Unsupervised extraction of semantic relations using discourse information

Juliette Conrath

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Présentée et soutenue le *14/12/2015* **par :**

JULIETTE CONRATH

Unsupervised extraction of semantic relations using discourse information

Directeurs de Thèse :

Nicholas ASHER
Directeur de Recherche
IRIT-CNRS

Philippe MULLER
Maître de Conférences
Université Paul Sabatier

Stergos AFANTENOS
Maître de Conférences
Université Paul Sabatier

Rapporteurs :

Johan BOS
Professeur d'Université
Rijksuniversiteit Groningen

James PUSTEJOVSKY
Professeur d'Université
Brandeis University

Examinateur :

Pascal DENIS
Chargé de Recherche
INRIA Lille

École doctorale et spécialité :

MITT ; Domaine STIC : Intelligence Artificielle

Unité de Recherche :

Institut de Recherche en Informatique de Toulouse (IRIT - UMR 5505)

Abstract

Natural language understanding often relies on common-sense reasoning, for which knowledge about semantic relations, especially between verbal predicates, may be required. This thesis addresses the challenge of using a distibutional method to automatically extract the necessary semantic information for common-sense inference. Typical associations between pairs of predicates and a targeted set of semantic relations (causal, temporal, similarity, opposition, part/whole) are extracted from large corpora, by exploiting the presence of discourse connectives which typically signal these semantic relations.

In order to appraise these associations, we provide several significance measures inspired from the literature as well as a novel measure specifically designed to evaluate the strength of the link between the two predicates and the relation. The relevance of these measures is evaluated by computing their correlations with human judgments, based on a sample of verb pairs annotated in context. The application of this methodology to French and English corpora leads to the construction of a freely available resource, Lecsie (Linked Events Collection for Semantic Information Extraction), which consists of triples: pairs of event predicates associated with a relation; each triple is assigned significance scores based on our measures.

From this resource, vector-based representations of pairs of predicates can be induced and used as lexical semantic features to build models for external applications. We assess the potential of these representations for several applications. Regarding discourse analysis, the tasks of predicting attachment of discourse units, as well as predicting the specific discourse relation linking them, are investigated. Using only features from our resource, we obtain significant improvements for both tasks in comparison to several baselines, including ones using other representations of the pairs of predicates. We also propose to define optimal sets of connectives better suited for large corpus applications by performing a dimension reduction in the space of the connectives, instead of using manually composed groups of connectives corresponding to predefined relations. Another promising application pursued in this thesis concerns relations between semantic frames (e.g. FrameNet): the resource can be used to enrich this sparse structure by providing candidate relations between verbal frames, based on associations between their verbs.

These diverse applications aim to demonstrate the promising contributions provided by our approach, namely allowing the unsupervised extraction of typed semantic relations.

Keywords: computational linguistics, distributional semantics, lexical semantics, discourse parsing.

Résumé

La compréhension du langage naturel repose souvent sur des raisonnements de sens commun, pour lesquels la connaissance de relations sémantiques, en particulier entre prédicats verbaux, peut être nécessaire. Cette thèse porte sur la problématique de l'utilisation d'une méthode distributionnelle pour extraire automatiquement les informations sémantiques nécessaires à ces inférences de sens commun. Des associations typiques entre des paires de prédicats et un ensemble de relations sémantiques (causales, temporelles, de similarité, d'opposition, partie/tout) sont extraites de grands corpus, par l'exploitation de la présence de connecteurs du discours signalant typiquement ces relations.

Afin d'apprécier ces associations, nous proposons plusieurs mesures de signifiance inspirées de la littérature ainsi qu'une mesure novatrice conçue spécifiquement pour évaluer la force du lien entre les deux prédicats et la relation. La pertinence de ces mesures est évaluée par le calcul de leur corrélation avec des jugements humains, obtenus par l'annotation d'un échantillon de paires de verbes en contexte discursif. L'application de cette méthodologie sur des corpus de langue française et anglaise permet la construction d'une ressource disponible librement, Lecsie (Linked Events Collection for Semantic Information Extraction). Celle-ci est constituée de triplets: des paires de prédicats associés à une relation; à chaque triplet correspondent des scores de signifiance obtenus par nos mesures.

Cette ressource permet de dériver des représentations vectorielles de paires de prédicats qui peuvent être utilisées comme traits lexico-sémantiques pour la construction de modèles pour des applications externes. Nous évaluons le potentiel de ces représentations pour plusieurs applications. Concernant l'analyse du discours, les tâches de la prédiction d'attachement entre unités du discours, ainsi que la prédiction des relations discursives spécifiques les reliant, sont explorées. En utilisant uniquement les traits provenant de notre ressource, nous obtenons des améliorations significatives pour les deux tâches, par rapport à plusieurs bases de référence, notamment des modèles utilisant d'autres types de représentations lexico-sémantiques. Nous proposons également de définir des ensembles optimaux de connecteurs mieux adaptés à des applications sur de grands corpus, en opérant une réduction de dimension dans l'espace des connecteurs, au lieu d'utiliser des groupes de connecteurs composés manuellement et correspondant à des relations prédéfinies. Une autre application prometteuse explorée dans cette thèse concerne les relations entre cadres sémantiques (semantic frames, e.g. FrameNet): la ressource peut être utilisée pour enrichir cette structure par des relations potentielles entre frames verbaux à partir des associations entre leurs verbes.

Ces applications diverses démontrent les contributions prometteuses amenées par notre approche permettant l'extraction non supervisée de relations sémantiques.

Mots-clés : traitement automatique du langage naturel, sémantique distributionnelle, sémantique lexicale, analyse du discours.

Acknowledgements

This thesis is the result of a three-year adventure which has been one of the most life-changing and rewarding challenges I have experienced.

I am deeply indebted to my three supervisors, Nicholas Asher, Philippe Muller and Stergos Afantenos, for their guidance and constant support, even when I chose to do the remaining work from abroad. Each of them brought the subject of this thesis into a different light and I am very grateful for the broad perspective on natural language processing and lexical semantics they allowed me to acquire.

I also want to thank Johan Bos and James Pustejovksi for taking the time to review my thesis and provide insightful feedback. I am also grateful to Pascal Denis for participating in my defense committee, for his interest in my project and his enlightening questions and suggestions.

Special thanks go to Chloé Braud for sharing her ideas and insights on discourse analysis, but also for making our first international conference such a pleasant experience, for being a friendly face in the crowd during those stressful times, and for sharing the adventures of writing a thesis. I sincerely hope I was as comforting to her as she was to me.

I wish to express my gratitude to Mathieu Morey for the insightful discussions we had as well as for his guidance and advice in finishing my thesis, and the steps that followed. Tim van de Cruys also deserves special mention for his help and suggestions, and for sharing his knowledge on distributional semantics. I also thank Eric Kow for the software that he developed for our team and which was of great help for my work.

I also wish to thank the entire Melodi team, and in particular the team leader Nathalie Aussenac-Gilles, for making our work environment so friendly and motivating. And of course my fellow "thésards" in the team: Julien Corman, Antoine Venant, Jean-Philippe Fauconnier, Antoine Bride, Jeremy Perret. It was a pleasure to share the struggles and excitement of entering the world of research with such an enthusiastic group. I will also not forget the "older" students who shaped such a friendly environment, to work but also play, and who warmly welcomed me and every newcomer in their joyful gang: first and foremost, my office-mates Anaïs Cadilhac, Nadine Guiraud and Seifeddine Kramdi, and of course also Anke Brock, Nicolas Schmidt, Chloé Cimpello, Camille Pradel, Damien Bigot, Simon Vernhes, Ezgi Iraz Su along with all the other students at IRIT who I had the pleasure to meet.

I also want to acknowledge my new colleagues at Textkernel for being so understanding and giving me space to finish writing when I first started working at the company. I have to say working on fresh new projects with brand new colleagues offered me a welcome change of pace and inspired to finish what I had started with renewed motivation.

Finally, and perhaps most importantly, I thank my parents, my brother and my sisters, and every member of my large family for caring for me, supporting me in the most difficult times, and giving me the thirst for knowledge and the open mind that led me to pursue a career in research. It goes without saying that without them, this accomplishment would not have been possible. I dedicate this thesis to my dear Opa, who always believed in me and gave me the will to always be at my best and make him proud. He was and always will be my best source of inspiration.

Of course, last but not least, I am forever indebted to my partner, Pierre, who was my strongest support in this endeavor and who deserves much more credit than he knows. He always knew when to push me a little further and when to remind me to take a breath, he patiently reassured me through my doubts and fears and was there to share the small victories along the way. His passion for research allowed me to never lose sight of the true purpose of working through a thesis, and he always offered an open ear when I did not know which direction to follow. I must also acknowledge his Latex wizardy, which saved me numerous times and without which this manuscript would not look half as good as it does.

As a final note, I thank the French National Research Agency (ANR) and the Polymnie project for giving the opportunity to pursue my thesis and providing funding.

Contents

Introduction

Systems of natural language understanding aim to model the processes which people perform when interpreting a natural language utterance, that is, understanding the information it conveys about the world. Particulary, natural language semantics aim to assign meanings to utterances, by combining the meanings of words and phrases they are composed of.

Formal approaches to semantic analysis, inspired by work in logic and philosophy of language, analyze natural language using formal methods, in particular the languages and concepts of first order and higher order logic. They focus on constructing a formal representation of the meaning of a sentence, and computing its truth conditions, that is, what the world would be like if the semantic content expressed by this sentence (its proposition) were true.

The world is usually represented as a mathematical abstraction, a model, to which linguistic expressions relate. For example, proper nouns refer to objects, and therefore relate to entities in the model, while verbs are functions from entities to truth values. The sentence *John sleeps* can be analyzed by applying the function *sleeps* to the entity *John*, and the resulting proposition will be true only in a model of the world where John indeed sleeps.

Then, through principled inference rules, the entailments of the proposition can be derived. Deriving entailments consists in determining what other propositions have to be true if this one is.

Such inferences can be drawn from the analysis of logical features expressed by function words, *e.g.*, logical connectives (such as *and*, *or*), or quantifiers (such as *some*, *all*). In Example 1, if the proposition expressed in sentence (a) is true, then the proposition expressed in sentence (b) is necessarily true as well: we can say that (a) entails (b). This inference can be drawn no matter what the words *Tina*, *tall* and *thin* mean, it only relies on the logical connective *and*.

(1) (a) Tina is tall and thin.

(b) Tina is thin.

Similarly, in Example 2, the inference that (a) entails (b) only relies on the analysis of the quantifiers *all* and *some*.

(2) (a) All men are mortal.
(b) Some men are mortal.

Such structurally-based inferences allow simple syllogistic reasoning, as in Example 3, where the conjunction of (a) and (b) entails (c).

(3) (a) Socrates is a man.
(b) All men are mortal.
(c) Socrates is mortal.

The role of lexical information in natural language interpretation

As we have seen in the previous examples, the range of inferences that can be drawn from semantic representations alone, without any additional knowledge, is very limited. Indeed, the vast majority of intuitively plausible inferences require additional background knowledge. In particular, some inferences crucially rely on knowledge about the meaning of individual lexical items.

For instance, in Example 4, the knowledge that *bachelor* and *unmarried man* have the same meaning allows to infer that (a) and (b) are equivalent: (a) entails (b) and (b) entails (a).

(4) (a) John is a bachelor.
(b) John is an unmarried man.

Such knowledge can be captured by representing the word *bachelor* as the composition of a set of primitives: the meaning of *man* combined with the negation of the meaning of *married*. This approach to the representation of word meaning is known as the decompositional approach, where a lexical item is represented as a set of atomic meaning descriptions (Katz and Fodor, 1963; Jackendoff, 1983). The relations between lexical items can then be indirectly induced by their semantics, and lexically-based inferences can be drawn. However, agreement on a basic inventory of primitives has been elusive (Winograd, 1978).

Another approach to draw such lexically-based inferences is to reason directly with relations between lexical items. This can be done easily based on theories of lexical semantics which represent the meaning of words by their relations with other words (Lyons, 1968).

In Example 5, the inference that (a) entails (b) requires to know the lexical relation between the nouns *cat* and *animal*: *cat* is a **hyponym** of *animal*.

(5) (a) Olivette is a cat.
(b) Olivette is an animal.

In Example 6, the knowledge that the adjectives *dead* and *alive* are **antonyms** is necessary to infer that (a) and (b) are equivalent, that is, (a) entails (b) and (b) entails (a).

(6) (a) John is dead.
(b) John is not alive.

To support such inferences, natural language understanding systems need to have access to lexical knowledge bases which contain such information. The amount of knowledge which may be relevant for inference is huge, and so hand-crafting comprehensive knowledge bases is a very expensive and cumbersome task. Some knowledge bases contain certain types of lexical-semantic knowledge, such as WordNet (Fellbaum, 1998). But these resources typically lack coverage.

To alleviate this problem, distributional semantic approaches aim to characterize the meaning of words by their usage in context (Firth, 1957), which can be automatically extracted from large corpora. However it is difficult to determine a precise connection between word meaning and word usage, and to articulate a proper notion of entailment based on distributional representations beyond the vague notion of relatedness, which can be computed via similarity metrics applied to these representations.

Inferences based on relations between verbs

So far we have considered entailments based on classical lexical relations between adjectives or nouns. Predicates, which are often expressed by verbs, function as the organizational core of sentences, by expressing relations over the arguments. Understanding the meaning of verbs is therefore key to understanding the general meaning of sentences.

Similarly to nouns and adjectives as we have seen before, lexical relations between verbs can also serve as the basis of inferences. In Example 7, inferring that (a) entails (b) requires knowing that *kill* is a **troponym** (the equivalent of a hypernym for verbs) of *murder*, and inferring that (b) entails (c) requires knowing that there is a causal relation between *killing* and *dying*.

(7) (a) John murdered Paul.
  (b) John killed Paul.
  (c) Paul died.

However, when considering contexts beyond the sentence level, more complex inferences need to be considered. A discourse is a coherent sequence of sentences, and therefore sentences or clauses composing a discourse must be related to each other in some way. Inferring these relations between clauses often relies at least partially on the relations between the main predicates of each clause.

In Example 8, the intuitive inference of a causal relation between the proposition expressed in (a) and that expressed in (b) is difficult to capture formally.

(8) (a) John stabbed Paul.
  (b) Paul died.

Contrary to the relation between the verbs *kill* and *die* in Example 7, *stabbing* does not necessarily cause *dying*. However, intuitively we understand in this discourse context that Paul died because he was stabbed by John. The association of these two verbs clearly suggests a possible causal link, and their presence in two adjacent discourse clauses is responsible to a large degree of the inference of a causal relation between these clauses.

Such inferences based on relations between verbs are more complex and not well studied. They are however crucial in many natural language processing tasks.

Applications in Natural Language Processing

The task of Recognizing Textual Entailment (RTE), for instance, is concerned with inferring the meaning of a piece of text from that of another one. Textual entailment between two sentences can be conveyed by using predicates linked by an entailment relation to describe the event in each sentence. For instance, in Example 9, an RTE system would need access to knowledge about the lexical relation between the predicates *assassinate* and *die* in order to recognize the textual entailment relation between sentence (a) and sentence (b).

(9) (a) President John F. Kennedy was assassinated in November 1963.
(b) President John F. Kennedy died.

Similarly, presupposition resolution relies on lexical relations. In Example 10, sentence (a) presupposes sentence (b), and this presupposition is implicitly encoded through the relation between the predicates *wake up* and *sleep*.

(10) (a) Mark woke up at 10.
(b) Mark was sleeping.

Another task in which lexical relations are relevant is discourse parsing, which aims at inferring the structure of a discourse. It requires to infer which clauses are related, and by which discourse relations they are related. In some cases, individual lexical items or multiword expressions which carry information about discourse relations in their meaning are used to convey these relations. In Example 11, the expression *as a result* explicitly conveys a causal relation between the two clauses.

(11) (a) Peter stayed up very late last night.
(b) As a result, he overslept this morning.

These expressions are referred to as discourse connectives, and are easily identifiable cues for the prediction of discourse relations. However, discourse relations can be conveyed by other means, and specifically lexically related items. In Example 12, the causal relation is still inferrable in the absence of the discourse connective, through the association of the predicates *stay up* and *oversleep*.

(12) (a) Peter stayed up very late last night.
(b) He overslept this morning.

These tasks, while interesting by themselves, are also able to provide very valuable information for more practical applications. For instance, in Question Answering, the aim is to build a system which is able to retrieve answers to specific questions when provided one or several relevant texts. Based on the examples given previously, such a task can be resolved with the help of textual entailment recognition, presupposition resolution, or discourse relation identification, as in Examples 13, 14, and 15, where the answer to question (b) can be derived from sentence (a).

(13) (a) President John F. Kennedy was assassinated in November 1963.

(b) When did Kennedy die ?

(14) (a) Mark slept until 10.
(b) At what time did Mark wake up ?

(15) (a) Peter stayed up very late last night.
(b) Why did Peter oversleep ?

Other tasks which can benefit from these systems are for instance Machine Translation, where the aim is to produce reliable translations of texts from one language to another, or Automatic Summarization, where the aim is to reduce a text in order to create a coherent summary that retains the most important points of the original text and organizes them in more general sentences.

Purposes and approach

It is clear that capturing lexical meaning conveyed by pairs of predicates is essential for event-based semantic processing tasks. The main purpose of this thesis is to design a method to automatically extract typical semantic relations between predicates.

We propose to achieve this by using insights from distributional and lexical semantics. Our method is distributional in the sense that pairs of predicates are represented by their usage in context, and so these representations are obtained in a data-driven way. It is inspired by relational approaches to lexical meaning, in that we aim to characterize predicates by their semantic relations to other predicates. By targeting a specific set of relations, we avoid the difficulties that beset distributional semantics with formulating entailments in a proper way.

More specifically, our approach is based on the co-occurrence between pairs of predicates and discourse connectives. Such connectives, which explicitly express discourse relations, can indeed be thought to be redundant with other information expressed in the related clauses, and in particular the association between pairs of predicates. By applying lexico-syntactic patterns, we automatically extract pairs of predicates linked by these connectives, and identify the relation which may be conveyed by their association.

The relations we extract, although inspired by discourse coherence, can also be seen as representative of the semantic association between predicates. We can therefore explore the regularities in our extracted data to draw data-driven insights about the usage of these semantic relations.

The information obtained with our method can also be used to evaluate the significance of association between pairs of predicates and semantic relation. In order to do so we define measures of association strength of triples composed of two predicates and a specific relation. We collect the resulting information in a knowledge base named *Lecsie*, for *Linked Event Collection for Semantic Information Extraction*[1].

We can then use *Lecsie* to induce a vector-based representation of pairs of predicates in the space of our set of relations, where the vector components are the scores of association with each relation. Such a semantically-informed representation of verb pairs can then be useful in external applications such as the recognition of textual entailment or discourse parsing.

In NLP, it is customary to distinguish between intrinsic evaluations, testing a system in itself, and extrinsic evaluations, measuring its performance in some task or application (Jones and Galliers, 1995). We first propose an intrinsic evaluation of the knowledge contained in *Lecsie* by comparing its association scores to human intuition, by way of manual annotations. Our extrinsic evaluation is done by assessing the potential of our vector-based representations of predicate pairs in the external application of discourse parsing.

We also propose to evaluate the potential of our resource for the automatic expansion of a manually-constructed lexical knowledge-base, namely FrameNet, by using the strength of association between verbs found in frames and our relations to induce new relations between frames.

Overview of the Thesis

In Chapter 1, we start by presenting some aspects of lexical semantics and the notion of lexical semantic relations. We also present some existing manually-constructed lexical resources and their uses. We then review existing approaches to the automatic extraction of semantic relations.

Chapter 2 is concerned with theories of discourse structure and the underlying motivations of our approach based on discourse cues. We also review methods for automatic discourse parsing found in the literature, showing the necessity of lexical information of the type we propose, to improve existing methods.

In Chapter 3, we detail our method for automatically extracting related pairs of predicates, and the reasoning behind it.

[1] The resource is freely available as a sqlite database at `https://dl.dropboxusercontent.com/u/78938139/LECSIE.zip`, along with a script for simple extractions, as well as the data and software used in our evaluations.

In Chapter 4, we explore the resulting semantic information, particularly in terms of the impact of negations and shared arguments on the expression of semantic relations.

Chapter 5 presents our significance measures and their manual evaluation.

In Chapter 6, we evaluate our resource extrinsically on the task of discourse parsing, and specifically on two sub-tasks: predicting the attachment of discourse clauses, and predicting the discourse relations linking them.

Finally, in Chapter 7, we present our preliminary results on other potential applications of our resource: automatically expanding FrameNet and automatically inducing optimal sets of connectives, instead of manually-defined ones, for better adapted, data-driven representations of pairs of predicates.

Chapter 8 summarizes the contributions of the presented work, and discusses promising avenues of research.

Chapter 1

Background: Lexical semantic relations

Contents

As we outlined in the introduction, lexical knowledge, and in particular knowledge of lexical semantic relations, is crucial for the interpretation of language. In this chapter, we aim to give a more detailed overview of this question and how it has been addressed in the literature.

Section 1.1 focuses on lexical semantic relations and their role in the representation of lexical meaning and the construction of complex meaning. It also presents existing manually-constructed lexical semantic resources, and their limitations. In Section 1.2, we give an overview of existing approaches to the automatic extraction of these relations from large corpora.

1.1 Lexical semantic relations in natural language interpretation

1.1.1 Symbolic models of natural language

According to Montague (1970), "*the basic aim of semantics is to characterize the notion of a true sentence (under a given interpretation) and of entailment.*"

Most formal approaches to the semantics of natural language are truth-conditional and model-theoretic: the meaning of a sentence is taken to be a proposition which will be true or false relative to some model of the world. The world is represented as a mathematical abstraction made up of sets and the semantic theory relates linguistic expressions to this model. Every type of linguistic expression must relate to something in the model. For example, proper nouns refer to objects, so they will relate to entities in the model.

Through predicate logic or dynamic semantic formalisms, and principled meaning construction methods, these theories aim to translate natural language expressions into a logical language with explicitly defined syntax and semantics.

Logical representation

Natural languages have a finite (if evolving) vocabulary, out of which the acceptable and truth evaluable sentences are constructed via syntactic rules of composition. The lexical meanings of words are combined to build the meanings of larger phrases, based on the compositionality principle according to which *the denotation of an expression is determined by the denotations of its parts and the ways they combine with each other*.

A sentence meaning can be represented by a logical representation, sometimes called its **logical form**, which is a formal representation of its logical structure derived from the corresponding surface form, and is taken to capture its truth conditions. Truth conditions correspond to what the world would be like if the proposition expressed were true.

For example, the sentence "*Shakespeare wrote a tragedy.*" can be assigned the following logical representation: $\exists t(\text{tragedy}(t) \wedge \text{write}(s,t))$. One assumption of formal semantics is that sentences which have the same or closely related meaning should be represented in the same way. Thus, the sentence "*A tragedy was written by Shakespeare.*" would be assigned the same representation as above.

Entailment criterion

In formal logic, a proposition P is said to entail another proposition Q when the truth of Q is a logically necessary consequence of the truth of P. Based on this principle, entailments between sentences can be derived from their truth values, to model natural language inferences. The meaning of a sentence is thus defined in terms of its potential entailments with respect to other sentences.

In this way, the truth value assignment mechanisms can be validated if they satisfy the entailment criterion. Indeed, semantic theories aim to capture the inferential structure of natural language: every inference that a competent speaker would regard as valid should be derivable in the theory.

Entailment is perhaps the most important of the semantic intuitions to capture in a semantic theory, since it is the basis of the inferences we make in language comprehension, and many other semantic notions reduce to entailment. For example, two propositions can be synonymous, as in Examples 1.1 and 1.2, but the notion of synonymy reduces to the notion of identity of entailments: if the proposition expressed by 1.1 is true then the proposition in 1.2 is true, and if the proposition in 1.2 is true then the proposition in 1.1 is true.

(1.1) *John is a bachelor.*

(1.2) *John is an unmarried man.*

Some inferences depend for their validity on the syntactic structure and on the logical properties of function words (*e.g.*, quantifiers or logical connectives), like the inference from "*Every man is mortal and Socrates is a man*" to "*Socrates is mortal*". Other inferences depend on properties of non-logical words that are usually regarded as semantic, like the inference from "*Kim is pregnant*" to "*Kim is not a man*".

Theories of formal semantics mainly concentrate on linguistic means of expressing logical properties of a natural expression. The meaning of the non-logical predicates (*e.g.*, *Shakespeare*, *write*, *tragedy* in our initial example) expressed by content words, as opposed to function words (*e.g.*, *and*, *if*, *a*), is irrelevant in the context of formal semantics. Thus, the sentences "*a cat eats a rat*" and "*a mat facilitates a nap*", which have the same syntactic structure, will be assigned the logical representations equal to $\exists x, y(P(x) \wedge Q(y) \wedge R(x, y))$[1]. Distinguishing between these sentences is then a matter of an interpretation function mapping P, Q and R to different sets. Formal semantics constrains this function on constants

[1]Example taken from Vieu (2009).

and predicates only as far as their logical type is concerned. To account for the semantic differences related to the meaning of words and their combinations, it is necessary to turn to lexical semantics.

Theories of **lexical semantics** aim to get at the meaning of individual words, and therefore serve as the basis for recognizing and modelling entailments due to individual lexical items. Theories of **compositional semantics** are concerned with the way these individual meanings are combined to form the meanings of phrases and sentences, and can therefore account for entailments due to the semantics of a grammatical construction, given that the meanings of words has been fixed from a particular method of composition.

1.1.1.1 Lexical semantics

Theories of lexical semantics propose various ways of representing the required lexical semantic knowledge. We will focus on three types of approaches: those based on **meaning postulates**, the **decompositional** approaches, and the **relational** approaches.

Meaning postulates

In traditional formal semantics, lexical meaning is defined by way of meaning postulates, *i.e.*, stipulations on the relations among the extensions of lexical items (Carnap, 1947). They consist of entailments where the antecedent is an open lexical proposition. Example 1.3 stipulates that any individual that is in the extension of *bachelor* is also in the extension of *man*, and Example 1.4 stipulates that any individual that is in the extension of *bachelor* is not in the extension of *married*.

(1.3) $\forall x[\text{bachelor}(x) \rightarrow \text{man}(x)]$

(1.4) $\forall x[\text{bachelor}(x) \rightarrow \neg\text{married}(x)]$

Meaning postulates can also express bidirectional entailments as in Example 1.5 where the biconditional expresses an equivalence between a word and its decomposition.

(1.5) $\forall x[\text{bachelor}(x) \leftrightarrow (\text{man}(x) \& \neg\text{married}(x))]$

Informally, one might say that bidirectional meaning postulates provide definitions, while monodirectional ones represent single aspects of word meaning in the form of relations to other semantic elements.

The theory of meaning postulates posits that the meaning of a lexical item is the set of meaning postulates necessary and sufficient to build a model adequate for the language in which the expression occurs: a model in which an arbitrary individual can be a *bachelor* without being *unmarried* is inadequate. Thus, by proposing analogous constraints for each of the lexical items, lexical semantics can be reduced to a list of sets of meaning postulates.

However, while words like *bachelor* can easily be represented by the set of links they hold with the rest of the lexicon, other words like *bird* or *water* are more difficult to define exhaustively, and it has been argued that human interpretation does not in fact rely on such exhaustive representations (Kamp and Partee, 1995; Fodor and Lepore, 1996).

Another issue with meaning postulates, is one of coverage: to be applicable in computational semantics, such an approach would need an extremely large list of lexical items linked with their meaning postulates, even if those could be made adequate for all types of words.

Decompositional approaches

Decompositional theories are based on the idea that words encode complex concepts that may be decomposed into a restricted set of simpler notions, or semantic components. These theories can thus be viewed as a restriction on meaning postulates. For instance, the meaning of the word *bachelor* could be described by using the primitives *man* and *unmarried*, as shown in Example 1.6.

(1.6) *bachelor*:$\lambda x[\text{man}(x)\&\text{unmarried}(x)]$

Katz and Fodor (1963) combined componential analysis with a mentalistic conception of word meaning and developed a method for the description of lexical phenomena in the context of a formal grammar. In their theory, word meanings are structured entities whose representations are called **semantic markers**. A semantic marker is a tree with labeled nodes whose structure reproduces the structure of the represented meaning, and whose labels identify the word's conceptual components. For example, Figure 1.1 illustrates the sense of *chase*.

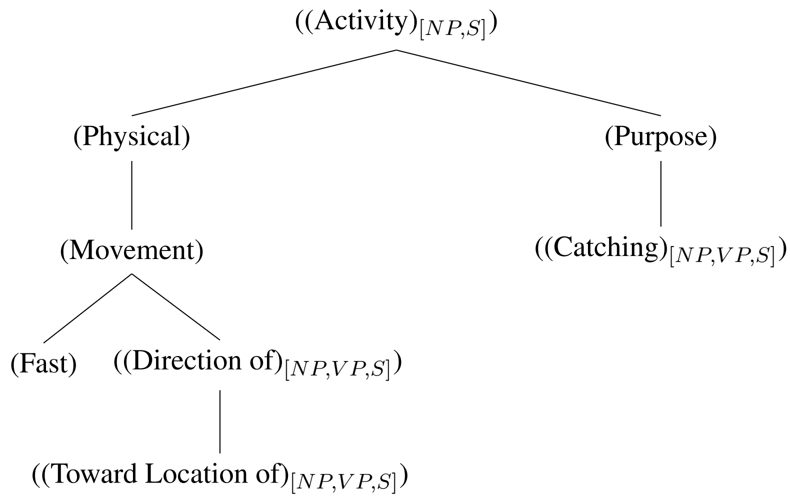


Figure 1.1: Componential analysis of *chase* (Katz and Fodor, 1963).

Katz claimed that this approach was superior to the kind of semantic analysis that could be provided via meaning postulates. For example, the validation of conditionals such as $\forall x, y(\text{chase}(x, y) \rightarrow \text{follow}(x, y))$ could be reduced to checking whether the semantic marker of *follow* was a subtree of the semantic marker of *chase*. However, this theory has no account for the way lexical expressions contribute to the truth conditions of sentences, because these primitives still need an interpretation (Lewis, 1972).

Following the idea of decompositional meaning, other frameworks aimed to formally implement the notion of entailment. Although they all shared the view that word meaning is composed of primitive semantic predicates, the aims and scope of decomposition diverge considerably accross decompositional theories, and there is no agreement on a basic inventory of atomic meaning descriptions (Winograd, 1978).

Dowty (1979) focused on explaining systematic meaning relations between diathesis alternations (*e.g.*, inchoative and causative readings of *open* or *close* using primitives like CAUSE and BECOME), and on the ability of these semantic relations to predict the range of possible constructions for different types of predicates. In his framework, decomposition is restricted to modelling the grammaticalized categories of lexical meaning, leaving the core lexical semantics of verbs largely unanalysed.

Jackendoff (1983) proposed to capture the lexical meaning of predicates in terms of a set of perceptually grounded primitives (*e.g.*, EVENT, STATE, THING, PATH, PLACE, PROPERTY, AMOUNT) which are assumed to be innate, cross-modal and universal categories of human cognition. Additionally, he used syntactic tags to represent the way a word interacts with the grammatical environment where it is used. The resulting **Conceptual Structures** are representations of those aspects of the meaning of words and phrases which are relevant to syntax. The decomposition can be rather fine-grained, as shown in Example 1.7, where the verb *drink* is represented as *cause a liquid to go into one's mouth*.

(1.7) *drink*:

$[_{\text{Event}}\text{CAUSE}([_{\text{Thing}}]_i,$

$[_{\text{Event}}\text{GO}([_{\text{Thing}}\text{LIQUID}]_j,$

$[_{\text{Path}}\text{TO}([_{\text{Place}}\text{IN}([_{\text{Thing}}\text{MOUTH OF}([_{\text{Thing}}]_i)])])])])]$

Inference rules can be defined which will be triggered by such structures:

(1.8) At the termination of $[_{\text{Event}}\text{GO}(X, [_{\text{Path}}\text{TO}(Y)])]$
it is the case that $[_{\text{State}}\text{BE}(X, [_{\text{Path}}\text{AT}(Y)])]$

This theory is thus adequate to be used in a computational framework, in the sense that it allows for inferences to be made.

Decompositional approaches aim at capturing as many generalities as possible about linguistic and semantic relationships among words. Their main drawback, on top of the coverage issue mentioned above, lies in the difficulty of choosing relevant primitives, and the lack of interpretation of these primitives.

Pustejovsky (1995) proposed to describe the generative capacity of lexical meaning from an opposite viewpoint: instead of concentrating on how a word meaning may be decomposed, he aimed to provide a computational semantics for the way words modulate their meaning in language use. In his theory, the Generative Lexicon, the contextual flexibility of lexical meaning is modelled as the output of formal operations defined over a generative lexicon. The computational resources available to a word are described in terms of lexical data structures consisting of four levels: a **lexical typing structure**, an **argument structure**, an **event structure**, and a **qualia structure**. In particular, the qualia structure captures how humans understand objects and relations in the world and provides a minimal explanation for the behavior of lexical items based on some properties of their referents (Pustejovsky, 1998). The Generative Lexicon thus does not enumerate senses, but allows meaning to be generated by context.

