAL}}$)**	$\boldsymbol{MV_{Sc}(Pr_e)}$	$\boldsymbol{MV_{Sc}}(Pr_e)'$
1	= X	= 1.0	X
2	= X	= 0.0	1-X
3	$>$ Y	$> 0.5 \wedge =$ Y	min(0.9, Y+0.2)
4	$<$ Y $\wedge >= 0.5$	$> 0.5 \wedge =$ Y	min(0.5,Y-0.2)
5	< 0.5	$> 0.5 \wedge =$ Y	1-Y
6	$>= 0.5$	$< 0.5 \wedge =$ Y	Y
7	< 0.5	$< 0.5 \wedge =$ Y	1- Y

Table 5.2: The composition of scalar modality values in the context of MODAL predicates.

When P is a scale-shifting predicate, the update procedure is guided by its semantic category, as illustrated in Table (5.3). The last column, again, represents the new scalar modality value of the embedded predication after the update.

	***Sem*(P)**	$\boldsymbol{MV_{Sc}(Pr_e)}$	$\boldsymbol{MV_{Sc}}(Pr_e)'$
1	NEGATOR	= 0.0	0.5
2	NEGATOR	$> 0.0 \wedge =$ Y	1-Y
3	INTENSIFIER	$(= 0.0 \vee = 1.0) \wedge =$ Y	Y
4	INTENSIFIER	$>= 0.5 \wedge =$ Y	min(0.9,Y+0.2)
5	INTENSIFIER	$< 0.5 \wedge =$ Y	max(0.1,Y-0.2)
6	DIMINISHER	$(= 0.0 \vee = 1.0) \wedge =$ Y	Y
7	DIMINISHER	$>= 0.5 \wedge =$ Y	max(0.5,Y-0.2)
8	DIMINISHER	$< 0.5 \wedge =$ Y	max(0.4,Y+0.2)
9	HEDGE	= 0.0	0.2
10	HEDGE	= 1.0	0.8
11	HEDGE	= Y	Y

Table 5.3: The composition of scalar modality values in the context of scale shifting predicates.

To illustrate, consider a sentence we discussed earlier in Section 3.3, duplicated below as Example (104). Note that the scalar modality values shown in Example (104b) are the final composed values.

(104) (a) Maybe people don't believe I want to give this money away. (wsj_1409)

(b) *people*(t_1) $\wedge$ *I*(t_2) $\wedge$ *this_money*(t_3) $\wedge$ *give_away*$(e_1, t_1, 0.0_{volitive}, t_2, t_3)$ $\wedge$

want:VOLITIVE$(em_2, t_1, 0.2_{epistemic}, t_2, e_1)$ $\wedge$

don't:NEGATOR$(em_3, t_1, 0.5_{epistemic}, em_2)$ $\wedge$

believe:EPISTEMIC$(em_4, WR, 0.5_{epistemic}, t_1, em_3)$ $\wedge$

maybe:EPISTEMIC$(em_5, WR, 1.0_{epistemic}, em_4)$

These final composed values are the result of scalar modality value composition, which occurs four times in this sentence, with the predicates *want*, *don't*, *believe*, and *maybe*:

- With the VOLITIVE predicate *want*, the embedded predication is e_1, which is by default on the EPISTEMIC scale with the value of 1. The prior scalar value of *want* is 1 (X). The effect of the composition is to associate the embedded predication with the VOLITIVE scale and prior scalar value of *want*, which is 1 (Table (5.2), row 1).

- Next, with the NEGATOR predicate *don't*, the effect is to reduce the scalar modality values of the predications in scope (e_1 and em_2), which are initially both 1 (Y), to 0 (1-Y). (Table (5.3), row 2).

- With the epistemic predicate *believe*, which has a prior scalar value of 0.8 (X), the effect is:
  - to lower the modality value of em_3 from 1 to 0.8 (X) (Table (5.2), row 1)
  - to increase the value of em_2 from 0 to 0.2 (1-X) (Table (5.2), row 2).

- With the epistemic predicate *maybe*, which has a prior scalar value of 0.5 (X), the effect is:
  - to lower the modality value of em_4 from 1 to 0.5 (X) (Table (5.2), row 1)
  - to lower the modality value of em_3 from 0.8 to 0.5 (Table (5.2), row 4).

  The value of em_2 remains unchanged at 0.2 (Table (5.2), row 6).

5.3.4 Polarity Composition

Polarity composition is only relevant in the context of polarity-shifting predicates and causal embedded predications. The effect of polarity composition is to combine the predication associated with the polarity-shifting predicate and the embedded predication (Pr and Pr_e, respectively) to yield a new composite predication (Pr') with a new semantic category, $Sem(Pr')$, as shown in Table (5.4).

Sem(P)	**Sem(Pr$_e$)**	**Sem(Pr$'$)**
POSITIVE_SHIFT	CAUSAL	CAUSE
NEGATIVE_SHIFT	CAUSAL	PREVENT

Table 5.4: The polarity composition operations.

We find the applicability of this type of composition largely limited to biomedical text. Consider the sentence in Example (105a), from the GENIA corpus. The result of the polarity composition is illustrated in collapsing two predications (em_2 and em_3) in Example (105b) into one (em_4) in Example (105c).

(105) (a) IL-4 has already been shown to negatively regulate the development of naive T cells ...

(b) $naive\_T\_cells(t_1) \wedge IL\text{-}4(t_2) \wedge development(e_1, WR, 1.0_{epistemic}, t_1) \wedge$
$regulate{:}\textsc{causal}(em_2, WR, 1.0_{epistemic}, t_2, e_1) \wedge$
$negatively{:}\textsc{negative\_shift}(em_3, WR, 1.0_{epistemic}, em_2) \wedge$

(c) $negatively\_regulate{:}\textsc{prevent}(em_4, WR, 1.0_{epistemic}, t_2, e_1)$

5.3.5 Argument Propagation

Argument propagation extends argument identification, discussed earlier in Section 5.3.1. In that discussion, we assumed that each node in the semantic embedding graph is semantically bound: either the surface element is associated with an embedding predicate sense from the dictionary or with a generic semantic object (an entity or a predicate) based on its structural and lexical properties[8]. In practical tasks, relevant semantic objects, entities and/or predicates, may be specified explicitly. For example, in biological event extraction tasks (see Section 6.3), the entities of PROTEIN type may be provided from a named entity recognition system. In such cases, we do not apply the Generic Semantic Object Binding transformation, allowing semantically free intermediate nodes in the embedding graph[9]. Argument propagation plays a role in such scenarios. It is concerned with determining whether a descendant of the current node in the semantic embedding graph can serve as its argument, when the intermediate nodes between them are *semantically free*.

Definition 19. Let A and C be semantically bound surface elements ($[\![A]\!] \neq \emptyset \wedge [\![C]\!] \neq \emptyset$), C an ancestor of A in the embedding graph, and $\mathcal{B}$ the set of nodes that form the path from C

[8]Recall the Generic Semantic Object Binding transformation described in Section 5.2.1.

[9]A question that may arise in this context is why this transformation is not always applied. The decision is largely pragmatic. In the context of tasks, such as biological event extraction, only relations involving entities and predicates of certain types are of interest. Argument propagation allows such underspecified interpretation more readily. Application of the transformation, on the other hand, would yield a fuller semantic interpretation, but would also require a larger post-processing effort to extract the relations of interest.

to A ($\mathcal{B} \neq \emptyset$). $[\![A]\!]$ can be an argument to $[\![C]\!]$, if all nodes on the path are semantically free and there is an embedding relation E such that $E = T(C,B_i)$, where $B_i \in \mathcal{B}$, and an argument identification rule R applies to the *(C,E)* pair:

$$E = T(C, B_i) \wedge B_i \in \mathcal{B} \wedge \forall B : (B \in \mathcal{B} \wedge [\![B]\!] = \emptyset) \wedge \exists R : applies\_to(R, C, E)$$

Consider the sentence in Example (106a), again taken from the GENIA corpus. The PROTEIN entities associated with the fragment are underlined, the embedding relations are given in (106b), and the result of argument propagation in (106c).

(106) (a) . . . no NF-κB **bound**$_C$ to the main NF-κB-binding site$_B$ 2 of the <u>IL-10</u>$_A$ promoter . . . (PMID_10089566)

(b) *PREP_TO(bound:*BINDING*,site)*
PREP_OF(site,promoter)
*NN(promoter,IL-10:*PROTEIN*)*

(c) *bind:*BINDING(e_1,t_1) $\wedge$ *IL-10:*PROTEIN(t_1)

When traversing the embedding graph, checking the daughter nodes of the node *bound* (corresponding to C in Definition 19 and associated with an atomic predicate of type BINDING) for arguments invokes an argument identification rule, which stipulates that *bind* can link to an argument of Object type via an embedding relation of *PREP_TO* type, which in this case is *site* (B), a semantically free node. At this point, argument propagation makes the nodes in scope of the daughter node accessible, which results in finding the node *IL-10* (A), corresponding to a PROTEIN term. Thus, *IL-10* is allowed as an Object argument of *bound*.

As a result of these compositional operations, a directed acyclic graph of fully composed predications is constructed. While the resulting predications are usable on their own, for practical tasks, they may need to be mapped to task specifications. We will illustrate some of these mappings in Chapter (6), where we discuss the application of the embedding framework to particular tasks.

5.4 Coreference Resolution

We posited that coreference resolution is a crucial aspect in moving beyond the sentence level in semantic interpretation. However, the focus of the current work has not been to develop state-of-the-art coreference resolution algorithms. We have only explored coreference resolution in a limited domain, namely, the biomedical research literature. The principles we outline in this section are largely domain-independent; however, we have only attempted to resolve certain issues within this domain and not in general. For example, the semantic categories of predications, and not only embedding, but also atomic predications, need to be known. However, in the general domain, we do not have this information. Furthermore, in coreference resolution research, the focus is mostly on coreference involving entities rather than relations. In other words, for a more general solution, we need to integrate general-domain entity recognition tools into the framework. In this section, our examples and discussion are based on the biomedical research literature and event extraction tasks due to these practical issues. Nonetheless, they suggest a promising approach to coreference resolution, if sufficient domain information is available.

The inability to resolve coreference has been demonstrated as a factor that hinders biological event extraction (Kim *et al.*, 2009). Coreference resolution is essentially a recall-increasing measure: in the following fragment, recognizing that *Eotaxin* is the antecedent of the pronominal anaphor *Its*, would allow our system to identify this term as the Theme participant of the GENE_EXPRESSION event triggered by the nominalization *expression*, which remains unidentified otherwise. This example also illustrates how coreference resolution serves in moving towards discourse interpretation, since it allows us to identify an event argument in a previous sentence.

(107) (a) <u>Eotaxin</u> is an eosinophil specific beta-chemokine assumed to be involved in eosinophilic inflammatory diseases such as atopic dermatitis, allergic rhinitis, asthma and parasitic infections. <u>Its</u> **expression** is stimulus- and cell-specific.

(b) *expression:*GENE_EXPRESSION$(e_1,t_1) \wedge$ *eotaxin:*PROTEIN(t_1)

In the description of the Protein Coreference Task (Nguyen *et al.*, 2011), four main classes of coreference are identified:

RELAT Coreference indicated by relative pronouns and adjectives (e.g., *that, which, whose*)

PRON (*pronominal anaphora*) Coreference indicated by personal and possessive pronouns (e.g., *it*, *its*, *they*, *their*)

DNP (*sortal anaphora*) Coreference indicated by definite and demonstrative noun phrases (NPs that begin with *the*, *these*, *this*, etc.)

APPOS Coreference in appositive constructions

Our embedding framework accommodates RELAT and APPOS classes naturally, since they are intra-sentential and they can largely be identified based on syntatic dependencies alone. The more complex anaphoric classes (PRON and DNP) are accommodated by a module, partially inspired by the deterministic coreference resolution system described in Haghighi and Klein (2009). To summarize, for each anaphoric mention identified in the text, their system selects an antecedent among the prior mentions by utilizing syntactic constraints and assessing the semantic compatibility between mentions. Of the remaining possible antecedents, the one with the shortest path from the anaphoric mention in the parse tree is selected as the best antecedent. The syntactic constraints used by their system include number, person, and entity type agreement as well as recognition of appositive constructions. On the other hand, their semantic compatibility filter aims to pair hypernyms, such as *AOL* and *company*. They extract such pairs from their corpus using bootstrapping. We provide more details about our treatment of the four coreference classes below.

5.4.1 RELAT and APPOS type

The RELAT type is the most frequent type of coreference annotated for the Protein Coreference Task (56% of all training instances), while the APPOS type is rarely annotated. The APPOS type is illustrated in Example (108) below, where the underlined expressions corefer.

(108) . . . upregulation of <u>the lung vascular adhesion molecule</u>, <u>intercellular adhesion molecule-1</u>, was greatly reduced by . . .

To determine the antecedent *ANT* of a relative pronoun *RP*, we use the following transformation rule, where *rel* denotes a *relative* dependency, and *rcmod* a *relative clause modifier* dependency. This rule simply states that the antecedent of a relative pronominal anaphora is the noun phrase head it modifies.

$$\textit{rel(X,RP)} \wedge \textit{rcmod(ANT,X)} \Rightarrow \textit{COREF(RP,ANT)}$$

On the other hand, coreference in appositive constructions is handled with the following rule, where $\textit{APPOS} \in \{\textit{appos}, \textit{abbrev}, \textit{prep\_including}, \textit{prep\_such\_as}\}$.

$$\textit{APPOS(ANT,ANA)} \vee \textit{APPOS(ANA,ANT)} \Rightarrow \textit{COREF(ANA,ANT)}$$

5.4.2 PRON and DNP type

PRON type of coreference is the second most frequent type annotated for the Protein Coreference Task (35% of all training instances), while the DNP type corresponds to 9% of the training instances. With respect to the PRON type, we only consider personal and possessive pronouns of the third person (*it/its*, *they/their*) as anaphors, since others do not seem relevant to the event extraction task (e.g., *Our results*). For sortal anaphora, the DNP type, we require that the anaphoric noun phrases are not associated with entities, allowing expressions such as *these factors* as anaphora while ruling out those like *the TRADD protein*.

Coreference resolution begins by identifying the set of candidate antecedents. We define the candidate antecedent set for a given anaphor as the set of embedding graph nodes which appear in the discourse prior to the anaphor and which are either semantically bound or involve hypernyms or conjunctions. The prior discourse includes the sentence that the anaphora occurs in as well as those preceding it in the paragraph.

The candidate antecedents are then evaluated for their syntactic and semantic compatibility. PRON requires person and number agreement, while DNP requires number agreement and one of the following constraints:

The head word constraint The head of the anaphoric NP and the antecedent NP are the same. This constraint allows "*CD4 gene*" as an antecedent for the anaphor "*the gene*".

The singular hypernymy constraint The head of the anaphoric NP is a hypernym of the antecedent, which involves an entity. This constraint accepts any Protein term as an antecedent for the anaphoric NP "*this protein*".

***The plural hypernymy constraint* (*set-instance anaphora*)** The head of the anaphoric NP is a plural hypernym of the antecedent, which corresponds to a conjunction of entities. This constraint accepts "*CD1, CD2, and CD3*" as antecedent for "*these factors*".

The meronymy constraint The head of the anaphoric NP is a meronym and the antecedent corresponds to a conjunction of entities. This constraint allows "*IBR/F*" as antecedent for the anaphoric NP "*the dimer*".

The predicate constraint The head of the anaphoric NP is associated with a predicate, P_1, and the antecedent with another predicate, P_2, where P_1 and P_2 belong to the same semantic category. This constraint aims to capture the coreference between, for instance, the anaphor *the stimulation* and the antecedent *stimulated*.

We induced the hypernym list from the training corpus automatically by considering the heads of the NPs with entities in modifier position. Such words include *gene*, *protein*, *factor*, and *cytokine*. Similarly, we induced the meronym list from the training data of the Static Relations supporting task (Pyysalo *et al.*, 2011a). These words essentially correspond to triggers for SUBUNIT-COMPLEX relations in that task, and include words such as *complex*, *dimer*, and *subunit*. These word lists remain one aspect of the coreference resolution module that is domain-dependent.

Several structural constraints over the semantic embedding graph block some of the possible antecedents for both coreference types:

- The antecedent directly embeds or is directly embedded by the anaphor.
- The antecedent is the subject and the anaphor is the object of the same relation. In addition, the anaphor is not reflexive (e.g., *itself*).
- The anaphor is in an adjunct position and the antecedent is in subject position of the same relation.

The candidate that is closest to the anaphor in the embedding graph is selected as the antecedent and a *COREF* embedding relation is created between the anaphor and the antecedent. For plural anaphora, multiple entities or triggers may be considered as antecedents, and thus multiple *COREF* relations may be created.

The integration of coreference information into the compositional process is trivial for all coreference types. In the composition phase, when an anaphoric expression appears in the argument position of a predication, it is naturally substituted by its antecedent(s) through argument propagation.

5.5 Conclusions

In this section, we outlined the main computational core of the embedding framework. Taking syntactic dependencies and a dictionary of embedding predicates, we illustrated how the semantic dependencies in a document and the scope relations are captured via transformation rules. The resulting embedding graph in combination with specific compositional operations, in particular, source propagation and scalar modality value composition, and argument identification rules allows us to compose predications in a bottom-up manner. Two other compositional operations, polarity composition and argument propagation, on the other hand, apply in more limited contexts. We also described our coreference resolution module, currently geared towards the biomedical literature. The dependence of the coreference resolution module is mainly due to lack of ontological knowledge and hypernymy and meronymy word lists, which have been extracted from a biomedical corpus.

Chapter 6

Evaluation

We presented the core embedding framework, from a theoretical as well as at a computational perspective in Chapters 3 and 5, respectively. The question we aim to address in this chapter is the evaluation of this framework. A full evaluation would require annotation of a gold standard embedding predication corpus according to our proposed categorization scheme and comparison of the system output to the gold annotations using standard information extraction evaluation metrics, such as precision, recall, accuracy, and F-measure. However, such annotation is clearly a time consuming, labor-intensive task that requires major resources. Large-scale annotation studies, even when they focus on more straightforward and limited phenomena, require tens of annotators. For example, in PropBank annotation (Palmer *et al.*, 2005), only verbal predicate-argument structures are annotated and it is the product of over thirty annotators and over three years of annotation. Moreover, the proposed model requires a finer-grained and more comprehensive annotation scheme than generally assumed in similar existing corpora. Given these challenges, we did not attempt to develop an annotated corpus. It seemed more sensible to focus on task-based evaluation of the framework and consider its easy adaptability to these tasks as a measure of its success. We participated in several shared task competitions that focus on various embedding phenomena and consider our results in these tasks as a proxy for overall evaluation of our framework.

In the following sections, we describe these relevant tasks and discuss our approach. Since some of the systems developed for these tasks do not reflect the full extent of the described framework, we note the differences between the implementation that was used for the given task and how the same task can now be performed within the full embedding framework, when necessary.

6.1 Uncertainty/Hedge Detection

The goal of the uncertainty or hedge detection task is to determine whether a given sentence describes a factual statement or is speculative or uncertain. Distinguishing facts from uncertain statements in biomedical research articles has been considered an important biomedical text mining task since it was proposed by Light *et al.* (2004). Early studies focused on using supervised machine learning techniques with variants of the "bag-of-words" approach (Light *et al.*, 2004) or weakly supervised techniques (Medlock and Briscoe, 2007; Szarvas, 2008). These last two studies were based on the Hedge Classification Dataset (Medlock and Briscoe, 2007), the earliest uncertainty corpus, in which sentences were classified as being speculative or not. Szarvas (2008), in addition to using this dataset, also extended his approach to the clinical domain (radiology reports). He reported relatively poor results on a corpus of biomedical articles from a different source, concluding that the portability of hedge classifiers (even in-domain) is limited.

In our early work, we approached this classification task from a more linguistically-oriented, less domain-dependent perspective. To summarize, we used an earlier version of the embedding dictionary in which 190 lexical hedging markers (largely belonging to HEDGE, DIMINISHER, INTENTIONAL, EPISTEMIC, and EVIDENTIAL embedding predicate types) were categorized and weighted by their hedging strength (1-5, 5 being the strongest). With hedging strength, we aimed to capture the central (strong) versus peripheral (weak) nature of hedging markers, described in Hyland (1998). Modal auxiliaries such as *may* were assumed to be strong hedge markers, while attribute hedges such as *approximately* were considered weak. The hedging strength of certain lexical items was increased further or reduced, based on the existence or absence of certain syntactic patterns. For example, the strength of an epistemic verbal marker (such as *believe*) was increased by 2, if it took a *that*-complement, by 1 if it took an infinitival complement, and was reduced by 1, otherwise. An overall hedging score was assigned to each sentence, calculated as the strength sum of all markers in the sentence. If the hedging score of a sentence exceeded a predefined threshold, the sentence was considered speculative. As an alternative to assigning hedging strengths semi-automatically in this manner (SA), we also experimented with inducing these weights automatically from the training set using information gain measure (IG) (Mitchell, 1997). The results of our experiments are shown in Table (6.1).

These results show a relatively stable performance across different datasets when the hedging weights are semi-automatically assigned (F_1-score of 0.85 vs. 0.82), in contrast with the finding in Szarvas (2008) that hedging markers are task-specific and not portable. The portability of the

Dataset	**Weighting method**	**Threshold**	**Precision**	**Recall**	**Accuracy**	**F_1-score**
HCD	SA	3	0.85	0.86	0.93	0.85
HCD	IG	1.5	0.81	0.79	0.90	0.80
BMC	SA	3	0.80	0.85	0.94	0.82
BMC	IG	1.75	0.75	0.67	0.90	0.71

Table 6.1: Results for hedge detection, HCD: Hedge Classification Dataset, BMC: BMC Bioinformatics Dataset, SA: Semi-automatic weighting, IG: Information Gain weighting

system with IG weighting is less pronounced; the F_1 score decreases from 0.80 vs. 0.71 when IG weighting is used, indicating lesser portability. The results on these datasets were state-of-the-art at the time of the experiments.

The BioScope corpus (Vincze *et al.*, 2008) has since become the standard annotated corpus for speculation and negation and their scope. One of the tasks in the CoNLL'10 Shared Task competition (Farkas *et al.*, 2010) focused on identifying sentences with uncertainty in this corpus. We essentially used the same system for this task (with SA weighting). Training on the BioScope corpus was limited to determining several classes of existing markers that were not considered to indicate uncertain sentences in this corpus. These classes were the HEDGE, DIMINISHER, and INTENTIONAL categories. With this minimal tuning, we obtained the official results in row 1 of Table (6.2). Our precision (92.07%) was highest among all participating systems. After the shared task, simply by lowering the hedging score threshold to 3, we obtained results that ranked higher overall. The threshold of 4 for the official submission was determined from the training data.

Threshold	**Precision**	**Recall**	**F_1-score**	**Rank**
4	92.07	74.94	82.62	12/24
3	83.43	84.81	84.12	8/24

Table 6.2: Detecting sentences with uncertainty in the BioScope corpus.

One of the important things to note about these results is that, by setting the threshold to 3, we essentially get the same performance on the BioScope corpus as on the Hedge Classification and BMC datasets used in earlier experiments (F_1 score of approximately 0.85), demonstrating the generalizability of the approach. It should also be noted that the systems with higher performance all used the training dataset to its full extent, while our approach utilized it minimally, indicating the robust baseline nature of our system in hedge detection[1]. Tang *et al.* (2010) achieved the highest

[1] A recent study in academic discourse pragmatics (Gross and Chesley, 2012) in fact used our system to compare the language of medical research articles sponsored by industry and those that are not, finding an interesting correlation

F_1-score of approximately 86.36% in this task using a cascade of conditional random field and large margin-based models, essentially computationally intensive machine learning techniques.

The hedge detection system briefly described above is a precursor to the current embedding framework. All the hedging markers are included in the embedding predicate dictionary and the syntactic patterns that play a role in increasing or decreasing the contribution of hedging markers are captured in the semantic embedding graphs corresponding to individual sentences. One thing that is currently not part of the core embedding framework is the notion of hedging strength associated with individual predicates and used in calculating hedging score. How can we calculate hedging scores for individual sentences in the absence of hedging weights for relevant predicates? One straightforward way to address this issue would be to extend the embedding dictionary and make *hedging strength* a *sense* feature. However, this may not be necessary, since the hedging marker categories largely correlate with hedging strengths. For example, the weak hedging markers (strength values of 1-2) often belong to HEDGE or DIMINISHER categories, while the strong markers (values 3-5) belong to EPISTEMIC and EVIDENTIAL categories. Therefore, it may be sufficient to assume that if a sentence includes one or more predicates from the EPISTEMIC or EVIDENTIAL categories, it is *speculative*. Similarly, if a sentence includes at least two predicates from the VALENCE_SHIFTER categories, it can be also considered *speculative*. This assumption seems reasonable, considering that thresholds 3 and 4 yielded the best performance in hedge detection tasks.

6.1.1 Vagueness Detection in Wikipedia Articles

Another task in the CoNLL Shared Task challenge in 2010 (Farkas *et al.*, 2010) was concerned with detecting uncertainty in Wikipedia articles. Uncertainty in this context refers more or less to *vagueness* indicated by *weasel words*, an undesirable feature according to Wikipedia policy. Weasel words in Wikipedia articles are tagged by contributors, making Wikipedia a readily annotated corpus. Ganter (2010) proposed n-gram and POS-based features for detecting weasel words. The best performing system in the CoNLL Shared Task challenge obtained an F_1-score of 60.17%, using SVM classifiers and simple lexical features (Georgescul, 2010).

Analysis of Wikipedia training data provided by the organizers revealed that there is overlap between *weasel words* and our hedging markers described earlier. Therefore, in the shared task, we adapted our hedging dictionary to the task of detecting vagueness in Wikipedia articles. Similar to the hedge detection task in the BioScope corpus (described in the previous section), changes involved eliminating several categories as markers (including INTENTIONAL). In addition, we also added a

between the level of certainty in the articles and industry sponsorship.

previously unconsidered category of markers, due to the prominence of its members in Wikipedia data as *weasel words*. This category of *vagueness quantifiers* (Lappin, 2000) includes words, such as *some*, *several*, *many* and *various*, which introduce imprecision when in modifier position. For instance, in the example below, both *some* and *certain* contribute to vagueness of the sentence.

(109) Even today, *some cultures* have *certain instances* of their music intending to imitate natural sounds.

Except for these small changes in the dictionary of markers, the methodology for uncertainty detection in Wikipedia articles was essentially the same as that for biological text in the BioScope corpus. For the official submission, the threshold was set to 4, determined from the training data. We obtained the results in row 1 of Table (6.3) officially.

Threshold	**Precision**	**Recall**	**F_1-score**	**Rank**
4	67.90	46.02	54.86	10/17
3	63.21	53.67	58.05	3/17

Table 6.3: Results for vagueness detection in Wikipedia articles

The results show that we were able to exploit the overlap between our hedging markers and the weasel words. The major difference we noted between hedging in these two genres was the class of *vagueness quantifiers*, and, with little effort, we extended our hedging dictionary to consider them. We also note that setting the threshold to 3 after the shared task, our recall and F_1-score improved significantly (row 2). The fact that we were able to achieve competitive results by using the training data minimally but carefully further indicates the domain-independent nature of our hedging markers and ease of extending our approach.

Vagueness detection within the embedding framework would not be much different than hedge detection within the framework, described earlier. The current embedding dictionary does not include vagueness quantifiers; they would have to be added, and some new syntactic dependency transformations may be needed.

6.2 Speculation Scope Resolution

A more challenging task than determining whether a sentence is speculative (or hedged) is identifying the scope of the speculation precisely. This task has been proposed with two different *scope* definitions. In the BioScope definition, the scope of the speculation is modeled as a textual span that includes the speculation marker and the biggest syntactic unit that involves the marker. In contrast, in the GENIA and the subsequent event-based schemes (Kim *et al.*, 2008; Thompson *et al.*, 2011), the scope is an event, essentially an abstract semantic object. We will discuss the latter in the next section in the context of biological event extraction. In this section, we focus on speculation scope resolution, where the scope corresponds to a textual span. An example is given below, where the speculation cue is underlined, and the textual span corresponding to the scope is in square brackets. This particular task was one of the subtasks in the CoNLL'10 Shared Task on Hedge Detection (Farkas *et al.*, 2010).

(110) This result [<u>suggests</u> that the valency of Bi in the material is smaller than +3].

Taking the speculation markers as input, our approach to this task involved a two-pronged methodology. The first method (constituency-based heuristics) exploited a constituent parse of the sentence. We simply identified the phrasal node that dominates the speculation marker and considered the tokens within that phrase as being within the scope of the marker, unless they meet one of the following exclusion criteria:

1. Exclude tokens within post-marker sentential complements (indicated by S and SBAR nodes) introduced by a small number of discourse markers (*thus*, *whereas*, *because*, *since*, *if*, and *despite*).

2. Exclude punctuation marks at the right boundary of the phrase

3. Exclude pre-marker determiners and adverbs at the left boundary of the phrase

For example, in the sentence below, the verbal phrase that included the modal auxiliary *may* (the speculation marker) also included the subordinating clause introduced by *thereby*. Using the exclusion criteria 1 and 2, we excluded the tokens following *SPACER* from the scope:

(111) ... motifs [<u>may</u> be easily compared with the results from BEAM, PRISM and SPACER], thereby extending the SCOPE ensemble to include a fourth class of motifs.

The second method (dependency-based heuristics) involved identifying the head of the scope using syntactic dependencies and then considering all the syntactic dependencies that the scope head was involved in. To determine the scope head, the syntactic dependency types found to be relevant in earlier work were used (Kilicoglu and Bergler, 2008, 2011b). For example, the scope head when the marker is a modal auxiliary is the token that is the governor of the syntactic dependency of type *aux* where the dependent is the marker (*aux(Head,Marker)*). The set of dependencies was then used in *right expansion* and *left expansion* of the scope. *Right expansion* involves finding the rightmost token that is in a dependency relation with the *scope head*. Consider the sentence below:

(112) The surprisingly low correlations between Sig and accuracy may [indicate that the objective functions employed by motif finding programs are only a first **approximation** to biological **significance**].

The epistemic verb *indicate* has as its *scope head* the token *approximation*, due to the existence of a clausal complement dependency (*ccomp*) between them. On the other hand, the rightmost token of the sentence, *significance*, has a prepositional modifier dependency (*prep_to*) with *approximation*. It is, therefore, included in the scope of *indicate*. Two dependency types, adverbial clause modifier (*advcl*) and conjunct (*conj*), were excluded from consideration when the rightmost token is sought, since they are likely to signal new discourse units outside the scope.

In contrast to *right expansion*, which applies to all hedging marker categories, *left expansion* applies only to a subset. Left expansion involves searching for a subject dependency governed by the *scope head*. The dependency types descending from the subject (*subj*) type in the Stanford dependency hierarchy are considered: *nsubj* (nominal subject), *nsubjpass* (passive nominal subject), *csubj* (clausal subject) and *csubjpass* (passive clausal subject). In the following example, the first token, *This*, is added to the scope of *likely* through left expansion (*cop*: copula).

(113) (a) [**This** is most likely a conservative estimate] since a certain proportion of interactions remain unknown ...

(b) *nsubj(likely,This)*
cop(likely,is)

Left expansion was limited to the following marker categories, with the additional constraints given:

1. *Modal auxiliaries*, only when their *scope head* takes a passive subject (e.g., *they* is added to the scope of *may* in *they may be annotated as pseudogenes*).

2. Adjectival markers, when they are in copular constructions (e.g., Example (113)).

3. Several markers in adjectival and verbal categories, when they take infinitival complements (e.g., *this* is added to the scope of *appears* in "However, this appears to add more noise to the prediction without increasing the accuracy").

After scope tokens are identified using the parse tree as well as via left and right expansion, the algorithm simply sets as scope the continuous textual unit that includes all the scope tokens and the hedging cue. Since, *likely* is the hedging cue and *This* and *estimate* are identified as scope tokens in Example (113), the scope associated with *likely* becomes "This is most likely a conservative estimate". We found that citations, numbers and punctuation marks occurring at the end of sentences caused problems in scope resolution, specifically in biomedical full text articles. Since they are rarely within any scope, we implemented a simple stripping algorithm to eliminate them from scopes in such documents. With this methodology, we obtained the results in Table (6.4). Our precision was highest in the competition. Morante *et al.* (2010) used a memory-based learning system that relies on syntactic dependencies to obtain the best results in this task (F_1-measure of 57.32%).

Precision	**Recall**	**F_1-score**	**Rank**
62.47	49.47	55.21	4/15

Table 6.4: Hedge scope resolution results in CoNLL Shared Task 2010.