In agreement with this idea against enumerating senses, Asher (2011) proposed to model polysemy and coercion phenomena using a typed approach. While in the Generative Lexicon, a particular coercion predicate selects a fixed lexical association provided by the qualia structure, in Asher (2011)'s framework meaning modification is modelled as a compositional problem: entities of different types are combined to build types of complex entities. Inference is thus guided by the relation structure defined over the set of types.

Relational approaches

While the approaches presented above require a formal definition for each lexical entry, other theories aim to build formalized models of lexical knowledge in which the lexicon is seen as a structured system of entries interconnected by sense relations such as synonymy, antonymy, and meronymy.

According to Lyons (1968), "*the sense of a lexical item may be defined to be, not only dependent upon, but identical with, the set of relations which hold between the item in question and the other items in the same lexical system.*"

An example of such theories is that of Collins and Quillian (1969). They proposed a hierarchical network model, in which words are represented as entries in a network of nodes comprising a set of conceptual features defining the conventional meaning of the word in question, and connected to other nodes in the network through semantic relations. Inferences between words are thus established as an explicit part of a network of word concepts.

1.1.2 Lexical relations

By assigning meaning to parts of a sentence, entailment can be defined not only among sentences but also among words and phrases. When applied to words, the entailment relation corresponds to logical inclusion (or hyponymy). The meaning of the word *dog* logically includes the meaning of *animal*, because to say that something is a dog implies that it is an animal. Similar reasoning applies to the pairs *waltz*, *dance*; *run*, *move*; *nightmare*, *dream*.

Synonymy correponds to logical identity (mutual inclusion): for example, the meaning of *couch* includes the meaning of *sofa*, and the meaning of *couch* includes the meaning of *sofa*, thus these two words can be said to be synonymous. Similarly, other lexical semantic relations can be derived, as seen below.

1.1.2.1 Classical lexical semantic relations

According to Cruse (1986), a lexical semantic relation is a relation between lexical units, where lexical unit is defined as a surface form along with a sense. As he points out, the number of semantic relations that bind concepts is innumerable; but certain relations, such as hyponymy, meronymy, antonymy, and troponymy, are more systematic and have enjoyed more attention in the linguistics community. These relations can be seen as revelatory of the semantic structure of the lexicon. There are both vertical and horizontal relations (or hierarchical and non-hierarchical).

The two principal vertical relations are hyponymy and meronymy. **Hyponymy** (also referred to as **is-a**) is the most fundamental lexical relation, in terms of which the lexicon is structured (Lyons, 1968). As seen above, hyponymy is when the meaning of one lexical element, the hyponym, is more specific than the meaning of the other, the hyperonym (*e.g.*, *dog—animal*). Lexical items that are hyponyms of the same lexical element and belong to the same level in the structure are called co-hyponyms (*dog*, *cat*, *horse* are cohyponyms of *animal*). **Meronymy** (also referred to as **part-whole**) refers to the relation between a concept/entity and its constituent parts: the meaning of one lexical element specifies that its referent is 'part of' the referent of another lexical element (*e.g.*, *hand—body*).

Horizontal relations include synonymy and opposition. **Synonymy** is defined as the sameness of meaning of different linguistic forms. Two expressions are absolutely synonymous if all their meanings are identical in all linguistic contexts. There are very few absolute synonyms, but words may be synonymous in given contexts. **Opposition** is a general notion, which groups various relations (Cruse, 1986). These include **antonymy** (gradable, directionally opposed, *e.g.*, *big—small*), **complementarity** (exhaustive division of conceptual domain into mutually exclusive compartments, *e.g.*, *aunt—uncle*, *possible—impossible*), **conversity** (static directional opposites: specification of one lexical element in relation to another along some axis, *e.g.*, *above—below*), and **reversity** (dynamic directional opposites: motion or change in opposite ways, *e.g.*, *ascend—descend*).

1.1.2.2 Relations between verbs

The characterisation of meaning by way of synonymy, hyponymy and other meaning relations works particularly well for lexical items that refer to entities, as most nouns do.

Predicates denoting events or states, such as verbs and deverbal nouns, have a more complex structure. Syntactically, they combine with arguments, which requires a semantic characterisation of the arguments in terms of their inherent relation to the event or state (their semantic role, such as agent, patient or experiencer) as well as their linking to surface positions. Also, events and states are often internally structured in terms of aspectual properties.

This makes a simple is-a hierarchy insufficient to express semantically relevant relations between events and states, for example to draw inferences about the result states of events or the involvement of participants.

Schemata and frame analysis

To describe the verb lexicon more adequately, theories based on cognitive frames or knowledge schemata have been proposed, aiming to provide a conceptual foundation for the meaning of words. In these frameworks, words and the concepts they stand for are not directly interrelated, but share membership in common frames or schemata (Schank and Abelson, 1977).

Fillmore (1982) introduced the theory of Frame Semantics, which assumes concept-specific semantic roles of predicate classes, defined in terms of semantic frames and their frame-specific roles. A frame is a script-like conceptual structure that describes a particular type of situation, object, or event along with the participants involved, denoted by their semantic roles. As an example, Figure 1.2 displays the main information contained in the COMMERCIAL TRANSACTION frame. The verb *buy* requires obligatorily a *buyer*, *goods* and optionally a *seller* and a *price*. Verbs with related meanings such as *sell* are expected to have the same meaning slots but in a syntactically different order.

Verb	**Buyer**	**Good**	**Seller**	**Money**	**Place**
buy	subject	object	from	for	at
sell	to				
cost	indirect object	subject		object	at
spend	subject	on		object	at

Figure 1.2: COMMERCIAL TRANSACTION frame (partial).

Semantic relations and syntactic regularities

Levin (1993) proposed a classification of verbs based on the assumption that the semantics of a verb and its syntactic behavior are regularly related. She defined 191 verb classes by grouping 4183 verbs which pattern together with respect to their diathesis alternations, *i.e.* alternations in the expressions of arguments. Levin also provided both positive and negative examples for each verb class to illustrate the legal or illegal syntactic patterns associated with the class.

Though there are criticisms on Levin's way of classification (Baker and Ruppenhofer, 2002), the central thesis of her work is validated via the experiment by Dorr and Jones (Dorr and Jones, 1996). They show that the syntactic signature of each verb class is a valid cue to differentiate each of them semantically.

Classifying relations between verbs

According to Fellbaum (1998), if an entailment relation holds between two sentences such as "Someone V_1" entails "Someone V_2", it can be said that there is an entailment relation between their licensing verbs: V_1 entails V_2. Similarly to nouns, when two verbs can be said to be mutually entailing, they must be synonyms.

Negation reverses the direction of entailment: if V_1 entails V_2, then *not* V_2 entails *not* V_1, but *not* V_1 does not necessarily entail *not* V_2. The converse of entailment is contradiction: if the sentence "*He is snoring*" entails "*He is sleeping*", then "*He is snoring*" also contradicts the sentence "*He is not sleeping*".

Fellbaum (1998) divides entailment relations between verbs into four types. These types can be classified by looking at the temporal relation between the events denoted by the entailing verb V_1 and the entailed verb V_2, as shown in Figure 1.3.

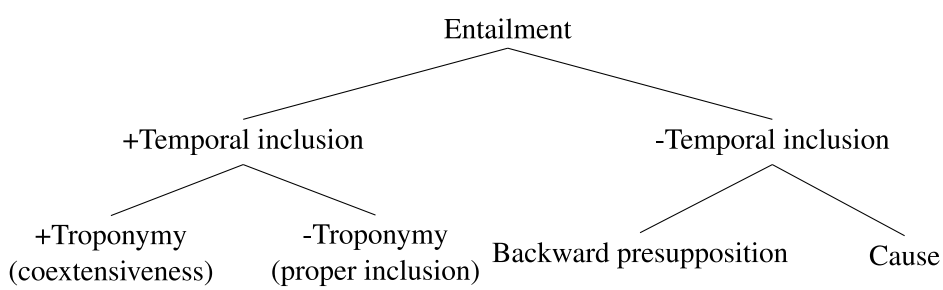


Figure 1.3: Classification of entailment relations between verbs proposed by Fellbaum (1998).

Temporal inclusion refers the notion that one verb includes the other. A verb V_1 includes a verb V_2 if there is some stretch of time during which activities denoted by the two verbs co-occur, but no time during which V_2 occurs and V_1 does not. **Proper inclusion** of V_2 in V_1 indicates that there is a time during which V_1 occurs but V_2 does not (*e.g.*, *snore* $\rightarrow$ *sleep*). **Troponymy** represents the hyponymy relation between verbs. Fellbaum (1998) notes that troponymy is a special case of entailment, where V_1 entails V_2 and V_1 and V_2 are always temporally co-extensive, that is, when the actions described by V_1 and V_2 always begin and end at the same times (*e.g.*, *limp* $\rightarrow$ *walk*).

Backward-presupposition stands when the entailed verb V_2 happens before the entailing verb V_1 and is necessary for V_1. For example, *win* entails *play* via backward-presupposition as it temporally follows and presupposes *play*. Finally, in **causation** the entailing verb V_1 necessarily causes V_2. In this case, the temporal relation is thus inverted with respect to backward-presupposition, since V_1 precedes V_2. In causation, V_1 is always a causative verb of change, while V_2 is a resultative stative verb (*e.g.*, buy $\rightarrow$ own).

As a final note, it is interesting to notice that the Subject-Verb structure of V_1 is preserved in V_2 for all forms of lexical entailment. The two verbs have the same subject. The only exception is causation: in this case the subject of the entailed verb V_2 is usually the object of V_1 (*e.g.*, X *give* Y $\rightarrow$ Y *have*), as in most cases the subject of V_1 carries out an action that changes the state of the object of V_1, that is then described by V_2.

Lexical relations and discourse

In the context of discourse semantics, Lascarides and Asher (1993) proposed to integrate lexical and discourse processing to catpure intuitive generalizations. Indeed, they argued that discourse interpretation requires one to make inferences about the relations holding between events reported in different sentences composing the discourse, and that lexical knowledge about the verbs denoting these events allows to infer a most plausible type of relation.

(1.9) Max stood up. John greeted him.

(1.10) Max fell. John pushed him.

Intuitively, a causal relation holds between the first and the second event in Example 1.10, while it does not in Example 1.9. The interpretation of Example 1.10 relies, among other things, on knowledge about the meanings of the verbs *push* and *fall*, as well as the fact that the event of *pushing* immediately precedes the event of *falling*. The lexical association of these two verbs brings a clue indicating a possible causal link (which can be referred to as a **permissible cause** relation), which can be used in context to infer a causal link between events.

These notions pertain to discourse structure theories, which will be described further in Section 2.2.

1.1.3 Manually constructed resources

As we will see in Section 1.1.4, broad-coverage semantically informed NLP systems need large coverage semantic lexicons. In order to attempt to provide the lexical information required by such tasks, a number of manually constructed lexicons have been produced, based on different theories of lexical semantics such as those presented previously. They are thus structured according to different aspects of meaning and differ considerably in their descriptive devices. Here we present some of the most widely used resources.

1.1.3.1 WordNet

WordNet (Fellbaum, 1998) is a very large lexical knowledge base. Its design is inspired by the psycholinguistic theories of human lexical memory. It is based on word meanings: all of the words that can express a given sense are grouped together in a set of synonyms, or a *synset*. Each synset represents one underlying concept. Synsets can then be linked by different kinds of semantic relations.

The lexicon is divided intro three categories: nouns, verbs and adjectives (function words are not represented explicitly). Each category is structured differently. For each part of speech, different relations play a major role.

Regarding nouns, WordNet includes 24825 noun synsets and 32364 different nouns with a total of 43146 senses. Nouns are organized in lexical memory as topical hierarchies. Synsets contain pointers to others representing concepts that are hyper-/hyponyms, mero-/holonyms, antonyms, and coordinate terms (sharing the same hypernym).

WordNet includes 10653 adjective synsets containing 12909 different adjectives organied into 1006 clusters. In the adjective lexicon, antonymy and similarity organize the various lexical items.

There are far fewer verbs than nouns in the language and verbs are more polysemous than nouns. There are some 8,500 verb forms organized into about 5.000 synsets and divided into 14 semantically distinct groups: verbs of bodily care and functions, change, cognition, communication, competition, consumption, contact, creation, emotion, motion, perception, possession, social interaction and weather verbs. The major division into 14 semantically coherent groups reflects the division between the major conceptual categories EVENT and STATE. Verbs are organized according to the classification of entailment relations presented earlier in Figure 1.3 on page 19.

1.1.3.2 VerbNet

VerbNet (Schuler, 2005) is a hierarchical verb lexicon based on Levin's classes (Levin, 1993). VerbNet refines Levin classes into intersective Levin classes, which are more fine-grained and which exhibit more coherent sets of syntactic frames and associated semantic components. It is structured as a hierarchy of verb behavior, from groups of verb classes that share similar semantics to the tree structure of the verb classes themselves. For example, the verb type **Verbs of Change of State** includes the verb classes *break*, *bend*, *cooking*.

Each class is characterized extensionally by its set of verbs, and intensionally by a list of the arguments of those verbs and syntactic and semantic information about the verbs. The argument list consists of thematic roles (23 in total) and possible selectional restrictions on the arguments expressed using binary predicates. The syntactic information maps the list of thematic arguments to deep-syntactic arguments (*i.e.*, normalized for voice alternations, and transformations). The semantic predicates describe the participants during various stages of the event described by the syntactic frame.

As an example, Figure 1.4 displays the class BREAK.

Class	BREAK
Members	break
	break apart
	break down
	chip
	crack
	fragment
	...
Roles	Agent[+INT_CONTROL]
	Patient[+SOLID]
	Instrument[+SOLID]
	Result

Figure 1.4: Class BREAK in VerbNet (Schuler, 2005).

VerbNet covers 4526 senses for 3769 lexemes. A primary emphasis for VerbNet is the grouping of verbs into classes that have a coherent syntactic and semantic characterization, that will eventually facilitate the acquisition of new class members based on observable syntactic and semantic behavior. The hierarchical structure and small number of thematic roles is aimed at supporting generalizations.

1.1.3.3 FrameNet

FrameNet (Baker et al., 1998) is a manually constructed database based on Frame Semantics theory (Fillmore, 1982). The goal of FrameNet is to produce a semantic dictionary that documents the combinatorial properties of English lexical items in semantic and syntactic terms based on examples in a very large corpus. Similarly to VerbNet, it focuses on argument patterns, although in a more specific approach, and it is not limited to verbs.

A frame in the FrameNet project is a schematic representation of a particular type of situation involving various participants. A frame contains various Frame Elements (FEs), kinds of entities that can participate in a frame. They correspond roughly to thematic roles in an eventuality.

A frame description lists lexical units that can evoke the described situation. A lexical unit is defined as the pairing of a word with a sense. FrameNet contains nearly 800 frames with over 10,000 lexical units. Lexical units are mainly verbs but may also be nouns, adjectives, or other parts of speech.

Frames can be related to each other via inter-frame relations, forming a complex hierarchy which situates frames in a semantic space. FrameNet lists several inter-frame relation types, including **Inheritance** (is-a relation), **Subframe** (subevent), **Precedence**, **Usage** (presupposition), **Perspective** and **Causation**. A relation also provides a partial or complete mapping between the frame elements of the two linked frames.

FrameNet contains more than 135,000 annotated sentences that represent typical lexical unit occurrences in texts. Each annotation refers to one lexical unit in a specific frame and includes the lexical unit itself and the frame elements of the frame that occur in the sentence. Not all lexical units have been annotated. For example, of the approximately 4,100 verb lexical units in FrameNet, only about 2,800 have annotated sentences.

Each sentence is also annotated with a valence pattern, which provides grammatical functions of the core frame elements with respect to the lexical unit.

As an example, Figure 1.5 displays the Frame CAUSE TO FRAGMENT.

Frame	CAUSE TO FRAGMENT
Frame Elements	Agent (Semantic type: Sentient)
	Cause
	Pieces
	Whole_patient
Frame-Frame Relations	
Inherits from	Transitive_action
Uses	Destroying
Is causative of	Breaking_apart
Lexical Units	break apart
	break down
	break up
	break
	chip
	dissect
	fragment
	...

Figure 1.5: CAUSE TO FRAGMENT Frame in FrameNet (Baker et al., 1998).

1.1.3.4 Limitations

These resources have been used in many applications, as we will detail in the following section. However, they suffer from important limitations.

Their construction requires human experts and it is time intensive. As a result, they tend to have very limited coverage.

Additionally, due to their theoretical assumptions, they aim to only cover the prescriptive cases, where the relations they identify are guaranteed to hold. For instance, while the backward presupposition relation between *win* and *play* is covered in WordNet, the relation between *sell* and *buy* is outside of its scope, since it is defeasible (*sell* could also be related to *manufacture* or *steal*) (Chklovski and Pantel, 2004). Such relations are however present in FrameNet, but its severe lack of coverage poses more important difficulties.

Another problem with these resources lies in the fact that they revolve around word senses: in order to apply them on corpora, word sense disambiguation is thus necessary, which is a notoriously hard challenge. Indeed, lexical ambiguity comprises a large and extremely heterogeneous class of individual phenomena.

1.1.4 Applications

1.1.4.1 Large coverage computational semantics

A few systems for wide coverage deep semantic analysis have been implemented, for instance *Boxer* (Bos, 2005, 2008a), for written Engish and *Grail* (Moot, 2010) for spoken Dutch and written French. *Boxer* uses Discourse Representation Theory as its semantic representation. The semantic representations output by the system can be used to perform inferences.

However, without any background knowledge, such a system would only be able to generate a representation of the explicit content of the sentence, that mirrors the syntactic dependency structure. In order to constrain the process of semantic interpretation and enrich those interpretations with further inferred content, additional background knowledge is necessary.

For this purpose, Bos (2005) extended *Boxer* with lexical knowledge automatically generated in the form of first-order axioms, using WordNet. However, since WordNet contains information that is specific to word senses, a disambiguation would be needed. In *Boxer*, this is approximated by taking the most frequent sense.

Bos (2008a) showed examples of inferences that are supported by the system and discusses where the system makes mistakes. The inferences mostly involve simple hyponymy, synonymy relations and the mistakes mostly involve discourse interpretation (pronouns, presuppositions). Aside from errors in anaphora and presupposition resolution, the issues with WordNet mentioned above (*e.g.*, inconsistency, incompleteness, and the need for word sense disambiguation) are an important source of errors. This motivates the need for better and higher coverage sources of lexical knowledge, which can be learned directly from text as we will detail in Section 1.2.

1.1.4.2 NLP tasks

Logical inferences such as those derived by *Boxer* can be used in NLP applications such as information retrieval, information extraction, spoken dialogue systems and open-domain question answering (Bos, 2008b). For instance, in information retrieval, it is necessary to infer whether a document satisfies the information need which is conveyed by a natural language query. In information extraction, a potential occurrence of an event in the text should imply the meaning of the event. And in question answering, an answer is expected to be inferred from candidate retrieved passages.

They can also be used for the task of Recognizing Textual Entailment (RTE) (Bos and Markert, 2006; Tatu and Moldovan, 2006). The concept of **textual entailment** is used to indicate the state in which the semantics of a natural language written text can be inferred from the semantics of another one, *i.e.*, when the truth of an enunciation entails the truth of another enunciation. However, in comparison with the notion of entailment in formal semantics which we studied previously, the textual entailment relation is not expected to be satisfied in each and every case. In other words, the inference needs not be unquestionably certain. It is rather a somewhat relaxed relation which is assumed true if "*a human reading T would infer that H is most likely true*" (Dagan et al., 2006).

Aside from the approaches mentioned previously, most practical NLP systems perform inferences at a "shallow" semantic level, due to the difficulty of producing robust semantic interpretation into logic-based meaning-level representations. Such approaches propose an inference model that approximates entailment without any explicit interpretation into meaning representations, but rather operating directly over lexical-syntactic units.

In order to do this, such systems often require information about lexical relations, particularly between verbs. Many systems have thus used manually constructed resources such as those we presented. We present some examples of these uses:

- In Question Answering, Novischi and Moldovan (2006) use verb argument structures derived from VerbNet, which are then propagated along lexical chains consisting of WordNet relations. Webber explores using lexical resources such as FrameNet or VerbNet to generate potential answer-containing sentences.

- In Text Summarization, Barzilay and Elhadad (1999) construct lexical chains based on WordNet to identify significant sentences. Copeck et al. (2009) propose to represent sentences as FrameNet types in a conceptual graph, and use pairwise similarity to identify the most central sentences.

- In Document Classification, Klavans and Kan (1998) use WordNet and Levin classes to assess the event profile of texts as a function of verb type, and demonstrate the document type is correlated with the presence of many verb of a certain Levin class.

These approaches are however limited by the shortcomings of the resources they are based on.

With the advent of mass access to textual information, many attempts at automatically extracting the necessary lexical information have emerged. This is the subject of the next section.

1.2 Semantic relation extraction: Existing methods and resources

Most research on automatic extraction of semantic relations focuses on exploiting large amounts of unannotated corpora, which have become increasingly available for many languages and domains, often by harvesting from the web.

Such approaches are based on the distributional hypothesis, which suggests that word meanings can be derived in part from their distribution across different linguistic envionments. This hypothesis was notably inspired by the maxim: "*You shall know a word by the company it keeps*" (Firth, 1957), and also suggested by Harris (1968): "*the linguistic meanings which the structure carries can only be due to the relations in which the elements of the structure take part.*"

This idea is at the basis of the field of distributional semantics, in which the aim is to get at the meaning of words by analyzing the contexts in which they are used, and observing distributional regularities. In practice, the representation of a word is derived by building a vector composed of the frequency of co-occurrence of that word with every other words in a corpus.

Here we focus on a related but distinct goal: instead of word meaning representation, we are interested in the relations holding between words. Two types of approaches are based on the distributional hypothesis. The first type directly exploits word representation from distributional semantics to infer semantic relatedness between words. These **distributional approaches** are described in Section 1.2.1. The second type of approaches exploit the context in a different way, by computing co-occurrence frequencies between word pairs and specific patterns supposed to be representative of a specific semantic relation. We present these **pattern-based approaches** in Section 1.2.2. Some research has been dedicated to combining both types of techniques, which have been proven to be quite complementary. We discuss this strand of work in Section 1.2.3. Finally, we focus on the usage of discourse information for the extraction of semantic relations between verbs in Section 1.2.4.

1.2.1 Distributional approaches

Distributional approaches have the advantage of being purely unsupervised. They are based on the vector representations of lexical items proposed in distributional semantics. As previously stated, these representations are obtained by recording frequencies of co-occurrence with other words in contexts extracted from large corpora. Co-occurence can either be considered on the basis of a fixed window of surrounding words, or on the basis of syntactical dependencies. In Figure 1.6, we show examples of co-occurence counts for several words occurring with different context words.

	red	delicious	fast
apple	2	1	0
wine	2	2	0
car	1	0	1
truck	1	0	1

Figure 1.6: Examples of co-occurence counts for several lexical items (taken from (Van de Cruys and Mertens, 2006), translated from French).

Based on these distributional representations, the hypothesis behind distributional approaches to relation extraction is that two words can be considered to be linked by a semantic relation if they have many co-occurring words (*e.g.*, *car* and *truck* have identical vectors in our very simple representation in Figure 1.6, while *car* and *wine* have very different vectors). Distributional measures such as cosine can be used to quantify the amount of common co-occurring words between two considered words.

1.2.1.1 Semantic similarity and semantic relatedness

Budanitsky and Hirst (2006) noted that there is an important distinction between semantic similarity and semantic relatedness, with the former being a subset of the latter. Semantic similarity denotes relations of synonymy, hyponymy (and hypernymy), antonymy, or troponymy, while semantic relatedness denotes any semantic relation existing between two words. In practice, Budanitsky and Hirst (2006) formalize this distinction according to the syntactic relations between each word and its co-occurring words: two words are distributionally similar if they have the same syntactic relation with their co-occurring words, but only distributionally related if they do not.

Thus, a manner of measuring similarity instead of general relatedness is to consider the context of a word in terms of its syntactic relations with words in the same sentence, instead of a fixed window of surrounding words. The former method is applied in most approaches to semantic relation extraction as described below, although they tend to use the terms similarity and relatedness interchangeably.

1.2.1.2 Inducing semantic classes

The general idea of these distributional approaches is to compute the distributional similarity between words extracted from large corpora, and to apply clustering algorithms to induce semantic classes grouping words with similar semantic properties.

One of the earliest approaches of this kind was proposed by Hindle (1990). His study aimed at producing a classification of nouns according to the predicates they occur with, based on predicate-argument relations obtained by applying a deterministic parser on their 6 million word corpus of news stories. He proposed a metric of similarity based on mutual information of verbs and arguments. Using this metric, his method outputs a ranked list of similar nouns for each considered noun. This approach is however limited beause it does not distinguish between different senses of a polysemous word.

Lin (1998) followed a similar methodology, and represented each word by features corresponding to its co-occurring words and the dependency link between them. He also defined his own similarity measure, perhaps the most widely used in later approaches. This measure is defined to be *the amount of information in the commonality between the objects divided by the amount of information in the descriptions of the objects*. The similarity between two words w_1 and w_2 is thus expressed as follows:

$$\mathrm{sim}(w_1, w_2) = \frac{2 \times I(F(w_1) \cap F(w_2))}{I(F(w_1)) + I(F(w_2))},$$

where $F(w)$ is the set of features possessed by w and $I(S)$ is the amount of information contained in a set of features S. After applying a clustering algorithm similarly to Hindle (1990), he proposed an additional step of pruning the clusters based on the detection of meaning shifts, to account for different senses of the extracted words.

Pantel and Lin (2002) observed that this method proposed by Lin (1998), and the earlier ones, did not provide a satisfying solution for the treatment of polysemous words. They designed an algorithm specifically aimed at discovering word senses, called *Clustering by Committee* (CBC). The idea is to first construct so-called committees, clusters composed of the most similar words, and to find the centroid of each cluster. Then, each word is assigned to its most similar cluster. The discovery of other less frequent senses of this word is done by removing the overlapping features between the word and the centroid of its cluster, and using the remaining featuresto assign the word to another cluster. They performed an evaluation based on comparison to WordNet senses, and obtained an F-score of 0.55, while their best baseline obtained 0.49.

1.2.1.3 Asymmetric similarity

Aside from constructing classes of semantically similar words, distributional methods have also been used in the task of lexical entailment acquisition. Lexical entailment corresponds to several lexical semantic relations such as synonymy, hyponymy, and in some cases meronymy (Geffet and Dagan, 2005). For instance, the word *company* can substitute the words *firm* (synonym), *automaker* (hyponym) and *division* (meronym), and as such, *company* is entailed by these words.

One important characteristic of entailment is that it is directional: if one word entails another, the converse does not necessarily hold. In order to capture such a relation, symmetric measures of similarity such as those used in the approaches we described previously are not sufficient. Instead, several distributional approaches to lexical entailment recognition are based on the concept of distributional inclusion. The idea is that if one word entails another, it can be expected that all context features of the entailing word are also features of the entailed word, so that the vector representation of the former is included in that of the latter.

Weeds et al. (2004) showed that there is indeed a high correlation (of about 70%) between the inclusion of one distributional vector into another and human intuition about a general/specific relation between these words.

Geffet and Dagan (2005) applied this idea specifically for noun entailment recognition. They found that although it is usually the case that if a word entails another, the feature vector of the first is usually included in that of the second, it is not necessarily the case that when a feature vector is included in another, their corresponding words are in an entailment relation. However, their method yielded a significant improvement of 0.15 of the F-score over a baseline using a symmetric similarity measure.

1.2.1.4 Limitations of distributional approaches

An important limiting factor of distributional approaches is that distributional measures do not permit to distinguish between different semantic relations, they only quantify a level of relatedness between lexical items. Baroni and Lenci (2011) performed a manual study of highly ranked pairs of similar words, and identified a variety of semantic relations, among which synonyms, hyper-/hyponyms, holo-/meronyms, as well as a large proportion of pairs annotated as "topically related, but none of the above".

Similarly, Lin et al. (2003) and Geffet and Dagan (2005) that distributionally similar words tend to include not only synonyms but also antonyms, and co-hyponyms, *i.e.*, words that share a common hypernym. In directional distributional similarity, high scores are often given to rules with obscure entailing verbs (Kotlerman et al., 2010).

In order to get at more precise semantic relations, another type of approaches, based on lexico-syntactic patterns, has been investigated by a number of researchers. We present these techniques in the following section.

1.2.2 Pattern-based approaches

Pattern-based approaches are based on a different take of the distributional hypothesis. In distributional approaches, relatedness between two lexical items is indicated by their co-occurrence in similar contexts with no restriction on the type of contexts. In contrast, pattern-based approaches target specific relations which are expressed by explicit expressions.

Pairs of lexical items linked by these relations are identified by their co-occurrence with patterns defined as indicators of the targeted relations. These approaches can thus been qualified as "weakly supervised", since a manual input is necessary for the specification of patterns (either fully manually defined or automatically discovered with the help of preliminary seeds).

Accross the existing approaches, the patterns are sometimes defined only by string-based regular expressions, or also include syntactic constraints. Some approaches also add semantic constraints obtained from external resources.

1.2.2.1 Lexico-syntactic patterns

One of the first studies to exploit patterns for automatic lexical acquisition was led by Hearst (1992). This study specifically focuses on the extraction of lexical items linked by a hyponymy relation. They define lexico-syntactic patterns based on regular expressions and part-of-speech constraints, for instance "*such* NP *as* NP, * *or* | *and* NP".

(1.11) *... works by such **authors** as **Herrick**, **Goldsmith**, and **Shakespeare**.*

Applied to the phrase in Example 1.11, this pattern would trigger the extraction of the following relations in the associated illustrative sentence:

- hyponym ("author", "Herrick"),
- hyponym ("author", "Goldsmith"),
- hyponym ("author", "Shakespeare").

Although the strict restrictions of these patterns and the unavailability of truly large corpora at the time allowed for the extraction of only a small number of related items (152 pairs of items linked by a hyponymy relation were obtained by applying the pattern in Example 1.11 on a 8.6M words corpus of encyclopedia texts), this approach inspired many researchers to develop similar methods.

For instance, Berland and Charniak (1999) followed the same approach for the extraction of meronymy relations (part-whole) on a much larger corpus and obtained better results, although they observed some ambiguity in their patterns, which could be used to express other relations in different contexts, leading to some precision problems.

1.2.2.2 Discovering new patterns

Both of these methods were limited by the small number of manually defined patterns they considered. To address this issue, Hearst (1992) proposed the idea of automatically discovering new patterns in a recursive manner. Pairs of lexical items are extracted with the predefined patterns and can thus be considered to be linked by the corresponding relation. Then if such a pair is found in corpus, linked by another expression, this expression could be a potential new pattern expressing the same relation. This idea was however not implemented by Hearst (1992).

Such a bootstrapping algorithm has later been used to discover patterns aimed at other kinds of semantic relations. For instance, Ravichandran and Hovy (2002) reported good results on specific relations such as **birthdates**, but low precision on more generic relations such as **is-a** and **part-of**.

Pantel et al. (2004) focused on **is-a** relations, and showed that a pattern-based approach enriched with the bootstrapping technique, when applied to very large corpora, yields comparable performance to linguistically-rich approaches to relation extraction, such as that of Pantel and Ravichandran (2004) described below in Section 1.2.3.

This technique for automatic pattern discovery was also applied by Girju et al. (2006) for the extraction of part-whole relations, who also proposed an algorithm to counteract the ambiguity of some patterns and discriminate meronymic instances from erroneous extractions. They evaluated their results against manual annotations, and obtained an F-score of 0.83, while their best baseline obtained 0.37. They also mentioned the perspective of using the same method for other semantic relations, such as **kinship** and **purpose**.

1.2.2.3 Limitations of pattern-based approaches

Although pattern-based approaches have been shown to achieve high precision and to allow for the identification of particular relations and their distinction, they tend to have very low recall scores. Their limited coverage is due to the fact that they can only recover relations that are explicity expressed by strict patterns, which are very rare in corpora (Cimiano et al., 2005).

1.2.3 Combining distributional and pattern-based approaches

As noted by Pennacchiotti and Pantel (2009), distributional and pattern-based extraction algorithms capture aspects of paradigmatic and syntagmatic dimensions of semantics, respectively. The former allow for high coverage but lack precision and are not suitable for the identification of specific relations, while the latter tend to yield high precision but low recall due to the restrictive nature of patterns. Those approaches are thus believed to be quite complementary. Following this observation, a number of studies have attempted to combine both types of approaches.

Lin et al. (2003) proposed to separate the two tasks of extracting lexical items and discovering the specific relation linking them, using the advantages of distributional methods for the first task and patterns for the second. Aiming to distinguish noun synonymy and antonymy, they first collected distributionnally similar nouns, and then used lexical patterns applied to Internet queries to identify the type of relatedness between pairs of nouns.