We also measured the relative contribution of the enhancements to scope resolution. The results are presented in Table (6.5). The baseline was taken as the textual span that covers the hedging marker as well as the scope head (somewhat analogous to GENIA scope definition).

	Precision	**Recall**	**F_1-score**
Baseline	3.29	2.61	2.91
Baseline+ Left/ right expansion	25.18	20.03	22.31
Parse tree	49.20	39.10	43.58
Baseline+ Parse tree	50.66	40.27	44.87
All	62.47	49.47	55.21

Table 6.5: Effect of scope resolution enhancements

These results show that: (a) the scope definition of GENIA is essentially incompatible with the BioScope definition, (b) constituency-based heuristics (simply taking the phrase that the hedging cue belongs to as the scope) provides relatively good results, and (c) left and right expansion heuristics are needed for increased precision and recall.

How can the BioScope notion of scope, defined as a textual span, be adapted to the embedding framework? Let us consider that the hedging marker corresponds to a node N in the semantic embedding graph of a sentence. It seems reasonable to hypothesize that the marker scope is the continuous textual unit that corresponds to the descendants of N, since the graph essentially captures scope relations and dependency-based heuristics. While examining the examples given earlier in this section seems to largely confirm this, it needs to be evaluated further. Concerning right expansion, in Example (112), the surface element *significance* is a descendant of *indicate* and, therefore, would be included in its scope. With regard to left expansion, constraints concerning adjectival and verbal markers would be satisfied (constraints b and c above), due to embedding transformations. The constraint concerning the modal auxiliaries is not satisfied completely; however, the effectiveness of this particular constraint would need to be evaluated further. More questionable is whether the constituency-based heuristics would be captured within the embedding framework naturally. It seems that it can to a large extent. In fact, in the relevant example, duplicated below, the eventual syntactic embeddings (shown in (114b)) would predict that the fragment beginning with *thereby* would be excluded from the scope of *may*.

(114) (a) . . . motifs [may be easily compared with the results from BEAM, PRISM and SPACER], thereby extending the SCOPE ensemble to include a fourth class of motifs.

(b) *thereby* $>_d$ *may*

thereby $>_d$ *extending*

Punctuation and pre-marker adverb constraints are also satisfied. The pre-marker determiner constraint, on the other hand, may be violated; however, again, its effectiveness should be further evaluated.

6.3 Biological Event Extraction

The major focus of information extraction in the biomedical domain has been on extracting relevant relations and events from scientific literature. The type of information extracted includes protein-protein interactions (see Kabiljo *et al.* (2009) for a relatively recent assessment), gene-disease relations (Rindflesch *et al.*, 2003), regulatory events (Tsujii, 2009), and pharmacogenomic relations (Ahlers *et al.*, 2007). The GENIA event corpus (Kim *et al.*, 2008), in which biological events concerning transcription factors are annotated, formed the basis for much recent work in this area. Two shared task competitions based on this corpus and its extensions were organized.

The first competition, BioNLP'09 Shared Task on Event Extraction (Tsujii, 2009), consisted of three subtasks. Given protein entity annotations as input, the mandatory *core event extraction* task involved detection of event triggers (predicates), their semantic categories, and primary arguments (Theme and Cause). The optional *event enrichment* task involved recognition of entities other than proteins and assigning these as secondary event arguments (e.g., Site, toLoc). The optional *speculation and negation detection* task involved identifying events that are speculated or negated (called *event modifications* in shared task parlance). The core event types addressed can be grouped into three categories.

1. Simple event types involving a single Theme argument of protein type (Gene_expression, Transcription, Protein_catabolism, Phosphorylation, and Localization)

2. Binding event type is in its own class and may involve one or more Theme arguments of protein type.

3. Complex regulatory event types (Regulation, Positive_regulation, and Negative_ regulation) may involve a Cause argument, in addition to a Theme. They can also take other events as arguments in either role, making the task of identifying the arguments of such events more challenging.

Speculation and Negation are event modifications. They take an event argument, modifying its meaning[2]. Accurately identifying event modifications is even more challenging, because this depends on correctly finding the base events in their scope.

Using the terminology we used throughout this document, complex regulatory events largely correspond to embedding predications. Speculation and Negation annotations, on the other hand, always correspond to embedding predications. Meanwhile, simple event types and Binding events are atomic predications. An example is given below, where a sentence with its gold standard event annotations are provided. The protein annotations with their character offsets in (115b) are provided as input, while the event trigger (115c) and event and event modification (115d) annotations were expected as output.

(115) (a) Together these data suggest that ETS1 may be involved in mediating the increased GM-CSF production. (PMID 7478534)

(b) T16 Protein 1478 1482 ETS1
T17 Protein 1526 1532 GM-CSF

(c) T34 Regulation 1490 1498 involved
T35 Positive_regulation 1502 1511 mediating
T36 Positive_regulation 1516 1525 increased
T37 Gene_expression 1533 1543 production

[2]This is how the GENIA scope definition differs from the BioScope scope definition.

(d) E17 Regulation:T34 Theme:E18 Cause:T16

M3 Speculation E17

E18 Positive_regulation:T35 Theme:E19

M4 Speculation E18

E19 Positive_regulation:T36 Theme:E20

E20 Gene_expression:T37 Theme:T17

The theme of the second competition, BioNLP-ST'11 (Kim *et al.*, 2011a) was generalization. In this spirit, there were four event extraction tracks: in addition to the GENIA track that again focused on transcription factors (Kim *et al.*, 2011b), the epigenetics and post-translational modification track (EPI) focused on events relating to epigenetic change, such as DNA methylation and histone modification, as well as other common post-translational protein modifications (Ohta *et al.*, 2011), whereas the infectious diseases track (ID) focused on bio-molecular mechanisms of infectious diseases (Pyysalo *et al.*, 2011b). Both GENIA and ID tracks included data from full-text scientific articles in addition to abstracts, which were the focus of the earlier competition. Detection of event modifications (SPECULATION and NEGATION) is an optional task in all three tracks. The fourth track, Bacteria, consisted of two sub-tracks: Biotopes (BB) (Bossy *et al.*, 2011) and Interactions (BI) (Jourde *et al.*, 2011).

We participated in both competitions. In the first competition, we participated in the mandatory *core event extraction* task and the optional *speculation and negation detection* task. We did not participate in the *event enrichment* task, since this was essentially a named entity recognition (NER) task, which has not been the focus of our research. In the second competition, we participated in three tracks, GENIA, ID, and EPI. While our focus has always been on embedding predications, these competitions (particularly the second) also allowed us to test how our semantic interpretation approach extends to atomic predications, as well.

In the first competition, our approach was more shared task oriented and simpler, even though the system contains the seeds of the embedding framework, such as the use of a trigger dictionary and argument identification rules based on syntactic dependency relations. On the other hand, in the second competition, our system was more or less based on the current version of the embedding framework, even though not all components of the framework were found to be relevant to shared task specific concerns. As such, here, regarding the first competition, we will only present our official results (Table 6.6) and note that our system was the top rule-based system[3]. Below, we focus on

[3]For more information on the first shared task system, we refer the reader to Kilicoglu and Bergler (2011b).

Event Type	**Recall**	**Precision**	**F_1-score**	**Rank**
Localization	35.63	92.54	51.45	5/24
Binding	20.46	40.57	27.20	11/24
Gene_expression	55.68	79.45	65.47	5/24
Transcription	15.33	60.00	24.42	13/24
Protein_catabolism	64.29	56.25	60.00	5/24
Phosphorylation	69.63	95.92	80.69	2/24
EVT-TOTAL	43.10	73.47	54.33	5/24
Regulation	24.05	45.75	31.53	1/24
Positive_regulation	28.79	50.45	36.66	2/24
Negative_regulation	26.65	51.53	35.13	3/24
REG-TOTAL	27.47	49.89	35.43	2/24
Negation	14.98	50.75	23.13	1/6
Speculation	16.83	50.72	25.27	1/6
MOD-TOTAL	15.86	50.74	24.17	1/6
ALL-TOTAL	32.68	60.83	42.52	3/24

Table 6.6: Official BioNLP'09 Shared Task on Event Extraction evaluation results (EVT-TOTAL: Atomic predications), REG-TOTAL: Complex regulatory events (mostly embedding predications), MOD-TOTAL: Speculation and negation (all embedding predications), ALL-TOTAL: All annotations)

how embedding framework was used in the second competition.

Our approach in the second competition incorporated the embedding framework. We viewed event extraction as a two-phase procedure. The first phase essentially corresponds to the semantic predication composition discussed in detail in Chapter 5. There were three shared task-specific aspects of this phase:

- Entities of PROTEIN type were provided as input, and therefore, construction of entity objects based on structural constraints (as described earlier in Section 5.2) was not performed.
- A dictionary of atomic predicates was constructed from the training data using maximum likelihood estimation (for details, see the earlier shared task article (Kilicoglu and Bergler, 2011b)). Therefore, each atomic predicate was associated with its most likely sense, alleviating the need for disambiguation.

- If an embedding predicate was ambiguous between an EPISTEMIC sense and another sense, the EPISTEMIC sense was preferred, due to the nature of scientific writing.

After the shared task competition, we also incorporated the coreference resolution module described earlier into the compositional phase, with the aim of moving to the inter-sentential level in semantic interpretation. Recall the earlier sentence, whose gold annotations were provided and which is duplicated below. The result of the first phase for this sentence is given in (116b).

(116) (a) Together these data suggest that ETS1 may be involved in mediating the increased GM-CSF production. (PMID 7478534)

(b) *ETS1*:PROTEIN$(t_{16}) \wedge$ *GM-CSF*:PROTEIN$(t_{17}) \wedge$
these_data$(DET_1) \wedge$ *production*:GENE_EXPRESSION$(e_1, DET_1, 1.0_{epistemic}, t_{17}) \wedge$
increase:CAUSE$(em_2, DET_1, 1.0_{epistemic}, e_1) \wedge$
mediate:ENABLE$(em_3, DET_1, 1.0_{epistemic}, em_2) \wedge$
involve:CAUSAL$(em_4, DET_1, 0.7_{epistemic}, t_{16}, em_3) \wedge$
may:SPECULATIVE$(em_5, DET_1, 0.8_{epistemic}, em_4) \wedge$
suggest:DEDUCTIVE$(em_6, WR, 1.0_{epistemic}, DET_1, em_5)$

Note that the atomic predication (e_1) is given a shared task semantic category (GENE_EXPRESSION), based on the atomic predicate dictionary.

The second *mapping* phase was concerned with imposing shared task definitions and constraints on the partial semantic interpretation obtained in the previous phase. This was achieved in three steps. The first step is to convert embedding predication types to event (or event modification) types. This step is guided by constraints on embedding predication type and scalar modality information (the relevant scale and the value), as presented in Table (6.7). In this way, the semantic categories of em_2 and em_3 above are substituted for POSITIVE_REGULATION (row 2) and that of em_4 for REGULATION (row 1). Furthermore, em_5 and em_6 are converted into SPECULATION instances (row 6).

Next, we convert logical arguments to semantic roles. A small number of mappings, illustrated in Table (6.8), are defined for this purpose. These are similar to *argument identification rules*, in that the mapping can be constrained to certain event types or event types can be excluded from it. For example, the first two mappings (row 1-2) allow the Object and Subject arguments of em_4 in Example (116b) to be converted to THEME and CAUSE semantic roles, respectively.

Finally, we prune event participants that do not conform to the event definition as well as the predications whose types could not be mapped to a shared task event type. Thus, a CAUSE

Track	Pred. Type	Scale (Sc)	MV_{Sc}	Event (Mod.) Type
GENIA,ID	CAUSAL	*	*	REGULATION
GENIA,ID	CAUSE, ENABLE	*	*	POSITIVE_REGULATION
GENIA,ID	PREVENT	*	*	NEGATIVE_REGULATION
GENIA,ID,EPI	*	SUCCESS	$= 0.0$	NEGATION
EPI	CAUSE, ENABLE	*	*	CATALYSIS
GENIA,ID,EPI	*	EPISTEMIC	$> 0.0 \wedge < 1.0$	SPECULATION
GENIA,ID,EPI	*	EPISTEMIC	$= 0.0$	NEGATION
GENIA,ID,EPI	*	INTERROGATIVE	*	SPECULATION
GENIA,ID,EPI	*	INTENTIONAL	*	SPECULATION

Table 6.7: Constraints used in mapping from embedding predication types to event and event modification types in BioNLP-ST'11.

Logical Argument	Semantic Role	Constrained To	Exclusions
Object	THEME	-	PROCESS
Subject	CAUSE	-	BINDING
Subject	THEME	BINDING	-
Object	PARTICIPANT	PROCESS	-
Object	SCOPE	SPECULATION, NEGATION	-

Table 6.8: Logical argument to semantic role mappings in BioNLP-ST'11.

participant for a GENE_EXPRESSION event is pruned, since only THEME participants are annotated as relevant for the shared task; likewise, a predication of DEONTIC type is pruned, because such predications are not considered for the shared task. In (116b), the embedding predication em_6 is pruned, because its argument corresponds to an event modification (em_5), rather than an event as expected for event modification annotations. This concludes the progressive transformation of the semantic predications to event and event modification annotations. All the gold event and gold event modification annotations shown in (115d) were extracted as the result of the mapping phase.

6.3.1 BioNLP-ST'11: Results and Discussion

With the two-phase methodology presented above, we obtained the official results, shown in Tables (6.9) and (6.10), for GENIA and EPI, ID tracks, respectively. Overall, we were ranked 5th in the GENIA track (5/15), 7th in the EPI track (7/7) and 4th in the ID track (4/7). There were only two submissions for the GENIA speculation/negation task and our results in this task were

Event Class	**Recall**	**Precision**	**F_1-score**	**Rank**
Localization	39.27	90.36	54.74	7/15
Binding	29.33	49.66	36.88	7/15
Gene_expression	65.87	86.84	74.91	5/15
Transcription	32.18	58.95	41.64	9/15
Protein_catabolism	66.67	71.43	68.97	2/15
Phosphorylation	75.14	94.56	83.73	4/15
EVT-TOTAL	52.67	78.04	62.90	6/15
Regulation	33.77	42.48	37.63	3/15
Positive_regulation	35.97	47.66	41.00	7/15
Negative_regulation	36.43	43.88	39.81	5/15
REG-TOTAL	35.72	45.85	40.16	5/15
Negation	18.77	44.26	26.36	2/2
Speculation	21.10	38.46	27.25	1/2
MOD-TOTAL	19.97	40.89	26.83	2/2
ALL-TOTAL	43.55	59.58	50.32	5/15

Table 6.9: Official GENIA track results in BioNLP-ST'11

comparable to those of the other participating group (Björne and Salakoski, 2011); our system performed slightly better with speculation, and theirs with negation. We note that their system was ranked higher than ours in core event extraction (third vs. fifth), which suggests that our system performance on speculation/negation task *alone* is probably a bit better than theirs. In the GENIA and ID tracks, our system was the top rule-based system. Our poor performance in the EPI track was largely due to the fact that non-core entity types (Site, toLoc, etc.) were included as core event participants and were expected to be extracted automatically, and our core system did not attempt to extract them.

Development Set vs. Test Set

A particularly encouraging outcome for our system was that our results on the GENIA development set versus on the test set were very close (an F_1-score of 51.03 vs. 50.32), indicating that our general approach avoided overfitting, while capturing the linguistic generalizations, as we intended. We

Track-Eval. Type	Recall	Precision	F_1-score	Rank
EPI-FULL	20.83	42.14	27.88	7/7
EPI-CORE	40.28	76.71	52.83	6/7
ID-FULL	49.00	40.27	44.21	4/7
ID-CORE	50.91	43.37	46.84	4/7
ID-FULL-T	45.26	53.18	48.90	4/7
ID-CORE-T	46.75	56.94	51.34	4/7

Table 6.10: Official evaluation results for EPI and ID tracks in BioNLP-ST'11. The primary evaluation criteria underlined. ID-FULL-T and ID-CORE-T refer to the post-shared task scenario where ID triggers are drawn only from ID training data.

observe similar trends with the other tracks, as well. In the EPI track, development/test F_1-score results were 29.1 vs. 27.88; while, in the ID track, interestingly, our test set performance was better (39.64 vs. 44.21). We also obtained the highest recall in the ID track (49), despite the fact that our system typically favors precision. We attribute this somewhat idiosyncratic performance in the ID track partly to the fact that we did not use a track-specific trigger dictionary for the official submission. Instead, we constructed a trigger dictionary based on all training datasets at once. All but one of the ID track event types are the same as those of the GENIA track, which led to identification of some ID events with triggers consistently annotated only in the GENIA corpus and to low precision particularly in complex regulatory events. A post-shared task re-evaluation confirms this: the F_1-score for the ID track increases from 44.21 to 48.9 when only triggers extracted from the ID track corpus are used; recall decreases from 49 to 45.26, while the precision increases from 40.27 to 53.18. It is unclear to us why a reliable trigger in one corpus is not reliably annotated in another, even though the same event types are considered in both corpora. One possibility is that different annotators annotating different corpora may have a different conceptualization of the same event types. Consider the following sentences: Example (117a) is from the GENIA corpus and Example (117b) from the ID corpus. Even though the verbal predicate *lead* appears in similar contexts in both sentences, it is annotated as an event trigger only in Example (117a)[4]

(117) (a) Costimulation of T cells through both the Ag receptor and CD28 <u>leads</u> to <u>high level</u> IL-2 **production** ...

[4]It is also worth noting that the predicate in question here, *lead*, indicates a high level causal relation, which corresponds to a RELATIONAL predication, rather than an observable event. That is, it functions at the discourse level. This perhaps causes confusion among annotators as to whether it indicates a biological event.

*lead:*POSITIVE_REGULATION(em_1,em_2)

*high_level:*POSITIVE_REGULATION(em_2,e_3)

*production:*GENE_EXPRESSION(e_3,t_1) $\wedge$ *IL-2:*PROTEIN(t_1)

(b) ...the two-component regulatory system PhoR-PhoB leads to increased hilE P2 **expression** ...

*increased:*POSITIVE_REGULATION(em_1,e_2,t_1) $\wedge$ *PhoR-PhoB:*PROTEIN(t_1)

*expression:*GENE_EXPRESSION(e_2,t_2) $\wedge$ *hilE:*PROTEIN(t_2)

We refer to the results concerning the post-shared task re-evaluation as ID-T in Tables (6.10) and (6.13).

Full-text Articles vs. Abstracts

One of the interesting aspects of BioNLP-ST'11 was its inclusion of full-text articles in training and evaluation. Cohen *et al.* (2010a) show that the structure and content of biomedical abstracts and article bodies differ markedly and suggest that some of these differences may pose problems in processing full-text articles. Since one of our goals was to determine the generality of our system across text types, we did not perform any full text-specific optimization. Our results on article bodies are notable: our system had stable performance across text types (in fact, we had a very slight F_1-score improvement on full-text articles: 50.28 to 50.4). This contrasts with the drop of a few points that seems to occur with other high performance systems. Taking only full-text articles into consideration, we would be ranked 4th in the GENIA track. Furthermore, a preliminary error analysis with full-text articles indicates that parsing-related errors are more prevalent in the full-text article set than in the abstract set, consistent with the findings of Cohen *et al.* (2010a). At the same time, our results confirm that we were able to abstract away from such errors by a careful, selective use of syntactic dependencies and correcting them with heuristic transformation rules, when necessary.

Speculation and Negation

Among shared task concerns, speculation and negation are perhaps most closely associated with our embedding focus. Therefore, we examined our results on the GENIA development set with respect to speculation and negation detection more closely. We determined that the majority of errors were due to misidentified or missed base events (70% of the precision errors and 83% of the recall errors). An even bigger percentage of speculation/negation-related errors in the EPI and

ID tracks were due to the same problem, as the overall accuracy in those tracks is lower. When we use the gold standard GENIA event annotations as input to the system and, thus, eliminate core event extraction-related errors and evaluate speculation/negation detection *alone*, we obtain the results shown in Table (6.11). These results constitute a more accurate characterization of the system in speculation/negation detection than the official results, which do not account for core event extraction related errors.

Event Modification Type	**Recall**	**Precision**	**F_1-score**
NEGATION	49.31 (18.77)	87.70 (44.26)	63.13 (26.36)
SPECULATION	65.70 (21.10)	73.27 (38.46)	69.28 (27.25)
MOD-TOTAL	57.95 (19.97)	78.47 (40.89)	66.67 (26.83)

Table 6.11: Speculation and negation detection based on gold event annotations. Official results are duplicated in parentheses for reference.

A closer look at precision errors in speculation and negation detection reveals cases in which speculation or negation is debatable, as the examples below show. In Example (118a), our system detected a SPECULATION instance, due to the verbal predicate *suggesting*, which scopes over the event indicated by *role*. In Example (118b), our system detected a NEGATION instance, due to the verbal predicate *lack*, which scopes over the events indicated by *expression*. Neither were annotated as such in the shared task corpus. Annotating negation and speculation is clearly nontrivial, as there seems to be some subjectivity involved, and such errors seem acceptable to a certain extent.

(118) (a) . . . <u>suggesting</u> a **role** of these 3' elements in beta-globin gene expression.

(b) . . . DT40 B cell lines that <u>lack</u> **expression** of either PKD1 or PKD3 . . .

Some of the recall errors were due to lack of an appropriate argument identification rule, as it is currently implemented. One recall problem involved copular constructions, which we had not sufficiently addressed within our framework. Therefore, we miss the relatively straightforward SPECULATION instance in the following example, indicated by the verb *appear* affecting the event indicated by *active*.

(119) . . . the A3G promoter <u>appears</u> constitutively **active**.

Similarly, the lack of an embedding predicate with the appropriate sense in the embedding dictionary causes recall errors. The embedding predicate *characterize* below has two senses in the embedding dictionary: MANNER and CORRELATIVE, neither of them relevant for speculation. It seems that an additional sense with the semantic category INTERROGATIVE might be appropriate[5]

(120) To further characterize **altered** expression of TCRζ, p56(lck) ...

Our system also missed an interesting, domain-specific type of negation, in which the minus sign acts similar to a negative determiner (e.g., *no*) and indicates negation of the event that the entity participates in.

(121) ... CD14- surface Ag **expression** ...

Coreference Resolution

Post-shared task, we integrated the coreference resolution module into the BioNLP-ST'11 event extraction pipeline. While the earlier Protein Coreference Supporting Task (Nguyen *et al.*, 2011) focused on coreference resolution as an isolated task, we find it more important to evaluate coreference resolution within the context of semantic interpretation, rather than in isolation. We measured the effect of each type of coreference resolution (RELAT, APPOS, PRON and DNP), discussed earlier in Section 5.4, on event extraction over the GENIA development set. The results, presented in Table (6.12), show that improvement in event extraction performance due to our current coreference resolution algorithm is modest. We observe that there is a consistent recall increase, while the precision suffers slightly in all cases. Resolving all four classes of coreference simultaneously seems to have a synergistic effect on the performance. On the test sets of the three tracks we participated in, we see minor improvements due to coreference resolution in GENIA and EPI tracks, but not in the ID track (Table (6.13)).

System	**Recall**	**Precision**	**F_1-score**
Base	46.32	56.81	51.03
Base + RELAT	46.57	56.52	51.06
Base + APPOS	47.07	56.40	51.32
Base + PRON	46.76	56.28	51.08
Base + DNP	46.85	56.26	51.13
Base + ALL	47.98	55.77	51.62

Table 6.12: Effect of different types of coreference resolution on event extraction performance on GENIA development set.

It is interesting to note that while APPOS type coreference was rarely annotated in the Protein Coreference Task corpus, resolving it had the biggest effect on event extraction. This is in contrast

[5]It may be also argued that infiniteness of the clause contributes to its speculativeness.

System	Recall	Precision	F_1-score
GENIA	43.55	59.58	50.32
GENIA + COREF	44.45	58.92	50.67
- Abstracts	44.31	59.82	50.91
- Full-text	44.78	56.82	50.09
EPI	20.83	42.14	27.88
EPI + COREF	21.48	40.63	28.10
ID	49.00	40.27	44.21
ID + COREF	49.97	38.81	43.69
ID-T	45.26	53.18	48.90
ID-T + COREF	46.37	50.95	48.55

Table 6.13: Event extraction performances on test sets after coreference resolution.

to the RELAT type, which had the highest percentage of instances in the corpus but had little effect on event extraction. We were particularly interested in the results involving PRON and DNP types, since the participants of events resulting from resolving these types can potentially span multiple sentences, playing a role in our higher level goal of discourse interpretation. We manually analyzed the events extracted through resolution of PRON and DNP types of coreference. We found that 32.5% of such events were correct, however the positive effect was largely limited to intra-sentential coreference resolution (43.2% vs. 16%). Among the events correctly identified due to intra-sentential coreference resolution, 56% involved coreference of PRON type. On the other hand, among those due to inter-sentential coreference resolution, 84% involved the DNP type. In the following example, the possessive adjective *their* (PRON type) refers to the proteins *GATA3* and *FOXP3* and we extract the relevant events shown in (122b).

(122) (a) Thus, although GATA3 and FOXP3 showed similar kinetics, their **expression** polarizes at the end ...

(b) *expression:*GENE_EXPRESSION(e_1, t_1) ∧ *GATA3:*PROTEIN(t_1)
*expression:*GENE_EXPRESSION(e_2, t_2) ∧ *FOXP3:*PROTEIN(t_2)

In Example (123), we correctly identify the event in (123b) from the sentence in (123a) by resolving the inter-sentential coreference between *this restriction factor* and *APOBEC3G*:

(123) (a) APOBEC3G (A3G), a member of the recently discovered family of human cytidine deaminases, is expressed in peripheral blood lymphocytes and has been shown to be active

against HIV-1 and other retroviruses. To gain new insights into the **transcriptional regulation** of <u>this restriction factor</u>, ...

(b) *transcriptional_regulation:*REGULATION(e_1,t_1) ∧ *APOBEC3G:*PROTEIN(t_1)

Among the misidentified events, we observe that some are due to shortcomings of the event extraction algorithm, rather than coreference resolution. In the following example, the coreference between the expression *these receptors* and the entities *CD3, CD2, and CD28* is correctly identified; however, we extract the event annotation in (124b), since we ignore the quantifier *any*. The gold standard annotations are as given in (124c).

(124) (a) <u>CD3, CD2, and CD28</u> are functionally distinct receptors on T lymphocytes. **Engagement** of any of <u>these receptors</u> induces the rapid tyrosine phosphorylation of a shared group of intracellular signaling proteins, ...

(b) *engagement:*BINDING(e_1, t_1,t_2) ∧ *CD2:*PROTEIN(t_1) ∧ *CD28:*PROTEIN(t_2)

(c) *engagement:*BINDING(e_1, t_1) ∧ *CD2:*PROTEIN(t_1)
*engagement:*BINDING(e_2, t_2) ∧ *CD28:*PROTEIN(t_2)

We also noted cases in which the events that our system identifies due to coreference resolution seem correct, even though they are not annotated as such in the gold standard, as exemplified below. In this example, the anaphoric expression *their* is found to corefer with *IL-2 and IFN-*γ, and therefore, the event annotations in (125b) are extracted, whereas the gold standard only includes the event annotation in (125c).

(125) (a) Runx1 activates <u>IL-2 and IFN-γ</u> gene expression in conventional CD4+ T cells by **binding** to <u>their</u> respective promoter ...

(b) *binding:*BINDING(e_1, t_1,t_2) ∧ *Runx1:*PROTEIN(t_1) ∧ *IL-2:*PROTEIN(t_2)
*binding:*BINDING(e_2, t_1,t_3) ∧ *Runx1:*PROTEIN(t_1) ∧ *IFN-*γ*:*PROTEIN(t_3)

(c) *binding:*BINDING(e_1, t_1) ∧ *Runx1:*PROTEIN(t_1)

However, the shortcomings of the coreference resolution are evident in most error cases. The fact that we only consider semantically bound elements (surface elements corresponding to entities or triggers) as potential antecedents leads to a considerable number of errors. In such cases, the actual antecedent closer to the anaphoric expression may be ignored, in favor of a more distant entity or predication. In the following example, we identify as antecedent *PKD1, PKD2, and PKD3* for the pronoun *they*, because the actual antecedent, *PKD enzymes*, is semantically free. This leads to three false positive errors shown in (126b).

(126) (a) The protein kinase D (PKD) serine/threonine kinase family has three members: PKD1, PKD2, and PKD3. Most cell types express at least two PKD isoforms but PKD enzymes are especially highly expressed in haematopoietic cells, where they are activated in response to antigen receptors stimulation.

(b) *activated:*POSITIVE_REGULATION(e_1, t_1) $\wedge$ *PKD1:*PROTEIN(t_1)
*activated:*POSITIVE_REGULATION(e_2, t_2) $\wedge$ *PKD2:*PROTEIN(t_2)
*activated:*POSITIVE_REGULATION(e_3,t_3) $\wedge$ *PKD3:*PROTEIN(t_3)

6.4 Attribution Resolution

In the Penn Discourse TreeBank (PDTB) (Prasad *et al.*, 2008), attribution information, in addition to discourse relations, is annotated. Each discourse relation as well as individual arguments (Arg1 and Arg2) are associated with attribution information. For each discourse element, four attribution features are annotated in PDTB:

1. **Source** identifies the agent associated with the discourse element (Wr (writer), Ot (an agent introduced in the text), Arb (non-specific individuals)). By default, a discourse relation has Wr as its Source feature and the arguments have Inh, which indicates that the arguments inherit the Source value from the discourse relation. This feature is largely analogous to the *source* feature of our predications, even though the level of annotation in PDTB is coarse-grained.

2. **Type** indicates the relation between the agent (Source) and the abstract object indicated by the discourse relation or its argument (Comm (assertions), PAtt (beliefs), Ftv (presupposed facts), and Ctrl (intended eventualities)). By default, the Type feature of a discourse relation is Comm, while that of the arguments are Null.

3. **Scopal polarity** indicates transferred negation (see Section 2.2.1) of attribution predicate (Neg or Null). By default, this feature is Null.

4. **Determinacy** signals a context that cancels the entailment of the attribution itself (Indet or Null). By default, this feature is also Null.

Consider the sentence in Example (127) from the PDTB corpus, whose attribution annotation is represented below in Table (6.14). The discourse connective in the sentence is underlined. The discourse segment corresponding to Arg1 is italicized, while that corresponding to Arg2 is in bold. The discourse relation indicates CONTRAST between the fragments *Having the divident increases is* *a supportive element in the market outlook* and *it's a main consideration*. The fragments relevant to attribution are shown in boxes.

(127) [*"Having the dividend increases is a supportive element in the market outlook,*]$_{Arg1}$ but $\boxed{\text{I don't}}$ $\boxed{\text{think}}$ [**it's a main consideration**]$_{Arg2}$," $\boxed{\text{he says}}$.

	REL	Arg1	Arg2
[Source]	Ot	Inh	Ot
[Type]	Comm	Null	PAtt
[Polarity]	Null	Null	Neg
[Determinacy]	Null	Null	Null

Table 6.14: The attribution annotation associated with the sentence in Example (127).

Regarding the annotations in Table (6.14), the following points can be made:

- The discourse relation in this example as a whole (REL) has Ot value for its Source feature and Comm value for its Type feature. Both these values are due to the matrix clause *he says*, which indicates that someone other than the author of the text (Ot) is *asserting* (Comm) the statement corresponding to the discourse relation.

- One of the arguments of the discourse relation (Arg1) *inherits* its Source value (Inh) and its Type value (Null) from the discourse relation, since there is no attribution specific to this argument.

- The other argument (Arg2) has its own Source and Type values, since its attribution is indicated by the matrix clause *I don't think*. The argument is attributed to someone other than the author (Ot) and it is stated as a *belief* (PAtt). Its Scopal Polarity feature also has a non-default value Neg, indicating the existence of scopal polarity, due to the verbal predicate of the matrix clause, *think*.

Attribution resolution is the task of automatically identifying these features.

Attribution is an important component in discourse interpretation, However, attribution resolution is not a task that is generally considered in the context of discourse analysis, as indicated by Prasad *et al.* (2007) and Danlos and Rambow (2011). Furthermore, despite being annotated in PDTB, we are not aware of any work that specifically attempted to resolve attribution in this corpus. Instead, attribution has been discussed as a source of non-alignment between syntactic arguments and discourse arguments (Dinesh *et al.*, 2005), presumably a factor which makes it harder to accurately extract discourse elements accurately. As discussed in Section 3.4.1 and illustrated with the example above, MODAL predications contribute significantly to attribution of discourse elements. Therefore, attribution resolution can be viewed as a task that links lower level extra-factual phenomena to higher level discourse structure.