In order to predict **is-a** relations, Caraballo (1999) combined clustering with a pattern-based approach following Hearst (1992). He first used distributional techniques to construct clusters of semantically similar nouns. Then he discovered potential hypernyms with the help of the pattern "*X, Y and other Zs*", where *X* and

Y are nouns appearing in the same cluster, and *Z* is likely to be a hypernym for them. After extracting all possible hypernyms for all nouns in a cluster, the best hypernym was then chosen as the most frequent one. By manual evaluation, he obtained a precision of 0.33 with his algorithm.

Pantel and Ravichandran (2004) followed a similar approach as an extension of the CBC algorithm in Pantel and Lin (2002). They used syntactic patterns like apposition (*e.g.*, "***Oracle**, a **company** known for ...*"), nominal subject (*e.g.*, "***Apple**, was a hot young **company** ...*") as well as lexico-syntactic patterns, for example *X such as Y* (*e.g.*, "***companies** such as **IBM***"), in order to label the clusters with corresponding hypernyms. Their manual evaluation, although following a slightly different protocol from that of Caraballo (1999) and thus not truly comparable, yielded a precision of 0.42.

Beside such two-step procedures to combine both types of approaches, another possibility is to encode both types of information as features for supervised classification.

Following this idea, Mirkin et al. (2006) introduced a system for predicting entailment between nouns. They first used separately a distributional algorithm and a pattern-based one (relying on the patterns proposed by Hearst (1992) and Pantel et al. (2004)) to extract candidate entailment pairs. After manually annotating these candidates, they trained an SVM classifier on 700 of these candidates, using features based on both types of approaches. To estimate the contribution of such a method, they compared their results to those obtained with each system used for candidate extraction. Their method yielded an F-score of 0.62, compared to 0.40 for their distributional baseline and 0.51 for their pattern-based baseline.

Pennacchiotti and Pantel (2009) applied a similar approach for the task of entity extraction, while extending it by including additional web-based features, hypothesizing that distributional and pattern-based features do not exhaust the semantic space. They reported an improvement of 22% in mean average precision over state-of-the-art methods.

1.2.4 Extracting relations between verbs

Most of the approaches we have mentioned focus on extracting semantic relations between nouns. Indeed, nouns are naturally ordered in a hierarchy (Fellbaum, 1998), and co-occur more often within the same text fragment than verbs. Lexico-syntactic patterns are thus better adapted for the extraction of related nouns.

One approach that has attempted to use such strict lexico-syntactic patterns to extract semantic relations between verbs is that of Chklovski and Pantel (2004).

They aimed to distinguish between five fine-grained relations: **similarity**, **strength**, **antonymy**, **enablement** and **temporal happens-before**. Their motivation for focusing on verb pairs instead of noun pairs is that "verbs are the primary vehicle for describing events and expressing relations between entities."

Like the combined approaches we mentionned in the previous section, they started by extracting highly associated verb pairs using a paraphrasing algorithm designed by Lin and Pantel (2001). Interestingly, this algorithm is itself based on syntactic patterns learned with a distributional approach: if the same pairs of words occur frequently in different syntactic dependency paths involving other words, then these paths can be considered to indicate relatedness between these other words. For example, the paths "*NP1 <-subject- solves -object-> NP2*" and "*NP1 <-subject- tackles -object-> NP2*" are often used with the same pairs of nouns represented by *NP1* and *NP2*, and the verbs *solve* and *tackle* are thus considered to be related. Chklovski and Pantel (2004) used only dependency paths of the type *subject-verb-object*, thus limiting their approach to transitive verbs.

After extracting related verb pairs, they then designed string-based patterns representing each of their five relations. For instance, the relation **happens-before** is represented by the pattern "*to V1 and then V2*", among others. Using these patterns, they first collected co-occurrence data on pairs of verbs from the web. They then assessed the strength of the associations by evaluating their mutual information, and manually set a threshold to determine whether each association between a verb pair and a relation is valid and should be entered in their output resource, *VerbOcean*. No precise evaluation methodology was provided for the results obtained.

The fact that they worked on the web, although granting access to very large amounts of data, constrained their extraction to strict, intrasentential patterns. However we can note that these patterns were mostly constructed around **discourse markers**, or cues for discourse relations. For instance their patterns for antonymy include words like *but*, *either-or*, and their patterns for happens-before relations include words like *then*, *eventually*, or *later*. Indeed, as we saw in Section 1.1.2, semantic relations between verbs are tightly connected to discourse relations between the events they denote. Other approaches have therefore made explicit use of discourse cues.

Inui et al. (2005) focused on acquiring causal relations between events, and sub-classified them into four types: cause, effect, precondition and means. Their extraction method is based on the identification of causal connectives. They exploit the syntactical dependency structure to identify the events linked by these connectives. In order to classify the extracted pairs of events into the aforementioned four types of causal relations, they used linguistic tests based on templates to determine the volitionality of each event in a pair. According to these tests, an event can be recognized as either an action or a state of affairs. Depending on the type of each event in a pair, the subclass of causal relation associated with them can be inferred.

However this approach is limited to the extraction of events occurring in the same sentence, due to the fact that it is based on syntactical dependencies. Other approaches have made use of discourse information in broader contexts.

Instead of using strict intra-sentencial patterns, Pekar (2008) proposed to use local discourse coherence to identify related verbs, for the task of predicting entailment relations between verbs. Based on theories of discourse structure and discourse coherence (Hobbs, 1985; Asher and Lascarides, 2003), and notably Centering theory (Grosz et al., 1995), whereby local coherence is a function of the amount of inferences required in the processing of the text, he proposed the assumption that "*the stronger two events are related by entailment, the more frequently the two corresponding verbs should be appearing within a segment of local discourse.*"

His method identifies discourse related clauses by indicators such as textual proximity and shared arguments between events. Then, templates representing the shared arguments between each pair of verbs are produced. A measure of asymmetric association (based on the method for learning selectional preferences of verbs proposed by Resnik (1993)) is then applied on these templates to determine the direction of the entailment relation existing between the pairs of verbs.

A manual evaluation showed that altough he managed to correctly predict the direction of the entailment relation more often than not, more than half of their evaluated pairs were annotated as not related by entailment. He found that many of these pairs were in fact related by an antynomy relation. This observation seems to indicate that although discourse relatedness does help in the recognition of relatedness between pairs of verbs, it is not limited to lexical entailment relations, and further distinction between the types of extracted relations needs to be made.

Chambers and Jurafsky (2008) applied a similar hypothesis on discourse coherence, but based on shared arguments between events, to extract narrative chains of events, which they define as partially ordered sets of events centered around a common protagonist. The notion is similar to the concept of **scripts**, structured sequences of participants and events, proposed by Schank and Abelson (1977). Essentially, the number of times two verbs share an argument in a document is used to determine whether they are involved in a narrative relation, using a PMI measure. They compared these results with a baseline considering only the co-occurrence of verbs in a document, without taking their arguments into account, and found an improvement of 36.4% over this baseline. This study thus showed that shared arguments are indeed a good indicator of narration relations between verbs.

Weisman et al. (2012) focused on predicting entailment between verbs with a classification approach, using a large set of linguistically motivated features specific to verbs, including some adapted from prior work as well as some novel features. Notably, discourse markers corresponding to four discourse relation (contrast, cause, condition, and temporal) are used as features, with the assumption that they indicate a relation between the clauses' main verbs, similarly to the work of Pekar (2008). They also designed a feature representing the dependency relation between clauses. Additionally, they used features based on shared arguments, inspired by Chambers and Jurafsky (2008), the polarity of the verbs (with the idea that change in polarity can be an indicator of non-entailment between the two verbs), distributional similarity (including Lin (1998)'s measure), among other features.

Using a similar experimental protocol as that of Mirkin et al. (2006), they obtained an F-score of 0.51, with a good recall of 0.71 and moderate precision of 0.40. They compared these results to a baseline using only pattern-based and distributional features to produce a model similar to the one proposed by Mirkin et al. (2006) (but applied to verbs instead of nouns as in the original model), and with this baseline model they obtained an F-score of 0.41.

Do et al. (2011) focused on identifying causality between event pairs triggered by verb-verb, verb-noun and verb-verb associations in context. They proposed to combine discourse relation predictions and distributional similarity methods in a minimally-supervised approach. Distributional similarity features are collected from unannotated corpora, while discourse relation predictions are obtained by a model trained on explicit discourse annotations in the PDTB (Prasad et al., 2008) (such models are discussed in Section 2.4).

Through a global inference procedure, they defined constraints aimed at capturing the interactions between event causality and discourse relations. For example, they enforced the constraint that when an event pair is judged to be causal, the discourse relation linking the two clauses they are extracted from need to belong to the following set: {Cause, Condition, Temporal, Asynchronous, Synchrony, Conjunction}. Conversely, an event pair should be judged non-causal if the clauses are linked by one of the following relations: {Comparison, Concession, Contrast, Pragmatic-concession, Pragmatic-contrast, Expansion, Alternative, Exception, Instantiation, List, Restatement}.

To evaluate their system, they produced manual annotations on 25 news articles. With this method combining distributional similarity and discourse information, they obtained an F-score of 41.7 for event-pair causality detection.

In contrast to these approaches, Tremper and Frank (2013) aimed to distinguish between several fine-grained relations linking verbs: presupposition, entailment, temporal inclusion and antonymy. They analyzed discriminating properties of these relations, on the basis of temporal sequence and negation properties, and designed features representing these properties, as well as features of contextual relatedness, based on indicators such as embedding or coordinating conjunctions.

In order to obtain labeled instances for their supervised classification, they designed annotation tasks. The first task relied on candidate verb pairs without context, for which they obtained an inter-annotator agreement value of 0.47, indicating a highly difficult task. As a second, potentially easier task, they provided verb pairs in their original contexts, and obtained an inter-annotator agreement value of 0.44. They analyzed that the difficulty was due to determining whether or not the pair was related, and not to distinguishing between the possible relations. In order to obtain more reliable annotations, they designed a third task based on a set of questions. With this more involved task they obtained an inter-annotator agreement value of 0.64 for expert annotators, and used the results as their gold standard annotation set.

By training their model on a small training set of 48 verb pairs, they obtained an overall F-score of 0.55. They compared their results for the antonymy class to those found in VerbOcean (Chklovski and Pantel, 2004), which yielded a precision of 0.71 and a recall of 0.35, while with their own system, Tremper and Frank (2013) obtained a precision of 0.72 and a recall of 0.74, thus improving significantly on the recall value.

Roth and im Walde (2014) proposed a vector-based representation of word pairs using as features the discourse markers that occur between them, with the goal of identifying relations of antonymy, synonymy and hyponymy. They formulated the hypothesis that "*if two phrases frequenty co-occur with a specific marker, then the discourse relation expressed by the discourse marker should also indicate the relation between the words in the affected phrases*", with the intuition that contrast relations might indicate antonymy, whereas elaboration may indicate synonymy or hyponymy. They considered all pairs of nouns, verbs and adjectives in related discourse segments, hypothesizing that although the relation would not hold between all words, high relative frequency would guide the identification of correct instances.

They collected intra-sentential pairs of phrases related by markers from a set of 61 markers obtained from the annotations in the PDTB for English, and a set of 155 markers translated from the English markers for German, creating single-word patterns with an arbitrary number of wild-card tokens such as "*X * though * Y*". They represented the word pairs from each pair of related phrases as vectors containing 61 and 255 features for English and German, respectively.

They experimented with different vector values such as absolute frequency, log frequency, euclidean distance or PMI, and found PMI combined with euclidean distance to work best, although they do not detail their computation. They implemented three-way and two-way classification models, which they compared against a baseline using more classical lexico-syntactic patterns such as "*X affect how you Y*" taken from an approach by Im Walde and Köper (2013). Their results showed that pattern features in fact perform better than marker features for German noun pairs, while markers yield better results for verb and adjective pairs. For English, marker features perform better for all syntactic categories. However their best results were obtained by combining both types of features, yielding F-score improvements ranging from 1 to 7.5 points over the best of the two models for all syntactic categories.

The results of these different studies demonstrate the usefulness of discourse information for the extraction of relations between verbs.

1.3 Conclusion

This chapter has been concerned with lexical semantic relations. We first studied the theoretical aspects of lexical semantic relations and their role in the representation of the lexical meaning of individual words, at the basis of complex meaning construction. Then we presented data-driven approaches aimed at identifying these relations in large corpora, enabling the extraction of broad-coverage lexical information.

Formal and distributional approaches to natural language interpretation can be seen as complementary. However, distributional methods are typically not related to logic and are often not even based on a full syntactic parse of the sentence. This shallow view on language from a linguistic point of view makes the interface with symbolic systems difficult. Targeting a specific set of relations can help alleviate this problem.

More complex entailments concerning discourse coherence can be derived from semantic relations between verbs. As we have seen, the extraction of such relations between verbs requires the consideration of broader contexts to account for relations between events in discourse. Discourse analysis is therefore a key element for the extraction of such relations between verbs, and in turn can benefit from such information. This is the focus of the next chapter.

CHAPTER 2

BACKGROUND: DISCOURSE

Contents

We have seen in the previous chapter that discourse semantics is strongly intertwined with lexical semantics and compositional semantics. As noted by Asher and Pustejovsky (2000), a lexical theory should be sensitive to facts about discourse interpretation as well as sentential composition, since discourse structure and discourse context can affect lexical meaning. Asher and Lascarides (2003) demonstrate the benefit of modelling the structure of a discourse for the semantic analysis of the sentences composing it, in particular for problems such as anaphora, ellipsis, temporal structure, or lexical ambiguity, among others.

Modelling discourse can also benefit a variety of NLP applications. For example, in the task of Question Answering, causal relations between pairs of text spans can help locate the answers for a *why* question (Higashinaka and Isozaki, 2008), for example. In Automatic Summarization, discourse relations can help identify which text should be included in a summary (Marcu, 1997), and produce appropriate ordering of sentences in a summary (Barzilay and Elhadad, 2002). Discourse models can also be used to evaluate summary quality (Barzilay and Lapata, 2008; Louis and Nenkova, 2009). The task of Information Retrieval requires to identify complex events involving subevents, often expressed in separate sentences: the ability to infer how these subevents are related is thus needed, and requires an analysis of the discourse they appear in (Gaizauskas and Wilks, 1998).

This chapter focuses on models of discourse interpretation and computational approaches to discourse parsing. In Section 2.1, we start by underlining some aspects of discourse coherence, and in particular the role of lexical knowledge. We then focus on some of the most noteworthy theories of discourse structure and their differences in Section 2.2. In Section 2.3 we present manually annotated corpora based on these theories. Finally, in Section 2.4 we review existing approaches to automatic discourse parsing and especially the treatment of lexical information in such approaches.

2.1 Discourse interpretation

Generally speaking, a discourse is a sequence of natural language utterances. However, in order for a discourse to be intelligible, an important refinement is necessary. As stated by Jurafsky and Martin (2009), "*language does not normally consist of isolated, unrelated sentences, but instead of collected, structured, coherent groups of sentences. We refer to such a coherent structured group of sentences as a discourse.*"

The notions of *coherence* and *structure* are thus crucial properties of discourse. Indeed, understanding a discourse requires to find some meaningful structure among its sentences. A new sentence can only be interpreted by understanding how it relates to the information introduced by the previous discourse.

2.1.1 Discourse relations

Discourse relations are meaning relations which connect two discourse units. This notion was first introduced by Hobbs (1979) under the name **coherence relations** (as it is those relations that make a discourse coherent), following various similar proposals that used different names (Fillmore; Grimes, 1975; Halliday, 1976). In other theories they have been referred to as **rhetorical relations** (Mann and Thompson, 1988). Although originally conceived in distinct ways, these notions can be considered equivalent. In this thesis, following Asher (1993), we will use the term **discourse relations**.

Studies of discourse structure have proposed a variety of discourse relations. There appears to be an open-ended list of these relations (cause, goal, sequence, elaboration...). The entities corresponding to discourse units also vary across discourse theories. These different choices and their motivations will be presented in Section 2.2.

In Example 2.1, the second sentence is interpreted as being causally related to the first one: it can be inferred that John's state of *feeling full* is caused by the event of *eating fish*. In Example 2.2, the two events reported by the two sentences appear to be parallel parts of a broader event.

(2.1) John ate fish. He felt full.

(2.2) John ate fish. Mary ate beef.

The interpretation of the relations between discourse events relies on various sources of knowledge. In some cases, discourse relations are inferred on the basis of knowledge about specific lexical items: a discourse relation may be directly indicated by a **discourse connective**, as in Example 2.3 with the phrase *so that*. We present this notion in Section 2.1.2.

(2.3) John opened the door so that he could go out.

(2.4) John opened the door. He went out.

However discourse relations may be interpreted even in the absence of such connectives, as in Example 2.4. Other cohesive means can be used to indicate that a relation exists between two events, for example the reference to the same entities in both sentences (*e.g.*, *He* and *John* in Example 2.4). The identification of such cohesive ties requires mechanisms such as anaphora resolution. We discuss the link between coreference and coherence in Section 2.1.4.

In Example 2.4, another clue to the interpretation of the relation is the adjacency and ordering of the sentences.

However, the syntax and logical form of the sentences in the discourse do not always provide sufficient information to assign coherence. This is illustrated in Examples 2.5 and 2.6, which have the same syntax, but which receive different interpretations in two aspects.

(2.5) Max stood up. John greeted him.

(2.6) Max fell. John pushed him.

In terms of discourse relations, a causal relation is inferred in Example 2.6, while a narrative relation connects the events in Example 2.5. The temporal order of the events matches the order of the sentences in Example 2.5, whereas in Example 2.6 it doesn't: the event of *falling* happens after the event of *pushing*. In this case, knowledge about lexical associations is required. We turn to this question in Section 2.1.3.

2.1.2 Discourse connectives

Discourse connectives are the most obvious signals of discourse relations. For this reason, in the literature as well as in this thesis, relations that are signaled by a discourse connective are referred to as **explicit relations**, while all others are called **implicit relations**.

The notion of connective does not correspond to a single syntactic category but is primarily characterized by its discourse function, and therefore it is not easy to define precisely. Syntactically, connectives can be adverbs (*however*, *yet*), subordinating conjunctions (*when*, *if*) or coordinating conjunctions (*and*, *but*). They can also be multiword expressions, such as prepositional phrases (*due to*) or adverbial phrases (*for instance*). Discourse relations can also be conveyed by more complex, less frozen expressions. For instance, "*for this reason*" can be used to express a causal relation. Such expressions may allow for syntactic modifications, as in "*for these reasons*".

A major issue associated with discourse connectives is their potential ambiguity. Stede (2011) identifies three different kinds of ambiguity:

- Words interpreted as connectives in some contexts (*e.g.*, *since* in Examples 2.8 and 2.9) may also be used in contexts where they don't serve the purpose of conveying a discourse relation between discourse units (*e.g.*, *since* in Example 2.7); these two cases can be called **discourse reading** and **sentential reading**, following Hirschberg and Litman (1993);

- Some connectives can appear in multiple syntactic forms: *since* is used as a subordinating conjunction in Example 2.10, and as an adverb in Example 2.11, with a discourse reading in both cases;

- Some connectives can be used to signal distinct discourse relations in different contexts: in Example 2.8, *since* conveys a temporal relation, while in Example 2.9 it conveys a causal relation.

(2.7) John has been living in Florida since 2008.

(2.8) Since he arrived in Florida, John has set up two new companies.

(2.9) Since John comes from Chicago, he is well accustomed to lots of snow.

(2.10) John likes living in Florida, though he misses Chicago.

(2.11) John likes living in Florida. He misses Chicago, though.

The problem of ambiguity between discourse and non-discourse usage has been addressed in several studies and can be robustly resolved by classification models: Pitler and Nenkova (2009) report a 95% accuracy for their disambiguation model.

The ambiguity between syntactic forms can be resolved by simple part-of-speech tagging, and is not in itself a problem, since the relation holds in any case (although in some frameworks it is considered differently, for example the PDTB considers a difference between structural versus anaphoric relations, see Section 2.3.2).

The last type of ambiguity, which concerns connectives signalling different relations depending on the context, is in fact a question of underspecification. A relation holding between discourse units is sometimes just reinforced by the presence of a discourse connective. The disambiguation then relies on other aspects of the discourse units. In Example 2.8, the tense of the events indicate a temporal relation between them, while in Example 2.9, a causal relation can be inferred from the association between *coming from* and *being accustomed to*. We discuss the question of such associations in the next section.

2.1.3 Lexical associations

Discourse relations can also be conveyed by more subtle lexical means. For example, Danlos (2001) studied causal relations conveyed by causative verbs. Causative verbs are generally analyzed as complex predicates involving a causing sub-event which brings about a new state (Dowty, 1979; Chierchia, 1989; Levin and Rappaport, 1995). The effect of such verbs in discourse is illustrated in Example 2.12[1], which has a natural causal interpretation in which the action of *hitting* directly caused the resulting *cracked* state.

(2.12) Fred hit the carafe against the sink. He cracked it.

Kehler et al. (2008) also investigated Implicit Causality verbs, and performed psycholinguistic experiments aiming to find out which kind of causal relation are more likely to be expected by participants when prompted for continuation after such verbs. For instance, with the verb *fascinate*, continuing with an Explanation relation, such as Example 2.13[2] was found to be more likely than with a Result relation as in Example 2.14.

(2.13) Mary fascinated John. She always knew what to say.

(2.14) Mary fascinated John. He decided to ask her out.

Danlos also studied what she refers to as “discourse verbs”, such as *precede* or *cause*, which take as arguments eventualities or facts. She refers to these verbs as “discourse verbs”, by analogy with discourse connectives. In Example 2.15[3], a Narration relation is conveyed by the verb *precede*, between the event of *leaving* in the first sentence, and the event of *arriving* as argument of the discourse verb. Similarly, in Example 2.16, a Result relation is conveyed by the verb *cause*.

(2.15) Ted left. This preceded Sue’s arrival.

(2.16) Ted didn’t stop joking. This caused hilarity among his friends.

Psychological verbs (*i.e.* verbs denoting a mental state) have also been considered in terms on their involvement in discourse causal relations (Duque, 2014; Tantos, 2004). These verbs can be considered as a type of causative verbs, denoting complex events with their first subevent representing a process,and the second

[1] Example taken from Danlos (2001)
[2] Example taken from Kehler et al. (2008)
[3] Example taken from Danlos

representing a result connected with a causal link. Pustejovsky (1995) assumes that a specific kind of causation exists for psychological verbs, called Experienced Causation. Within the framework of his Generative Lexicon, he establishes the causative connection between the two subevents through semantic elements of the arguments represented in the lexical data structure of the verb. The causing and resulting subevents are represented in the agentive and formal roles, respectively, of the qualia structure of the verb, and the aspectual connection between the subevents is represented in the event structure.

Duque (2014) noted that when the resulting state and its causing event are presented in different sentences, a causal relation is created between the two sentences. In Examples 2.17 and 2.18[4], the first sentences appear as causes of the resulting state of *anger*.

(2.17) Marta failed the exam. Miguel got angry.

(2.18) Marta passed the exam. Miguel got angry.

However, while in Example 2.17, the causal relation relies on the intuiton that *failing causes anger*, the causal relation in Example 2.18 relies on a more complex inference: "*Miguel wanted Marta to fail*". The relation is not signalled by a single lexical item, but rather by the combination of the two predicates. This is also the case in Example 2.19, as noted by Wellner (2009).

(2.19) Max fell. John pushed him.

The standard discourse interpretation (*i.e.* before(*fall*, *push*)) is overridden by the lexical (or perhaps common sense) properties of the events in the two sentences. From the Generative Lexicon view, the lexical information required to infer the relation is represented in the qualia structure of the two lexical items. The structure for *push* includes lexical knowledge to the effect that "pushing results in some kind of movement". Coupled with similar lexical knowledge about *fall*, this can allow for the proper inference to be made.

2.1.4 Coreference

Cohesive phenomena like repeated reference to the same entities, *i.e* coreference, play a key role in discourse coherence (Halliday, 1976). Examples 2.20 and 2.21 show that the presence or absence of coreference in similar texts can change the inferred relation between two discourse clauses.

[4]Example taken from Duque (2014)

(2.20) Mark spent the evening with Jane. She felt lonely.

(2.21) Mark spent the evening with Jane. Anna felt lonely.

Indeed, the coreference between *Jane* and *She* in Example 2.20 implies that the event denoted by the first sentence is caused by the state denoted in the second sentence (*i.e.*, *Jane feeling lonely*), while the introduction of a new entity, *Anna*, in Example 2.21, implies that the state denoted in the second sentence, *Anna feeling lonely*, is caused by the event denoted in the first sentence.

While the fact that predicates denoting events share an argument can be an indication of their relation, the specific relation between this argument and each predicate, *e.g.*, as subject or object, can also have an impact. This is illustrated in Examples 2.22 and 2.23.

(2.22) Sarah hit John. She was furious.

(2.23) Sarah hit John. He was furious.

In Example 2.22, the subject of the second predicate refers to the subject of the first, and the first event is interpreted as a consequence of the second. In Example 2.23, the subject of the second predicate refers to the object of the first, and this time the second event is interpreted as a consequence of the first. We study the impact of different predicate-argument mappings in Section 4.3.

We have presented different notions which play a role in the mechanisms for natural interpretation of discourse. In the next section we present formal frameworks for discourse analysis and discourse structure representation, which need to account for all these notions.

2.2 Theories of discourse structure

A number of theories for the construction of discourse structure, that is a structure organizing the units a discourse contains and identifying the discourse relations connecting them, have emerged with the aim of defining frameworks for discourse analysis. Among these theories, the most noteworthy, in chronological order of their appearance, are the Linguistic Discourse Model (Polanyi and Scha, 1984), Grosz and Sidner (1986)'s approach, the theory of discourse coherence developed by Hobbs (1990), Rhetorical Structure Theory (RST, Mann and Thompson (1988)), the cognitive account of discourse relations by Sanders et al. (1992), Segmented Discourse Representation Theory (SDRT, Asher and Lascarides (2003)), and D-LTAG, a lexicalized Tree Adjoining Grammar for discourse (Forbes et al., 2003).

Current work on practical discourse analysis is mostly based on three of these approaches, as they offer fully-specified and well-defined sets of discourse relations: RST, SDRT, and the annotation framework of the Penn Discourse TreeBank (PDTB, Prasad et al. (2008)) which is not a model *per se* and does not offer full discourse structures or any constraints thereon, but is loosely based on D-LTAG. In this section we thus focus mainly on presenting the models defined in RST and SDRT and how they relate to other theories. We also present some aspects of D-LTAG adopted in the PDTB annotation framework.

The existing theories of discourse structure are based on different motivations. As a consequence, they differ with respect to several aspects, including:

- the definition of a discourse unit;
- the definition of a discourse relation;
- the specific set of relations and its granularity;
- the way of representing the structure (*e.g.*, tree vs. graph).

2.2.1 Theoretical motivations

Mann and Thompson (1988) present RST as a descriptive framework that "identifies the hierarchical structure of a text" and "describes the relations among text parts in functional terms". Thus, it does not attempt to have an explanatory or predictive value. Similarly to Grosz and Sidner (1986)'s theory, RST is based on the perspective of the speaker's communicative intentions and how the discourse was constructed according to these intentions.

On the contrary, SDRT aims to explain and not only describe discourse structure. It takes basis on the informational content of the clauses, and was defined in semantic terms. SDRT extends dynamic approaches like Discourse Representation Theory (DRT, Kamp and Reyle (1993)), and adds a separate nonmonotonic logic, the *glue logic*, to connect the different knowledge sources necessary to reason with discourse relations, which are defeasibly inferred. Among the various different knowledge sources taken into account in this theory are formally analyzed linguistic input and prior context, contextual knowledge specific to the discourse situation, information about discourse participants and their intentions and goals, and general knowledge about the world.

D-LTAG (Forbes et al., 2003) focuses on the syntax-semantic interface at the discourse level. It is based on the resemblance between the sources of discourse meaning and the sources of sentence meaning, and thus makes an analogy to syntactical structure. Formally, it is an extension of Lexicalized Tree-Adjoining Grammar (LTAG) to discourse (Webber and Joshi, 1998).

2.2.2 Defining discourse units

In order to model the structure of a discourse, it first needs to be segmented in elementary discourse units (EDUs). Theories provide different definitions of these units, according to their motivations.

RST is independent from any semantic formalism. Thus, discourse units are simply uninterrupted sequences of text, referred to as text spans.

In SDRT, an EDU corresponds to the semantic representation of the corresponding text. EDUs can be clausal or subclausal, *e.g.*, appositions or frame adverbials.

Similarly, in D-LTAG, EDUs are defined as the semantic representations of text, or abstract objects (*e.g.*, events, states, or propositions), although these units can represent clausal, sentential or multi-sentential utterances.

2.2.3 Defining discourse relations

There is no consensus on the set of discourse relations used in discourse, and the granularity varies widely between theories.

Cognitively inspired approaches, like that of Sanders et al. (1992), posit a strong connection between coherence structures and the cognitive processes involved in the production and processing of discourse. Sanders et al. (1992) proposed a taxonomy based on four dichotomous dimensions ("cognitive primitives") which combine to produce a full set of discourse relations:

- additive versus causal;
- basic versus non-basic order of segments (for causal relations only);
- positive versus negative polarity;
- semantic versus pragmatic source of coherence.

In particular, the distinction between semantic and pragmatic relations concerns the type of content they link. Semantic relations involve the semantic content of the units they link, while pragmatic relations involve beliefs or dialogue acts taking scope over the semantic content of one or both units they link.

In RST, the definitions of relations revolve around speaker intentions: the aim is to reconstruct the goals and plans that the writer followed in composing the text. This intended effect can be the mere recognition of the discourse relation by the addressee, or an influence on his beliefs, desires, and intentions. This distinction in terms of the intended effect yields two groups of relations: **subject-matter** versus **presentational** relations.

In the original formulation, Mann and Thompson (1988) proposed a set of 23 relations, motivated by empirical analysis of a variety of texts. However they considered the inventory of relations to be an open set, and there have been many variations to this set in later work grounded in RST, depending on the intended application. For example, the annotated RST Discourse Treebank (Carlson et al., 2007), described in Section 2.3.1, was built using a large set of 78 relations grouped into 16 coarse-grained classes.

In SDRT, discourse relations are defined on the basis of truth conditional semantics and tend to be less fine-grained than those in RST. Additional syntactic and semantic constraints, specific to each type of relation, are defined. For instance, a Contrast relation between two discourse units entails that they have parallel syntactic structures that induce contrasting themes, and a Result relation entails that the event described in one unit causes the event in the other unit.

In D-LTAG, discourse relations are not defined explicitly but rely only on the discourse connectives which trigger them. Implicit relations, conveyed by other means, are thus not considered in the original framework. This is not the case in the PDTB annotation framework, as we will discuss in Section 2.3.2.

2.2.4 Hierarchy of relations

Some discourse units present foreground information while others present background information. This leads to a difference between relations which link two units presented on the same level of importance, and relations which link one unit of foreground content and one of background content.

In RST, this translates as so called multinuclear and mononuclear relations, and the notions of Nucleus and Satellite. In mononuclear relations, one discourse unit is more important and referred to as the Nucleus, while the other, representing less important information, is referred to as the Satellite. The criterion to distinguish the Satellite from the Nucleus in mononuclear relations is that Satellite can be deleted without harming the coherence of a text. Multinuclear relations link two or more Nuclei of equal importance. The relations of this latter type are Sequence, Contrast, Joint and List, while all other relations are mononuclear.

In SDRT, these different behaviors roughly translate as a distinction between coordinating and subordinating discourse relations. These relation types were first introduced by Hobbs (1979), and also used by Polanyi (1988). Intuitively, some discourse units have the purpose of elaborating or supporting the ideas presented in other discourse units: they play a subordinate role to these other units. On the other hand, some discourse units come at the same level of detail and coordinate together to express an idea.

By default, Narration, Continuation and Result are coordinating relations, while Elaboration, Evidence, Explanation, Background and Purpose are subordinating relations. However, in the formulation of SDRT, these relation types are in fact not an intrinsic property of discourse relations, they can be overriden in certain contexts. The reasoning behind this, discussed by Asher and Vieu (2005), is due to the structure proposed by the theory and the constraints it follows. This is explained below.

2.2.5 Discourse structures

Complex discourse units (CDUs) are formed by linking EDUs with discourse relations. A hierarchical text structure arises by recursively applying discourse relations between CDUs and EDUs. The way in which discourse units are combined and the resulting structure differ between theories, however.

In RST, discourse structure is represented as a tree. In the original graphical representation proposed by Mann and Thompson (1988), a Nucleus in a mononuclear relation is identified by a vertical line above the corresponding unit. In a multinuclear relation, oblique lines start from each Nucleus and are joined at the top. This type of representation is illustrated in Figure 2.1 for the discourse presented in Example 2.24.

(2.24)
1. Bob shot the sheriff,
2. because he treated him badly
3. and tried to put him in.
4. He didn't shoot the deputy, though.