In this section, we describe our preliminary experiments in attribution resolution. To this end, we tested whether we can extract the attribution information from the predications extracted by the framework, without any changes to the framework itself. In the current work, we specifically focus on Source and Type features. The other two features, Polarity and Determinacy, are rarely annotated with non-default values in PDTB. The corpus statistics for particular feature-value combinations are given in Table (6.15).

Feature	**Value**	**REL**	**Arg1**	**Arg2**
[**Source**]	Wr	28532	126	12
	Ot	6539	4308	3868
	Arb	65	46	37
	Inh	-	30656	31219
[**Type**]	Comm	34544	3958	3451
	Ctrl	252	91	35
	Ftv	86	94	47
	PAtt	254	337	294
	Null	-	30656	31219
[**Polarity**]	Neg	-	45	36
	Null	35136	35091	35100
[**Determinacy**]	Indet	84	24	7
	Null	35052	35112	35129

Table 6.15: PDTB attribution statistics.

Despite the fact that we do not consider the Polarity and Determinacy features, it is worth noting that transferred negation associated with certain predicates is encoded explicitly in the embedding dictionary, and this is reflected in the semantic embedding graph. Therefore, it is only a matter of using this information already encoded by semantic predications to determine the Polarity of discourse elements, even though we have not so far attempted this. For example, in Example (127) above, we already know that the predication associated with the predicate *think* has scope over another with the type NEGATOR, which would be sufficient to conclude that the Polarity value of

Arg2 is Neg.

So far, we have only considered Explicit type of discourse relations and their arguments. However, the same principles can easily be extended to Implicit and AltLex type of relations. One difficulty with regard to Implicit relations is that since their discourse connective is not explicitly stated in the text, additional means are necessary to determine the attribution of the discourse relation itself. For example, consider the example below:

(128) A Lorillard spokewoman said, "This is an old story. [*We're talking about years ago before anyone heard of asbestos having any questionable properties*]$_{Arg1}$. [**There is no asbestos in our products now**]$_{Arg2}$."

There is an Implicit discourse relation of COMPARISON type between the sentences marked as Arg1 and Arg2. The Source of this discourse relation is Ot (*A Lorillard spokewoman*), due to the fact that the discourse relation in its entirety is embedded in a quotation context and the source of the quote is *A Lorillard spokewoman.* The embedding framework currently does not incorporate this type of multi-sentential quotation contexts; however, the ability to refer to such contexts as special nodes in the semantic embedding graph (similar to nodes representing entire sentences) is a natural extension of the framework and is currently being addressed in ongoing work. We detail how we derive attribution information for discourse elements below.

6.4.1 Deriving Attribution From Embedding Predications

We model attribution resolution as a post-processing step to compositional semantic interpretation, in a similar manner to biological event extraction. The composition phase is essentially the same as detailed earlier. The main difference is that gold PDTB discourse relation and argument annotations are provided to the system as input. In other words, we assume that the discourse relation and its arguments are known, and our task is to determine their attribution only, namely, the Source and Type features. In addition, while we identify anaphoric expressions, we do not perform coreference resolution, since PDTB does not annotate the actual agent associated with the element, but instead, identifies three types of abstract agents (writer (Wr), generic (Arb), and other (Ot)). The argument annotations are treated as atomic predications, whose inner structures are ignored, and the discourse connectives indicating the relations are treated as embedding predicates in the composition phase. The predications constructed at composition phase are already associated with a source feature, through source propagation procedure discussed in Section 5.3.2. Accordingly, at the end of the compositional process, predications corresponding to discourse relations and their arguments are all assigned a source.

Extracting the attribution Source feature from these annotations is a post-processing step. We use the following simple mapping rules:

1. If the source of the predication is assigned as WR, the Source feature of the associated discourse element is also Wr.

2. If the source of the predication is assigned a semantic term (Entity or Anaphora), the Source feature of the associated discourse element is Ot.

3. If the source of the predication is assigned as GEN, the Source feature is Arb.

With regard to the Type feature, the computation, which is also performed as a post-processing step, is more involved. By default, each discourse relation is given the Type value Comm, and the arguments the value Null. The exception is that the Arg1 argument is assigned the value Comm when its span corresponds to a sentence different than the one with the discourse connective. Then, we examine the extracted predications with regard to the scope relations and predication semantic categories to determine the Type features. The constraints for the Type features are given below. We assume that Pr_e is a predication that corresponds to an embedded predication, and Pr corresponds to another predication which has Pr_e in its immediate scope ($Pr > Pr_e$). Pr is indicated by the predicate P and Pr_e with the predicate P_e. The discourse element whose Type feature is under consideration corresponds to the predication Pr_e.

1. If the embedding predication is of a subtype of DEONTIC or has the type VOLITIVE or INTENTIONAL, the Type feature of the discourse element is Ctrl.

   $Sem(P) \in \{\text{OBLIGATIVE}, \text{PERMISSIVE}, \text{VOLITIVE}, \text{INTENTIONAL}\} \Rightarrow Type(Pr_e) = Ctrl$

2. If the embedding predication belongs to a subtype of the EPISTEMIC category or has the type SENSORY and the scalar value associated with its predicate is 1.0 or the predication is of the EVALUATIVE type, the Type feature of the discourse element is Ftv.

   $Sem(P) \in \{\text{EPISTEMIC}, \text{ASSUMPTIVE}, \text{SENSORY}, \text{EVALUATIVE}\} \wedge MV(P) = 1.0 \Rightarrow Type(Pr_e) = Ftv$

3. If the embedding predication is EPISTEMIC or EVIDENTIAL and the scalar EPISTEMIC value associated with it is between 0.5 and 1.0 AND its source is not Wr, the Type feature of the discourse element is PAtt.

   $Sem(P) \in \{\text{EPISTEMIC}, \text{EVIDENTIAL}\} \wedge MV(P) > 0.5 \wedge MV(P) < 1.0 \wedge Source(Pr) \neq Wr \Rightarrow Type(Pr_e) = PAtt$

4. If the embedding predication is of EPISTEMIC or EVIDENTIAL subtypes and its source is not Wr, the Type feature of the discourse element is Comm.

   $Sem(P) \in \{\text{EPISTEMIC}, \text{EVIDENTIAL}\} \wedge Source(Pr) \neq Wr \Rightarrow Type(Pr_e) = Comm$

6.4.2 Results and Discussion

We took the majority class values as the baseline for attribution resolution. In other words, the Source and Type features for a discourse relation are taken to be Wr and Comm, respectively. On the other hand, the features for a discourse argument are Inh and Null. We present the results of this experiment in Table (6.16). The evaluation metric we use is *accuracy*, that is, the percentage of correctly identified features. The last column refers to the case where the relevant feature of both the discourse relation and the arguments are considered jointly.

	REL	Arg1	Arg2	All
Baseline	74.77	67.61	70.01	65.25
Current System	80.84	82.28	81.04	75.97

Table 6.16: PDTB attribution Source resolution results.

The table shows that our system clearly improves over the already high baseline in extracting the Source feature. Interestingly, our system performs best in identifying the Source of Arg1 arguments, which yield the lowest baseline. In the opposite direction, we obtain the smallest improvement (6%) in identifying the Source of discourse relations, which yield the highest baseline. Arg1 generally lies in previous discourse and, thus, is frequently not syntactically linked to the discourse connective. Therefore, identifying its Source can be considered more challenging, as the likelihood of complicating factors such as multi-sentential arguments or quotation contexts is higher.

In comparison, the system performs more or less similarly to the baseline method, which clearly performs very well as shown in Table (6.17), in identifying the Type feature of discourse elements. In recognizing the Type feature of Arg1 arguments, we obtain a slight improvement over the baseline. On the other hand, with the discourse relation, there is a small performance drop in identifying the Type feature, whereas with Arg2, the performance is essentially same as the baseline.

	REL	Arg1	Arg2	All
Baseline	96.92	90.31	95.10	84.86
Current System	95.79	91.05	95.08	84.81

Table 6.17: PDTB attribution Type resolution results.

These results show that the post-processing rules devised to identify the Type features of the discourse elements may be a bit simplistic, and that more nuanced rules based on further analysis of the corpus are required. Since recognizing the Type of the abstract object corresponding to the discourse element may have an effect on determining the discourse relation, as shown by Asher and Lascarides (2003) and Danlos and Rambow (2011) and illustrated in Section 3.4.1, we plan to continue our work in attribution Type recognition.

As the discussion so far indicates and the post-processing rules above clearly show, identification of Type features is dependent on identifying the Source feature accurately. Therefore, it is informative to consider the results for both features jointly. We present the joint results in Table (6.18), which illustrates clear improvement over the baseline for all discourse elements and their combination.

	REL	Arg1	Arg2	All
Baseline	74.71	67.57	69.96	58.59
Current System	76.29	77.35	77.05	72.23

Table 6.18: Joint PDTB attribution resolution results.

Analyzing the error cases in recognizing Source and Type features, we identified several major classes of errors. One problem in identifying Source is the complex nature of quotation contexts. Consider the fragment in Example (129).

(129) "[*There's no question that some of those workers and managers contracted asbestos-related diseases,*"]$_{Arg1}$ said Darrell Phillips, vice president of human resources for Hollingsworth & Vose. "<u>But</u> [**you have to recognize that these events took place 35 years ago**.]$_{Arg2}$ ..."

For the discourse relation signaled by *But*, the Source feature is annotated as Ot, and for its arguments Inh, indicating that the arguments inherit the Source feature from the relation. The Source feature of the discourse relation is indicated with the fragment in box. Our system identifies the Source feature of Arg1 correctly, since the predication indicated by the source-introducing predicate *say* has in its scope the predication corresponding to Arg1. However, our system is unable to recognize that the second sentence, corresponding to the discourse connective and Arg2, belongs in the same quotation context as Arg1 and, therefore, both the discourse relation and Arg2 should have Ot as their Source feature. As mentioned earlier, extending the framework to consider such contexts and integrate them into the semantic embedding graph is ongoing work. Simple heuristics, such as simply counting the double quotes to identify the quotation contexts, are often useful in identifying the full extent of a quotation. In Example (130), counting from the beginning of the document, it is easy to find out that the first sentence is in the quotation that ends in the next sentence. On the other hand, other quotations are more complex and challenging, such as the gapped quotation in Example (129).

(130) ... "When [*we evaluated raising our bid*]$_{Arg2}$, [**the risks seemed substantial and persistent over the next five years, and the rewards seemed a long way out**]$_{Arg1}$. That got hard to take," $\boxed{\text{he added}}$.

With respect to identifying the Type feature, the rules based on embedding categories appear to be too coarse-grained in some instances. In Example (131), Arg1 is annotated as having Type value of Ctrl. On the other hand, our system annotates it as having the value of PAtt, due to the embedding predicate *predict*, which is encoded as a predicate of SPECULATIVE type in the embedding dictionary.

(131) "Although [*stocks have led bonds this week*]$_{Arg2}$, $\boxed{\text{some traders predict}}$ [**that relationship will reverse during the next few weeks**]$_{Arg1}$."

Most of the errors were associated with the shortcomings of the composition phase, however. In the sentence in Example (132a), our system was able to identify that the discourse relation has Ot as its Source feature (due to the noun phrase *One station manager* being the subject of the reporting predicate *say*), and that the arguments inherit this value. However, it does not identify the Type features correctly, due to the scope relations illustrated in (132b). Since the head of Arg1 (*strike*) is dominated by the belief predicate *believe* according to these scope relations, our system identifies the Type feature of Arg1 as being PAtt, whereas the discourse relation, indicated by *because* is assigned the value of Comm, since *because* is dominated by the reporting predicate *say.* The accurate scope relations and inferences would be as in Example (132c).

(132) (a) $\boxed{\text{One station manager says}}$ $\boxed{\text{he believes}}$ [*Viacom's move is a "pre-emptive strike"*]$_{Arg1}$ because [**the company is worried that "Cosby" ratings will continue to drop in syndication over the next few years**]$_{Arg2}$.

(b) $says >_d because >_d believes >_d strike \Rightarrow Type(because) = Comm, Type(strike) = PAtt$

(c) $says >_d believes >_d because >_d strike \Rightarrow Type(because) = PAtt, Type(strike) = Null$

6.5 Conclusions

In this chapter, we presented the practical tasks to which we applied various versions of the embedding framework. We noted the lack of an appropriate, single task on which all aspects of the framework can be evaluated and the difficulty of annotating a corpus based on the proposed framework as a motivation for our task-based evaluation. The tasks on which we evaluated the systems were uncertainty/hedge detection, speculation and negation scope resolution, biological event extraction, and attribution resolution. We described each evaluation scenario, noting the differences between the current framework and the precursor systems that were evaluated, if any. We also discussed how the same task can be accommodated within the current framework, when appropriate. Most of the evaluation has so far taken place in the biomedical domain, while we have recently been more actively evaluating our work in discourse-oriented tasks. Some of the discussion in this chapter was adapted from earlier or forthcoming articles (Kilicoglu and Bergler, 2008, 2009, 2010, 2011a,b, 2012).

Chapter 7

Conclusions and Future Directions

In the current work, we proposed a linguistically grounded framework, aimed at characterization and interpretation of semantic phenomena beyond simple, categorical assertions. Its main accomplishments are the comprehensive, domain-independent embedding categorization and the embedding predicate dictionary based on this categorization, as well as the bottom-up, compositional approach to semantic interpretation. The embedding categorization provides a comprehensive coverage of extra-propositional semantic content and, in doing so, unifies two levels of text understanding, namely, propositional meaning and discourse levels, often distinguished in CL/NLP research. Our rule-based semantic interpretation methodology provides an automatic, fine-grained analysis of extra-propositional semantic content, which forms the basis for addressing specific semantics-oriented tasks at both levels.

As theoretical contribution, we proposed *embedding* as the core notion in moving from shallow semantics towards extra-propositional meaning and discourse interpretation. By making a structural distinction between *embedding predications* and *atomic predications* and incorporating the notions of *scope*, *scale* and *source*, we demonstrated how extra-propositional meaning can be modeled in a compositional manner. With the proposed embedding categorization, we consolidated a variety of semantic phenomena discussed and studied in disparate research streams and illustrated in various corpora. With the embedding predicate dictionary, we brought together various kinds of linguistic classifications and resources regarding relevant semantic phenomena and defined them in a unified, consistent way, which, we believe, can prove beneficial as a robust lexical semantic resource for future research in a variety of research areas.

Our computational contribution consists of a compositional approach to semantic interpretation, which relies on lexical semantic information from the embedding predicate dictionary and a selective use of syntactic dependency relations. While the idea of compositional semantic interpretation is not new, it has hitherto been only explored in restricted domains and tasks in CL/NLP research. And even in those cases, compositionality has mostly been considered within a constituency perspective (Moilanen and Pulman, 2007; Choi and Cardie, 2008). Taking a semantic dependency perspective to compositionality, we modeled semantic content of a document as an embedding graph, on which we defined compositional operations, some of which were partially inspired by other work (polarity composition (Moilanen and Pulman, 2007), source propagation (Saurí, 2008)). Other compositional operations, scalar modality value composition and argument propagation, are the result of our own work, the former allowing a fine-grained semantic analysis and the latter accommodating semantic underspecification to some extent. Our selective, linguistically-principled use of syntactic dependencies as the basis for semantic interpretation contrasts with how they are used in most semantics-oriented research, such as relation extraction and semantic role labeling, where they are treated non-discriminatively as features for machine learning based approaches. Our results suggest that such selective use of syntactic dependencies provides robustness, while avoiding the computational complexity of such approaches. Our analysis suggested coreference resolution as a major component in moving towards discourse interpretation beyond the sentence level and we explored this task, integrating ideas from an existing, deterministic coreference resolution system (Haghighi and Klein, 2009) into our computational framework. The current coreference resolution module is limited in its scope and it was only fully evaluated in biological event extraction task, where the performance improvement due to resolution of coreference was very modest. However, the analysis of the coreference resolution results suggest that a major problem in this regard is the lack of ontological knowledge in the current framework. Integrating a named entity recognizer into our system would allow us to impose more semantics on the embedding graph, and thus, could improve coreference resolution performance. Our modular, incremental approach ensures that such new capabilities can be added and their effect on overall system performance in practical tasks can be measured.