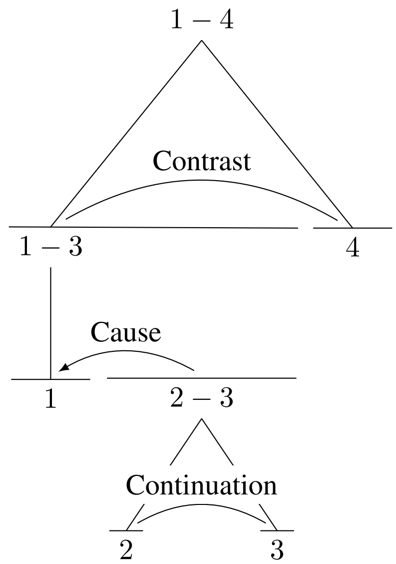


Figure 2.1: RST representation of Example 2.24, as originally proposed by Mann and Thompson (1988).

In a further development of RST, Marcu (1996) posited the Nuclearity Principle, which is related to the embedding of subtrees representing mononuclear relations in a complete tree. According to this principle, a relation between any discourse unit U_1 and a CDU U_2, itself formed by a mononuclear relation, also holds between U_1 and the Nucleus of U_2. In his formalization of RST, discourse structures are represented as binary trees, where the non-terminal nodes are discourse relations, and the leaves are EDUs. CDUs are sub-trees which also represent text spans, that is they are made of adjacent units only. In Figure 2.2 we show such a representation for Example 2.24. The Nuclearity Principle enables for instance **(1)** to be the argument of two discourse relations.

Another example of theory representing discourse structure as a tree is the Linguistic Discourse Model (Polanyi, 1988). In this proposal, a discourse tree is constructed by recursively applying a set of discourse construction rules (*e.g.*, coordination or subordination) to a sequence of basic discourse units. The choice of a particular rule is determined by information contained in the surface structure of the unit. Units are attached to the existing discourse tree on its right edge. This constraint, also known as the **Right Frontier Constraint** (RFC), is a very important constraint in discourse interpretation and anaphora resolution.

SDRT, in accordance with other theories (Wolf and Gibson, 2005), argue that

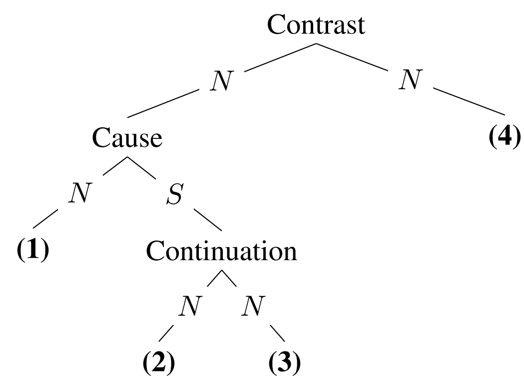


Figure 2.2: RST representation of Example 2.24, in the tree formalization proposed by Marcu (1996).

a much less constrained type of structure is needed to represent discourse coherence. Therefore, SDRT structures are represented as oriented graphs, where nodes can have multiple parents. Therefore, more than one discourse relation may hold between two discourse segments.

Graphically, a coordinating relation is represented by an horizontal arrow: both units linked by such a relation are on the same level in the structure. A subordinating relation is represented by a vertical arrow: the least important unit is placed one level down from the most important one. Figure 2.3 displays the SDRT representation for Example 2.24 above.

SDRT also implements the Right Frontier Constraint originally proposed by Polanyi (1988). In the SDRT framework, the RFC imposes contraints on the nodes, already present in the graph, to which a new constituent may attach. Only the nodes on the right frontier of the graph are accessible for new attachments, that is, the last node entered in the graph or one of the nodes that dominate this last node. This means that a coordinating relation pushes the right frontier to the right and only allows attachment to the last unit of the relation, whereas both units of a subordinating relation are accessible nodes. Indeed, the last node (representing the subordinate unit) is the most recent which confers it a high focus, and the dominant node contains more important information and is also in high focus. This notion of focus is also addressed in other theories, *e.g.*, Centering (Grosz and Sidner, 1986) which considers attentional focus, implying that some topics are more active than others.

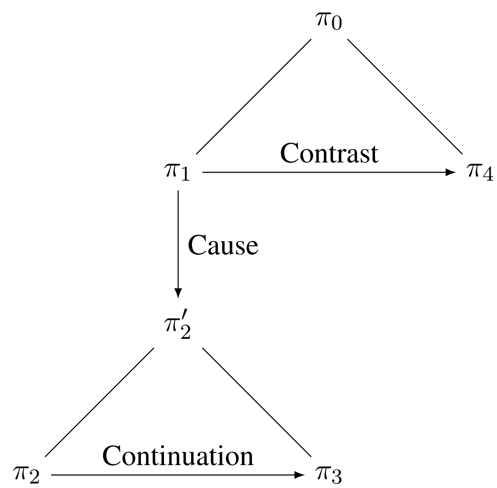


Figure 2.3: SDRT representation of Example 2.24.

Due to the RFC, the subordinating or coordinating quality a relation has by default needs to be overriden in certain contexts, as was shown by Asher and Vieu (2005) in a study based on linguistic tests to determine whether a relation is coordinating or subordinating. For instance, the Result relation is coordinating by default, but can behave like a subordinating relation in particular instances, as in Example 2.25[5].

(2.25) a. Lea bought a new car. b. As a result, she'll be able to go to Mexico this Christmas, c. and she will get to work quickly. d. It's a Subaru.

There is a Result relation between (a) and (b), as indicated by the marker *as a result*. The marker *and* indicates a Continuation between (b) and (c). The pronoun in (d), which is clearly coreferential with the *car* introduced in (a), can only be resolved if the Result relation is assumed as being subordinating. If it were coordinating, then segment (a) would not be on the right frontier of the discourse, and hence the *car* would not be an available antecedent for the pronoun in (d).

In D-LTAG, no such distinction between coordinating and subordinating relations is made. In this framework, elementary trees are constructed in the same way as syntactical predicate-argument trees in LTAG, with discourse connectives

[5]Example taken from (Asher and Vieu, 2005).

as predicates and two discourse units as arguments. It focuses on low-level discourse structures, constructed using only tree adjunction or insertion. This results in essentially linear structures, except for the presence of complex sub-structures. Because of this, it is not capable of resolving semantic and pragmatic phenoma like anaphora. In the PDTB, complex discourse units are not considered, since the aim is not to arrive at a complete discourse structure.

In this section we have seen that discourse structure can be formalized in different ways, depending on the underlying assumptions of each theory and its motivations. These different theories have been used as the basis for annotation frameworks applied to large corpora. We present three of these projects in the next section.

2.3 Annotated corpora

In discourse parsing experiments, the two most widely used English language corpora with manual discourse annotations are the RST Discourse Treebank (Carlson et al., 2007) and the Penn Discourse Treebank (PDTB, Prasad et al. (2008)), which are both based on the Wall Street Journal corpus. They differ in the discourse formalism they are grounded in: the RST Discourse Treebank offers annotated structures in the RST framework, detailed in the previous section, while the PDTB is based on Discourse Lexicalized TAG (D-LTAG, Webber and Joshi (1998); Webber et al. (2003)). In French, a similar discourse corpus is Annodis (Afantenos et al., 2012), based on SDRT, which was also introduced in the previous section. These three resources and their specificities are presented below.

2.3.1 RST Discourse Treebank

The RST Discourse Treebank (Carlson et al., 2007) is a resource containing discourse annotations in the RST framework on 385 Wall Street Journal articles from the Penn TreeBank. Following the Rhetorical Structure Theory, the annotation process, described in the RST Discourse tagging manual (Carlson and Marcu, 2001), consists in several steps leading to the building of an RST tree for each text.

The first step is the identification of the EDUs as building blocks of a discourse tree. Most EDUs correspond to clauses and are identified using lexical and syntactic cues. Example 2.26 contains a superordinate clause and subordinate clause containing a discourse marker as lexical cue, whearas in Example 2.27 the subordinate clause does not contain a discourse marker[6]. However clauses that are subjects

[6]All examples in this section are taken from the RST Discourse tagging manual (Carlson and Marcu, 2001)

(as in Example 2.28), objects (as in 2.29), or complements of a main clause (as in Example 2.30) are not usually EDUs.

(2.26) [Such trappings suggest a glorious past] [but give no hint of a troubled present.]

(2.27) [Xerox Corp.'s third-quarter net income grew 6.2% on 7.3% higher revenue,] [earning mixed reviews from Wall Street analysts.]

(2.28) [*Deciding what constitutes "terrorism"* can be a legalistic exercise.]

(2.29) [So far, it appears cautious about *taking the big step*.]

(2.30) [Ideally, we'd like *to be the operator of the project and a modest equity investor*.]

An exception is made in the case of clausal complements of attribution verbs, including both speech acts (Example 2.31) and other cognitive acts (Example 2.32): these are considered as separate EDUs (annotated with "attribution" as rhetorical relation) . This is a departure from the original theory: Mann and Thompson (1988) reject attribution as a rhetorical relation, and state that a reporting clause functions as evidence for the attributed material and thus belongs with it. The question of attributions will be elaborated further in Section 3.1.2.6.

(2.31) [The legendary GM chairman *declared*] [that his company would make "a car for every purse and purpose."]

(2.32) [Analysts *estimated*] [that sales at U.S. stores declined in the quarter, too.]

The next step of the annotation process consists in determining which adjacent discourse units are connected and assigning their nuclearity status. As stated in Section 2.2, in RST a relation can be mononuclear (containing two units, a nucleus and a satellite) or multinuclear (containing two or more units, all nuclei). For mononuclear relations, the nucleus (unit representing the more salient piece of information in the relation) and satellite (unit representing supporting information) have to be identified.

This step is often performed simultaneously with the last step: assigning a rhetorical relation. These discourse relations are identified empirically, based on evidence from the corpus. The original 23 discourse relations defined by Mann and Thompson (1988) are further divided into 78 finer-grained relations: 53 mononuclear and 25 multinuclear relations. These can be grouped into 16 coarse grained classes. These classes are listed in Table 2.1, with representative members of each class. Only one relation is allowed for each node. In ambiguous cases, the most general relation is selected.

Class	Representative fine-grained relations
Attribution	attribution, attribution-negative
Background	background, circumstance
Cause	cause, result, consequence
Comparison	comparison, preference, analogy, proportion
Condition	condition, hypothetical, contingency, otherwise
Contrast	contrast, concession, antithesis
Elaboration	elaboration-general-specific, elaboration-part-whole, example, definition
Enablement	purpose, enablement
Evaluation	evaluation, interpretation, conclusion, comment
Explanation	evidence, explanation-argumentative, reason
Joint	list, disjunction
Manner-Means	manner, means
Topic-Comment	problem-solution, question-answer, topic-comment, rhetorical-question
Summary	summary, restatement
Temporal	temporal-before, temporal-after, temporal-same-time, sequence
Topic Change	topic-shift, topic-drift

Table 2.1: Rhetorical relation classes and representative members used in the RST Discourse Treebank annotations.

The second and third step produce nodes linking units, creating larger text spans. They are repeated until a full RST tree representation is built for the whole text.

2.3.2 Penn Discourse TreeBank

The Penn Discourse TreeBank (PDTB, Prasad et al. (2008)) is another discourse annotated corpus, based on 2159 Wall Street Journal articles (it is a superset of the RST discourse treebank). Unlike the previous corpus, the PDTB is not based on the RST framework. As we mentioned previously, it is grounded in a lexicalized approach to discourse, namely Discourse Lexicalized TAG. The PDTB focuses primarily on identifying discourse connectives and their arguments, and does not necessarily create a structure covering the whole text. Therefore it differs from RST and other theories in which these connectives are only considered as possible lexical cues of rhetorical relations, and which focus on producing a complete discourse structure.

In the PDTB, explicit discourse connectives include subordinating conjunctions (*e.g.*, *when*, *because*, *although*) and coordinating conjunctions (*e.g.*, *and*, *or*, *so*, *nor*), as well as discourse adverbials (*e.g.*, *then*, *however*, *as a result*). The connectives can be modified by adverbs and focus particles. There are also paired connectives (*e.g.*, *on the one hand/on the other hand*), which are treated as complex connectives.

Discourse units are defined as abstract objects such as events, states, or propositions. These units can be clausal, sentential or multi-sentential. Relations can only link two abstract object arguments, labeled Arg1 and Arg2, where Arg2 is the clause with which the connective is syntactically associated, and Arg1 is the other argument. Unlike in RST, no difference is made with regard to the saliency of the arguments (that is, there is no notion of nucleus and satellite arguments). There are also no constraints on relative order of the arguments.

In most cases, the arguments are local and adjacent to the discourse connective: this is the case for subordinating and coordinating connectives, which are considered as "structural" in that their arguments are identified structurally from adjacent units of discourse, usually in the same sentence. In contrast, discourse adverbials are considered as "anaphoric" connectives: their first argument can be long-distance and must be identified anaphorically.

In practice, as detailed in the PDTB annotation manual (Prasad et al., 2007), the annotation process for relations explicitly expressed by discourse connectives consists in selecting the text spans corresponding to the connective and its arguments. In Example 2.33 the connective is structural and both arguments are located in the same sentence, while in Example 2.34 the connective is anaphoric and the first argument is found in the previous sentence[7].

(2.33) [The federal government suspended sales of U.S. savings bonds] <u>because</u> [Congress hasn't lifted the ceiling on government debt.]

(2.34) [Such problems will require considerable skill to resolve.] <u>However</u>, [neither Mr. Baum nor Mr. Harper has much international experience.]

Aside from considering discourse connectives as lexical triggers of discourse relations, D-LTAG also uses structure as a possible trigger. Indeed, following the

[7] All examples in this sections are taken from the the PDTB annotation manual (Prasad et al., 2007)

analogy with syntax, where adjacency between words can convey relations implicitly (*e.g.*, noun-noun modifiers), they posit that discourse relations can be implied between adjacent clauses.

Therefore, in PDTB annotations, each pair of adjacent sentences (separated by a period, a colon or a semi-colon) is considered as possible arguments of an implicit relation and must be annotated. If such an implicit relation can be inferred, a discourse connective best conveying the relation must be inserted (referred to as "Implicit connective"). When multiple relations can be inferred, several connectives may be inserted, as shown in Example 2.35.

(2.35) [The third principal in the S. Gardens adventure did have garden experience.] **Implicit** = because, for example [The firm of Bruce Kelly/David Varnell Landscape Architects had created Central Park's Strawberry Fields and Shakespeare Garden.]

If no connective can be inserted, three cases are considered, as stated in the PDTB annotation manual:

(a) AltLex, where a discourse relation is inferred, but insertion of an Implicit connective leads to redundancy in its expression due to the relation being alternatively lexicalized by some other expression;

(b) EntRel, where no discourse relation can be inferred and where the second sentence only serves to provide some further description of an entity in the first sentence (akin to entity-based coherence (Knott et al., 2001));

(c) NoRel, where neither a discourse relation nor entity-based coherence can be inferred between the adjacent sentences.

Examples 2.36 and 2.37 contain AltLex annotations. In 2.36, the expression signifying the relation is not considered a connective because its first part refers to the relation while the second part refers anaphorically to the first argument.

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

(2.37) [Ms. Bartlett's previous work, which earned her an international reputation in the nonhorticultural art world, often took gardens as its nominal subject.] **AltLex** [Mayhap this metaphorical connection made the BPC Fine Arts Committee think she had a literal green thumb.]

For Explicit, Implicit and AltLex relations, “senses” corresponding to the conveyed relation are annotated. Multiple senses can be tagged in cases of several possible simultaneous interpretations. Similarly to the rhetorical relations used in the RST Treebank, a hierarchical organization of senses was designed for PDTB annotations. This hierarchy is composed of three levels. The top level contains four major semantic “classes”. The second level defines finer-grained “types”, of which there are 16 in total. The third level further divides some of the types into “subtypes”, 23 in total. The first two levels of the hierarchy are reproduced in Table 2.2. The three levels of senses are not necessarily used in each annotation: in cases of ambiguity or uncertainty, only coarse grained relations can be specified.

Classes	**Types**
Temporal	Asynchronous
	Synchronous
Contingency	Cause
	Pragmatic cause
	Condition
	Pragmatic condition
Comparison	Contrast
	Pragmatic contrast
	Concession
	Pragmatic concession
Expansion	Instantiation
	Restatement
	Alternative
	Exception
	Conjonction
	List

Table 2.2: First two levels of the relation hierarchy used in the PDTB.

As in the RST Treebank, attribution is considered in the PDTB. However, it is not regarded as a discourse relation, but rather as supplementary information included in the annotation. Indeed, the arguments of a relation in the PDTB must be abstract objects, while attributions are relations between an agent and a proposition. These different approaches are discussed in Section 3.1.2.6.

2.3.3 Annodis

Annodis (Afantenos et al., 2012) is a French language corpus enriched with manual annotations of discourse structures. It is based on texts from different sources representing four text genres: short news articles from the French daily *Est Républicain*, encyclopedia articles (from the French Wikipedia), linguistics research papers (from *Congrès Mondial de Linguistique Française*) and international relation reports (from the *Institut Français de Relations Internationales*).

Annodis contains two distinct types of annotations: a "bottom-up" approach which aims to construct the structure of a discourse from EDUs linked by coherence relations, and a "top-down" approach which focuses on annotation of multi-level discourse structures, specifically topical chains and enumerative structures. As our main interest in this section is in discourse parsing, we only focus on the first approach, which is conceptually similar to the annotations offered in the RST Treebank and the PDTB.

Annodis differs from the corpora presented previously in that is it grounded in the SDRT framework[8]. This choice is motivated by the assertion that graph-based discourse structure is more expressive than those proposed in other theories (Venant et al., 2013), and allows, among other differences, long distance attachments between discourse units.

The annotation process involves segmenting a text into EDUs. EDUs are mostly clauses, appositions, and some adverbials. They are defined to contain at least one eventuality description, and often only one.

The next step consists in identifying which EDUs are related. There are no constraints on the relative position of linked EDUs in a text: they can be in different sentences and several sentences apart in the text.

A relation is then assigned to each pair of linked EDUs. The relation set used in Annodis is a (simplified) subset of the coherence relations proposed in SDRT, chosen to be more or less common to all the discourse structure theories, with an intermediate level of granularity. It correspond to the second level chosen in the PDTB, and is coarser-grained than the RST Treebank relation set. The resulting list of 15 relations is presented in Table 2.3.

The complex discourse units resulting from linked pairs of EDUs can then recursively be linked to other EDUs or CDUs, until the text is represented as a complete structure.

[8]The DISCOR project (Baldridge et al., 2007) is another corpus annotated with discourse structures in the SDRT framework, based on English-language texts.

Relations
Explanation
Goal
Result
Parallel
Contrast
Continuation
Alternation
Attribution
Background
Flashback
Frame
Temporal-Location
Elaboration
Entity-elaboration
Comment

Table 2.3: Relation set used in Annodis.

2.3.4 Comparison between the annotated corpora

The three presented discourse annotated corpora have been constructed using very different approaches to the annotation of discourse relations.

Most significantly, the PDTB does not aim to build a structural representation of the entire text. Rather, it focuses on explicit and implicit discourse relations linking EDUs. There is no concern of relating complex discourse units to construct a recursive structure, contrary to the RST Treebank and Annodis. However, since the problem of automatically identifying relations between EDUs in a reliable fashion is still under intensive research, the PDTB offers valuable data for computational approaches to discourse parsing.

Another difference between all the corpora concerns the adjacency of arguments. In the RST Treebank, only adjacent EDUs can be related in the RST tree. However, depending on how the final trees are interpreted, and specifically if a dependency parsing approach is applied, attachment to distant EDUs can exist (Li et al., 2014b). In the PDTB, attachment to distant EDUs is only possible for explicit relations, and specifically those expressed with so-called anaphoric connectives, since it is assumed that implicit relations only hold because of the structure, *i.e.* the adjacency of the arguments. In Annodis, there is no constraint on the distance between related EDUs.

Finally, since manual annotation of discourse relations is a very difficult task, we compare the corpora in terms of agreement between annotators.

For the RST Treebank, the identification of EDUs yields an inter-annotator agreement value of 0.97, and discourse relation assignment yields an agreement value of 0.75 (Carlson et al., 2003). Indeed, the latter task is much more difficult. In the PDTB, since arguments are less constrained, their identification is actually the more difficult task, while identifying the relation is guided by the discourse marker and is therefore slighlty easier. According to Prasad et al. (2008), agreement on the identification of both arguments of a relation is 90.2% for the explicit connectives and 85.1% for implicit connectives., while first-level relation identification reaches 94%, and 80% for the third level of relations. In Annodis, Afantenos et al. (2012) report a F-score of 66% on the task of attaching EDUs (identifying related pairs of EDUs). For discourse relation assignment, they report an inter-annotator agreement value of 0.40 for the full set of relations, and 0.57 when relations are grouped into coarser-grained classes. Since the annotations in Annodis are much less constrained, lower agreement values are expected.

In fact, considering the significant differences in terms of annotation framework in each corpus, no quantitative comparison can be made. However it appears that all three corpora contain viable data for discourse parsing, although it should be taken into consideration that Annodis is a quite smaller corpus and also less developed, since it is much more recent. It is however the only resource of the kind in French, and is thus very relevant to our work.

Consistency in manual annotations is indeed very important when these corpora are used for computational approaches to discourse parsing. Indeed, automatic classifiers need reliable experimental datasets and cannot outperform them. In the next section, we review some of these approaches and the difficulties they meet.

2.4 Discourse parsing

Discourse parsing aims at automatically constructing a structural description for a text, in accordance with a theory. As we have seen with manually annotated corpora, this is a difficult task even for humans. Moreover, the recognition of a coherence relation requires world knowledge, which is very difficult to simulate in a computational model.

Just like for manual annotations, a discourse parsing system must perform several steps in order to build the full structural description of a text (Stede, 2011):

- Identify the EDUs (segmentation)
- Determine which EDUs are related (attachment)
- Identify the relation linking each pair of related EDUs (labelling)
- Recursively build the full structural representation of the text, in the chosen framework (applying the first three steps in a recursive manner over the full text)

Some models aim to perform the complete task, while other approaches focus on specific steps. Indeed, fully parsing a discourse into its structural representation has revealed to be rather elusive. Most approaches aiming to build a full discourse structure use data from the RST Treebank, since the PDTB contains only annotations between EDUs and no full structure is given.

Early attempts at modelling discourse structures focused on the sentence level (Soricut and Marcu, 2003). When considering only the sentence, syntactical features were shown to be very useful. Such features are however not applicable for relations occuring intersententially, preventing the generalization of these models to document-level disourse parsing. They also do not consider the interdependence of sentence-level subtrees.

There have however been improvements in document-level parsing. Early attempts include Subba and Di Eugenio (2009) and Hernault et al. (2010), who obtained modest results but developed frameworks and features which inspired later research. Hernault et al. (2010) employ two classifiers in a pipeline, one for attachment prediction (between adjacent units only, since their work is based on the RST Treebank where it is necessarily the case that adjacent units are attached in the tree structure), and one for relation labeling.

Muller et al. (2012) also used two such classifiers, but proposed to use them to make joint predictions instead of using them in a pipeline. This approach was also followed by Feng and Hirst (2012) and Joty et al. (2015). Their results indicate that joint models outputting both the structure and the relation labels outperform independent models. Joty et al. (2015) additionnally built specific models for intra- and intersentential instances, and then combined both to obtain the final trees. This allowed them to use more adapted features for each case (mostly syntactical features for intrasentential instances and lexical features for intersentential instances), and to take into consideration the discrepancy in relation distribution between those two types of instances.

2.4.1 Identifying related EDUs

While the first step of discourse parsing, segmenting the text in EDUs, is considered the easiest (Fisher and Roark (2007) reported an F-score of over 90%), the second task, attachment prediction, has been shown to be very challenging. It consists in identifying which discourse units are connected in a discourse. This problem is often overlooked in discourse parsing approaches who consider this as given. In the PDTB, attachment to distant units only exists in annotations of explicit relations, while annotations of other types of relations (*i.e.*, Implicit, AltLex, Entrel and Norel) can only occur between adjacent instances.

For explicit relations, the problem is to identify the first argument (since the second argument is syntactically connected to the connective). Wellner and Pustejovsky (2007) proposed an approach to identify just the lexical head of the first argument, arguing that this is in fact sufficient for many practical purposes. They used a maximum-entropy ranker to determine the most likely head, considering only verbs, common nouns and adjectives as candidates. With this method they obtained over 74% accuracy.

Later on, Lin et al. (2010) developed a full discourse parser for PDTB-style annotations. Unlike the previously mentioned approaches aiming to build full discourse parsers in other frameworks, their parser is only concerned with relations between EDUs, since this is what is represented in the PDTB. By locating discourse connectives in a text, they could determine where explicit relations occurred, in which case they had to make a decision as to which span of text to consider as the first argument. Using a classifier based mostly on the position of the connective in the sentence and its part-of-speech, they predicted whether it was located in the same sentence as the connective or in a previous sentence (Prasad et al. (2008) found only 8 instances in the whole corpus where the first argument was in a following sentence, this case was therefore not considered). However, for the case of previous sentence, they always picked the immediately previous sentence, since this is the majority of cases and yields an F-score of 77%. This provides satisfying performance, but does not solve the question of how to predict more distant attachment. For the other types of relations, their model predicts whether each pair of adjacent EDUs should be attached or not, in the sense that they predict whether it is a Norel instance, or some other type. They do not provide their results specifically for this differentiation but for all non-explicit instances, where they manage an F-score 6% over the baseline of always predicting an Entrel instance.

The approach by Muller et al. (2012), mentioned above, is based on the SDRT framework and uses data from the Annodis corpus. SDRT discourse representations are directed acyclic graphs imposing only the Right Frontier Constraint. As a result, attachment to more distant units is allowed, as long as they respect this constraint. This allows less constrained annotations, but makes attachment prediction more challenging. They proposed to transform the SDRT representations into dependency graphs, in order to apply techniques from syntactic dependency parsing to compute probabilities of attachment between pairs of EDUs. They obtained an improvement of F-score of more than 5 points over previous baselines.

Li et al. (2014b) proposed a similar approach to parsing RST trees. Previously, approaches in the RST framework used techniques from syntactic constituency parsing, since RST trees are analogous to constituency based syntactic trees. Li et al. (2014b) found several limitations to this analogy, mainly concerning the design of appropriate production rules, the design of uniform features for the different levels of discourse units, and avoiding local maxima with reasonable time complexity.

Therefore, they proposed to convert RST trees to dependency graphs between EDUs, taking the nuclearity principle into account (the nucleus of a relation is considered as the head of the complex unit formed with the satellite, and discourse units related to this complex unit in the tree are attached to this head in the dependency graph). With this view, attached EDUs are not necessarily adjacent. They were then able to use dependency parsing algorithms in combination with classifiers to predict the dependency structure of documents. They used syntactic and lexical features, including a feature representing the semantic similarity of two EDUs based on the similarity of word pairs from each EDU (using WordNet for similarity values). This idea of using lexical features based on word pairs has been investigated in many attempts at predicting implicit relations, as we will discuss in the next section, and specifically in 2.4.2.2.

2.4.2 Classifying implicit instances

In this section we focus on the third step, the problem of relation labelling, and more specifically between implicit instances, which is more challenging than explicit relation labelling, as discussed below.

The first attempts at modelling discourse parsers used unsupervised models making use of the presence of discourse markers and cue phrases (Marcu, 2000). However, this type of approach is limited in several ways. First, it does not take into account the possibility that these expressions identified as cues might in fact, in some contexts, not signal a discourse relation. There is also an ambiguity in the specific relation signalled by the connective: some connectives may be used to express different relations depending on the context. These two aspects of ambiguity of the connectives were later studied by Pitler and Nenkova (2009), who proposed an effective method of disambiguation using syntactical features.

More importantly, an approach based solely on discourse markers to identify relations passes over implicit relations which do not contain those lexical cues. Pitler et al. (2008) showed that explicit relations can be identified with high accuracy (93% on the first level of the PDTB relation hierarchy) using only these connectives, but implicit relations are much more challenging and require a more elaborate analysis of the context in which they are used.

Before the release of the PDTB, manually annotated data for implicit relations was not available in sufficient amount to build reliable supervised models. To address this problem and attempt to predict implicit relations as well as explicit relations, Marcu and Echihabi (2002) proposed an approach in which two novel ideas particularly stand out since they inspired later research, and are relevant to our work.

The first idea is to exploit explicit relations, which are more easily identified automatically, to help predict implicit relations. They used a restricted list of a few unambiguous discourse connectives each associated with one of four coarse-grained discourse relations (Contrast, Cause-Explanation-Evidence, Condition, Elaboration). For each connective they applied simple patterns to automatically extract related EDUs from a large non-annotated corpus. For example, the four patterns used to extract Contrast relations are the following:

- [BOS $\underbrace{\ldots}_{W_1}$ EOS] [BOS But $\underbrace{\ldots}_{W_2}$ EOS]
- [BOS $\underbrace{\ldots}_{W_1}$] [but $\underbrace{\ldots}_{W_2}$ EOS]
- [BOS $\underbrace{\ldots}_{W_1}$] [although $\underbrace{\ldots}_{W_2}$ EOS]
- [BOS Although $\underbrace{\ldots}_{W_1}$,] [$\underbrace{\ldots}_{W_2}$ EOS]

In these patterns, BOS and EOS denote BeginningOfSentence and EndOfSentence boundaries. These show that they extracted both inter- and intrasentential examples. W_1 and W_2 represent the words in the extracted spans. Note that the connective that served to identify the relation is discarded, since the aim is to allow the prediction of implicit relations, in which no connective is used to signal the relation. Instead, they paired the words from each span, and used this lexical information as features to train Naive Bayes binary classification models, one for each pair of relations. The second idea which significantly impacted later approaches lies in these word pair features. In fact it directly stems from the first idea, *i.e.*, using information from explicit relations to predict implicit relations: the connectives are ignored, and instead the focus is put on other lexical cues that may signal the relation.

We first focus on the implications of the first aspect we described, before examining the second one.

2.4.2.1 Redundancy hypothesis

The approach proposed by Marcu and Echihabi (2002) can be seen as creating artificial examples of implicit relations, since they use the connectives in explicit relations to identify the relation, and then extract other information from the arguments, not relying on the connective for predictions. This same idea was followed by Sporleder and Lascarides (2005), who successfully improved the model by including additional linguistic features. However, it is difficult to draw any conclusion from both of these studies, since they did not test their models on natural implicit examples.

In later work, Sporleder and Lascarides (2008) analyzed the approach of artificially producing implicit examples from explicit examples as making two assumptions:

- *There has to be a certain amount of redundancy between the discourse marker and the general linguistic context,* i.e.*, removing the discourse marker should still leave enough residual information for the classifier to learn how to distinguish different relations*;
- *Marked and unmarked examples have to be sufficiently similar that a classifier trained on the former generalises to the latter. In particular, properties which are predictive of a given relation in unmarked examples should also be predictive of the same relation in marked examples*.

Concerning the first assumption, since accurate relation prediction can be obtained when testing on artificial examples, it seems that there is indeed a certain amount of redundancy between the connective and the context: without the connective, sufficient information remains in the examples to identify the relation.

However, to study this question further, Sporleder and Lascarides (2008) tested the same models that were proposed by Marcu and Echihabi (2002) and Sporleder and Lascarides (2005) on natural implicit examples which were manually annotated. This resulted in a significant loss of accuracy (of more than 16%, from 42.34% to 25.92%, for the Naive Bayes model proposed by Marcu and Echihabi (2002), and of 35%, from 60.88% to 25.80%, for the model proposed by Sporleder and Lascarides (2005)). The fact that the models failed to generalize to natural data led Sporleder and Lascarides (2008) to infer that the second assumption may be too strong, and that explicit and implicit examples may in fact be too linguistically dissimilar to allow generalization.

The release of the PDTB allowed further testing of this method on natural examples, since it natively contains annotations for implicit relations clearly distinguished from explicit relations. This was exploited by Blair-Goldensohn et al. (2007), who reimplemented the approach proposed by Marcu and Echihabi (2002) and tested it on PDTB implicit instances and on artificial implicit instances for comparison. Like Sporleder and Lascarides (2008), they reported a significant drop in relation prediction accuracy for the natural implicit instances. To counteract the possible negative effect of training and testing on different corpora, they also attempted to test on artificial examples this time extracted from explicit instances of the PDTB, in which case they did not encounter a significant loss of performance. These experiments led them to the conclusion that artificial instances cannot be treated as fully equivalent to natural ones.

This result was confirmed by Pitler et al. (2009), who found that training on artificial data either from a large non-annotated corpus or from the PDTB explicit instances yielded less accurate results than training on natural data from the PDTB implicit instances. By computing the information gain of all word pairs produced by Blair-Goldensohn et al. (2007) and observing the most informative pairs, they also found that most of these pairs unexpectedly contain a discourse connective. Since the discourse connective which triggered the original extraction of the pair had been removed, the discourse connectives found in the word pairs must in fact have been used in addition to this triggering connective. From their Example 2.38 , which would be extracted as an instance of the Cause relation due to the connective *because*, Pitler et al. (2009) analyzed that the model learns that the presence of the pair *but-but* is an indicator of a Cause relation. This is one likely source of discrepancy between artificial and natural implicit instances, which would explain why a model trained on the first type of instances seems to be ill-adapted to the latter.

(2.38) The government says it has reached most isolated townships by now, *but because* roads are blocked, getting anything *but* basic food supplies to people remains difficult.

Because of this discrepancy, Pitler et al. (2009) and other research focused on attempting to use natural instances to build their models. However, they were confronted with a new challenge: the amount of annotated natural instances is very limited, leading to very sparse data. To address this limitation, a more advanced design of features is necessary.

2.4.2.2 Word pairs as features

As we have seen, the idea of using word pairs as features for the classification on implicit relations, put forward by Marcu and Echihabi (2002), was further studied in many later approaches.

In their first implementation, Marcu and Echihabi (2002) based their model on the complete list of words included in the two spans of each extracted instance, excluding the unambiguous discourse connective which allowed them to identify the instance, and operated no preprocessing on these (*e.g.*, stemming etc.).

They used the Cartesian product over the words ($W_1 = \{w_1...w_n\}, W_2 = \{w_{n+1}...w_m\}$) in the two spans to build their Naive Bayes model. The most likely relation is given by:

$\arg\max_{r_i} P(r_i|W_1, W_2)$, which equals to:

$\arg\max_{r_i} P(W_1, W_2|r_i)P(r_i)$.

By assuming that the word pairs are independent (which we should note is a strong assumption), $P(W_1, W_2|r_i)$ becomes equivalent to:

$\prod_{(w_i,w_j)\in W_1,W_2} P((w_i, w_j)|r_i)$,

and the values $P((w_i, w_j)|r_i)$ can be obtained by maximum likelihood estimation.

They also attempted to reduce the noise in their training data (*i.e.*, words that are not likely to be good predictors of discourse relations) in a second experiment, by filtering out these likely uninformative words based on part-of-speech information: only verbs, nouns and cue phrases (other than the discourse connective used for extraction) were retained. Although they obtained lower performance with this data, they hypothesized that it was due to the lesser amount of training data available after filtering and that more training instances would allow better results with their filtered data.

Lapata and Lascarides (2004) also used word pairs with specific parts-of-speech as features, although they aimed to predict which connective should be used to express a relation, and not to label the relation. They used features based on co-occurrence of pairs of lemmatised verbs, nouns and adjectives. They also included features obtained by mapping verbs and nouns to broad semantic classes to capture meaning regularities: for verbs, two separate features were created by mapping each element of a pair to its WordNet class (Fellbaum, 1998) for the first feature and its Levin class (Levin, 1993) for the second feature; for nouns, they also used the WordNet taxonomy. Additionally, other syntactic features were used. The features with the most impact turned out to be verb pairs, both in their lexical forms and with class mapping. Nouns and adjectives were however much less informative.

Blair-Goldensohn et al. (2007) experimented with several refinements of the word pair features. They reported that tokenizing the words, limiting the vocabulary size to the most frequent tokens (computed over the Gigaword corpus (Parker et al., 2011)) and applying a minimum frequency cutoff on wordpairs all improved the classification accuracy, while applying a stoplist to filter out the most frequent tokens degraded the results. These results imply that function words, which tend to be the most frequent words, have significant impact on the predictions. This finding, which corroborates the low performance obtained by Marcu and Echihabi (2002) with filtered data (which in effect excludes function words), seems counter-intuitive: indeed, one would expect these function words to be uninformative with regard to the signalled relation.

Aiming to improve the automatic extraction of spans involved in the explicit relations used to create their artificial implicit instances, they experimented with topic segmentation and syntactic heuristics. In topic segmentation, they marked topic boundaries to allow their patterns to select spans of multiple sentences if they concern the same topic, and to prevent them from identifying a relation between spans concerning different topics. This method did not yield significant improvement.

They also used syntactic parsing to cut out irrelevant constituents. In Example 2.39[9], the temporal relative clause, which is uninformative to the contrast relation, is cut out, and only the spans contained in square brackets are extracted as arguments. They remarked however that in some instances, this process cuts out informative content, and noted that designing appropriate heuristics is a difficult task. In Example 2.40, *drop* and *sold* would be used by humans to infer a causal relation but *sold* is excluded.

(2.39) For the past six months, [management has been revamping positioning and strategy,] but [also scaling back operations.]

(2.40) Wall Street investors, citing [a *drop* in oil prices] because [weakness in the automotive sector], *sold* off shares in GM today.

They tested whether these heuristics allowed them to select the correct spans when applied on PDTB explicit instances by comparing the resulting extracted spans with those selected by human annotation. They found positive results, with

[9] Examples taken from Blair-Goldensohn et al. (2007)

half of the sentences correctly processed by their heuristics, and the main source of problems in incorrect cases being caused by the presence of attribution (reported speech). When using these modified instances to build classification models, they found slightly lower accuracy compared to the original instances when testing on artificial data, but slightly better accuracy on natural data, although none of these differences were statistically significant. As in Marcu and Echihabi (2002)'s results with part-of-speech filtering, they posit that this may be due to the reduced amount of training data, and that if more data were available their method might yield significant improvement.

Again, the study of the information gain of word pairs led by Pitler et al. (2009) confirmed the finding that the most impactful word pairs contain function words, in addition to discourse connectives as mentioned before. This seems to suggest overfitting on the artificial training data.

After this analysis, they focused on defining features which could exploit the sparser but natural data offered by the PDTB. As stated earlier, they compared the original word pair features with those obtained from the PDTB instances (with a set of features from explicit instances and another from implicit instances), but also crossed the results from artificial and natural data by including only word pairs from PDTB implicit instances that were found informative in their initial study on word pairs from artificial instances. An analysis of the impact of each feature type concluded that the best word pair features were the latter type for Comparison and Contrast, and those from all PDTB implicit instances for Expansion and Temporal.

They also designed additional features based on the semantic properties of the spans, which were extracted from PDTB instances only. To deal with the sparsity problem, they used Levin verb classes in a feature representing the number of pairs of verbs from both spans belonging to the same class, among other verb pair features. They also included features based on word polarities in each span, obtained from the Multi-perspective Question Answering Opinion Corpus (Wilson et al., 2005), as well as other semantic or contextual features. These additional features did not have much impact on their results, allowing a gain of F-score of at most 2 points compared to the word pair features depending on the relation, and negative results for the Comparison relation model.

Park and Cardie (2012) applied a greedy feature selection based on the same features as Pitler et al. (2009), plus a few others from different approaches. In particular, they used lexico-semantic features inspired by Wellner et al. (2006) based on the first and last words appearing in each span, as well as syntactic features based on production rules, first introduced by Lin et al. (2009). Their feature selection from this diverse set of features achieved an even higher performance than previous studies. Surprisingly, they found the original word pair features to have a low impact on the results. They do not mention if they applied any preprocessing on the words (*e.g.*, stemming), but it seems that they did not, which could explain these results. However when clustering the verbs into their Levin verb classes, they found positive results, which supports the idea that raw word pair features are too sparse, and grouping them by similarity is beneficial.

Another approach to the sparsity problem was proposed by Biran and McKeown (2013), also based on word pair features, but involving the aggregation of these word pairs. They automatically extracted artificial instances from the Gigaword corpus, albeit with a different method than that of previous approaches. Instead of using patterns, they simply extracted the words occuring before and after each marker (from the list of 102 markers provided in the PDTB), within the same sentence, and formed the word pairs by taking the cross-product of these words. Instead of working with word pair frequencies, they aggregated the word pairs by the marker which triggered their extraction, and associated a weight to each word pair computed by TF-IDF . To train their classifier on the PDTB, they used one feature per marker (102 features) which was computed as the cosine similarity of the word pairs in each PDTB instance with the marker's set of word pairs. With these features, they obtained slightly better results than those of Pitler et al. (2009) based on word pairs only (no additional features), with an F-score gain of 2 to 3 points for all relations except Contingency where their model resulted in a loss of more than 1 point. When using the additional features proposed by Pitler et al. (2009), they were however outperformed by previous models.

Rutherford and Xue (2014) proposed to use a more compact representation of the words occuring in each word pair. To obtain this representation, they used Brown clustering (Brown et al., 1992), which allows to form clusters of words based on co-occurrences in a large corpus and thus reveals semantic commonalities between words. Using externally provided Brown clusters (Turian et al., 2010), they replaced each word in a pair by its assigned cluster. This denser representation of word pairs allowed them to achieve better classification performance than all previous attempts, when training and testing on the PDTB data. Addtionally, they proposed to use coreference-based features to represent another aspect of the semantics involved in span pairs. More specifically, they included binary features indicating whether the spans made reference to the same entities as subjects or objects of the main predicates. Previously, coreference-based features had been introduced by Louis et al. (2010) who did not obtain conclusive results but suggested it was an area worth exploring. With more complex coreference patterns, Rutherford and Xue (2014) found these features to be very effective.

Recently, Braud and Denis (2015) performed a detailed comparison of previous representations of word pairs and their effects on implicit relation classification, and also proposed new representations. They compared so-called *one-hot* representations, where each word pair corresponds to a single component of a very-high dimensional vector (leading to the previously mentioned sparsity issues), as well as the Brown cluster representation proposed by Rutherford and Xue (2014). They also investigated so-called *word embeddings*, where a word is represented as a low-dimensional vector with each dimension corresponding to a latent feature of the word. Word embeddings were previously used in discourse parsing by Li et al. (2014a). This type of representation has the limitation that the embeddings are not learnt specifically for the task, but are taken from other applications (*e.g.*, tagging tasks (Collobert et al., 2011)). Finally they explored distributional semantics, in which words are considered similar if they occur in similar contexts, and are represented in terms of their co-occurrence with other words. For all representations, they also used either all words from span pairs, or only verb pairs. In addition to these features based on different representations of word pairs, they included additional features introduced in the previous studies. They found that denser representations always outperform raw word pairs, but the best dense representation depends on which relation is being predicted. They also found that using all word pairs performs better than using verb pairs only.

2.5 Conclusion

We have seen in this chapter that lexical information plays a very important role in discourse interpretation, and that it can be used as a powerful tool for computational approaches to discourse parsing.

Many systems have attempted to find accurate representations of lexical information found in discourse clauses in order to predict the relations linking them. Some approaches use words or pairs of words from each clause as features for their models, which yield very sparse representation, ill-adapted for the construction of an accurate model. Therefore, other studies have aimed to find more abstract representations of such word pairs.

Instead, we propose to first collect word pairs linked by discourse connectives, and infer semantic relations linking them by making use of the redundancy hypothesis. These pairs can then be represented by the strength of their association with these relations. By encoding these associations in an external resource, we then have direct access to sparse representations of word pairs based on their association with different types of semantic relations, which can then be used for the prediction of implicit discourse relations.

These semantic links are not only relevant for discourse parsing, but can be useful for the definition of lexical entries in relational approaches to lexical semantics, which we have discussed in the previous chapter.

CHAPTER 3

EXTRACTING SEMANTIC RELATIONS USING DISCOURSE CUES

Contents

The previous chapters have been focused on the background and motivations behind the work presented in this thesis. In this chapter we describe the core of this work.

In Section 3.1 we detail the motivations underlying the design of our method for inferring regular semantic relations between pairs of predicates. In Section 3.2 we detail its practical implementation, and the resulting resource, which contains triples each composed of a pair of predicates and a relation, associated to the number of occurrences of these triples found in corpora.

The resource is called *Lecsie*, for *Linked Event Collection for Semantic Information Extraction*.

3.1 General concept

In this section, we detail the assumptions and principles motivating our method for extracting relevant predicates and infering lexical relations.

3.1.1 Existing approaches and motivations

Our work is at the interface of semantic relations and discourse analysis. We present existing approaches to the extraction of semantic relations between verbs and to the representation of lexical information contained in word pairs for discourse parsing, and the motivations at the basis of our choice of approach.

3.1.1.1 Extracting semantic relations between verbs

In Section 1.2.4, we have detailed how most approaches to the automatic extraction of semantic relations between verbs rely, at least in part, on discourse information and require to consider context above the sentence level.

Some approaches use patterns constructed around discourse connectives, such as that of Chklovski and Pantel (2004) in their work for the extraction of fine-grained semantic relations between verbs, or Inui et al. (2005) who aim to acquire causal relations between events.

Another manner to exploit discourse coherence is to identify verbs which share some arguments, as was done by Pekar (2008) to predict entailment relations between verbs, or by Chambers and Jurafsky (2008) to extract narrative chains of events.

Yet another approach, proposed by Tremper and Frank (2013) is to use temporal sequence, among other features, to distinguish between fine-grained semantic relations linking verbs, and in particular presuppostion and entailment relations.

Finally, we are closest in spirit to the approach of Roth and im Walde (2014), who used discourse connectives to construct vector-based representations of word pairs, and use those representations as features for the prediction of antonymy, synonymy and hyponymy relations.

However, while they use a supervized method to exploit their vector-based representations for the prediction of semantic relations, we propose a fully unsupervized approach, where the connectives are grouped according to the relations they tend to denote, and the association between the extracted pairs of predicates and these relations are quantified based on association measures. We also propose to use the information offered by shared arguments to better specify the predicates.

3.1.1.2 Representing lexical information for discourse parsing

As we have seen in Section 2.4.2, lexical features can be very useful to predict implicit discourse relations. Many approaches in the literature use word pairs extracted from each discourse segment as features. Representing these features in a dense fashion is however problematic. As a reminder, we briefly review the different solutions proposed in the literature.

Early approaches used so-called one-hot encoding, where each word pair corresponds to a single component of a very high dimensional feature vector (Marcu and Echihabi, 2002; Sporleder and Lascarides, 2005). This leads to sparsity issues since there is a very large amount of possible combinations of words.

To address this problem, different solutions have been proposed to obtain denser representations of word pairs. An approach explored by Lapata and Lascarides (2004) and Pitler et al. (2009) is to represent each word by its class in a taxonomy, for instance WordNet, or the Levin classification for verbs. Another possibility is to use a cluster representation of the word pairs, as was proposed by Rutherford and Xue (2014) with Brown clusters. Braud and Denis (2015) have proposed to use word-embedding representations. All these ideas, although successful in circumventing the issue of sparse representations, have the drawback of using representations not specifically designed for the task of discourse parsing, and therefore not necessarily adapted.

An approach to feature representation is that of Biran and McKeown (2013). As detailed in Section 2.4.2, they extracted words occurring before and after discourse connectives in a large annotated corpus (Gigaword), and aggregated the resulting word pairs occurring with the same connective. They then associated a weight to each word pair, defined as a function of the number of occurrences of the word pair and the total number of occurrences of word pairs with the considered connective. These weights are therefore computed based on considerations on the usage of each word pair in discourse, providing features adapted to the task. This method however yields only slightly better results than other approaches.

We postulate that this is due to the lack of syntactic and semantic consideration in the extraction of word pairs. It seems more sensible to target meaningful words in each segment instead of using all of them or focusing on words in the direct neighborhood of the connective. Indeed these words do not necessarily contribute much to the overall meaning.

Instead, we propose to focus on the event predicates contained in each segment, to obtain a better representation of their lexical content. We also propose to group connectives by the relation they lexicalize (the problem of ambiguity will be addressed in Section 3.1.2.2). Pairs of events can then be aggregated according to lexical relations. The resource we describe specifically contains association scores for pairs of events with diverse lexical relations, providing well-adapted features for the task of predicting discourse relations.

3.1.2 Principles

The general idea behind our method is to use discourse connectives to extract related event predicates, yielding triples consisting of the pair of events and the relation lexicalized by the discourse connective which triggered their extraction. We then assign each triple several scores aimed at representing the significance of their association.

3.1.2.1 Discourse markers and the redundancy hypothesis

The idea of using discourse connectives to extract lexically related events implicates assuming that the relation is not only lexicalized by the connective, and that the combination of events expressed in the segments it links also supports the relation. This is similar to the so-called redundancy hypothesis, as described by Sporleder and Lascarides (2008) for implicit discourse relation prediction. This hypothesis was formulated for approaches based on creating artificially implicit relations by removing the connective from explicit relations. Since they extract lexical features from these artificial instances, the linguistic context needs to contain information on the implied relation in addition to the discourse markers.

This hypothesis was proved to at least partially hold, since studies by Marcu and Echihabi (2002) and Sporleder and Lascarides (2005), among others, have found promising results when training models on artifical implicit instances (*i.e.*, explicit instances where the marker has been removed) and testing on similar data.

Sporleder and Lascarides (2008) state that a second assumption is implied by these methods: explicit and implicit instances need to be sufficiently similar to allow generalization to implicit instances when training on explicit instances. The soundness of this assumption has however been found to be questionable, since there is a significant drop in performance when the models described above are tested on natural instances instead of artificial instances similar to the training data (Sporleder and Lascarides, 2008).

In our application, the redundancy hypothesis seems to be the most important: we are only looking to use explicit instances to extract semantic relations between predicates that can also hold without a connective. These semantic relations can in turn potentially be used for the prediction of implicit discourse relations, where the resulting resource can provide features representing the lexical information contained in the pairs of predicates, but should/could be used in addition to other features based on the specific properties of implicit instances. In this fashion, the issues caused by the second assumption can be minimized.

3.1.2.2 Discourse relations and ambiguity

As stated previously, we group connectives by the relation they lexicalize in order to obtain associations between lexical relations and pairs of events. Therefore, we need to define a set of relations we want to represent, and to assign one of these relations to each connective.

Ambiguity in relations

This poses the issue of ambiguous connectives, which we described in Section 2.1.2. Connectives can be ambiguous in terms of their usage, to mark a discourse relation or solely to contribute to the sentence meaning. This ambiguity has been shown to be accurately resolvable by classification models (Pitler and Nenkova, 2009), and is not an issue in our method since we only consider connectives linking predicates, as we will detail in Section 3.2.2.

Another level of ambiguity concerns the discourse relations themselves: depending on the context, one connective can be used to establish different relations. In our approach, we chose to focus on non-ambigous connectives, keeping as a possible perspective the implementation of disambiguation techniques in order to include ambiguous connectives in our set.

Relation set

We focus on relations which are common to most theories of discourse analysis, in order to remain as theory-neutral as possible, and to produce results which could be compatible with several theories and taxonomies. We also aim to find a compromise between reasonably fine-grained relations and minimized relation ambiguity (with coarser grained relations, connectives are less ambiguous).

These relations can be grouped into four classes (Prasad et al., 2008): causal (*contingency*) relations, temporal relations, comparison relations (mainly contrast type relations), and expansion relations (*e.g.*, elaboration or continuation).

There are some variations between the sets of relations we use for French and English, mainly due to differences in usage of discourse connectives in each language, as well as differences on the available lexicons of connectives and associated relations. Our specific sets and their correspondence with the sets used in the external resources we worked with are presented in Section 3.2.1.3.

3.1.2.3 Syntactic parsing

Lexico-syntactic patterns have often been used to automatically extract elements of interest in large corpora. In Section 1.2.2 we discussed pattern-based methods for semantic relation extraction, such as the VerbOcean resource (Chklovski and Pantel, 2004). They used for instance the pattern "*whether to V1 or V2*" to extract *antonymy* relations between the verbs V1 and V2. Such a pattern would allow to extract the pair {*stay*, *go*} in Example 3.1, but would miss slightly more elaborate instances such as Example 3.2.

(3.1) But often, the question of *whether to* **stay** *or* **go** isn't simple.

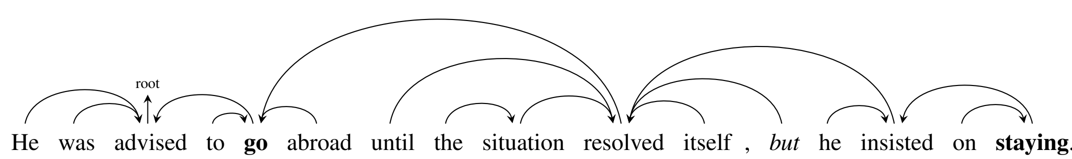
(3.2) She couldn't decide *whether to* **stay** until the end of the party *or* **go** home early.

Some of their patterns do allow for some variation with the inclusion of one additional word, such as "*to V1 * but V2*", where * matches any single word. This is helpful in instances such as Example 3.3.

(3.3) Many couples try *to* **stay** together *but* **go** to different colleges.

However such fixed patterns are very limited in terms of the variations they can allow, and for instance cannot handle appositions or subjunctive clauses placed between the two verbs. Instances such as Example 3.4 clearly need more flexible patterns.

(3.4)



In order to achieve precise identification of the relevant clauses, and precise extraction of the relevant verbs, our method relies on syntactic patterns based on dependency parsing. In Section 3.2 we present how we can exploit dependencies such as those displayed in Example 3.4 to extract the relevant predicates (shown here without dependency labels and parts of speech for more clarity, although we do use them as well).

3.1.2.4 Adjacent clauses as discourse arguments

Another hypothesis crucial to our extraction method is that adjacent clauses are often arguments of discourse relations. The syntactic dependencies allow us to infer whether the relation is intrasentential or intersentential, as we will demonstrate in Section 3.2.2. In intrasentential cases such as Example 3.4 above, the verbs in each connected clause can be identified thanks to the dependencies as well.

However, in intersentential cases such as Example 3.5, it is more difficult to identify the first clause (that is, the one that does not contain the connective). In the PDTB, Prasad et al. (2008) report that in 77% of explicit intersentential instances, the first clause is located in the immediately previous sentence.

(3.5) Ahmed had the flu and should have stayed home. However, he went to work anyway.

Therefore, considering the amount of data our method aims to extract by processing very large corpora, intersentential instances are always assumed to hold between consecutive sentences. This simplification introduces minor imprecisions, which can be accounted for by computing association measures instead of considering simple raw counts. We address the question of measuring association strength in Chapter 5.

3.1.2.5 Event predicates

In most cases event predicates are lexicalized by verbs. In other cases, they can be lexicalized by predicative adjectives following copulas, or nouns (mainly deverbal nouns).In our method, we focus on extracting verbs as well as predicative adjectives, although we intend to extend it to deverbal nouns in future work.

Refining verb meaning: Phrasal verbs

The context of each verb may contain relevant information to distinguish between different possible meanings of the verb. For this reason, our method not only extracts verb lemmas, but also aims to identify verbal phrases by exploiting information included in the dependencies of the verb, as well as some external resources when necessary. The practical means to do this will be described in Section 3.2.3.1, here we present the different aspects we are interested in.

We aim to identify prepositional verbs and particle verbs to disambiguate between different meanings of the same main verb. For example, the French verb *tenir* means ~*to hold* on its own, but *tenir de* means ~*to take after* while *tenir à* means ~*to care about*. Similarly, the English verb *to look* might take alternative meanings when used in combination with prepositions or particles, *e.g.*, *to look after* or *to look up*. It is however necessary to differentiate between prepositional verbs, where the preposition is part of the verb (*e.g.*, *look after* in Example 3.6), and prepositions used as part of temporal or spatial complements (*e.g.*, *after midnight* in Example 3.7: the extracted verb should be *leave* and not *leave after*).

(3.6) Mark looks after his brothers every afternoon.

(3.7) Mary left after midnight.

We also aim to identify verbal locutions, which are commonly used in French, as in *prendre garde* (~*to beware*), *faire référence* (~*to refer to*). Another specificity of French which is relevant for sense disambiguation is the use of reflexive articles: for example, *agir* means ~*to act*, while *s'agir* means ~*to concern*.

Predicative adjectives

In addition to verbs, we consider predicative adjectives following copulas. Indeed, in Example 3.8, the pair *stay calm* / *be unreasonable* is much more informative in terms of conveying a contrastive relation, in comparison to the pair *stay* / *be*.

(3.8) John stayed calm even though Sally was clearly being unreasonable.

In order to avoid unnecessary sparsity, since the relevant information is contained in the adjectives rather than in the associated copulas, we replace the latter by "[*state verb*] +" during the extraction: in our example, the retrieved pair would be [*state verb*] + *calm* / [*state verb*] + *unreasonable*.

3.1.2.6 Event factuality in discourse

Saurí and Pustejovsky (2012) define event factuality as *the level of information expressing the factual nature of eventualities mentioned in text*. They define the factuality value of an event in terms of modality and polarity. Modality is defined on a scale ranging from *certain* to *uncertain*, while polarity is defined as a binary distinction, positive vs. negative. Additionnally, the source assigning the factuality value to an event is also a component of the so-called factuality profile of this event.

They identify several types of factuality markers, *i.e.*, expressions denoting polarity and modality. Polarity is marked by particles of negation which switch the original polarity. Modality can be marked by modality particles, among which modal auxiliaries are of particular interest to our study. Modality can also be expressed through *Event-Selecting Predicates*, *i.e.*, predicates with an event-denoting argument, which qualify the factuality of that event as part of their meaning. These are, for instance, predicates of report, knowledge, belief, or volition. Finally, factuality information can also be conveyed through syntactic constructions involving subordination.

The degree of factuality of events plays a role in discourse interpretation. This is illustrated by Examples 3.9, 3.10 and 3.11 from Danlos and Rambow (2011), based on the SDRT framework. In Example 3.9, both segments are attributed to the writer, and are linked with an Evidence relation. In Example 3.10, the last segment is attributed to another source (Jane), and the Evidence relation also holds because the writer endorses what Jane said. On the contrary, in Example 3.11, the writer does not endorse Jane's statement and there is no Evidence relation.

(3.9) [The neighbors have gone on vacation]. [Newspapers are accumulating on their doorstep].

(3.10) [The neighbors have gone on vacation]. [*Jane told me that*] [newspapers are accumulating on their doorstep].

(3.11) [The neighbors have gone on vacation]. [*Jane claimed that*] [newspapers are accumulating on their doorstep], [*but that is wrong*].

Danlos and Rambow (2011) therefore postulate that discourse analysis requires taking into account the "propositional attitude" (degree of assertion) towards the content of discourse segments, and propose to analyze the factuality of events in discourse segments in a similar fashion to Saurí and Pustejovsky (2012).

The impact of reportative verbs and their different uses (evidential or intensional) on discourse semantics was also studied by Hunter et al. (2006), who suggest that there are complex interactions between discourse structure and the lexical content of reportative verbs.

The alterations caused by factuality in terms of semantic relations are however too complex and fine-grained to be treated in depth by our method. In fact, we do not know of any reliable and rigorous method for disentangling these uses. In the subsequent paragraphs we show how our method handles the different factuality factors described by Saurí and Pustejovsky (2012).

Polarity: Negations

Regardless of factuality, it seems intuitive that the semantic relation linking two verbs would change if one of them is used with a negation. In a manual study focused on the annotations of the relation **Expansion.Alternative. Chosen alternative** in the PDTB, Webber (2013) found that among other features, the presence of a negation in the scope of the verb in the first clause was the most common in instances of this relation. For explicit relations, they focused on instances containing the connective *instead*. Example 3.12, taken from their study (originally in the PDTB), is an instance of such a relation expressed with this connective and a negated verb in the first clause.

(3.12) If the flex is worn, [do not use insulating tape to repair it]. Instead, [you should replace it ...] .

In a follow-up study, Asr and Demberg (2015) computed the correlation between the polarity of the first clause of an instance in the PDTB and its annotated relation. They define negative polarity as the presence of a negation in the clause, and the relation is taken as is from the annotations, meaning that some only contain a relation class from the first level of the PDTB hierarchy while others are specified down to the third level. They found significant differences in the correlation scores associated with the different relations. Some relations, like **Temporal.Synchronous**, obtain a significantly negative correlation value, indicating that a negative polarity sentence in a text is very unlikely to be followed by this relation. Other relations, and particularly **Expansion.Alternative.Chosen alternative**, obtain a positive score, meaning that a negative polarity sentence is often involved in one of these relations.

Following these observations and intuition, our extraction method also takes into account the presence of a negation in the dependencies of a verb. In Example 3.12, the pair *not use* / *replace* would be extracted.

Attribution

As mentioned in Section 2.3, the question of attribution is not treated equally in the different discourse structure theories and annotation frameworks.

Originally, Mann and Thompson (1988) rejected attribution as a rhetorical relation within the RST framework, arguing that a reporting clause functions as evidence for the attributed material and thus belongs with it, like any clausal complements of verbs. On the contrary, Carlson and Marcu (2001) made an exception for clausal complements of attribution verbs, and included attribution as a relation in the RST Treebank annotation manual, as illustrated in the RST tree displayed in Figure 3.1 corresponding to Example 3.13. The attribution relation is defined not only for speech verbs, but also for cognitive predicates, *e.g.*, *to believe*, *to expect*.

(3.13) [The legendary GM chairman declared]$_{\textbf{(A)}}$ [that his company would make "a car for every purse and purpose."]$_{\textbf{(B)}}$

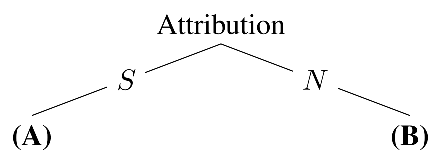


Figure 3.1: RST tree corresponding to Example 3.13.

In the PDTB, attribution is not treated as a discourse relation, but is represented by four features associated with each annotated relation: *Source*, *Type*, *Scopal polarity* and *Determinacy*. Notably, the *Type* feature is meant to encode the degree of factuality of a clause, based on the attributive verb which is used. Attributive verbs can be of different types: communication, belief, factive and control verbs.

Similarly, we do not consider attribution to be a relation. We make the simplifying hypothesis that a relation denoted by a discourse connective holds even when one or both events are in fact reported by other sources (introduced by predicates of report, knowledge or belief), or when they represent future events introduced by predicates of volition (*e.g.*, *want*) or command (*e.g.*, *require*).

In Example 3.14, following our method without more specifications, the identified pair of verbs would be *say/not return*, however in this case it seems more relevant to retrieve the pair *send/not return*.

(3.14) Officials said an envoy had been sent for "consultations" with the rebels on November 21 but had not yet returned to Ziguinchor.

The *Scopal polarity* feature in the PDTB identifies "*cases where verbs of attribution are negated on the surface – syntactically (*e.g., didn't say, don't think*) or lexically (*e.g., denied*), but where the negation in fact reverses the polarity of the attributed relation or argument content*" (Prasad et al., 2007).

We also take scopal polarity into account by reporting negations applied on attributive verbs to the attributed event. Additionally, when an event is introduced by a downward-entailing predicate like *reject* or *refuse*, the implied negation is also reported on the attributed event. Indeed, Webber (2013) suggest that these predicates have similar effects as negations on discourse relations. In Example 3.15, the negation implied in the verb *refuse* should be retained in the resulting contrastive pair as *recognize/not establish*.

(3.15) Turkey recognized Armenia in 1991 but refuses to establish diplomatic ties.

Similarly to attribution, for modal auxiliaries (*e.g.*, *can, could, may, must*) and diverse support verbs introducing events (*e.g.*, *help*, *like*), we consider that the relation, although indirect, holds with the introduced event.

These different considerations imply that a predicate syntactically linked to a discourse connective might not be the relevant event that we wish to extract, requiring further analysis of the dependency links. These considerations will be described in Section 3.2.3. In Example 3.16, instead of retrieving the pair *not extend/want*, it would be more useful to retrieve the pair *not extend/relocate*.

(3.16) Since they cannot extend Highbury, Arsenal want to relocate to a new state-of-the-art stadium.

3.1.2.7 Shared arguments

As discussed in Section 2.1.4, coreference between entities is a cohesive device which plays a key role in discourse structure. Predicates sharing one or several arguments is an important clue that they are linked by a semantic relation. The specific mapping of these arguments can also affect the type of relation linking them, as we showed with our Examples 2.22 and 2.23 in Section 2.1.4, repeated here as Examples 3.17 and 3.18.

(3.17) Sarah hit John. She was furious.

(3.18) Sarah hit John. He was furious.

For this reason, coreference has been used as a feature for discourse relation prediction in the literature (Corston-Oliver, 1998; Schauer and Hahn, 2001; Duque, 2014; Bärenfänger et al., 2008).

Including information about shared arguments is also a means of further refining event predicates. We consider the following possible cases:

- The *subject of the second predicate* refers to the same entity as the *subject of the first predicate* (as in Example 3.17),
- The *subject of the second predicate* refers to the same entity as the *object of the first predicate* (as in Example 3.18),
- The *subject of the second predicate* represents a *different entity* from the subject and object of the first predicate,
- The *object of the second predicate* refers to the same entity as the *subject of the first predicate*,
- The *object of the second predicate* refers to the same entity as the *object of the first predicate*,
- The *object of the second predicate* represents a *different entity* from the subject and object of the first predicate,
- The *second predicate* is an *intransitive* verb (it does not have an object).

Now that we have laid out the principles of our approach, we turn to the practical details of our implementation.

3.2 Practical method

In this section we present the external resources on which our method is based and the practical ways to identify related clauses, and to identify and refine the involved predicates.

3.2.1 Resources

The resource is built without supervision. It thus does not require a discourse annotated corpus. Three elements are required: a large corpus, a syntactical parser, and finally a lexicon of discourse connectives annotated with corresponding relations.

Additional external resources are used to enable finer specifications of the predicates. These resources will be presented in Section 3.2.3 along with their usage. The following sections describe our choices for the three main elements, for French and English.