One major challenge for our framework has been its evaluation. We adopted a task based evaluation approach, in which the embedding framework or its precursors were empirically tested on corpora and shared task competitions that annotated semantic phenomena relevant in the embedding framework. We considered easy adaptability and extensibility of our framework to specific tasks, while avoiding task- or domain-specific optimizations, as an indication of its soundness. We demonstrated success in hedge/speculation detection, as well as scope resolution and biological event extraction tasks, in which we achieved competitive results. We also improved on a majority class baseline in attribution resolution task. These tasks allowed us to test MODAL and VALENCE_SHIFTER embedding categories to a large extent, while testing of RELATIONAL and PROPOSITIONAL categories was less extensive. The biological event extraction task also allowed us to explore identification of atomic predications, in addition to embedding ones, and the results for this task indicate that our semantic interpretation approach extends naturally to extraction of such predications, as well. Our results in these tasks demonstrate the viability of our general proposal and suggest that our computational approach suffers little from the brittleness often attributed to rule-based systems and from the presumed domain dependence of extra-factual phenomena, such as hedging (Szarvas, 2008). We consider this robustness a result of the generality and comprehensiveness of the underlying rules and the use of syntactic information in a linguistically-principled manner.

Much of the evaluation focused on the biomedical research literature, since interest in some of the relevant semantic phenomena (modality, negation, certainty, etc.) has been most significant in bioNLP research, and various biomedical corpora as well as shared task competitions have been available (Kim *et al.*, 2008; Vincze *et al.*, 2008; Tsujii, 2009; Farkas *et al.*, 2010; Kim *et al.*, 2011a; Thompson *et al.*, 2011). On the other hand, our framework has benefitted from analysis of several corpora focusing on news articles in the Wall Street Journal portion of the Penn TreeBank (Prasad *et al.*, 2008; Palmer *et al.*, 2005; Meyers *et al.*, 2004b). Furthermore, vagueness detection in Wikipedia articles allowed us to demonstrate the domain-independent nature of our approach to hedging/speculation detection. With our more recent work in attribution resolution, we are also moving in the direction of empirically evaluating the embedding framework on news articles and discourse-oriented tasks more comprehensively. Potential future work in this regard would be to extend the compositional interpretation approach to discourse segmentation and parsing tasks, which would allow us to evaluate discourse-oriented RELATIONAL categories more comprehensively. The embedding framework in its current state can accommodate these tasks to some extent, as discussed in Section 3.4. In the context of PDTB, the current framework can be used to address Explicit discourse relations. The discourse connectives have already been extracted from the PDTB corpus and integrated into the embedding predicate dictionary with RELATIONAL senses. When it comes to identifying discourse segments (arguments of discourse connectives), the semantic embedding graph can assist significantly in accurately identifying them. The approach would be similar to that proposed for speculation scope resolution in the context of the BioScope corpus, as discussed in Section 6.2. One complication in identifying discourse segments is that they often span multiple sentences. Such discourse segments generally seem to occur in quotation contexts. We are currently exploring extensions to the semantic embedding graph that are concerned with quotation contexts and multi-sentence units, as we stated earlier in Section 6.4. Extending the framework to Implicit discourse relations between sentences is more challenging, since a dictionary-based methodology for identifying connectives and their senses would not work. Prasad *et al.* (2010) propose a combination of heuristics based on quotation contexts and coreference resolution to identify the sentence in prior discourse that is in a discourse relation with the current sentence, which, we believe, would assist in resolving such implicit discourse relations. We provide a level of semantic interpretation of discourse segments themselves (if they involve embedding), and this, in combination with attribution source and type, which we also identify to some extent, may also prove useful in resolving implicit discourse relations. Another less evaluated portion of the embedding framework is PROPOSITIONAL categories. The most appropriate task for SEMANTIC_ROLE subtypes would be semantic role labeling, particularly nominal semantic role labeling, since Meyers *et al.* (2004b) discuss the relevant phenomena in the context of NomBank. For ASPECTUAL categories, an appropriate evaluation would be to identify the ALINKs (aspectual links) defined in the TimeBank corpus (Pustejovsky *et al.*, 2003) for events.

In the biomedical domain, relevant biomedical corpora that can be used to validate the embedding framework further include the multi-dimensional, qualitatively annotated corpus of Wilbur *et al.* (2006) and the very recent meta-knowledge annotation of GENIA corpus (Thompson *et al.*, 2011). These two corpora annotate very similar phenomena, even though the former annotates sentence fragments, while the latter annotates biological events with the meta-knowledge dimensions, including *certainty level*, *evidence*, and *polarity*. The approach toward these tasks would be very similar to that taken toward biological event extraction: that is, semantic composition followed by post-processing rules. Since we can already extract GENIA-based events, the meta-knowledge corpus seems to be the most natural evaluation platform, in this respect.

One of the main shortcomings (although intentionally so) of the embedding framework is that it does not incorporate any domain and ontological knowledge, with respect to world objects, entities and domain-specific events. The main reason for this domain knowledge-poor approach was our hypothesis that embedding was a domain-independent, linguistic notion and that domain knowledge could be plugged into the framework as the bottom layer, such that embedding-related mechanisms sit on top of and use domain knowledge. We incorporated such knowledge into the framework successfully in the biological event extraction task, using PROTEIN entities and biological event types, such as GENE_EXPRESSION and PHOSPHORYLATION, as they are defined in the GENIA ontology. While this provides proof-of-concept for our hypothesis, another future possibility is integrating the embedding framework into a large-scale information extraction system which uses domain knowledge to a fuller extent. One such application is SemRep (Rindflesch and Fiszman, 2003) in the biomedical domain, which extracts propositional content from biomedical text in the form of semantic predications: subject-predicate-object triples. The elements of the predications are drawn from the UMLS knowledge sources (Bodenreider, 2004); the subject and object pair corresponds to UMLS Metathesaurus concepts and the predicate to a relation type in the UMLS Semantic Network. SemRep extracts a wide range of predicates relating to clinical medicine (e.g. TREATS, DIAGNOSES, ADMINISTERED_TO, PROCESS_OF), substance interactions (e.g., INTERACTS_WITH, INHIBITS, STIMULATES), genetic etiology of disease (e.g., ASSOCIATED_WITH, CAUSES, PREDISPOSES), and pharmacogenomics (e.g., AFFECTS, AUGMENTS, DISRUPTS). For example, SemRep identifies the predications in Example (133b) from the sentence in Example (133a). Arguments, drawn from the UMLS, have the form *ConceptIdentifier: ConceptName (ConceptSemanticType)*. The predications are indicated by underlined words, and the arguments are in square brackets. In integrating with the embedding framework, we would consider SemRep relations as atomic predications.

(133) (a) [MRI] <u>revealed</u> a [lacunar infarction] <u>in</u> the left [internal capsule].

(b) C0024485: Magnetic Resonance Imaging (Diagnostic Procedure)-DIAGNOSES-C0333559: Infarction, Lacunar (Disease or Syndrome)
C0152341: Internal Capsule (Body Part, Organ, or Organ Component)- LOCATION_OF-C0333559: Infarction, Lacunar (Disease or Syndrome)

The appeal of SemRep from our perspective is four-fold:

- It normalizes textual mentions of entities by mapping them to ontological concepts in the UMLS Metathesaurus, which contains more than 1.2 million concepts.

- The Medline database of article titles and abstracts (the main resource for biomedical literature) has been preprocessed with SemRep within the context of Semantic Medline (Kilicoglu *et al.*, 2008) project. This corresponds to over 20 million abstracts and approximately 60 million predications, which covers a very wide range of biomedical information. This is in contrast to biological event corpora and extraction systems, which cover and are trained on very narrow subdomains, respectively.

- The relation types SemRep uses are mostly domain-specific, allowing the possibility of using them as atomic predications in a complementary fashion to the embedding framework. Despite its breadth of coverage, SemRep aims at shallow semantics, ignoring phenomena such as modality, coreference, and discourse structure, which would be contributed by the embedding framework.

- In the opposite direction, SemRep could also contribute to more accurate coreference resolution. As we noted above, the underspecified nature of our coreference resolution approach (that we do not perform additional named entity recognition) lowers precision. By integrating with SemRep, the embedding framework would have access to rich entity and relational semantics encoded in UMLS, which, we expect, would improve coreference resolution.

One of the future directions for the current research is then to integrate SemRep and the embedding framework to the fullest extent. The most immediate consequence of such integration could be determining the epistemic status of the predications in the Semantic Medline database. Are they speculations, facts or counter-facts? What is the level of confidence associated with a predication? The success of the framework in this task can be evaluated on the small corpus of SemRep relations recently annotated (Kilicoglu *et al.*, 2011) or in a post-hoc analysis of a small set of randomly selected relations from the database. This integration can also serve literature-based discovery and hypothesis generation tasks, for which SemRep relations have been exploited previously (Hristovski *et al.*, 2006; Cohen *et al.*, 2010b; Miller *et al.*, 2012). For example, based on analysis of SemRep relations, Miller *et al.* (2012) posited cortisol as the mechanistic link for the generally assumed but unexplained link between testosterone and sleep. Our framework can enhance the value of semantic predications that contribute to these tasks by determining whether they are supported by strong, compelling evidence, based on their epistemic status and explicitness of evidence. Since the Medline database covers more or less all biomedical and life sciences research from mid-20th century on, this also gives us the ability of tracking how the scientific knowledge changes diachronically. One can, for example, assume that when a particular piece of biomedical information first appears in the literature (captured as a predication by SemRep), its factual status is more tentative, and in later periods, the same information is supported by more evidence or perhaps refuted by counter-evidence, which we believe can be captured via embedding framework and aggregation over the entire predication database. This is essentially similar to the idea of capturing *paradigm shifts*, proposed by Lisacek *et al.* (2005), albeit at a much larger scale.

The work presented in this thesis, in some sense, runs against the current trends in CL/NLP research, which generally focus on various kinds of statistically-based, machine learning approaches with differing levels of supervision on well-defined tasks, for which annotated corpora exist or can be induced with automatic or semi-automatic means. Our work argues that linguistic principles can be brought more comprehensively into the picture for a finer-grained text understanding and in moving towards discourse interpretation. However, we have not discussed how our approach would work within the current trends. We believe that while the embedding framework provides a solid foundation for semantic interpretation, machine learning techniques can be useful for more focused, smaller tasks within the framework. One area in which machine learning techniques could potentially be useful would be automatic learning of relevant predicates and their lexical semantic and syntactic features from corpora or unrestricted text, such as Web. We explored this to a small extent in earlier work on a corpus using standard supervised machine learning techniques (Kilicoglu and Bergler, 2011b). While the results did not indicate much improvement over a careful selection of predicates based on relatively simple heuristics, whether more sophisticated techniques could in fact improve results should be further investigated. On the other hand, with the current work, we provide robust, fine-grained semantic interpretation and we believe that using the resulting semantic interpretation as the basis for advanced feature engineering within a machine learning framework can be beneficial for such applications.

Bibliography

Agarwal, S. and Yu, H. (2009). Automatically classifying sentences in full-text biomedical articles into Introduction, Methods, Results and Discussion. *Bioinformatics*, **25**(23), 3174–3180.

Agarwal, S. and Yu, H. (2010). Detecting hedge cues and their scope in biomedical text with conditional random fields. *Journal of Biomedical Informatics*, **43**(6), 953–961.

Ahlers, C. B., Fiszman, M., Demner-Fushman, D., Lang, F. M., and Rindflesch, T. C. (2007). Extracting semantic predications from Medline citations for pharmacogenomics. *Pacific Symposium in Biocomputing*, pages 209–220.

Aikhenvald, A. Y. (2004). *Evidentiality*. Oxford University Press, Oxford, UK.

Andreevskaia, A., Bergler, S., and Urseanu, M. (2007). All Blogs are Not Made Equal: Exploring Genre Differences in Sentiment Tagging of Blogs. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM-2007)*, pages 195–198.

Asher, N. (1993). *Reference to Abstract Objects in Discourse: A Philosophical Semantics for Natural Language Metaphysics*. Kluwer Academic Publishers.

Asher, N. and Lascarides, A. (2003). *Logics of Conversation*. Cambridge University Press.

Asher, N. and Vieu, L. (2005). Subordinating and coordinating discourse relations. *Lingua*, **115**, 591–610.

Auwera, J. V. D. and Plungian, V. A. (1998). Modality's semantic map. *Linguistic Typology*, **2**(1), 79–124.

Averbuch, M., Karson, T., Ben-Ami, B., Maimon, O., and Rokach, L. (2004). Context-sensitive medical information retrieval. In *Proc MEDINFO-2004*, pages 1–8.

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The Berkeley FrameNet project. In *Proceedings of the 17th International conference on Computational Linguistics*, pages 86–90.

Baker, K., Bloodgood, M., Dorr, B., Filardo, N. W., Levin, L., and Piatko, C. (2010). A Modality Lexicon and its use in Automatic Tagging. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*.

Baldridge, J. and Lascarides, A. (2005). Probabilistic Head-Driven Parsing for Discourse Structure. In *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*, pages 96–103, Ann Arbor, Michigan. Association for Computational Linguistics.

Baldridge, J., Asher, N., and Hunter, J. (2007). Annotation for and Robust Parsing of Discourse Structure on Unrestricted Texts. *Zeitschrift fur Sprachwissenschaft*, **26**, 213–239.

Bale, A. (2011). Scales and comparison classes. *Natural Language Semantics*, **19**(2), 169–190.

Bergler, S. (1992). *Evidential analysis of reported speech*. Ph.D. thesis, Brandeis University.

Bethard, S., Yu, H., Thornton, A., Hatzivassiloglou, V., and Jurafsky, D. (2006). Extracting opinion propositions and opinion holders using syntactic and lexical cues. In J. G. Shanahan, Y. Qu, and J. Wiebe, editors, *Computing Attitude and Affect in Text: Theories and Applications*, volume 20, pages 125–141. Springer, 1nd edition.

Björne, J. and Salakoski, T. (2011). Generalizing Biomedical Event Extraction. In *Proceedings of BioNLP Shared Task 2011 Workshop*, pages 183–191. Association for Computational Linguistics.

Björne, J., Pyysalo, S., Ginter, F., and Salakoski, T. (2008). Extracting protein-protein interactions from text using rich feature vectors and feature selection. In *Proceedings of the Third International Symposium on Semantic Mining in Biomedicine (SMBM 2008)*, pages 125–128.

Blake, C. (2009). Beyond genes, proteins and abstracts: Identifying scientific claims from full-text biomedical articles. *Journal of Biomedical Informatics*, **43**, 173–189.

Blühdorn, H. (2008). Subordination and coordination in syntax, semantics, and discourse: Evidence from the study of connectives. In C. Fabricius-Hansen and W. Ramm, editors, *Subordination versus coordination in sentence and text: A cross-linguistic perspective*, pages 59–85. John Benjamins Publishing Company.

Bodenreider, O. (2004). The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Research*, **32**(Database issue), 267–270.

Bossy, R., Jourde, J., Bessières, P., van de Guchte, M., and Nédellec, C. (2011). BioNLP Shared Task 2011 - Bacteria Biotope. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task*, Portland, Oregon. Association for Computational Linguistics.

Bratman, M. (1987). *Intentions, Plans, and Practical Reason*. Harvard University Press, Cambridge, MA.

Buchholz, S. and Marsi, E. (2006). Conll-x shared task on multilingual dependency parsing. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)*, pages 149–164, New York City. Association for Computational Linguistics.

Bybee, J. L., Perkins, R. D., and Pagliuca, W. (1994). *The evolution of grammar: tense, aspect, and modality in the languages of the world*. University of Chicago Press.

Carlson, L. and Nirenburg, S. (1990). World modeling for NLP. Technical Report CMU-CMT-90-121, Carnegie Mellon University, Center for Machine Translation.

Carreras, X. and Màrquez, L. (2005). Introduction to the CoNLL-2005 shared task: Semantic role labeling. In *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*, pages 152–164, Ann Arbor, Michigan. Association for Computational Linguistics.

Chafe, W. (1986). Evidentiality in english conversation and academic writing. In W. Chafe and J. Nichols, editors, *Evidentiality: The Linguistic Coding of Epistemology*, volume 20, pages 261–272. Ablex Publishing Corporation, Norwood, NJ, 1nd edition.

Channell, J. (1994). *Vague Language*. Oxford University Press, Oxford, UK.

Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F., and Buchanan, B. G. (2001). A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*, **34**(5), 301–310.

Charniak, E. and Johnson, M. (2005). Coarse-to-fine n-best parsing and maxent discriminative reranking. In *Proceedings of the 43rd Meeting of the Association for Computational Linguistics*, pages 173–180.