3.2.1.1 Large unannotated corpus

The choice of corpus is guided by two main requirements. First, it needs to be as large as possible in order to allow the retrieval of many instances. Indeed, without supervision we can expect a lot of noise to arise, originating from erroneous syntactical parsing, ambiguity in the discourse usage of the identified markers, and ambiguity in the relation they can trigger. Secondly, it needs to come from as many different domains as possible, in order to be usable in a range of applications and domains. Another more practical requirement for our purpose is that it needs to be within the resources we have access to.

French

The French language corpus best responding to these requirements is the **frWaC** corpus, from the WaCky set of corpora (Baroni et al., 2009). frWaC contains about 1.6 billion words and was collected on the Web on the .fr domain. It is thus indeed very large and covers very diverse domains. This includes texts extracted from blogs, which implies that some parts contain many spelling and grammatical errors. This in turn causes difficulties and errors in the syntactical parsing.

English

The corpus of choice for English is the **Gigaword** corpus (Parker et al., 2011). It contains about 1.8 billion words from four international sources of English newswire. This avoids the troubles that were encountered in terms of language correctness with frWaC, with the drawback of being admittedly more domain-specific.

3.2.1.2 Dependency parser

Our method to retrieve predicates connected by semantic relations relies on syntactic patterns involving predicates and discourse connectives. Therefore, it is necessary to retrieve the syntactic structure of the sentences composing the texts. Our method is based on syntactic dependencies. In order to obtain these, three operations need to be performed: part-of-speech tagging, lemmatization and dependency parsing. Dependency structures consist of lexical items, linked by binary asymmetric relations called dependencies.

French

The dependency trees for the frWaC corpus were obtained using the Bonsai tool (Candito et al., 2010), which includes a part-of-speech tagger and lemmatizer, MElt

(Denis and Sagot, 2012) and the MaltParser (Nivre et al., 2007), trained on the French Treebank (Abeillé et al., 2003), for syntactic parsing. The resulting dependency trees are in the CoNLL format.

English

The Gigaword corpus comes syntactically parsed via the Stanford CoreNLP pipeline (Manning et al., 2014), which outputs parsed sentences in the Stanford Dependencies format.

3.2.1.3 Lexicon of connectives

Our approach is based on extracting predicates linked by connectives signalling discourse relations. Therefore it requires a set of connectives each linked with one discourse relation, or a semantic label.

As detailed in Section 2.1.2, discourse connectives belong to various syntactical categories. Most of them are adverbs (*however*, *yet*), subordinating conjunctions (*when*, *if*) or coordinating conjunctions (*and*, *but*). They can also be multiword expressions, such as prepositional phrases (*due to*) or adverbial phrases (*for instance*). Discourse relations can also be triggered by more complex, less frozen expressions. For instance, "*Two days later*" can be used to express a temporal relation. In the PDTB, these expressions are refered to as "alternative lexicalizations", which are considered separately from explicit connectives.

French

Such a lexicon is readily available for French connectives: the *Lexconn* resource (Roze et al., 2012)[1]. This resource aims to provide an exhaustive collection of discourse connectives for French, and is grounded in the SDRT framework (Asher and Lascarides, 2003). It was manually constructed on the base of the *FRANTEXT*[2] corpus. It includes 358 connectives and gives their syntactic category as well as associated discourse relations inspired from SDRT.

Lexconn uses a list of 24 relations, which we collapsed into 12 relations. Table 3.1 presents the original list of relations in *Lexconn* and their correspondence with our relations for the French version of *Lecsie*.

[1]Freely available at: `https://gforge.inria.fr/frs/download.php/31052/lexconn.tar.gz`.

[2]*FRANTEXT* is a textual base of French Literature (ATILF - CNRS & Université de Lorraine), available at: http://www.frantext.fr.

Lexconn	***Lecsie-fr***
result result* goal explanation explanation* consequence	cause
narration flashback	narration
contrast parallel opposition concession condition	contrast
elaboration	elaboration
alternation	alternation
continuation	continuation
background background-inverse	background
temploc	temploc
commentary	commentary
summary rephrasing evidence detachment	

Table 3.1: Correspondence table between Lexconn relations and *Lecsie-fr* relations.

Figure 3.2 displays an extract of the entry in Lexconn for the connective *pourtant* (*however*). This connective is presented as having two discourse uses: one for signalling Opposition relations and one for signalling Concession relations.

Finer distinctions in the discourse usage of connectives which can be attained by manual analysis are out of the scope of our unsupervised extraction method, and are also not necessarily desirable in the context of infering lexical relations between predicates. In particular, some SDRT relations are distinguished between pragmatic and non-pragmatic usage (pragmatic usage is referred to as "meta-talk"

Connective	**Pourtant**
Syntactic Category	Adverb
Relation 1	Opposition
Relation 2	Concession

Figure 3.2: Partial example of entry in Lexconn: *Pourtant*.

in SDRT, identified with "*" in Table 3.1). We hypothesize that these two aspects of a relation often occur in similar lexical contexts, and therefore group them together.

According to our own set, the connective *pourtant* shown in Figure 3.2 becomes unambiguously associated with a Contrast relation, since Opposition and Concession are grouped in this category. In practice, this grouping achieves two positive effects on our resulting resource:

- By minimizing the issue of ambiguity in relation described in Section 3.1.2.2, the list of non-ambiguous connectives is larger, allowing for the **extraction of more instances**;
- By grouping more connectives associated with each relation (more non-ambiguous connectives which we can consider, and less relations to associate them with), it **reduces the sparsity** of our resulting resource.

As we specified earlier, we only consider non-ambiguous connectives. With respect to our own list of relations, there are 263 such connectives in *Lexconn*. Table 3.2 lists the number of non-ambiguous connectives per relation.

English

For English, we relied on the list of connectives provided in the PDTB annotation manual (Prasad et al., 2007). As detailed in Section 2.3.2, the PDTB offers discourse annotations in the form of relations between elementary discourse units, either explicitely signalled by a connective, or implicitely signalled. For the purpose of creating a connective lexicon for English, we took advantage of the annotations marked as "explicit", "implicit", and "alternative lexicalization". The PDTB annotation manual (Prasad et al., 2007) provides the list of explicit connectives that was proposed to the annotators, along with the number of times they were annotated

Relation	**Number of connectives**
Cause	129
Contrast	91
Continuation	18
Narration	17
Condition	17
Detachment	8
Summary	7
Background	6
Alternation	5
Evidence	4
Elaboration	3
Temploc	3
Commentary	3
Rephrasing	2
Unknown	27

Table 3.2: Counts of non-ambiguous connectives per relation used in *Lecsie-fr*.

as signalling each relation in explicit usage or as best representative of implicitly signalled relations.

From this list we extracted a lexicon of connectives and their usage counts with respect to the different relations was recovered directly from the PDTB corpus, without considering the distinction between explicit usage (where the connective is present in the text) and implicit usage (where the connective is added in the annotation process).

In the PDTB, connectives are positioned into a 3-level hierarchy, with 4 and 16 groups at level 1 and 2, which provide useful sets of discourse functions corresponding to semantic relations between eventualities introduced in the discourse. An example extracted from the PDTB with a connective and the 3 levels of relation it signals is shown in 3.19.

(3.19) Those looking for real-estate bargains in distressed metropolitan areas should lock in leases *or* buy now.

(EXPANSION:Alternative:disjunctive)

For the same reasons as above, we grouped some of the level 2 relations used in the PDTB, as described in the correspondence Table 3.3, and used the following

PDTB		***Lecsie***
Level 1	Level 2	
Temporal	Asynchronous Synchronous	Temporal
Contingency	Cause Pragm. cause Condition Pragm. condition	Cause
Comparison	Contrast Pragm. contrast Concession	Contrast
Expansion	Instantiation Restatement	Elaboration
	Alternative	Alternation
	Exception	Contrast
	Conjonction List	Continuation

Table 3.3: Correspondence table between PDTB relations levels 1 and 2 and *Lecsie* relations.

set : {Temporal, Cause, Contrast, Elaboration, Alternation, Continuation}.

Temporal typically binds verbs appearing in a common chronology, either before each other or during each other, **Cause** relates events standing in a causal relation. **Continuation** is a vaguer kind of relation between events occuring in a common scenario, **Alternation** and **Contrast** denote semantic similarity or opposition, and **Elaboration** is a kind of sub-event relation.

As with the French lexicon, this resulted in less ambiguity for our list of connectives: we otained a total of 104 non-ambiguous connectives. In Table 3.4, we show the number of non-ambiguous connectives associated with each relation.

3.2.2 Identifying the clauses

In order to design a method to extract triples automatically with minimal error, it is necessary to first observe how explicit relations are expressed, in particular where the connective is located with respect to the clauses it connects, and how this translates in the dependency parsing.

Relation	**Number of connectives**
Temporal	30
Contrast	24
Cause	20
Continuation	14
Elaboration	9
Alternation	7

Table 3.4: Counts of non-ambiguous connectives per relation used in *Lecsie-en*.

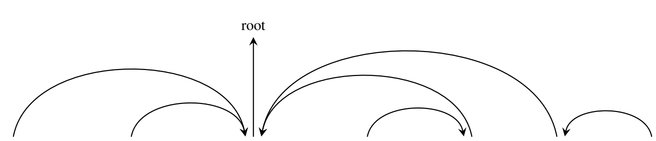
In some frameworks, for instance in the PDTB, the connective is not considered as part of either clause linked by an explicit relation and is treated separately. Here we consider that the connective is part of the clause which contains it, which is referred to as the "host clause", in accordance with the terminology defined by Danlos (2011) and followed in *Lexconn*. The second clause concerned by the relation is refered to as the "mate clause".

Two aspects are to be considered about the position of the connective in an explicit relation between two clauses:

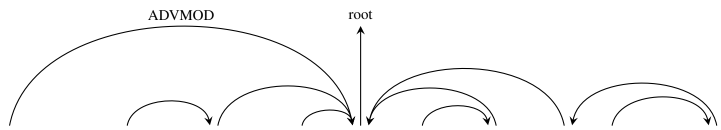
- its position within the host clause,
- the position of the mate clause with respect to the host clause.

The clauses can be in the same sentence, in which case we refer to the relation as being "intrasentential", or in two different sentences, where the relation is then "intersentential". This is determined by the properties of the connective regarding the two aforementioned aspects, and the position of the host clause in the sentence containing it. For some connectives, these aspects are fixed, while for others they can vary. In most cases, this difference can be attributed to their syntactic category.

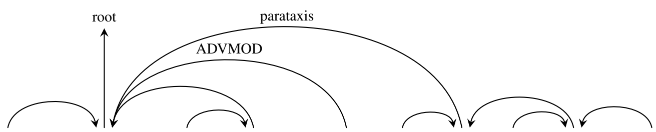
In the case of adverbs, the dependent clause is always on the left of the host clause, but the connective can appear at various positions in the host clause, as exemplified in 3.20, 3.21 and 3.22. As seen in these examples, relations triggered by adverbs can be both intrasentential and intersentential.



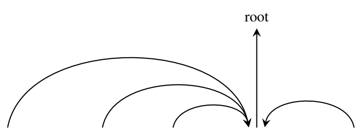
(3.20) George usually watches rugby games on sundays .



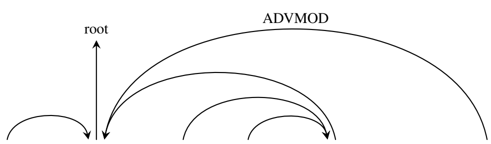
However , last weekend he took his family to the countryside .



(3.21) Mary enjoyed the party , yet she had to leave early .



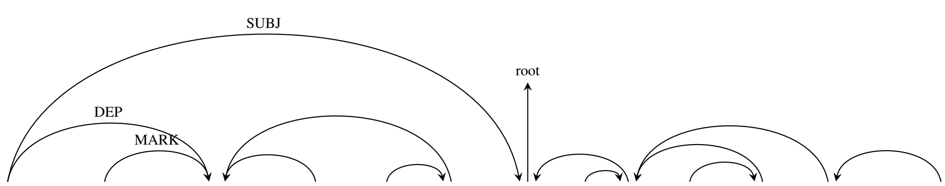
(3.22) Charlie did not study much .



He passed all his exams , though .

Some constructions, such as that presented in Example 3.23, are however more problematic, since the two predicates *work* and *need* are not directly related, but linked through the main subject *Peter*. As of yet, we do not implement solutions for such cases as they are not frequent. More precise rules aimed at these types of construction could be designed in the future to enable their processing.

(3.23)



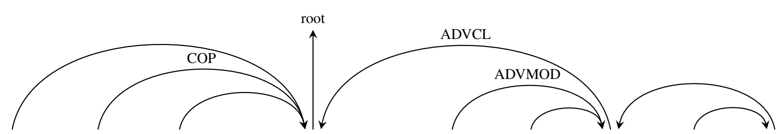
Peter, although working harder this term, needs to put more work into mathematics.

In terms of dependency relations, different cases can occur, mostly depending on whether the relation is intrasentential or intersentential. In intrasentential cases (as in example 3.21), the connective is usually a dependent of the verb in the mate clause (which is in fact the main clause of the sentence), with **ADVMOD** or

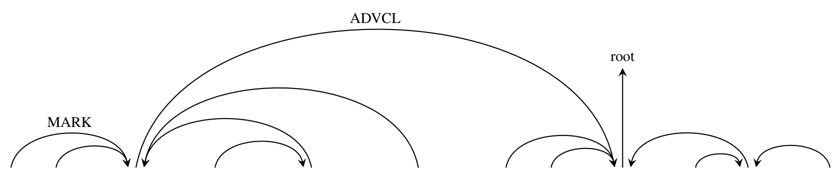
MARK as dependency relation depending on the specific connective. The verb in the host clause is also a dependent of the verb in the mate clause.

In intersentential cases (as in examples 3.20 and 3.22), the connective is a dependent of the verb in the host clause, again with **ADVMOD** or **MARK** as dependency relation, while the verb in the mate clause is the root of the sentence containing it.

In the case of subordinating conjunctions, the connective is always at the beginning of the host clause, but the mate clause can be either before or after the host clause, as seen in examples 3.24 and 3.25. In the majority of cases, the two clauses are in the same sentence.



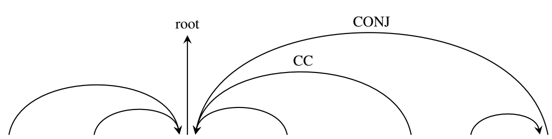
(3.24) Peter was very disappointed when he heard the news .



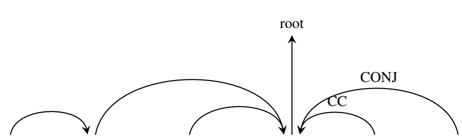
(3.25) If she finishes her homework early , she is allowed to play games.

In the dependency parsing, the connective is a dependent of the verb in the host clause with ‘**ADVMOD** or **MARK** as dependency relation, which itself is a dependent of the verb in the mate clause usually with **ADVCL** as dependency relation. The mate clause is in fact usually the main clause in the sentence.

In the case of coordinating conjunctions, they are always at the beginning of the host clause, and the mate clause is always before the host clause.



(3.26) That may sound crazy , but it works .



(3.27) The child never smiles or laughs .

In intrasentential cases, the connective is a dependent of the verb in the mate clause, with dependency relation **CC**, and the verb in the host clause is also a dependent of the verb in the mate clause, with dependency relation **CONJ**. In intersentential cases, the connective is a dependent of the verb in the host clause with dependency relation **CC**.

This preliminary study reveals that there are patterns that can be easily identified, although there are many outliers, and many variations can be found in the labels describing the dependency relations. Our method for automatic extraction is based on the following observations, which seem to be true most of the time.

Intersentential cases:

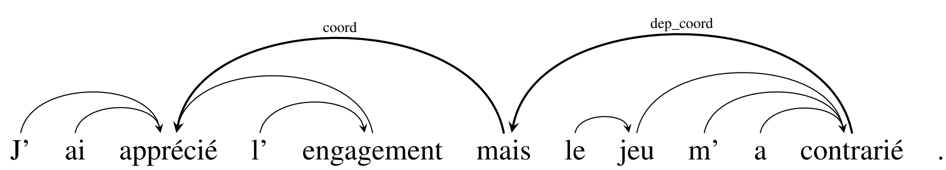
- The connective is a dependent of the main verb in its host clause,
- The main verb in the mate clause is usually the root of the sentence which contains it (that is, the mate clause is usually the main clause of the sentence).

Intrasentential cases:

- The connective is a dependent of either the main verb in its host clause, or the main verb in the mate clause,
- The two verbs are directly linked by a dependency relation, which can be in either direction.

Similar observations can be made in the French dependency parsing using the MaltParser, with the difference that in intrasentential cases, the verb from the mate clause is usually a dependent of the connective, and not of the verb from the host clause, as seen in Example 3.28.

(3.28)



I appreciated the kickoff but the game bothered me.

Therefore, for the extraction of semantic relations between French predicates, we observe the following rules, with only the last one for intrasentential cases being different from the rules for English.

Intersentential cases:

- The connective is a dependent of the main verb in its host clause,
- The main verb in the mate clause is usually the root of the sentence which contains it (that is, the mate clause is usually the main clause of the sentence).

Intrasentential cases:

- The connective is a dependent of either the main verb in its host clause, or the main verb in the mate clause,
- **The second verb is directly linked by a dependency relation to the connective.**

3.2.3 Extracting the relevant verbs and their properties

According to our analysis in Section 3.1.2, we are looking to extract more than simple verb lemmas. Our method implements automatic rules based on the dependency relations to extract phrasal verbs, predicative adjectives, and to access attributed clauses and otherwise supported clauses for cases where the main verb is not considered to be the most relevant.

3.2.3.1 Phrasal verbs

As specified in Section 3.1.2.5 we aim to extract additional elements from the context in order to obtain a finer description of the events linked by connectives. Here we present how this is done in practice. There are some differences in the distinctions we make in each language, due to the resources and information available.

French

Verbs can take very different meanings in French when they are combined with prepositions. However, the dependency links given by the MaltParser provide no direct way to distinguish between phrasal verbs and the use of prepositions as part of prepositional objects, for instance in temporal or spatial modifiers. In order to make this distinction, we use the *Dicovalence* resource (Van Den Eynde and Mertens, 2010). In this resource, which contains more than 3700 simple French verbs, each verb is associated with one or more valency frame characterising the number and type of the syntactic arguments expected by this verb. For each verb included in Dicovalence we extract the prepositions which can be of idiomatic usage. Then when a preposition (lemma with part-of-speech **P**) is the direct dependent of a verb (with dependency relation **OBJ** or **MOD**) , it is retrieved as an expression if the association exists in our list. In Example 3.29, this method enables the extraction of the phrasal verb *chercher à*, meaning *try*, rather than *chercher*, meaning *search*.

(3.29) Nous avons décidé de chercher à mieux comprendre leurs difficultés.
We decided to try to better understand their difficulties.

We also compiled a list of idiomatic verbal locutions from the Lefff resource (Lexique des Formes Fléchies du Français, Sagot (2010)), which is used to retrieve those. For example, Table 3.5 displays examples of locutions with the verb *avoir* (*have*) and their approximate translations. The full list of idiomatic verbal locutions which we considered can be found in Appendix A, Table A.1.

Verbal locution		**Translation**
avoir	honte	be ashamed
	hâte	be in a hurry
	faim	be hungry
	horreur	hate
	raison	be right
	lieu	take place
	conscience	be aware of
	tendance	tend to
	envie	want
	droit	be entitled to
	besoin	need
	tort	be wrong
	vent	hear about
	peur	be afraid
	foi	believe in

Table 3.5: Examples of verbal locutions with the verb *avoir*, extracted from the *Lefff* resource (Sagot, 2010) and their approximate translations.

As stated in Section 3.1.2.5, in French the presence of a reflexive particle also triggers differents meanings. These particles are identified by the **AFF** dependency relation and are thus easily recovered. In Example 3.30 we can thus extract the predicate *se trouver*, meaning *be located* rather than the verb *trouver* (*find*).

(3.30) La capitale se trouve au Sud du pays .

The capital city is located in the South of the country.

The MaltParser dependencies do not allow for direct identification of copulas followed by predicative adjectives. Therefore this case was not treated in French, although the implementation of a method to retrieve these instances would be an interesting perspective.

Other information, that does not lead to distinct lexical entries, are kept in memory in a secondary database: tense, and voice. This secondary database is currently unused but could prove useful in future applications.

English

As in French, identifying phrasal verbs and prepositional verbs is useful for disambiguation in English. In the Stanford dependencies, the dependency relation **PRT** (phrasal verb particle) allows to identify these particles directly. This is illustrated in Example 3.31, from which the phrasal verb *carry out* can be extracted, allowing for a more accurate representation of the denoted event than the simple verb *carry*.

PRT

(3.31) The government has carried out a national census .

Predicative adjectives following copulas are identified by the dependency relation **COP** (copula) in the Stanford dependencies, as shown in Example 3.32. As stated previously, the copula is replaced by the expression "[state verb] +", and the extracted predicate from this example would thus be represented as "[state verb] + likely".

COP

(3.32) Argentina is likely to end its currency peg to the dollar.

However, in order to avoid retrieving rare or obscure adjectives, we performed a preliminary selection of relevant adjectives by extracting all occurrences of **COP** relations and retaining only the adjectives occurring more than 100 times in the corpus. Table 3.6 displays the ten most frequent predicative adjectives and their number of occurrences in the corpus. The complete list we retained is presented in Appendix A, Table A.2.

Adjective	Count
able	160252
likely	104990
available	82094
good	78002
ready	76212
important	74320
clear	68566
sure	63306
due	56412
hard	55346

Table 3.6: Ten most frequent predicative adjectives following copulas in the Gigaword corpus, and number of occurrences.

3.2.3.2 Negations

The treatment of negation particles is easy in both languages thanks to their identification in the syntactic dependencies.

French

Negated forms are identified by the feature **s=neg** in one of the dependents of the verb. Negative expressions identified by this feature include *ne...pas*, *ne...plus*, *ne...rien*, *ne...jamais*. Regardless of the actual expression used in the text, our method always replaces it by the expression “ne pas” (“not”) prepended to the verb, in order to avoid additional sparsity. Thus, our extraction method would output the expression *ne pas cesser* (*not cease*) from Example 3.33.

s=neg s=neg

(3.33) Il n’ a jamais cessé de soutenir sa communauté.

He never ceased to support his community.

We make an exception for the expression *ne...que*, which is also parsed as a negation, although it usually translates to *only*, as in Example 3.34. In this case, the extracted predicate is not negated.

(3.34) Jean <u>n’</u>aime <u>que</u> les sucreries.

Jean <u>only</u> likes sweets.

English

The presence of a negation is identified by a lemma with dependency relation **NEG** to the verb. Similarly, all occurrences of this dependency relation are treated by prepending "not" to the extracted predicate. Thus, *not think* would be extracted from Example 3.35.

NEG

(3.35) I did n't think we could do it.

3.2.3.3 Predicates introduced by attributive and support verbs

The two verbs retrieved through our method are not necessarily the most relevant ones, in the sense that they do not carry the most relevant lexical information supporting the coherence relation. This is the case for attributive verbs, modal auxiliaries, and other support verbs.

French

In the French dependencies, modal verbs appear as the governor of their supported verbs. As in Example 3.36, the relevant verb to retrieve is the supported verb *demander* instead of the modal verb *pouvoir*. Thus we compiled the following list of modal verbs in French: {*pouvoir*, *devoir*, *vouloir*, *falloir*}. When one of these modal verbs is encountered, we select the supported verb (identified by verbal part-of-speech and **OBJ** as dependency relation to the modal verb), if present.

obj

(3.36) Les étudiants peuvent demander une aide financière.

Students may request financial support.

On a side note, the modal verbs are also encoded in the secondary database.

The treatment of attribution was not implemented in French, so that in most cases of attributions, we extract the attributive predicate and not the event it introduces. The resource would certainly benefit from the analysis of these cases in future work.

English

In the Stanford dependencies, modal verbs appear as dependent of their supported verbs with the dependency relation **AUX**. Hence there is no processing required in this case. As for French, the presence of modal verbs is also recorded for possible future use.

There are however further cases to be analysed in the Stanford dependencies. The first case we consider is attributions. As detailed in Section 3.1.2.6, we make the simplifying hypothesis that a relation between two events holds even when one or both events are in fact reported by other sources. Therefore when an attributive verb is encountered, we need to extract the attributed event.

There is however no unique dependency label between attributive verbs and their supported verbs, it may vary depending on the context. For this reason, in a preliminary step we extracted communication verbs from the corresponding class in VerbNet (Schuler, 2005). This class contains verbs such as *say*, *report*, *announce*, etc. Then, when a communication verb from this list is identified, we check whether it has a subordinate clause and retrieve its main verb instead if it exists. Otherwise, the communication verb is kept as the relevant verb.

Similarly, when an event is introduced by a very frequent support verb like *want*, *like*, *help*, we assume that the identified relation to another event holds, and wish to extract the supported event itself. However when the support verb is a downward-entailing predicate (*e.g.*, *refuse*), a negation should be introduced in the supported event. Supported verbs are usually linked to their support verb by the dependency relation **XCOMP** (open clausal complement).

In order to identify the most frequent support verbs (less frequent verbs are more likely to be of interest and we do not wish to discard them), and also distinguish those implying a negation, we extracted all occurrences of **XCOMP** relations in the corpus and compiled a list of the support verbs occurring more than 1000 times. Then we manually annotated them for positive or negative meaning. Some examples of negative verbs are *fail*, *refuse*, *decline*. In the triple extraction step this list can thus be used to correctly identify supported verbs.

As a recapitulative example, our algorithm would retrieve the triple *{not wait, [state verb] + punctual, contrast}* from the sentence in 3.37.

(3.37) She *refused* to <u>wait</u> <u>although</u> he *said* he *would* <u>be punctual</u>.

3.2.4 Extracting triples

In the previous sections we presented our method for selecting the related clauses thanks to the dependencies to the connectives, and for retrieving the relevant predicates from these clauses. Now we present to overall process for extracting triples composed of two predicates and a connective linking them in the text.

The first step is to identify connectives occurring in each sentence of each document in the corpus. We search for the occurrence of one or several lemmas or sequences of lemmas from our list of non-ambiguous connectives in each sentence.

Then for each occurring connective we exploit the dependency links according to the observations described in Section 3.2.2. Since the labels of the dependency relations can vary depending on the specific connective and on the context of the sentence in which it is used, we do not restrict on the dependency label but only on the part-of-speech to recover the verbs. In addition to verbs, we also allow adjectives linked to copulative verbs to be considered, under the specific restrictions defined in Section 3.2.3.1.

In case only one verb can be found, either as the governor or as the dependent of the connective, we consider that the relation is intersentential. As usual the host clause is then the clause containing the connective and the verb that could be identified.

To identify the mate clause, as stated in Section 3.1.2.4, we follow the simplifying hypothesis that it occurs in the sentence immediately before the sentence containing the connective in the document, if there is any (that is, if the sentence containing the connective is not the first sentence in the document).

In most cases, the verb selected in the previous sentence is the head verb of the sentence (usually the root of the sentence). The exception to this is when the connective is associated to the relation Narration in the French lexicon, in which case we select the last verb occurring in the sentence.

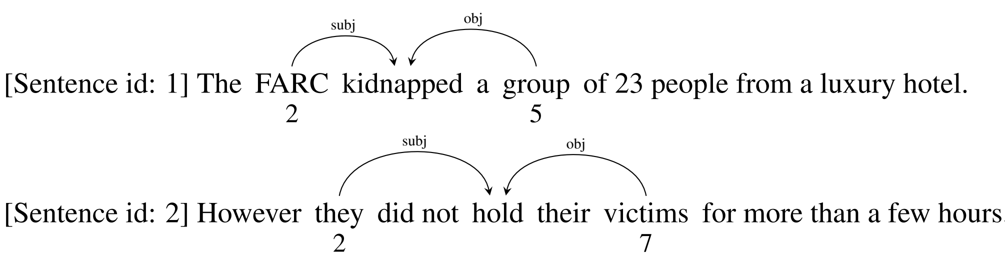
3.2.4.1 Shared arguments

As we stated in Section 3.1.2.7, information about shared arguments is a relevant addition to specify the relation between predicates.

This specification might not be useful for all applications, and creates a lot of additional sparsity for two reasons. First, each unique pair of predicates has many possible predicate-argument mappings, creating as many possible entries in the resource. Secondly, while the implementation without shared arguments orders the predicates alphabetically, this second implementation taking shared arguments into account requires to register the predicates in the order they appeared in the text. This potentially doubles the number of unique pairs of predicates. We therefore decided to keep both implementations separate, and to aggregate the results with shared arguments in a distinct resource.

Due to the lack of a reliable tool for coreference resolution in French, the treatment of shared arguments has only been implemented in our method for English.

The Stanford Parser includes a coreference resolution tool. For each document, it provides a list of lemmas referencing the same entity (referred to as the "coreference list" in the following paragraph). In order to recover shared arguments, we first identify the subject and object (if there is one) of each argument, using the **SUBJ**, **OBJ** and **COMP** dependency relations. Then, by matching the subject and object of each predicate to item mentions in the coreference list, we determine which arguments refer to the same entities, if any. Figure 3.3 illustrates this process.



Coreference list:

		Sentence id	Head token id	Tokens	Representative
Coreference item 1	Mention 1	1	2	1,2	True
	Mention 2	2	2	2	False
Coreference item 2	Mention 1	1	5	4,5,6,7,8	True
	Mention 2	2	7	6,7	False

Figure 3.3: Simplified example of two consecutive sentences and their coreference list.

From this example, we can extract the information that the subjects (*The FARC*, *they*) of the two predicates refer to the same entity represented by *The FARC*, and that the objects (*a group of 23 people*, *their victims*) refer to the same entity represented by *a group of 23 people*.

Shared arguments between predicates are identified by two features in *Lecsie*:

- **S2match**, representing the coreference value of the subject of the second predicate: its possible values are "S1" (the subject of the first predicate), "O1" (the object of the first predicate), or "New" (no coreference), depending on which element it corefers with;
- **O2match**, representing the coreference value of the object of the second predicate: its possible values are "S1", "O1" , "New" or "None" (no object).

For the example given in Figure 3.3, we would obtain the entry displayed in Figure 3.4.

Verb1	Verb2	S2match	O2match	relation
kidnap	not hold	S1	O1	Contrast

Figure 3.4: Resulting entry from Figure 3.3.

Matching entities also enable us to identify reflexive verbs, by identifying cases where the subject and object of a predicate refer to the same entity. This information enables us to include two additional binary features in *Lecsie*:

- **V1reflexive**, taking value “1” if the object of the first predicate refers to the same entity as its subject, “0” otherwise ;
- **V2reflexive**, with similar value definitions for the second predicate.

3.2.5 Collecting the results

3.2.5.1 Extracted triples

The application of our method to the French and English corpora yielded the following results: 2 million occurrences of triples **{verb 1, verb 2, connective}** from the French corpus, and 53 million occurrences from the English corpus.

The difference has several possible explanations. On a practical level, the results of the dependency parsing seem to be more reliable and better specified in English allowing more precise identification of related verbs, while many instances might be missed in French. On a linguistic level, it seems likely that explicit connectives are more often used to express relations in English, and that connectives are less ambiguous with respect to the relation they trigger.

3.2.5.2 Grouping by relation

Given that we have used only unambiguous connectives, the association of a relation with a connective is immediate. Thus our triples can directly be transformed into the form **{verb 1, verb 2, relation}**. Since we grouped together asymmetrical relations (*e.g.*, Cause, Condition), and since the host and mate clauses are not identified in our method, the order of the verbs in the triples is not meaningful. For this reason, the verbs are reordered alphabetically.

As a last step of the construction of *Lecsie*, we then collected the occurrences of triples in distinct triple types with their corresponding frequency. The French version, *Lecsie-fr*, contains 1 million distinct triples, while the English version, *Lecsie-en*, contains 8 million. In the English version which includes information on shared argument, *Lecsie-coref*, we obtained more than 18 million distinct triples, from the same 53 million occurrences found in the English corpus. Indeed, all the different possible cases of shared arguments for the same two predicates induce more distinctions, and thus additional sparsity. Although this type of information is potentially helpful in the characterization of events, the increased sparsity is problematic in external applications. The problem of imprecise or incorrect anaphora resolution also makes this additional information difficult to use. We will show this in Section 6.1 which concerns the usage of our resource for attachment prediction models.

3.3 Conclusion

We have presented our method to automatically collect predicates linked by semantic relations from a large corpus. By applying this process to a large corpus, valuable quantitative information about semantic relations between predicates can be extracted. We analyze these results in the following chapter.

CHAPTER 4

EXPLORING *Lecsie*

Contents

We applied the process described previously on the French-language corpus *frWaC* (Baroni et al., 2009) and on the English-language corpus *Gigaword* (Parker et al., 2011), and encoded the collected information in two resources, *Lecsie-fr* and *Lecsie-en*, respectively. In this chapter we aim to demonstrate the interest of our extraction method and explore the semantic information which can be derived from such a resource. We illustrate our observations with examples extracted from *Gigaword* rather than invented examples, in order to provide realistic contexts.

In Section 4.1, we first explore the general results obtained by our extraction method in terms of relation distribution and the typicality of the associations between pairs of predicates and relations. Then, in Section 4.2, we analyze the influence of negations on the relations we observe. Finally, Section 4.3 looks at the impact of coreference patterns on the relations.

4.1 Overall results

4.1.1 Relation distribution

First, we look at the relation distribution in terms of occurrences in the corpus. Tables 4.1 and 4.2 present these results for the French and English versions of *Lecsie* respectively, as well as the ratio of intra- and intersentential instances in the corpus for each relation.

Relation	Count	Distribution	Ratio intra	Ratio inter
Contrast	1038701	50.1	94.8	5.2
Cause	683709	33.0	94.4	5.6
Continuation	170222	8.2	99.5	0.5
Narration	131370	6.3	91.2	8.8
Background	38273	1.8	99.7	0.3
Temploc	3661	0.2	89.6	10.4
Detachment	3067	0.1	86.0	14.0
Unknown	2616	0.1	62.9	37.1
Summary	33	0.0	100.0	0.0
Alternation	33	0.0	100.0	0.0
Rephrasing	24	0.0	62.5	37.5
Elaboration	13	0.0	76.9	23.1
Evidence	2	0.0	100.0	0.0
Overall	2071724	100	94.9	5.1

Table 4.1: Relation distribution and ratios of intra- and intersentential occurrences in *Lecsie-fr*, in %.

Relation	Count	Distribution	Ratio intra	Ratio inter
Contrast	17597299	33.0	80.1	19.9
Temporal	14540660	27.3	88.2	11.8
Cause	13900091	26.1	88.3	11.7
Continuation	5052350	9.5	61.3	38.7
Alternation	1796703	3.4	95.4	4.6
Elaboration	381591	0.7	76.0	24.0
Overall	53268694	100	83.2	16.8

Table 4.2: Relation distribution and ratios of intra- and intersentential occurrences in *Lecsie-en*.

In the two languages, large differences can be observed for similar relations. A striking example is the difference between the temporal-location relation for French, which is very rare, and the Temporal relation for English, which is the second most frequent relation. This difference can be explained by the fact that French temporal connectives are very ambiguous.

As for proportions of intra- and intersential instances, for French we find that 95% of all instances are intrasentential, compared to 83% for English. Although it is slightly more elevated for English, in both cases, the low proportion of intersentential instances comes from our conservative scheme for finding these occurrences, which uses only those connectives at the beginning of the second sentence. Other schemes are possible but would, we fear, introduce too much noise into the data.

In Table 4.3 we compare the relation distribution obtained in *Lecsie-fr* with the distribution of expert annotations in the Annodis corpus reported by Afantenos et al. (2012). Their relation set is based on SDRT like ours, but they also made some simplifications, resulting in a different set from that used in *Lecsie-fr*, as well as that used in the *Lexconn* resource. Indeed, the relation set is not fixed and needs to be adapted depending on the intended application. Therefore a one-to-one comparison cannot be made, but some interesting observations can still be drawn from Table4.3.

Relations	***Lecsie-fr***	**Annodis**
Contrast	50.1	6.1
Cause	33.0	11.6
Continuation	8.2	20.3
Narration	6.3	11.2
Background	1.8	4.6
Temploc	0.2	0.5
Detachment	0.1	
Unknown	0.1	
Summary	0.0	
Alternation	0.0	0.5
Rephrasing	0.0	
Elaboration	0.0	34.3
Evidence	0.0	
Total count	2071724	3355

Table 4.3: Comparison of relation distributions in *Lecsie-fr* and annotations from the Annodis corpus.

Important differences arise from the fact that the reported distribution for Annodis annotations includes both explicit and implicit instances, while our distribution is based on instances specifically expressed by connectives. The results show that Elaboration, Continuation and Narration instances are considerably more frequent in the Annodis annotations. Indeed, the expression of these relations in discourse often does not require a connective. On the other hand, Contrast and Cause relations are much more frequently extracted by our method than they are manually annotated. These relations typically require to be explicitly signalled.

In Table 4.4 we compare the relation distribution obtained in *Lecsie-en* with the one reported by Prasad et al. (2008) for the explicit annotations in the PDTB. Since they report only the results for the first level in their hierarchy of relations, we make an approximative correspondence for the Expansion relation (our correspondence table with PDTB relations, as presented in Section 3.2.1.3 requires Level 2 relations to be complete).

***Lecsie-en* relations**	**Distribution**	**PDTB relations**	**Distribution**
Contrast	0.33	Comparison	0.29
Temporal	0.27	Temporal	0.19
Cause	0.26	Contingency	0.19
Continuation-Elaboration-Alternation	0.14	Expansion	0.33
Total count	53268694		19449

Table 4.4: Comparison of relation distributions in *Lecsie-en* and explicit annotations from the PDTB.

The frequency of Contrast relation is comparable in both distributions. However we extract significantly more Temporal and Cause relations, and the frequency of Expansion relations is much lower in our results. This observation is potentially due to the ambiguity of Expansion connectives, which means that they must be discarded from our list of unambiguous connectives, while Temporal and Cause connectives tend to be more typical of just one relation.

4.1.2 Typicality of associations

With our method, the same pairs of predicates are often found with markers of different relations. By looking at how the relations are distributed for each pair of predicates, we aim to answer the following questions:

- Can we identify typical links between pairs of predicates and relations? In this case, are the predicates a representative lexicalization of the relation?

- Do some relations often occur with the same pairs of predicates? Which elements then allow to differentiate them in context?

Regarding the first question, if a relation is prominent for a pair of predicates, it might indeed be tempting to conclude that the association of these predicates with this relation is representative. However it is important to recall that we are only considering explicit expressions extracted with our method. Therefore, we must consider the possiblity that the discourse marker which triggered the extraction of the pair in different contexts from our corpus might still be needed to lexicalize the relation.

Regarding the second question, if two relations often occur with the same predicates, some other elements must be at play to enable the differentiation between these relations in context. We can hypothesize that the difference stems from other lexical elements, for instance the marker, which might be necessary to trigger either relation. It could also be the case that predicate-argument mappings are responsible for the different possible interpretations of a pair of predicates.

In order to attempt to answer these questions, we first extract the number of occurrences each pair in *Lecsie* has with the different relations, and classify them according to the distribution of relations they occur with. Then, for each class we select illustrative pairs of predicates and extract corresponding instances from our corpus, and analyze them.

4.1.2.1 Relation distribution by pair of predicates

In order to filter out very unfrequent pairs of predicates, we only take into account pairs which occur at least 10 times overall in *Lecsie* (adding up occurrences for all relations with the pair). In *Lecsie-en*, there are 4617217 distinct pairs in total, among which 604129 have a count of at least 10 and are thus considered in this study.

For each sufficiently frequent pair, we compute the proportion of occurrences of each relation to the overall count of occurrences for all relations with the pair.

For each verb pair existing in *Lecsie*, we extracted information as shown in Table 4.5 for some examples of verb pairs from *Lecsie-en*: we combined the counts of triples composed of the verb pair and each relation, and computed the proportion of each relation to the overall count of the pair with any relation.

The pairs are then grouped according to several distinct conditions:

Verb 1	**Verb 2**	Overall	contrast	cause	continuation	temporal	alternation	elaboration
Translate	Understand							
Counts		47	6	19	2	18	6	0
Proportions		1	0.13	0.40	0.04	0.38	0.13	0
Arrest	Charge							
Counts		3398	483	226	741	1655	206	87
Proportions		1	0.14	0.07	0.22	0.49	0.06	0.02
Calm down	Wait							
Counts		145	2	100	3	40	0	0
Proportions		1	0.01	0.69	0.02	0.28	0	0
Acquit	Convict							
Counts		1922	1381	60	90	167	109	115
Proportions		1	0.72	0.03	0.05	0.09	0.06	0.06
Believe	Know							
Counts		2854	1756	524	292	166	108	8
Proportions		1	0.61	0.18	0.10	0.06	0.04	0.00
Return	Stay							
Counts		1490	456	260	28	535	210	1
Proportions		1	0.31	0.17	0.02	0.36	0.14	0.00

Table 4.5: Examples of verb pairs and frequencies for each relation from *Lecsie-en*.

1. The most frequent relation for the pair accounts for more than 60% of the overall occurrences of the pair and all other relations each account for less than 30% of the overall occurrences,

2. The most frequent relation for the pair accounts for more than 30% but less than 60% of the overall occurrences of the pair and all other relations each account for less than 30% of the overall occurrences,

3. The most frequent relation for the pair accounts for more than 60% of the overall occurrences of the pair and the second most frequent relation accounts for more than 30% of the overall occurrences,

4. The most frequent relation for the pair accounts for more than 30% but less than 60% of the overall occurrences of the pair and the second most frequent relation accounts for more than 30% of the overall occurrences,

5. All relations each account for less than 30% of the overall occurrences.

Group 1 indicates that one particular relation is typical of the verb pair in explicit expressions. For example, in Table 4.5, the pair {*Calm down,Wait*} belongs to this group: causal markers account for almost 70% of the instances of this pair extracted from the corpus. The pairs {*Acquit, Convict*} and {*Believe, Know*} also belong to this group, with contrastive relations as most frequent.

Group 2 indicates that one relation is more frequent than the others, although not as typical as in Group 1. The pair {*Arrest, Charge*} belongs to Group 2 with temporal relations as most frequent.

Group 3 indicates that two relations make up for almost all occurrences of the pair, with one more typical than the second.

Group 4 indicates that two relations out of the six appear more frequently than the others, although it is not clear cut. The pair {*Translate, Understand*} belongs to this Group, with causal and temporal relations significantly more frequent than the others, and the pair {*Return, Stay*} is another example with contrast and temporal relations.

Finally, Group 5 gathers pairs which show no substantial difference between relations.

Table 4.6 shows the resulting groups for each relation with proportions of the overall number of distinct pairs in *Lecsie-en*.

Groups 1, 2 and 4 each represent about a third of the data, with Group 2 slightly more important. The high number of pairs belonging to Groups 1 and 2 shows that in most cases, one relation is dominant while all other relations have low occurrences in comparison, thus the best relation can be considered typical of the pair. This occurs most often with Contrast, Temporal and Cause, while associations with other relations appear to be seldom typical. In Group 4, we can see that common co-occurring relations are {Contrast, Cause}, {Temporal, Contrast} and {Temporal, Cause}. This is to be expected since these three relations have been found to be more commonly explicitly expressed in the application of our extraction method.

Groups 1 and 4 are the most relevant to our questions. Group 1 gathers pairs of predicates which are potentially typical of a relation. In order to have a qualitative look at this hypothesis, we illustrate the contents of this group for Contrast and Cause relations in Tables 4.7 and 4.8.

In Table 4.7, it appears very clearly that there is a strong link between the presence of a negation and the Contrast relation. Indeed, almost all of these pairs which are always found with a contrastive connective include a negation in one of the predicates. The two exceptions appearing in the table in fact contain a predicative adjective implying a negation (*unhurt*, *imperfect*). The impact of negations is studied in the subsequent Section 4.2.

Group 4 gathers pairs of predicates occurring with two main relations. In order to get a better understanding of these cases, we present some examples of contexts

Best relation	contrast	cause	continuation	temporal	alternation	elaboration	Total
(1) BR >60% & SR <=30%	12.27%	6.38%	0.71%	10.63%	0.56%	0.02%	30.56%
(2) BR >30% & SR <=30%	13.80%	8.90%	2.16%	10.97%	0.65%	0.04%	36.52%
(3) BR >60% & SR >30%							1.85%
Second best relation							
contrast	-	0.31%	0.03%	0.32%	0.01%	0.00%	
cause	0.41%	-	0.01%	0.18%	0.00%	0.00%	
continuation	0.07%	0.03%	-	0.03%	0.00%	0.00%	
temporal	0.25%	0.14%	0.01%	-	0.01%	0.00%	
alternation	0.01%	0.01%	0.00%	0.01%	-	0.00%	
elaboration	0.00%	0.00%	0.00%	0.00%	0.00%	-	
(4) BR >30% & SR >30%							29.00%
Second best relation							
contrast	-	4.03%	0.54%	5.03%	0.08%	0.03%	
cause	6.46%	-	0.53%	3.43%	0.05%	0.04%	
continuation	1.31%	0.36%	-	0.64%	0.03%	0.00%	
temporal	3.28%	2.15%	0.29%	-	0.09%	0.00%	
alternation	0.18%	0.14%	0.04%	0.17%	-	0.00%	
elaboration	0.03%	0.02%	0.00%	0.01%	0.00%	-	
(5) BR <=30% & SR <=30%							2.07%
Second best relation							
contrast	-	0.10%	0.04%	0.47%	0.01%	0.01%	
cause	0.24%	-	0.10%	0.21%	0.01%	0.00%	
continuation	0.20%	0.03%	-	0.16%	0.01%	0.00%	
temporal	0.13%	0.12%	0.06%	-	0.02%	0.00%	
alternation	0.03%	0.03%	0.02%	0.03%	-	0.00%	
elaboration	0.00%	0.00%	0.00%	0.01%	0.00%	-	

Table 4.6: Most frequent relations in *Lecsie-en*. BR=Best relation, SR=Second best relation.

for two verb pairs belonging to this group: {*Translate, Understand*} and {*Return, Stay*}, for which the relation distribution are displayed in Table 4.5 above.

Cause - Temporal

A highly rated pair of verbs instantiating these two relations is {*Translate, Understand*}, for which we present some examples of contexts below.

(4.1) Interpreting the complicated document still boggles many employers who must first **understand** the regulations and then **translate** the changes into new job classifications for employees.

Verb 1	Verb 2	Count
answer	not honor	222
propose	spoil	189
not [state verb] + enough	welcome	95
appear	not verify	85
contain	not rate	77
gauge	not measure	73
establish	not assert	72
not block	plan	60
[state verb] + unhurt	wound	58
not rule	reign	58
not [state verb] + clear	speculate	56
[state verb] + imperfect	work	53

Table 4.7: Most frequent pairs of predicates appearing with Contrast exclusively.

(4.2) And once I **understand**, I can act like a transducer and **translate** it in plain English.

(4.3) I'm **translating** the music so people can **understand** what I did.

(4.4) Japanese government plans to begin **translating** its laws into English, in hopes that they can be better **understood** by foreign companies.

Even though Examples 4.1 and 4.2 include a temporal cue (*first*, *then*, *once*), in these contexts they introduce a condition (the event of *understanding* is conditionned by the *translating* event).

A similar relation is introduced by causal connectives (*so*, *in hopes that*) in Examples 4.3 and 4.4.

Causal and Temporal relations are indeed often intertwined and signalled in similar ways, making them difficult to distinguish computationnally.

(4.5) With my rusty synagogue Hebrew, I could**n't** fully **understand** the headline, so Barghouti happily **translated** it for me.

(4.6) Females seem to **understand** it innately – so innately that they can **not translate** it for males.

In Examples 4.5 and 4.6, a causal relation holds although one of the events is negated.

Verb 1	Verb 2	Count
not reach	sever	88
bribe	shop	70
[state verb] + excited	not imagine	42
cram	go down	42
desire	substitute	42
glide	shove	42
stick around	treat	42
[state verb] + powerful	destroy	40
not run	organise	40
not [state verb] + secret	sell	39
[state verb] + dry	not eat	38
exacerbate	get out	37
break up	not meet	36

Table 4.8: Most frequent pairs of predicates appearing with Cause exclusively.

However, the nature of these links is different than in the previous examples. In Example 4.5, the ordering of the events is not the same: the event of *not understanding* causes the event of *translating*.

In Example 4.6, the modifying adverb *innately* must be taken into account when interpreting the relation. The inability to *translate* is caused by the manner in which the object is *understood*.

As previously mentioned, the impact of negations on semantic relations is further explored in Section 4.2.

Temporal - Contrast

The pair {*Return, Stay*} is found in similar proportions in temporal and contrastive contexts. We show some examples of these contexts below.

(4.7) She **stayed** from February until July, then **returned** Nov. 21, staying at the Bakersfield apartment while she was in the country.

(4.8) The group is scheduled to **stay** in Urumqi for one night and then **return** to Taiwan via Xian July 8.

(4.9) 1995 - Russian President Boris Yeltsin **returns** to the Kremlin after **staying** in the hospital for two months.

(4.10) Zinni is **staying** on, but Burns will **return** to Washington after a few stops in a handful of Arab countries.

(4.11) The city had **returned** to calm, but women still **stayed** inside.

In Examples 4.7, 4.8 and 4.9, a Temporal relation holds between the events of *staying* and *returning*, in this order, with the two predicates sharing the same subject.

In Example 4.10, a Contrast relation holds between the two events, signalled by the connective *but*. However in this example, the two predicates have different subjects, suggesting a contrast between one event and the other. This points to the importance of coreference patterns, which are studied in Section 4.3.

Finally, in Example 4.11, additionnally to the events having different subjects, it is the complete predicate *return to calm* which is involved in a contrastive relation with the predicate *stay*. This illustrates the importance of considering more than simple verbs when aiming to interpret semantic relations.

4.2 Negations

Here we aim to study the impact of the polarity of the clauses on the relation. More specifically, we attempt to answer the following questions:

1. Are predicates of negative polarity more often used in the expression of some semantic relations than others?
2. How does switching the polarity of one or both predicates affect the semantic relation between the two clauses they appear in?

We consider a predicate to be of negative polarity simply if a negation could be identified by our method, either by a negation particle in the scope of the predicate, or by its introduction by a downward entailing support verb (as detailed in Section 3.1.2.6).

4.2.1 Relation distribution as a function of polarity

In order to answer the first question, we extract the number of occurrences of each relation in *Lecsie* with the following criteria:

1. Both predicates have positive polarity;
2. One predicate has positive polarity and the other has negative polarity;
3. Both predicates have negative polarity.

For more readability, the results are presented in two separate tables, which complement each other.

In Table 4.9, we present the distribution between the three different categories for each relation, as well as overall. We can see that Category 1 makes up most of the occurrences in the corpus, with almost 85% of occurrences overall, while Category 2 represents almost 15% of occurrences overall, and finally Category 3 is very rare: less than 1% of occurrences overall. In general, we can therefore conclude that positive polarity predicates are much more frequent.

Relation	**Category 1** Two positives	**Category 2** One positive – one negative	**Category 3** Two negatives	**Count**
Contrast	0.78	**0.21**	0.01	17597299
Temporal	**0.92**	0.08	0.00	14540660
Cause	0.83	**0.16**	0.01	13900091
Continuation	**0.92**	0.08	0.00	5052350
Alternation	0.86	0.13	0.01	1796703
Elaboration	0.82	**0.17**	0.01	381591
Overall	0.85	0.14	0.01	53268694

Table 4.9: Distribution between different pair polarity categories for each relation in *Lecsie-en*.

There are however some significant differences among the relations. Continuation and Temporal relations have a much higher proportion of occurrences with two positive predicates. To the contrary, Contrast relations, and to a smaller extent Cause and Elaboration relations have a higher proportion of occurrences with predicates of opposite polarities compared to the overall proportion.

In Table 4.10, we present the distribution between the relations for each category. The relation distribution of Category 1 is similar to the overall distribution, while the two other categories show interesting differences. In Category 2, Contrast relations represent almost half of the cases, and Cause relations are also more frequent than in Category 1, while Temporal relations are much less frequent than in Category 1. In Category 3, Cause relations are this time more prominent than Contrast relations.

The results for Category 2 are in agreement with the study of Asr and Demberg (2015), who reported negative correlation values between a first clause of negative polarity and the presence of a Temporal relation, and positive correlation values for Comparison and Contingency relations.

Relation	**Category 1** Two positives	**Category 2** One positive – one negative	**Category 3** Two negatives	**Overall**
Contrast	0.30	**0.48**	0.38	0.33
Temporal	0.30	**0.15**	0.08	0.27
Cause	0.26	**0.28**	**0.46**	0.26
Continuation	0.10	0.05	0.04	0.09
Alternation	0.03	0.03	0.03	0.03
Elaboration	0.01	0.01	0.01	0.01
Count	45233034	7694609	341051	53268694

Table 4.10: Relation distribution for each pair polarity category in *Lecsie-en*.

Our results suggest that negative polarity in one of the predicates might be an interesting cue for the prediction of contrastive relations, while negative polarity in both predicates might be helpful to predict causal relations.

4.2.2 Switching polarity

We now focus on the second question which concerns the impact of switching the polarity of one or both predicates of a pair on the relation they are associated with. With our method, the same predicate with and without a negation particle are considered as two distinct predicates. For the purpose of this study, we group together pairs of the same predicates where both are positive, where one is positive and one is negative, and where both are negative.

To get representative results, we filter out pairs occurring less than 10 times, and in the remaining pairs we only consider those which have a typical link with a relation (more than 50% of occurrences with their best relation). Note that we simplify the cases with one negative form by summing occurrences for both possible forms (*not Verb 1/Verb 2* or *Verb 1/not Verb 2*) and taking the resulting best relation. As an example, Table 4.11 shows the results for the pair {Know,Read}, where Cause is the most frequent relation for Category 1 and Category 3 forms, while Contrast is the most frequent relation for Category 2 forms.

To analyze the effects of a polarity switch in the pairs of predicates on the relation occurring with them, we present in Table 4.12 the following results: for each relation, we show the proportion of pairs of predicates representative of this relation in their Category 1 form, which are representative of each relation in their Category 2 form. We do the same analysis for Category 1 and Category 3. For

Verb 1	**Verb 2**	contrast	cause	continuation	temporal	alternation	elaboration
Know	Read						
Category 1 Counts		264	**606**	42	366	66	5
Not know	Read						
Counts		143	150	8	193	21	8
Know	Not read						
Counts		106	53	8	4	17	0
Category 2 Counts		**249**	203	16	197	38	8
Not know	Not understand						
Category 3 Counts		13	**69**	4	3	16	0

Table 4.11: Example of counts for each category for the verb pair {*Know, Read*} in *Lecsie-en*.

example, 55% of pairs representative of a Temporal relation in their Category 1 form (both predicates of positive polarity) switch to a Contrast relation when one of their predicates switches to negative polarity (Category 2).

Due to the heavy filtering we operated in the selection of candidates (only pairs with more than 10 occurrences, and only pairs representative of a relation), and the fact that not all pairs exist in all categories, the number of triples eligible for this analysis is significantly reduced compared to the number of available triples in *Lecsie-en*: 10339 triples for cross results in Categories 1 and 2, and 592 in Categories 1 and 3 (as seen in the overall analysis on negations, pairs of two negative polarity predicates are very rare).

In this table we can observe that for most relations, the most common effect of switching polarity from Category 1 to Category 2 results in a Contrast relation in Category 2. We see however that causal pairs in Category 2 remain causal in the majority of cases when this switch is operated, although Contrast is also very frequent in Category 2 for these pairs. With a switch to Category 3, we see that Cause is more frequent, as predicted from our first analysis.

In order to examine these results qualitatively, we present a series of examples of verb pairs where a polarity switch between different contexts occurs and analyze the effects on the expressed relation.

Category 1: Contrast - Category 2: Contrast

{*Condemn, Rule out*}:

	Category 1						
	Contrast	Temporal	Cause	Continuation	Alternation	Elaboration	Overall
Category 2							
Contrast_neg	**0.87**	**0.55**	**0.41**	**0.92**	0.37	1.00	0.66
Temporal_neg	0.02	0.26	0.04	0.02	0.02	0.00	0.10
Cause_neg	0.10	0.18	**0.54**	0.02	0.34	0.00	0.22
Continuation_neg	0.00	0.01	0.01	0.02	0.01	0.00	0.01
Alternation_neg	0.00	0.01	0.01	0.03	0.26	0.00	0.01
Elaboration_neg	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Count	4325	3540	2054	230	189	1	10339
Category 3							
	Contrast	Temporal	Cause	Continuation	Alternation	Elaboration	Overall
Contrast_doubleneg	**0.78**	0.29	0.20	0.50	0.11	0.00	0.46
Temporal_doubleneg	0.01	0.03	0.02	0.00	0.11	0.00	0.02
Cause_doubleneg	0.20	**0.66**	**0.77**	0.50	0.11	0.00	0.50
Continuation_doubleneg	0.01	0.02	0.00	0.00	0.56	0.00	0.02
Alternation_doubleneg	0.00	0.01	0.01	0.00	0.11	0.00	0.01
Elaboration_doubleneg	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Count	244	153	182	4	9	0	592

Table 4.12: Best relation cross-tabulation for Category 1 (two positives) and Category 2 (one positive - one negative) of each pair in *Lecsie-en* (first part), and Category 1 and Category 3 (two negatives) (second part).

(4.12) (a) President Boris Yeltsin and other officials have strongly **condemned** the NATO action, but **ruled out** any Russian military intervention. (*Category 1*)

(b) Some **condemned** the document, contending it implied that Baptists could no longer try to convert Catholics, and vice versa.
But two Baptists who signed the document [...] said the agreement does**n't rule out** sharing one's faith. (*Category 2*)

{*Admit, Blame*}:

(4.13) (a) Survivors **admitted** they had ignored safety standards and crowded onto the pier, but still **blamed** authorities for not maintaining order at the landing. (*Category 1*)

(b) Although the Russians initially **blamed** the Nazis, Moscow finally **admitted** responsibility in 1990 after the end of the Cold War. (*Category 1*)

(c) But you ca**n't blame** Jennifer Shea for succumbing to the temptation – though she **admits** she may have gone a bit far. (*Category 2*)

In these examples, we observe that while in Category 1 the subjects of the predicates refer to the same entity, this is not the case for Category 2. This points to the importance of the arguments: a relation holding in different contexts with and without a negation on one of the predicates might indicate that the arguments imply different presuppositions about the events.

The effects of arguments are examined in Section 4.3.

Category 1: Cause - Category 2: Cause

{*Convict, Execute*}:

(4.14) (a) Simpson could be **executed** if he is **convicted** of a double murder. (*Category 1*)

(b) If Catalan Roman and Medina Villegas are **convicted** and sentenced to death, they will **not** be **executed** on the island. (*Category 2*)

{*Believe, Vote*}:

(4.15) (a) McCallister **voted** for Bush because she **believed** his warnings that 2004 Democratic nominee John Kerry would weaken the nation. (*Category 1*)

(b) "I did**n't vote** for anyone in the elections, because I **believe** in independence, not in federalism." (*Category 2*)

In Examples 4.14 and 4.15, the subjects of both predicates refer to the same entities in both categories.

The difference seems to stem from modifiers in the rest of the context: in Example 4.14(b), the modifier *on the island* seems to affect the factuality of the predicate *execute*: it is not actually said that the people in question will *not be executed*. Therefore the negation does not introduce a contrastive relation, and a causal relation still holds between the two predicates.

{*Release, Speak*}:

(4.16) (a) He claimed to have several guns and a bomb and said he would **release** the hostages from his room at the Boardwalk Inn if he could **speak** to his wife. (*Category 1*)

(b) The Defense official, who has reviewed the report, **spoke** on condition of anonymity because the findings have **not** been **released**. (*Category 2*)

{*Afford, Provide*}:

(4.17) (a) Congress should **provide** consistent funding to the immigration agency so it can **afford** its necessary upgrades without pricing any immigrants out of the American dream. (*Category 1*)

(b) He said he can **not afford** to get sick because he must **provide** financially for his wife and three children, including an autistic son. (*Category 2*)

In Examples 4.16 and 4.17, changes in coreference patterns seem to be at play. While subjects of both predicates in Example 4.16(a) refer to the same entity, they do not in Example 4.16(b).

Conversely, in Example 4.17(a) the subject of the second predicate corefers with the object of the first one, while in Example 4.17(b), it is the subjects that refer to the same entity.

Category 1: Cause - Category 2: Contrast

{*Qualify, Win*}:

(4.18) (a) In the event of no result being achieved, Pakistan would **qualify** for the final in Johannesburg Monday because they have **won** more matches in the tournament. (*Category 1*)

(b) I did **qualify** twice, but then did**n't win**. (*Category 2*)

{*Accuse, Blame*}:

(4.19) (a) Dunne said Sullivan was **accused** of murdering his wealthy wife in 1987 because he **blamed** her for their failure to fit in with the country club set. (*Category 1*)

(b) Palestinian leader Yasser Arafat has **not** directly **blamed** Israel for Sharif's death.

However, a senior Palestinian Cabinet minister has **accused** Israel of having assassinated Sharif. (*Category 2*)

In Examples 4.18(b) and 4.19(b), the Contrast relations appearing in Category 2 appear to be lexicalized mainly by the presence of a negation in one of the predicates, as well as by the connective. In Example 4.18, the temporality of the events also seems to play a role: in 4.18(a), the *winning* happens before the potential *qualifying*, while in 4.18(b), the *not winning* event happens after the *qualifying* event: different presuppositions come into play in these examples.

Category 1: Temporal - Category 2: Contrast

{*Elect, Win*}:

(4.20) (a) I know that Wayne Lukas was **elected** to the Hall of Fame last year and <u>then</u> **won** the Kentucky Derby a few days later. (*Category 1*)

(b) A couple of Republican women with previous political experience were **elected**.
<u>But</u> as Emily's List saw it, **no** Democratic woman had **won** a Senate seat completely on her own merits . (*Category 2*)

Similarly as the previous examples with a Cause relation in Category 1, it seems that with a Temporal relation in Category 1, Contrast relations in Category 2 are triggered by the negation.

We now turn to the study of the impact of shared arguments and coreference patterns.

4.3 Coreference patterns

Similarly to the previous study on polarity, we examine coreference patterns with respect to the following questions:

- Are some coreference patterns more often used in the expression of some semantic relations than others?
- How do different mappings of the arguments affect the semantic relation between two predicates?

4.3.1 Relation distribution as a function of coreference patterns

In their work on the relationship between coherence and coreference, Kehler et al. (2008) conducted psycholinguistic experiments aiming to evaluate several previously proposed hypotheses on bias in pronoun interpretation and anaphora resolution. We focus on two of these hypotheses: grammatical subject preference and grammatical role parallelism preference.

Grammatical preference (Crawley et al., 1990) denotes a bias towards referents that occupy the grammatical subject position of the previous clause. Grammatical role parallelism preference (Sheldon, 1974; Smyth, 1994) denotes a preference for referents which occupy the same grammatical role as the pronoun. Kehler et al. (2008) suggest that the expectation of the entities to be mentioned in a clause is influenced by its coherence relation with the previous clause.

Here we examine their hypothesis quantitatively, by studying whether we can find correlations between coreference patterns and coherence relations. In a way, we are exploring the reverse hypothesis: does the predicate-argument mapping of two clauses influence the semantic relation that exists between them?

We must however be careful in our interpretations, since the automatic coreference resolution tool we used in our extraction method is likely to make mistakes in the most ambiguous cases, which might be the most interesting ones. In order to attempt to answer this question, we examine the distribution between the different predicate-argument mappings for each relation.

As a reminder from Section 3.2.4.1, the coreference patterns are recorded in *Lecsie-coref* as two features corresponding to the matching entities for the subject and the object of the second predicate with respect to the first predicate: **S2match** and **O2match**. In the following, we use the following notations:

- **S1**: subject of the first predicate;
- **S2**: subject of the second predicate;
- **O1**: object of the first predicate;
- **O2**: object of the second predicate;
- For instance, **S2=S1** means that the subjects of the two predicates refer to the same entity;
- **S2=New** means that the subject of the second predicate does not refer to either **S1** or **O1**, and similarly for **O2=New**;
- **O2=None** means that the second predicate does not have an object (intransitive verb).

In Table 4.13 we report separately the distribution between values of **S2match** and **O2match**, respectively, for each relation.

Relation	**Count**	**Match (ratio)**						
		S2=S1	**S2=O1**	**S2=New**	**O2=S1**	**O2=O1**	**O2=New**	**O2=None**
Contrast	17858331	0.33	0.31	0.36	0.02	0.24	0.36	0.38
Temporal	14696414	0.34	0.27	0.39	0.03	0.20	0.33	0.44
Cause	14118194	0.36	0.27	0.37	0.03	0.21	0.32	0.44
Continuation	5198046	0.33	0.30	0.37	0.02	0.28	0.40	0.30
Alternation	1832459	**0.61**	**0.22**	**0.17**	0.02	**0.10**	0.43	0.45
Elaboration	387128	0.35	0.25	0.40	0.02	0.24	0.30	0.44
Overall	54090572	**0.35**	0.29	0.36	**0.02**	0.22	0.35	0.41

Table 4.13: Distribution between values of **S2match** (first part) and **O2match** (second part) for each relation in *Lecsie-coref*. **S2=New** and **O2=New**': the argument is a new entity, **O2=None**: the second predicate is intransitive.

Overall, we remark that **S2=S1** is more frequent than **S2=O1**, which might indicate a preference for the subject of the first predicate as referent. However, **O2=S1** is very rare. With **O2=O1** being relatively frequent, parallelism preference is also likely.

Most relations show similar distributions, although Continuation seems to be correlated with a reference to **O1**, either by **S2** or **O2**. The most striking difference, however, is for Alternation, for which there is a very high proportion of cases where **S2=S1**, almost twice as much as in other relations. **O2=O1** is much less present than in other relations, which seems to indicate that subject preference is much more prominent in Alternation instances.