Chierchia, G. and McConnell-Ginet, S. (1990). *Meaning and Grammar: an introduction to semantics*. MIT Press, Cambridge, MA.

Choi, Y. and Cardie, C. (2008). Learning with compositional semantics as structural inference for subsentential sentiment analysis. In *Proceedings of the 2008 Conference on Empirical Methods in*

Natural Language Processing, pages 793–801, Honolulu, Hawaii. Association for Computational Linguistics.

Clegg, A. and Shepherd, A. J. (2007). Benchmarking natural-language parsers for biological applications using dependency graphs. *BMC Bioinformatics*, **8**, 24.

Coates, J. (1983). *The Semantics of The Modal Auxiliaries*. Croom Helm, London, UK.

Cohen, K. B., Johnson, H. L., Verspoor, K., Roeder, C., and Hunter, L. E. (2010a). The structural and content aspects of abstracts versus bodies of full text journal articles are different. *BMC Bioinformatics*, **11**, 492.

Cohen, T., Whitfield, G. K., Schvaneveldt, R. W., Mukund, K., and Rindflesch, T. C. (2010b). EpiphaNet: An interactive tool to support biomedical discoveries. *Journal of biomedical discovery and collaboration*, **5**, 21–49.

Comrie, B. (1985). Causative verb formation and other verb-deriving morphology. In T. Shopen, editor, *Grammatical categories and the lexicon. Language typology and syntactic description*, volume 3, pages 309–348. Cambridge University Press, New York.

Cosme, C. (2008). A corpus-based perspective on clause linking patterns in English, French and Dutch. In C. Fabricius-Hansen and W. Ramm, editors, *Subordination versus coordination in sentence and text: A cross-linguistic perspective*, pages 89–114. John Benjamins Publishing Company.

Cristea, D., Ide, N., and Romary, L. (1998). Veins Theory: a model of global discourse cohesion and coherence. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1*, ACL '98, pages 281–285, Stroudsburg, PA, USA. Association for Computational Linguistics.

Cristea, D., Ide, N., Marcu, D., and Tablan, V. (2000). An Empirical Investigation of the Relation Between Discourse Structure and Co-Reference. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING2000)*.

Cristofaro, S. (2003). *Subordination*. Oxford University Press, Oxford, UK.

Danlos, L. (2006). "Discourse verbs" and discourse periphrastic links. In C. Sidner, J. Harpur, A. Benz, and P. Kühnlein, editors, *Second Workshop on Constraints in Discourse (CID06)*, pages 59–65. Maynooth, Ireland.

Danlos, L. and Rambow, O. (2011). Discourse Relations and Propositional Attitudes. In *Proceedings of the Fourth International Workshop on Constraints in Discourse (2011)*. Agay-Roches Rouges, France.

de Haan, F. (1997). *The Interaction of Modality and Negation: a Typological Study*. Garland, New York, NY.

de Haan, F. (1999). Evidentiality and Epistemic Modality: Setting Boundaries. *Southwest Journal of Linguistics*, **18**(1), 83–101.

de Marneffe, M.-C. and Manning, C. D. (2008). Stanford typed dependencies manual.

de Marneffe, M.-C., MacCartney, B., and Manning, C. D. (2006). Generating typed dependency parses from phrase structure parses. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*, pages 449–454.

de Waard, A., Buitelaar, P., and Eigner, T. (2009). Identifying the Epistemic Value of Discourse Segments in Biology Texts. In *Proceedings of the 8th International Conference on Computational Semantics*, pages 351–354.

Dinesh, N., Lee, A., Miltsakaki, E., Prasad, R., Joshi, A., and Webber, B. (2005). Attribution and the (Non-)Alignment of Syntactic and Discourse Arguments of Connectives. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*, pages 29–36, Ann Arbor, Michigan. Association for Computational Linguistics.

Dowty, D. R., Wall, R. E., and Peters, S. (1981). *Introduction to Montague Semantics*. D. Reidel Publishing Company.

Farkas, R., Vincze, V., Mora, G., Csirik, J., and Szarvas, G. (2010). The CoNLL 2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text. In *Proceedings of the CoNLL2010 Shared Task*.

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

Fiszman, M., Demner-Fushman, D., Lang, F. M., Goetz, P., and Rindflesch, T. C. (2007). Interpreting comparative constructions in biomedical text. In *Biological, translational, and clinical language processing*, pages 137–144, Prague, Czech Republic. Association for Computational Linguistics.

Frawley, W. (1992). *Linguistic Semantics*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Friedman, C., Alderson, P., Austin, J., Cimino, J. J., and Johnson, S. B. (1994). A general natural-language text processor for clinical radiology. *Journal of American Medical Informatics Association*, **1**, 161–174.

Friedman, C., Kra, P., and Rzhetsky, A. (2002). Two biomedical sublanguages: a description based on the theories of Zellig Harris. *Journal of Biomedical Informatics*, **35**, 222–235.

Ganter, V. (2010). Detecting vagueness in two languages and two domains. In *Proceedings of the Konferenz zur Verarbeitung Natürlicher Sprache (KONVENS)*, pages 141–145.

Ganter, V. and Strube, M. (2009). Finding Hedges by Chasing Weasels: Hedge Detection Using Wikipedia Tags and Shallow Linguistic Features. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 173–176.

Georgescul, M. (2010). A Hedgehop over a Max-Margin Framework Using Hedge Cues. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, pages 26–31, Uppsala, Sweden. Association for Computational Linguistics.

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

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

Givón, T. (2001). *Syntax: A Functional-Typological Introduction*. John Benjamins B.V., Amsterdam, Netherlands.

Goldin, I. M. and Chapman, W. W. (2003). Learning to detect negation with "not" in medical texts. In *Proc Workshop on Text Analysis and Search for Bioinformatics at the 26th ACM SIGIR Conference*.

Gross, A. and Chesley, P. (2012). Hedging, stance, and voice in medical research articles: A study of deviance. In K. Hyland and C. Sancho-Guinda, editors, *Stance and voice in academic discourse*. Palgrave Macmillan, Basingstoke. To appear.

Grosz, B. J. and Sidner, C. L. (1986). Attention, Intentions, and the Structure of Discourse. *Computational Linguistics*, **12**(3), 175–204.

Grosz, B. J., Weinstein, S., and Joshi, A. K. (1995). Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics*, **21**(2), 203–225.

Guo, Y., Korhonen, A., Liakata, M., Silins, I., Sun, L., and Stenius, U. (2010). Identifying the Information Structure of Scientific Abstracts: An Investigation of Three Different Schemes. In *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing*, pages 99–107, Uppsala, Sweden. Association for Computational Linguistics.

Guo, Y., Korhonen, A., Silins, I., and Stenius, U. (2011). Weakly supervised learning of information structure of scientific abstractsis it accurate enough to benefit real-world tasks in biomedicine? *Bioinformatics*, **27**(22), 3179–3185.

Haghighi, A. and Klein, D. (2009). Simple Coreference Resolution with Rich Syntactic and Semantic Features. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 1152–1161, Singapore. Association for Computational Linguistics.

Haghighi, A., Toutanova, K., and Manning, C. (2005). A joint model for semantic role labeling. In *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*, pages 173–176, Ann Arbor, Michigan. Association for Computational Linguistics.

Halliday, M. A. K. and Hasan, R. (1976). *Cohesion in English*. Longman, London, UK.

Harkema, H., Dowling, J. N., Thornblade, T., and Chapman, W. W. (2009). ConText: an algorithm for determining negation, experiencer, and temporal status from clinical reports. *Journal of Biomedical Informatics*, pages 839–851.

Hengeveld, K. (2004). Illocution, mood, and modality. In G. Booij, C. Lehmann, and J. Mugdan, editors, *Morphology: A handbook on inflection and word formation, Vol.2*, pages 1190–1202. Mouton de Gruyter.

Hobbs, J. R. (1985a). On the coherence and structure of discourse. Technical Report CSLI-85-37, Center for the Study of Language and Information, Stanford University.

Hobbs, J. R. (1985b). Ontological promiscuity. In *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*, pages 61–69.

Horn, L. R. (1972). *On the Semantic Properties of Logical Operators in English*. Ph.D. thesis, University of California Los Angeles.

Horn, L. R. (1989). *A Natural History of Negation*. University of Chicago Press, Chicago, IL.

Hristovski, D., Friedman, C., Rindflesch, T. C., and Peterlin, B. (2006). Exploiting semantic relations for literature-based discovery. *AMIA Annual Symposium proceedings*, pages 349–353.

Huddleston, R. D. and Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge University Press, Cambridge, UK.

Hughes, G. E. and Cresswell, M. J. (1996). *A New Introduction to Modal Logic*. Routledge, New York. NY.

Hyland, K. (1998). *Hedging in scientific research articles*. John Benjamins B.V., Amsterdam, Netherlands.

Jourde, J., Manine, A.-P., Veber, P., Fort, K., Bossy, R., Alphonse, E., and Bessières, P. (2011). BioNLP Shared Task 2011 - Bacteria Gene Interactions and Renaming. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task*, Portland, Oregon. Association for Computational Linguistics.

Kabiljo, R., Clegg, A. B., and Shepherd, A. J. (2009). A realistic assessment of methods for extracting gene/protein interactions from free text. *BMC Bioinformatics*, **10**(1), 233+.

Kamp, H. (1981). A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, and M. Stockhof, editors, *Formal Methods in the Study of Language*, volume 1, pages 277–322. Mathematisch Centrum, Amsterdam.

Kamp, H. and Reyle, U. (1993). *From Discourse to Logic*. Kluwer, Dordrecht.

Karttunen, L. (1971). Implicative Verbs. *Language*, **47**(2), 340–358.

Kennedy, A. and Inkpen, D. (2006). Sentiment Classification of Movie Reviews Using Contextual Valence Shifters. *Computational Intelligence*, **22**(2), 110–125.

Kennedy, C. and McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*, **81**(2), 345–381.

Khoo, C., Chan, S., and Niu, Y. (2002). The many facets of the cause-effect relation. In R. Green, C. A. Bean, and S. H. Myaeng, editors, *The Semantics of Relationships: An Interdisciplinary Perspective*, pages 51–70. Springer.

Kilicoglu, H. and Bergler, S. (2008). Recognizing speculative language in biomedical research articles: a linguistically motivated perspective. *BMC Bioinformatics*, **9 Suppl 11**, s10.

Kilicoglu, H. and Bergler, S. (2009). Syntactic dependency based heuristics for biological event extraction. In *Proceedings of Natural Language Processing in Biomedicine (BioNLP) NAACL 2009 Workshop*, pages 119–127.

Kilicoglu, H. and Bergler, S. (2010). A High-Precision Approach to Detecting Hedges and their Scopes. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, pages 70–77, Uppsala, Sweden. Association for Computational Linguistics.

Kilicoglu, H. and Bergler, S. (2011a). Adapting a general semantic interpretation approach to biological event extraction. In *Proceedings of BioNLP Shared Task 2011 Workshop*, pages 173–182, Portland, Oregon, USA. Association for Computational Linguistics.

Kilicoglu, H. and Bergler, S. (2011b). Effective Bio-Event Extraction using Trigger Words and Syntactic Dependencies. *Computational Intelligence*, **27**(4), 583–609.

Kilicoglu, H. and Bergler, S. (2012). Biological Event Composition. *BMC Bioinformatics*, **13 Suppl 8**, S7. to appear.

Kilicoglu, H., Fiszman, M., Rodriguez, A., Shin, D., Ripple, A., and Rindflesch, T. (2008). Semantic MEDLINE: A web application to manage the results of PubMed searches. In T. Salakoski, D. R. Schuhmann, and S. Pyysalo, editors, *Proceedings of the Third International Symposium on Semantic Mining in Biomedicine (SMBM 2008), Turku, Finland*, pages 69–76.

Kilicoglu, H., Rosemblat, G., Fiszman, M., and Rindflesch, T. (2011). Constructing a semantic predication gold standard from the biomedical literature. *BMC Bioinformatics*, **12**(1), 486+.

Kim, J.-D., Ohta, T., and Tsujii, J. (2008). Corpus annotation for mining biomedical events from literature. *BMC Bioinformatics*, **9**, 10.

Kim, J.-D., Ohta, T., Pyysalo, S., Kano, Y., and Tsujii, J. (2009). Overview of BioNLP'09 Shared Task on Event Extraction. In *Proceedings of Natural Language Processing in Biomedicine (BioNLP) NAACL 2009 Workshop*, pages 1–9.

Kim, J.-D., Pyysalo, S., Ohta, T., Bossy, R., and Tsujii, J. (2011a). Overview of BioNLP Shared Task 2011. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task*, pages 1–6.

Kim, J.-D., Wang, Y., Takagi, T., and Yonezawa, A. (2011b). Overview of the Genia Event task in BioNLP Shared Task 2011. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task*, Portland, Oregon. Association for Computational Linguistics.

Kim, S.-M. and Hovy, E. (2006). Identifying and analyzing judgment opinions. In *Proceedings of the HLT/NAACL Conference*, pages 200–207.

Kintsch, W. and Keenan, J. (1973). Reading rate and retention as a function of the number of the propositions in the base structure of sentences. *Cognitive Psychology*, **5**(3), 257–274.

Kintsch, W. and van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, **85**(5), 363–394.

Klein, D. and Manning, C. D. (2003). Accurate unlexicalized parsing. In *Proceedings of the 41th Meeting of the Association for Computational Linguistics*, pages 423–430.

Kratzer, A. (1981). The notional category of modality. In H. J. Eikmeyer and H. Rieser, editors, *Words, Worlds and Context. New Approaches in Word Semantics*, pages 289–323. Walter de Gruyter, Berlin.

Krestel, R., Witte, R., and Bergler, S. (2007). Processing of beliefs extracted from reported speech in newspaper articles. In *International Conference on Recent Advances in Natural Language Processing (RANLP 2007)*.

Krestel, R., Witte, R., and Bergler, S. (2008). A Belief Revision Approach to Textual Entailment Recognition. In *Proceedings of the Text Analysis Conference (TAC)*.

Krestel, R., Witte, R., and Bergler, S. (2009). Believe It or Not: Solving the TAC 2009 Textual Entailment Tasks through an Artificial Believer System. In *Proceedings of the Text Analysis Conference (TAC)*.

Lakoff, G. (1972). Hedges: a study in meaning criteria and the logic of fuzzy concepts. *Chicago Linguistics Society Papers*, **8**, 183–228.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar*. Stanford University Press, Stanford, CA.

Lappin, S. (2000). An Intensional Parametric Semantics for Vague Quantifiers. *Linguistics and Philosophy*, **23**(6), 599–620.

Liakata, M., Teufel, S., Siddhartan, A., and Batchelor, C. (2010). Corpora for conceptualisation and zoning of scientific papers. In *Proceedings of LREC 2010*, pages 2054–2061.

Light, M., Qiu, X. Y., and Srinivasan, P. (2004). The language of bioscience: facts, speculations, and statements in between. In *BioLINK 2004: Linking Biological Literature, Ontologies and Databases*, pages 17–24.

Lisacek, F., Chichester, C., Kaplan, A., and Sándor, A. (2005). Discovering Paradigm Shift Patterns in Biomedical Abstracts: Application to Neurodegenerative Diseases. In *Proceedings of the First International Symposium on Semantic Mining in Biomedicine (SMBM)*, pages 41–50, Hinxton, England.

Lyons, J. (1977). *Semantics*. Cambridge University Press, Cambridge, UK.

Mani, I., Pustejovksy, J., and Gaizauskas, R. (2005). *The Language of Time: A Reader*. Oxford University Press, Oxford, UK.

Mann, W. C. and Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, **8**(3), 243–281.

Marcu, D. (1999). A decision-based approach to rhetorical parsing. In *Proceedings of the 37th Meeting of the Association for Computational Linguistics (ACL'99)*, pages 365–372. Association for Computational Linguistics.

Marcu, D. and Echihabi, A. (2002). An unsupervised approach to recognizing discourse relations. In *Proceedings of the 40th Meeting of the Association for Computational Linguistics (ACL-2002)*, pages 368–375. Association for Computational Linguistics.

Marcus, M., B, S., and A, M. M. (1993). Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics*, **19**(2), 313–330.

Màrquez, L., Comas, P., Giménez, J., and Català, N. (2005). Semantic role labeling as sequential tagging. In *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*, pages 193–196, Ann Arbor, Michigan. Association for Computational Linguistics.

Matthiessen, C. and Thompson, S. A. (1988). The structure of discourse and 'subordination'. In J. Haiman and S. A. Thompson, editors, *Clause combining in grammar and discourse*, pages 275–329. John Benjamins.

McClosky, D. and Charniak, E. (2008). Self-training for biomedical parsing. In *Proceedings of the 46th Meeting of the Association for Computational Linguistics*, pages 101–104.

McCray, A. T., Srinivasan, S., and Browne, A. C. (1994). Lexical methods for managing variation in biomedical terminologies. In *Proceedings of the 18th Annual Symposium on Computer Applications in Medical Care*, pages 235–239.

McDonald, R., Lerman, K., and Pereira, F. (2006). Multilingual dependency analysis with a two-stage discriminative parser. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)*, pages 216–220, New York City. Association for Computational Linguistics.

Medlock, B. and Briscoe, T. (2007). Weakly supervised learning for hedge classification in scientific literature. In *Proceedings of the 45th Meeting of the Association for Computational Linguistics*, pages 992–999.

Mel'čuk, I. (2004). Actants in semantics and syntax I: Actants in semantics. *Linguistics*, **42**(1), 1–66.

Mel'čuk, I. A. (1988). *Dependency syntax: Theory and Practice*. State University Press of New York, NY.

Mendes, S. and Amaro, R. (2009). Modeling adjectives in GL: accounting for all adjective classes. In *Proceedings of the Fifth International Conference on Generative Approaches to the Lexicon*, Pisa, Italy.

Meyers, A., Reeves, R., and Macleod, C. (2004a). NP-External Arguments: A Study of Argument Sharing in English. In T. Tanaka, A. Villavicencio, F. Bond, and A. Korhonen, editors, *Second ACL Workshop on Multiword Expressions: Integrating Processing*, pages 96–103, Barcelona, Spain. Association for Computational Linguistics.

Meyers, A., Reeves, R., Macleod, C., Szekely, R., Zielinska, V., Young, B., and Grishman, R. (2004b). The NomBank Project: An Interim Report. In A. Meyers, editor, *HLT-NAACL 2004 Workshop: Frontiers in Corpus Annotation*, pages 24–31, Boston, MA. Association for Computational Linguistics.

Miller, C. M., Rindflesch, T. C., Fiszman, M., Hristovski, D., Shin, D., Rosemblat, G., Zhang, H., and Strohl, K. P. (2012). A closed literature-based discovery technique finds a mechanistic link between hypogonadism and diminished sleep quality in aging men. *Sleep*, **35**(2), 279–285.

Miller, G. A. and Johnson-Laird, P. (1976). *Language and Perception*. Harvard University Press.

Miltsakaki, E., Prasad, R., Joshi, A., and Webber, B. (2004). The Penn Discourse TreeBank. In *Proceedings of Language Resources and Evaluation Conference*.

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill, New York, NY.

Mizuta, Y. and Collier, N. (2004). Zone identification in biology articles as a basis for information extraction. In *Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and its Applications (JNLPBA)*, pages 29–35.

Moilanen, K. and Pulman, S. (2007). Sentiment composition. In *Proceedings of Recent Advances in Natural Language Processing (RANLP 2007)*, pages 378–382.

Moore, J. D. and Pollack, M. E. (1992). A problem for rst: the need for multi-level discourse analysis. *Computational Linguistics*, **18**, 537–544.

Morante, R. and Daelemans, W. (2009). Learning the Scope of Hedge Cues in Biomedical Texts. In *Proceedings of the BioNLP 2009 Workshop*, pages 28–36.

Morante, R. and Sporleder, C., editors (2010). *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*. University of Antwerp, Uppsala, Sweden.

Morante, R., Liekens, A., and Daelemans, W. (2008). Learning the scope of negation in biomedical text. In *Proc Conference on Empirical Methods in Natural Language Processing*, pages 715–724.

Morante, R., van Asch, V., and Daelemans, W. (2010). Memory-based resolution of in-sentence scopes of hedge cues. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, pages 40–47, Uppsala, Sweden. Association for Computational Linguistics.

Mutalik, P. G., Deshpande, A., and Nadkarni, P. M. (2001). Use of general-purpose negation detection to augment concept indexing of medical documents: A quantitative study using the UMLS. *Journal of Americal Medical Informatics Association*, **8**(6), 598–609.

Nauze, F. D. (2008). *Modality in Typological Perspective*. Ph.D. thesis, University of Amsterdam.

Nawaz, R., Thompson, P., McNaught, J., and Ananiadou, S. (2010). Meta-knowledge annotation of bio-events. In *Proceedings of LREC 2010*, pages 2498–2505.

Nguyen, N., Kim, J.-D., and Tsujii, J. (2011). Overview of the Protein Coreference task in BioNLP Shared Task 2011. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task*, Portland, Oregon. Association for Computational Linguistics.

Nirenburg, S. and Raskin, V. (2004). *Ontological Semantics*. The MIT Press, Cambridge, MA.

Nivre, J., Hall, J., Nilsson, J., Eryiğit, G., and Marinov, S. (2006). Labeled pseudo-projective dependency parsing with support vector machines. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)*, pages 221–225, New York City. Association for Computational Linguistics.

Ohta, T., Pyysalo, S., and Tsujii, J. (2011). Overview of the Epigenetics and Post-translational Modifications (EPI) task of BioNLP Shared Task 2011. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task*, Portland, Oregon. Association for Computational Linguistics.

Özgür, A. and Radev, D. R. (2009). Detecting Speculations and their Scopes in Scientific Text. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 1398–1407.

Palmer, F. R. (1986). *Mood and modality*. Cambridge University Press, Cambridge, UK.

Palmer, F. R. (2001). *Mood and modality*. Cambridge University Press, Cambridge, UK.

Palmer, M., Gildea, D., and Kingsbury, P. (2005). The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, **31**(1), 71–106.

Pang, B. and Lee, L. (2008). *Sentiment Analysis and Opinion Mining*. Now Publishers Inc, Boston, MA.

Polanyi, L. and Scha, R. (1984). A syntactic approach to discourse semantics. In *Proceedings of the 10th Conference on Computational Linguistics*, COLING '84, pages 413–419. Association for Computational Linguistics.

Polanyi, L. and Zaenen, A. (2006). Contextual valence shifters. In J. G. Shanahan, Y. Qu, and J. Wiebe, editors, *Computing Attitude and Affect in Text: Theories and Applications*, volume 20, pages 1–10. Springer, 1nd edition.

Polanyi, L., van den Berg, M., and Ahn, D. (1988). A formal model of the structure of discourse. *Journal of Pragmatics*, **12**, 601–638.

Polanyi, L., Culy, C., van den Berg, M., Thione, G. L., and Ahn, D. (2004). Sentential structure and discourse parsing. In *Proceedings of the 2004 ACL Workshop on Discourse Annotation*, pages 80–87.

Power, R. (2007). Abstract verbs. In *Proceedings of the Eleventh European Workshop on Natural Language Generation*, pages 93–96.

Prabhakaran, V., Rambow, O., and Diab, M. (2010). Automatic Committed Belief Tagging. In *COLING 2010: Posters*, pages 1014–1022, Beijing, China. Coling 2010 Organizing Committee.

Pradhan, S., editor (2011). *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*. Association for Computational Linguistics, Portland, Oregon, USA.

Prasad, R., Dinesh, N., Lee, A., Joshi, A., and Webber, B. (2007). Attribution and its annotation in the Penn Discourse TreeBank. *Traitement Automatique des Langues, Special Issue on Computational Approaches to Document and Discourse*, **47**(2), 43–64.

Prasad, R., Dinesh, N., Lee, A., Miltsataki, E., Robaldo, L., Joshi, A., and Webber, B. (2008). The Penn Discourse TreeBank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.

Prasad, R., Joshi, A., and Webber, B. (2010). Realization of discourse relations by other means: Alternative lexicalizations. In *Coling 2010: Posters*, pages 1023–1031.

Prasad, R., McRoy, S., Frid, N., Joshi, A., and Yu, H. (2011). The biomedical discourse relation bank. *BMC Bioinformatics*, **12**, 188+.

Pustejovsky, J. (1995). *The Generative Lexicon*. MIT Press, Cambridge, MA.

Pustejovsky, J., Castaño, J. M., Ingria, R., Sauri, R., Gaizauskas, R. J., Setzer, A., Katz, G., and Radev, D. R. (2003). TimeML: Robust specification of event and temporal expressions in text. In *New Directions in Question Answering*, pages 28–34.

Pustejovsky, J., Ingria, R., Saurí, R., Castaño, J., Littman, J., Gaizauskas, R., Setzer, A., Katz, G., and Mani, I. (2005a). The specification language TimeML. In I. Mani, J. Pustejovsky, and R. Gaizauskas, editors, *The Language of Time: A Reader*, pages 545–558. Oxford University Press.

Pustejovsky, J., Knippen, B., Littman, J., and Saurí, R. (2005b). Temporal and event information in natural language text. *Language Resources and Evaluation*, **39**(2), 123–164.

Pyysalo, S., Ohta, T., and Tsujii, J. (2011a). Overview of the Entity Relations (REL) supporting task of BioNLP Shared Task 2011. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task*, Portland, Oregon. Association for Computational Linguistics.

Pyysalo, S., Ohta, T., Rak, R., Sullivan, D., Mao, C., Wang, C., Sobral, B., Tsujii, J., and Ananiadou, S. (2011b). Overview of the Infectious Diseases (ID) task of BioNLP Shared Task 2011. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task*, Portland, Oregon. Association for Computational Linguistics.

Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. Longman.

Ratcliff, R. and McKoon, G. (1978). Priming in item recognition: Evidence for the propositional structure of sentences. *Journal of Verbal Learning and Verbal Behavior*, **17**(4), 403–417.

Rindflesch, T. C. and Fiszman, M. (2003). The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. *Journal of Biomedical Informatics*, **36**(6), 462–477.

Rindflesch, T. C., Libbus, B., Hristovski, D., Aronson, A. R., and Kilicoglu, H. (2003). Semantic relations asserting the etiology of genetic diseases. In *Proceedings of AMIA Symposium*, pages 554–558.

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*, pages 45–50, Uppsala, Sweden. Association for Computational Linguistics.

Saurí, R. (2008). *A Factuality Profiler for Eventualities in Text*. Ph.D. thesis, Brandeis University.

Saurí, R. and Pustejovsky, J. (2009). FactBank: a corpus annotated with event factuality. *Language Resources and Evaluation*, **43**(3), 227–268.

Saurí, R., Verhagen, M., and Pustejovsky, J. (2006a). Annotating and Recognizing Event Modality in Text. In *Proceedings of 19th International FLAIRS Conference*.

Saurí, R., Littman, J., Gaizauskas, R., Setzer, A., and Pustejovsky, J. (2006b). *TimeML Annotation Guidelines*, 1.2.1 edition.

Scha, R. and Polanyi, L. (1988). An augmented context free grammar for discourse. In *Proceedings of the 12th Conference on Computational Linguistics - Volume 2*, COLING '88, pages 573–577. Association for Computational Linguistics.

Schuman, J. and Bergler, S. (2006). Postnominal prepositional phrase attachment in proteomics. In *Proc BioNLP Workshop on Linking Natural Language Processing and Biology*, pages 82–89.

Seretan, V. and Cristea, D. (2002). The use of referential constrains in structuring discourse. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas, Spain.

Shatkay, H., Pan, F., Rzhetsky, A., and Wilbur, W. J. (2008). Multi-dimensional classification of biomedical text. *Bioinformatics*, **24**(18), 2086–2093.

Soricut, R. and Marcu, D. (2003). Sentence Level Discourse Parsing using Syntactic and Lexical Information. In *Proceedings of HLT-NAACL*, pages 149–156.

Sowa, J. (1984). *Conceptual Structures: Information Processing in Mind and Machine*. Addison-Wesley.

Sporleder, C. and Lapata, M. (2005). Discourse chunking and its application to sentence compression. In *Proceedings of the HLT/EMNLP*, pages 257–264.

Sporleder, C. and Lascarides, A. (2005). Exploiting linguistic cues to classify rhetorical relations. In *Proceedings of Recent Advances in Natural Language Processing (RANLP)*, pages 257–264.

Sporleder, C. and Lascarides, A. (2008). Using automatically labelled examples to classify rhetorical relations: A critical assessment. *Natural Language Engineering*, **14**(3), 369–416.

Stede, M. (2008). RST revisited: Disentangling nuclearity. In C. Fabricius-Hansen and W. Ramm, editors, *Subordination versus coordination in sentence and text: A cross-linguistic perspective*, pages 33–58. John Benjamins Publishing Company.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Szarvas, G. (2008). Hedge classification in biomedical texts with a weakly supervised selection of keywords. In *Proceedings of the 46th Meeting of the Association for Computational Linguistics*, pages 281–289.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, **37**(2), 267–307.

Tang, B., Wang, X., Wang, X., Yuan, B., and Fan, S. (2010). A Cascade Method for Detecting Hedges and their Scope in Natural Language Text. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, pages 13–17, Uppsala, Sweden. Association for Computational Linguistics.

Teufel, S., Carletta, J., and Moens, M. (1999). An annotation scheme for discourse-level argumentation in research articles. In *Proceedings of EACL*, pages 110–117.

Teufel, S., Siddharthan, A., and Batchelor, C. R. (2009). Towards Domain-Independent Argumentative Zoning: Evidence from Chemistry and Computational Linguistics. In *Proceedings of EMNLP*, pages 1493–1502.

The PDTB Research Group (2008). The Penn Discourse Treebank 2.0. Annotation Manual. Technical Report IRCS Technical Report IRCS-08-01, University of Pennsylvania, Institute for Research in Cognitive Science.

Thompson, P., Venturi, G., McNaught, J., Montemagni, S., and Ananiadou, S. (2008). Categorising modality in biomedical texts. In *Proceedings of LREC 2008 Workshop on Building and Evaluating Resources for Biomedical Text Mining*.

Thompson, P., Nawaz, R., McNaught, J., and Ananiadou, S. (2011). Enriching a biomedical event corpus with meta-knowledge annotation. *BMC Bioinformatics*, **12**, 393.

Tsujii, J., editor (2009). *Proceedings of the BioNLP 2009 Workshop Companion Volume for Shared Task*. Association for Computational Linguistics, Boulder, Colorado.

Uzuner, Ö. (2009). Recognizing obesity and comorbidities in sparse data. *Journal of American Medical Informatics Association*, **16**, 561–570.

van Dijk, T. A. and Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Vincze, V., Szarvas, G., Farkas, R., Mora, G., and Csirik, J. (2008). The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, **9 Suppl 11**, S9.

Watson, J. D. and Crick, F. H. C. (1953). Molecular Structure of Nucleic Acids. *Nature*, **171**, 737–738.

Webber, B. (2006). Accounting for discourse relations: Constituency and dependency. In M. Butt, M. Dalrymple, and T. King, editors, *Intelligent Linguistic Architectures*, pages 339–360. Stanford CSLI Publications.

Wellner, B. (2009). *Sequence Models and Ranking Methods for Discourse Parsing*. Ph.D. thesis, Brandeis University.

Wellner, B., Pustejovsky, J., Havasi, C., Rumshisky, A., and Saurí, R. (2006). Classification of discourse coherence relations: An exploratory study using multiple knowledge sources. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*, pages 117–125.

White, E. (2010). *Pattern-Based Extraction of Argumentation from the Scientific Literature*. Ph.D. thesis, University of Colorado, Boulder.

Wiebe, J., Wilson, T., Bruce, R., Bell, M., and Martin, M. (2004). Learning subjective language. *Computational Linguistics*, **30**(3), 277–308.

Wiebe, J., Wilson, T., and Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, **39**(2), 165–210.

Wiegand, M., Balahur, A., Roth, B., Klakow, D., and Montoyo, A. (2010). A survey on the role of negation in sentiment analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 60–68, Uppsala, Sweden.

Wilbur, W. J., Rzhetsky, A., and Shatkay, H. (2006). New directions in biomedical text annotations: definitions, guidelines and corpus construction. *BMC Bioinformatics*, **7**, 356.

Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In *Proceedings of the Human Language Technologies Conference/Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP-2005)*, pages 347–354.

Wolf, F. and Gibson, E. (2005). Representing Discourse Coherence: A Corpus-Based Study. *Computational Linguistics*, **31**(2), 249–287.

Wolff, P. (2003). Direct causation in the linguistic coding and individuation of causal events. *Cognition*, **88**(1), 1–48.