In Table 4.14, we give another representation of this data which helps identify some effects more clearly. Instead of looking at the distribution between the different schemes for each relation, we look at the reverse statistics: the distribution between relations for each scheme. To analyze this data, it is important to compare the values to the overall distribution, to take into account the fact that some relations are much more frequent than others in general and are thus likely to be more frequent with any coreference pattern.

	S2=S1	**S2=O1**	**S2=New**	**O2=S1**	**O2=O1**	**O2=New**	**O2=None**	Overall
Contrast	0.32	**0.36**	0.32	**0.25**	**0.36**	0.34	0.31	**0.33**
Temporal	0.26	0.26	0.29	**0.34**	0.25	0.26	0.29	**0.27**
Cause	0.27	0.25	0.26	**0.32**	0.24	0.24	0.28	**0.26**
Continuation	0.09	0.10	0.10	**0.06**	0.12	0.11	0.07	**0.10**
Alternation	**0.06**	0.03	0.02	0.03	0.02	0.04	0.04	**0.03**
Elaboration	0.01	0.01	0.01	0.01	0.01	0.01	0.01	**0.01**
Count	19048971	15400102	19641499	1329488	11927851	18714210	22119023	54090572

Table 4.14: Distribution between relations for each value of **S2match** (first part) and **O2match** (second part) in *Lecsie-coref*.

These results confirm the preference for **S2=S1** in Alternation relations (twice more prominent than in the overall distribution). We also find that Contrast relations with reference to **O1** are more frequent than overall (36% of instances with **S2=O1** or **O2=O1** are Contrast, compared to 33% overall). Conversely, **O2=S1** is much more frequent with Temporal and Cause relations: they seem to be more prone to subject preference.

In Table 4.15 we report the distribution between the complete coreference patterns for each relation.

	S2=S1 O2=O1	**S2=O1 O2=S1**	**S2=New O2=S1**	**S2=New O2=O1**	**S2=S1 O2=New**	**S2=S1 O2=None**	**S2=O1 O2=New**	**S2=O1 O2=None**	**S2=New O2=New**	**S2=New O2=None**
Contrast	0.07	0.01	0.01	0.17	0.13	0.14	0.17	0.13	0.06	0.11
Temporal	0.06	0.01	0.02	0.15	0.13	0.15	0.13	0.13	0.07	0.16
Cause	0.07	0.01	0.02	0.14	0.14	0.16	0.12	0.13	0.06	0.15
Continuation	0.08	0.01	0.01	0.20	0.15	0.10	0.19	0.11	0.07	0.09
Alternation	0.04	0.01	0.01	0.06	0.28	0.29	0.11	0.10	0.03	0.07
Elaboration	0.05	0.01	0.01	0.19	0.13	0.16	0.12	0.13	0.05	0.15
Overall	0.06	0.01	0.01	0.16	0.14	0.15	0.15	0.13	0.06	0.13

Table 4.15: Coreference pattern distribution for each relation in *Lecsie-coref*.

In order to compare these results to the different bias hypotheses, we group these results accordingly in Table 4.16. The values corresponding to “parallelism bias” are obtained by summing columns **S2=S1 / O2=O1**, **S2=New / O2=O1**, **S2=S1 / O2=New** and **S2=S1 / O2=None**. The values corresponding to “non parallel” are obtained by summing columns **S2=O1 / O2=S1**, **S2=New / O2=S1**, **S2=O1 / O2=New** and **S2=O1 / O2=None**. The values corresponding to “subject bias” are obtained by summing all columns including **S2=S1** or **O2=S1**. The values corresponding to “Reference to O1” are obtained by summing all columns including **S2=O1** or **O2=O1**.

Relation	**Parallelism bias**	**Non parallel**	**Subject bias**	**Reference to O1**
Contrast	0.51	0.32	0.35	0.38
Temporal	0.48	0.29	0.37	0.34
Cause	0.50	0.29	0.39	0.35
Continuation	0.53	0.32	0.34	0.40
Alternation	0.67	0.23	**0.63**	0.21
Elaboration	0.53	0.27	0.36	0.38
Overall	0.51	0.30	0.38	0.36

Table 4.16: Proportions for each bias type and its converse.

This table demonstrates a clear prominence of parallel schemes in our data. As we observed before subject bias appears to be much more present for Alternation relations.

4.3.2 Impact of changing coreference patterns

To analyze the effects of a change in coreference patterns, we focus on the impact of switching between the patterns *S2=S1* and *S2=New* on the most frequent relation with which a pair of predicates is associated. In Table 4.17, we display for each relation the number of pairs of predicates which are most frequently associated with it when they occur with the pattern *S2=S1*, and with each relation when they occur with the pattern *S2=New*. As before, to get representative numbers we only take into account representative associations, where the pair of predicates is found with the considered relation in more than 50% of the total number of occurrences of the pair. Since the results with coreference patterns are very sparse, significantly less triples are considered in this study.

	S2=S1					
	alternation	elaboration	temporal	cause	contrast	continuation
S2=New						
elaboration	1	1	6	4	11	0
temporal	**138**	7	2826	677	808	168
cause	107	8	629	1715	**905**	165
contrast	116	4	**710**	**955**	2672	**277**
alternation	52	0	10	12	11	4
continuation	46	2	168	203	357	318

Table 4.17: Best relation cross-tabulation for **S2=S1** versus **S2=New**.

We note that for most relations (except alternation, and elaboration which is too rare to analyze quantitatively), a change in coreference pattern does not seem to affect the relation for most pairs, since the highest counts are found when the relation is unchanged. However we can identify other regularities. Pairs of predicates occuring with the relations Temporal and Cause with the pattern **S2=S1** are frequently found with a Contrast relation with the pattern **S2=New**. For Contrast relations with **S2=S1**, the second most frequent change in relation with **S2=New** is Cause.

We illustrate these observations with contexts of the pair {*Return, Stay*} which is most frequently found in temporal relations with **S2=S1** and in contrastive relations with **S2=New**. We provide both explicit and implicit examples, in order to demonstrate that our observations can hold without connectives, and at least partially rely on the predicate-argument mappings.

(4.21) 1995 - Russian President Boris Yeltsin **returns** to the Kremlin after **staying** in the hospital for two months.

(4.22) McAllister told them she **stayed** in New Jersey with her parents most of the last week. She claimed she *returned* to Boston on Saturday.

(4.23) Zinni is **staying** on, but Burns will **return** to Washington after a few stops in a handful of Arab countries.

(4.24) Her teacher decided to **stay** in Italy; Talam had to **return** to Sarajevo as she had no connections in Italy, and her home and family were in Bosnia.

(4.25) After they get out of prison, very few convicted criminals **stay** clean and sober. Most relapse and *return* to crime.

4.4 Conclusion

In this chapter, we have studied different aspects pertaining to the semantic relations between verbs, based on broad-coverage data. Specifically, we have studied the representativity of the associations we extract, and the types of relations which tend to co-occur with the same predicates. We have also studied the impact of negations on semantic relations, as well as the effect of changing coreference patterns.

This exploratory study has enabled us to demonstrate that relevant semantic information can be extracted from our collected data. Deeper analysis and other aspects could be explored in the future.

CHAPTER 5

SEARCHING FOR SIGNIFICANT RELATIONS

Contents

In the previous chapter, we have used simple raw counts in order to analyze our data and shown it can be an effective means to compare association strengths in some cases. When looking at distributions with respect to some fixed item in the triples (in our case, either the relation, or the pair of predicates), raw counts allowed us to compare the number of occurrences for the variable items, yielding interesting observations.

However, when we need to compare the association strength of all triples in order to rank them or to represent them for external purposes, quantifying association through raw count or frequency is not satisfactory. Indeed, triples composed of one or several rare items will occur with low frequency, which does not necessarily mean that their association itself is weak. Therefore, it is necessary to design association measures which can counteract the bias towards high frequency triples.

In this chapter, we focus on searching for significant associations. In Section 5.1, we review potential association measures from the literature, designed for sightly different purposes and present our own, specifically targeted at measuring the association strength of our triples. In Section 5.2, we detail our evaluation of these measures based on manual annotations, in order to determine the most significant ones. Finally, in Section 5.3, we investigate a different way of representing the semantic relations between two predicates, by automatically inducing optimal sets of connectives based on their association with pairs of predicates.

5.1 Defining association measures

An association measure is a tool aimed at identifying the strength of association between several items. In our case, these items are two predicates and a lexical relation, forming a triple.

The score produced by an association measure for a particular triple can be used independently to represent the magnitude of the association of its items, or it can be used as a comparative value with respect to other triples, in order to obtain a ranking.

By fixing some items of the triples, the scores can also be useful as a means to represent a specific aspect. For instance, considering all existing triples for a specific pair of fixed predicates (that is, selecting all the lexical relations appearing with this predicate) can be a means to produce a representation of the pair of predicates in terms of its score with each relation. We show such a representation in Figure 5.1.

Verb1/Verb2	Cause	Temp.	Contrast	Contin.	Altern.	Elab.
Scores	value1	value2	value3	value4	value5	value6

Figure 5.1: Representation of a pair of predicates {Verb1/Verb2} by its association score with each relation.

This type of representation is useful for external applications of our resource, and will be discussed in more details in Chapter 6.

There is no standard way of measuring association strength. Most research devoted to this question considers associations between pairs of items. For instance, a well-researched application is finding collocations by measuring the association strength of pairs of words occurring together in context.

In our work, we need to solve the more complex issue of measuring the association strength between triples of items. We adapted versions of standard lexical association measures based on information theory, as well as some measures specific to the association of a causal relation between items Do et al. (2011). We also experimented with a new measure specifically designed for our knowledge base.

In the following sections, we detail the different association measures we designed, inspired from related research.

Notations

Let us first introduce a few notations:

The set of all observed predicates, regardless of their position in a triple, is denoted as $\mathcal{V}$, and the set of all considered semantic relations is denoted as $\mathcal{R}$.

A triple is referred to as (v_1, v_2, r), where $v_1 \in \mathcal{V}$ is the first predicate, $v_2 \in \mathcal{V}$ is the second predicate, and $r \in \mathcal{R}$ is the semantic relation.

The number of occurrences (*i.e.*, the raw count) of a triple is denoted as follows: $C(v_1, v_2, r)$.

The total number of occurrences in *Lecsie* is the sum of the occurrences of all extracted triples:

$$N = \sum_{\forall v_1 \in \mathcal{V}, \forall v_2 \in \mathcal{V}, \forall r \in \mathcal{R}} C(v_1, v_2, r).$$

The observed relative frequency (corresponding to joint probability) of a triple is then: $F(v_1, v_2, r) = \frac{C(v_1, v_2, r)}{N}$.

The total observed relative frequency of a specific predicate is the sum of the relative frequencies of all triples containing this predicate, regardless of its position in the triple (as V_1 or V_2). That is, the frequency of a verb $v \in \mathcal{V}$ is:

$$F(v) = \sum_{\forall v_2 \in \mathcal{V}, \forall r \in \mathcal{R}} F(v, v_2, r) + \sum_{\forall v_1 \in \mathcal{V}, \forall r \in \mathcal{R}} F(v_1, v, r).$$

Similarly, the total observed relative frequency of a relation $r \in \mathcal{R}$ is:

$$F(r) = \sum_{\forall v_1 \in \mathcal{V}, \forall v_2 \in \mathcal{V}} F(v_1, v_2, r).$$

The expected relative frequency of a triple (v_1, v_2, r), which corresponds to the joint probability of the triple if its elements v_1, v_2 and r were independent, is:

$$E(v_1, v_2, r) = F(v_1) \times F(v_2) \times F(r).$$

Note that this expected relative frequency is based on occurrences in our extracted knowledge base and not in corpus: an occurrence of a verb is always associated to another.

5.1.1 Mutual Information measures

In information theory, mutual information is defined as the amount of information provided by the occurrence of an event about the occurrence of another event (Fano and Hawkins, 1961). Mutual information measures have been studied and used extensively in research on lexical collocation, which aims at discovering typical lexical associations between words (Manning and Schütze, 1999; Lin and Pantel, 2002; Evert, 2005).

For a pair of co-occurring items x and y, Pointwise Mutual Information (PMI) is defined as the logarithmic ratio of their joint probability to the expected joint probability if x and y were independent:

$$I(x, y) = \log \frac{F(x, y)}{F(x)F(y)} = \log \frac{F(x, y)}{E(x, y)}.$$

Hence, it represents an estimation of whether the probability of the co-occurrence of two items is greater than the *a priori* probability of the two items appearing independently. To apply this notion to our work, the aim is to estimate whether the co-occurrence of two predicates with a particular semantic relation is higher than the *a priori* probability of the three items occurring independently.

In order to generalize this measure to triples of elements, we defined PMI as the following (a similar generalization has been used by Moirón (2005) and Van de Cruys (2011)):

$$PMI(v_1, v_2, r) = \log(\frac{F(v_1, v_2, r)}{E(v_1, v_2, r)}).$$

One weakness of PMI is that is prone to overestimating low-frequency data (Church and Gale, 1991). As a result, several variants have been proposed with the intent of reducing this bias.

Bouma (2009) proposed a normalization for bivariate PMI, which we adapt to our trivariate definition. When three items of a triple only occur together (complete co-occurrence), we have:

$$F(v_1) = F(v_2) = F(r) = F(v_1, v_2, r),$$

$$E(v_1, v_2, r) = F(v_1, v_2, r)^3,$$

$$PMI(v_1, v_2, r) = \log(\frac{1}{F(v_1, v_2, r)^2}) = -2\log(F(v_1, v_2, r)).$$

Thus, normalized PMI is defined as:

$$normalized\_PMI(v_1, v_2, r) = \frac{PMI}{-2\log(F(v_1, v_2, r))}.$$

The values of normalized PMI lie between -1 and 1, approaching -1 when the items never appear together, taking the value 0 in the case of independence, and the value 1 when they always appear together.

To address the same problem, Lin and Pantel (2002) defined a weighted variant of PMI by multiplying PMI by a discounting factor, which counteracts the bias towards infrequent items by accounting for the least frequent item in the triple.

$$weighted\_PMI = discount \times PMI,$$

$$discount = \frac{F(v_1, v_2, r)}{F(v_1, v_2, r) + 1} \times \frac{\min[F(v_1), F(v_2), F(r)]}{\min[F(v_1), F(v_2), F(r)] + 1}.$$

5.1.2 Other association measures

Specificity

We also considered a measure of the specificity of association, inspired from a study on the ability of stochastic parsers to take into account verbal subcategorization frames (Mirroshandel et al., 2013). In this study, the measure is aimed at evaluating whether the parser has a strong preference to associate a given subcategorization frame to a given verb, by comparing the frequency of this association to all others.

In the context of our work, specificity is defined for each item as the ratio of the observed frequency of the triple to the sum of frequencies of triples containing all other possibilities for this item, with the two other items fixed. The overall specificity of a triple is thus defined as:

$$specificity(v_1, v_2, r) = N \times \frac{1}{3}\left(\frac{F(v_1, v_2, r)}{\sum\limits_{v \in \mathcal{V}} F(v_1, v, r)} + \frac{F(v_1, v_2, r)}{\sum\limits_{v \in \mathcal{V}} F(v, v_2, r)} + \frac{F(v_1, v_2, r)}{\sum\limits_{r \in \mathcal{R}} F(v_1, v_2, r)}\right)$$

Causal association

Do et al. (2011) focused on a task quite similar to ours: identifying causality between event pairs in context. Since they defined an event as an action or occurrence that happens with associated arguments, they obtained very sparse counts for pairs of predicates with their lists of arguments, which required an association measure adapted to this type of information.

They proposed a measure for cause-effect association (CEA) between two events e_i and e_j, triggered by two predicates p_i and p_j:

$$CEA(e_i, e_j) = s_{pp}(e_i, e_j) + s_{pa}(e_i, e_j) + s_{aa}(e_i, e_j),$$

where s_{pp} measures the association between event predicates, s_{pa} measures the association between the predicate of an event and the arguments of the other event, and s_{aa} measures the association between event arguments. The association between event predicates is defined as follows:

$$s_{pp}(e_i, e_j) = PMI(p_i, p_j) \times \max(u_i, u_j) \times IDF(p_i, p_j) \times Dist(p_i, p_j).$$

In this definition, *IDF* denotes the inverse document frequency, which has no equivalent in the type of information we consider, and similarly $Dist(p_i, p_j)$, the distance between two events in a document, does not apply to our task. The component $\max(u_i, u_j)$ aims to take into account whether predicates p_i and p_j appear most frequently with each other. u_i and u_j are defined as:

$$u_i = \frac{F(p_i, p_j)}{\max\limits_k [F(p_i, p_k)] - F(p_i, p_j) + \varepsilon},$$

$$u_j = \frac{F(p_i, p_j)}{\max\limits_k [F(p_k, p_j)] - F(p_i, p_j) + \varepsilon}.$$

The idea of incorporating a measure of whether items appear most frequently with each other is very attractive for our work. Thus we adapt their measure as follows:

$$U_{\text{do}}(v_1, v_2, r) = PMI(v_1, v_2, r) \times max\{U_{v_1}, U_{v_2}, U_r\},$$

where:

$$U_{v_1} = \frac{F(v_1, v_2, r)}{\max_{v \in \mathcal{V}}[F(v, v_2, r)] - F(v_1, v_2, r) + \varepsilon},$$

$$U_{v_2} = \frac{F(v_1, v_2, r)}{\max_{v \in \mathcal{V}}[F(v_1, v, r)] - F(v_1, v_2, r) + \varepsilon},$$

$$U_r = \frac{F(v_1, v_2, r)}{\max_{r \in \mathcal{R}}[F(v_1, v_2, r)] - F(v_1, v_2, r) + \varepsilon}.$$

5.1.3 Designing a new association measure

Finally, we designed our own measure which aims to evaluate the contribution of each item in the triple to the overall triple (this measure is similar to specificity). For each item, we consider how frequently it is associated with the two other items compared to the most frequent association of any item with these two others. Formally, our measure is defined as:

$$W_{combined}(v_1, v_2, r) = \frac{1}{3}(w_{v_1} + w_{v_2} + w_r),$$

where:

$$w_{v_1} = \frac{F(v_1, v_2, r)}{\max_{v \in \mathcal{V}}[F(v, v_2, r)]},$$

$$w_{v_2} = \frac{F(v_1, v_2, r)}{\max_{v \in \mathcal{V}}[F(v_1, v, r)]},$$

$$w_r = \frac{F(v_1, v_2, r)}{\max_{r \in \mathcal{R}}[F(v_1, v_2, r)]}.$$

The contribution of each item is thus measured as the ratio of the observed relative frequency of the triple to the highest frequency of a triple composed of the two other fixed items and any other item in this position.

5.2 Evaluation of the association measures

In order to evaluate the extracted triples and make use of our resource in diverse applications, it is necessary to assess the validity of our measures and find out which ones are the most sensible. It is however not trivial to determine what makes a measure sensible, and how to evaluate this.

In Section 5.2.1, we present a study of the correlations between the different measures we implemented, in order to find out how they are related to each other. In Section 5.2.2, we then present two manual annotation tasks designed to evaluate each measure against human intuition. These tasks also serve as an intrinsic evaluation of our extraction method.

5.2.1 Correlations of the association measures

We start by studying the Pearson correlation between all of our measures. The results for French and English are shown in Tables 5.1 and 5.2.

	intra	**inter**	**count**	**PMI**	**discounted PMI**	**normalized PMI**	**specificity**	U_{do}	$W_{combined}$
intra	1	0.09	**0.99**	-0.05	0.05	-0.02	-0.04	0.14	0.09
inter	0.09	1	0.21	-0.03	0.01	-0.02	-0.01	0.03	0.02
count	0.99	0.21	1	-0.05	0.05	-0.02	-0.04	0.15	0.09
PMI	-0.05	-0.03	-0.05	1	**0.89**	**0.99**	0.73	0.03	0.77
discounted_PMI	0.05	0.01	0.05	0.89	1	**0.91**	0.58	0.04	0.76
normalized_PMI	-0.02	-0.02	-0.02	0.99	0.91	1	**0.73**	0.03	**0.78**
specificity	-0.04	-0.01	-0.04	0.73	0.58	0.73	1	0.02	**0.76**
U_{do}	0.15	0.03	0.15	0.03	0.04	0.03	0.02	1	0.04
$W_{combined}$	0.09	0.02	0.09	0.77	0.76	0.78	0.76	0.04	1

Table 5.1: Pearson correlation values between association measures on French data.

Since there are more intersentential instances in English, we find a higher correlation between **intrasentential counts**, **intersentential counts** and **overall raw counts** than in French, while the correlation between intrasentential counts and overall counts in French is very high. By definition, all measures based on **PMI** are highly correlated. The measure $\boldsymbol{U_{do}}$ doesn't show much correlation with any measure in either language, except with count values in the English version, comparatively to its correlation with other measures, indicating U_{do} might be more influenced by frequency values than measures like PMI, specificity or $W_{combined}$, and might be less adequate for the estimation of association strength for low-frequency

	intra	inter	count	PMI	discounted PMI	normalized PMI	specificity	U_{do}	$W_{combined}$
intra	1	0.69	**0.99**	-0.02	-0.01	-0.01	-0.02	0.54	0.04
inter	0.69	1	**0.80**	-0.02	-0.01	-0.02	-0.02	0.25	0.02
count	0.99	0.80	1	-0.03	-0.01	-0.01	-0.02	**0.51**	0.04
PMI	-0.02	-0.02	-0.03	1	**0.93**	**0.99**	0.65	0.02	0.66
discounted_PMI	-0.01	-0.01	-0.01	0.93	1	**0.95**	0.55	0.04	0.64
normalized_PMI	-0.01	-0.02	-0.01	0.99	0.95	1	0.64	0.03	0.67
specificity	-0.02	-0.02	-0.02	0.65	0.55	0.64	1	0.01	**0.79**
U_{do}	0.54	0.25	0.51	0.02	0.04	0.03	0.01	1	0.04
$W_{combined}$	0.04	0.02	0.04	0.66	0.64	0.67	0.79	0.04	1

Table 5.2: Pearson correlation values between association measures on English data.

triples. Finally, we find a quite elevated correlation between $W_{combined}$, **specificity**, and the **PMI** measures, which all show low correlation with measures directly based on raw counts. These measures might therefore be potentially appropriate for the treatment of high-frequency versus low-frequency data.

5.2.2 Manual evaluations

The method we designed to validate our association measures relies on comparison to human intuition: if humans judge that two verbs are linked by a semantic relation, we expect a sensible measure to produce a high association score for this triple.

We performed two types of intrinsic evaluations. We started with annotations for pairs of verbs coupled with specific relations without context. This evaluation was only performed for French, as the process proved difficult and the results were not conclusive. As a second step for intrinsic evaluation, we decided to work with verbs in contexts extracted from our corpora, in order to make the task more intuitive in terms of the expected judgments to be made, and to obtain real-valued associations corresponding to human judgment in different contexts, instead of binary values corresponding to a single judgment for each triple without context.

5.2.2.1 Selection of the candidates

In their study on the evaluation of association measures for the extraction of collocations from corpora, Evert and Krenn (2005) note that in most approaches, the evaluation is done by manually annotating true positives in small lists of the highest ranking candidates for each measure. The performance of a measure is then quantified by the precision values obtained from these annotations of true positives.

Our evaluation method proceeds differently. Instead of looking only at the highest ranking candidates (triples), we select candidates with random scores, in order to cover a wide range of values for each measure. This then allows us to assess its validity on the whole scale.

In order to make the task more manageable, we focus on three semantic relations among the most frequent ones found in *Lecsie*. For French, the three relations we investigated are Cause, Contrast and Narration. For English, we studied instances of Cause, Contrast and Temporal relations.

5.2.2.2 Experimental protocols

Out of context evaluation

The protocol for out of context annotations, done for French only, is the following:

- For each relation, 100 pairs of verbs are selected, with scores ranging on the whole scale for our various measures, as well as pairs that do not exist in *Lecsie-fr* (with null frequency);
- Annotators (3 experts and 2 non experts) are asked to make a binary judgment about each pair: is it possible to make up a plausible context where the two verbs would be linked by the considered relation ?
- Finally, an adjudication was performed by the 3 expert annotators for the 100 pairs selected for the Contrast relation.

The 300 pairs of verbs used for these annotations are listed in Appendix B.

In context evaluation

Context plays an important role in determining the relation linking two verbs. The same verb pair can be linked by a specific relation in one context, but by another relation in another context, or no relation at all. Thus it appears necessary to take context into account to evaluate the association of a triple.

Do et al. (2011) performed a related experiment to evaluate causal relations in context. They asked two annotators to evaluate causal event pairs in news articles, with no restriction of the distance between the clauses expressing these events.

We performed a similar experiment aiming to evaluate the association between verb pairs and relations. The protocol for in context evaluation, performed for both languages, is the following:

- For each relation, 5 pairs of verbs are selected, with scores ranging on the whole scale of our various measures;
- For each selected verb pair, 40 pairs of adjacent clauses containing the two verbs are extracted from the corpus used to build *Lecsie*: *FrWaC* for French and *Gigaword* for English;
- Annotators (3 experts) are asked to make a binary judgment about each pair in each context: can the pair be linked by the corresponding relation in the specific context it appears in ?
- Finally, an adjudication for all contexts is performed by the 3 expert annotators.

To extract relevant contexts for each pair of verbs, we search the corpus for instances of adjacent clauses each containing one of the verbs. In order to get representative examples of all configurations in which can extract the pairs, we select both explicit (containing a discourse marker corresponding to the relation being studied, using the same lexicon of connectives as in our triple extraction method) and implicit (containing no discourse marker) instances. Indeed, we are interested in finding out whether the relation holds similarly with and without connective. We also select both intrasentential and intersentential instances. The contexts are then made up of the whole sentences containing the relevant clauses (thus we have one- and two-sentence instances).

The 40 contexts for each pair are then extracted randomly from the instances corresponding to these specifications, with the following proportions:

- 10 intrasentential and implicit,
- 10 intrasentential and explicit,
- 10 intersentential and implicit,
- 10 intersentential and explicit.

This distribution of contexts is not representative of the whole distribution in our resource, but it allows us to evaluate each configuration separately.

Examples 5.1 and 5.2 show two contexts for the pair *work* / *succeed*, to be evaluated for the relation Cause. Example 5.1 is intrasentential and marked with the connective *because*, while Example 5.2 is intersentential and unmarked.

(5.1) I think some of them will succeed because we work very hard to remind them of what's important.

(5.2) Analysts and industry observers say Ovitz and Eisner will have to work in tandem if their relationship is to succeed. " For this to work, they have to operate as a true partnership, like Michael Eisner and Frank Wells did," said Peter Paterno, former president of Disney's Hollywood Records division.

We display the annotation results for these examples in Table 5.4 on page 148, in the next section concerning the results of our experiments.

The 5 pairs for each relation are shown in Table 5.5 for French and in Table 5.6 for English.

In this task, instead of producing a binary value representing whether the association of a triple is positive or negative as in the previous task, we aim to express the human association value of a triple as a real value representing the typicality of the link. For this purpose, we designed a measure of human association. Our measure, called HAR for Human Association Ratio, is expressed as the proportion of contexts in which the association is confirmed according to human judgment, for a specific triple.

More formally, it is defined as the cardinal of the set of contexts judged positive (noted $|\mathcal{P}|$) divided by the cardinal of the set of all contexts for the pair (noted $|\mathcal{C}|$):

$$\text{HAR} = \frac{|\mathcal{P}|}{|\mathcal{C}|}.$$

5.2.2.3 Annotation results: inter-annotator agreement and adjudication

Wide differences in the judgments made by annotators on each instance may arise, due to several possible reasons. These reasons include the difficulty of defining precise annotation guidelines, potential disagreement between annotators about their interpretation of these guidelines, differences in the intuition of each annotator, and in our case language comprehension difficulties (since we handle two languages and did not have access to enough native speakers with sufficient expertise, we had to rely on non-native but fluent speakers; however at least one native speaker with expertise in linguistics was part of our group of annotators for each language).

Therefore, the assessment of inter-annotator agreement for each task is important. A widely used agreement measure is the kappa statistic (Cohen, 1960), where the observed agreement between annotators is compared with chance agreement. Kappa values range from 0 (mere chance) to 1 (perfect agreement). There is however no definitive interpretation for intermediate values. In the field of NLP, a common and well-accepted interpretation of these values (first suggested by Krippendorff (1980)) is the following:

- Kappa ≤ 0.6: to be discarded
- $0.6 \leq$ Kappa ≤ 0.8: tentative agreement
- Kappa ≥ 0.8: definite agreement

Out of context evaluation

We applied Cohen's Kappa on the annotations obtained for the out of context evaluation task. The resulting values are presented in Table 5.3 for each pair of annotators, as well as mean values for each subtask. Expert annotators are represented by the symbols 1, 2 and 3, while A and B represent our non experts, who performed only the subtasks of Contrast and Narration.

Annotators	Cause	Contrast	Narration
1/2	0.16	0.55	0.43
1/3	0.22	0.57	0.46
2/3	0.13	0.56	0.37
A/B		0.46	0.48
A/1		0.55	0.47
A/2		0.44	0.32
A/3		0.56	0.23
B/1		0.39	0.47
B/2		0.41	0.35
B/3		0.32	0.16
average kappa	0.17	0.56	0.42

Table 5.3: Inter-annotator agreement: Cohen's Kappa for each pair of annotators, and average values by relation.

The average Kappa values are the following: 0.17 for Cause, 0.42 for Narration and 0.56 for Contrast. If a 0.6 Kappa serves a measure for a feasible semantic judgment task, out of context judgments appear very difficult and many differences arise between annotators. Indeed, it can be very difficult to think of a typical context in which the pair of verbs could be linked by the relation, and finding just one positive context does not necessarily mean that a positive judgment is justified in general for the triple.

Contrastive pairs seem to be a relative exception since we obtained a Kappa value close to 0.6. Indeed, Contrast seems to be an easier relation to identify without context, intuitively. For this reason, adjudication of the results was performed by the three expert annotators for the Contrast subtask only.

In context evaluation

The resulting annotations and adjudication for Examples 5.1 and 5.2 on page 146 are displayed in Table 5.4, as an example.

	Context 5.1	**Context 5.2**
Annotator 1	yes	yes
Annotator 2	yes	no
Annotator 3	yes	no
Adjudication	yes	no

Table 5.4: Examples of annotations and adjudication.

The average Kappa values for each subtask are presented in Tables 5.5 and 5.6.

In French, we obtained an average agreement of 0.75 for Cause, 0.75 for Contrast, and 0.82 for Narration, resulting in an overall average agreement of 0.78. In English, we obtained an average agreement of 0.82 for Cause, 0.75 for Contrast, and 0.90 for Temporal, resulting in an overall average agreement of 0.82.

All of these values are very close to or above the 0.8 threshold proposed by Krippendorff (1980) for definite agreement. As a comparison, Do et al. (2011) obtained an inter-annotator ratio of 0.67 for the task of annotating related but not necessarily causal pairs, and 0.58 for causal pairs.

After adjudication of the results, for each verb pair we computed HAR values. The second part of Tables 5.5 and 5.6 presents the overall results as well as the decomposition between explicit and implicit contexts. According to these results, exploiting our resource built on explicit contexts for computing implicit discourse relations may be more straightforward for some relations than others. Indeed, while similar association rates seem to hold with and without connectives for temporal relations, the expression of contrastive relations seems to require a connective in most contexts (the HAR values for implicit contexts always being very low, even when the overall HAR score is high). For the Cause subtask we obtain more intermediate values for implicit context, indicating that causal relations can be expressed without a connective in some contexts.

Verb pair	Translation	HAR		*Lecsie* scores		
		All	Impl.	Count	W_{comb}	norm. PMI
Cause		**avg. Kappa: 0.76**				
inviter/souhaiter	invite/wish	0.13	0.09	142	0.76	0.29
promettre/élire	promise/elect	0.26	0.26	62	1	0.48
aimer/trouver	like/find	0.38	0.32	109	0.81	0.04
bénéficier/créer	benefit/create	0.51	0.45	266	1	0.47
aider/gagner	help/win	0.54	0.47	336	1	0.52
Contrast		**avg. Kappa: 0.75**				
proposer/refuser	propose/refuse	0.59	0.21	123	0.84	0.25
augmenter/diminuer	increase/decrease	0.64	0.33	385	1	0.57
tenter/échouer	try/fail	0.64	0.29	69	1	0.52
gagner/perdre	win/lose	0.72	0.32	285	1	0.39
autoriser/interdire	authorize/forbid	0.75	0.26	94	1	0.42
Narration		**avg. Kappa: 0.83**				
parler/réfléchir	speak/think	0.42	0.26	42	0.77	0.37
acheter/essayer	buy/try	0.70	0.50	53	0.90	0.44
atteindre/traverser	reach/cross	0.78	0.68	50	1	0.37
commencer/finir	begin/end	0.80	0.76	21	0.51	0.28
envoyer/transmettre	send/transmit	0.83	0.70	65	1	0.43

Table 5.5: Annotation results for French: Average kappa values for each subtask; Human association ratio (HAR) values for all instances and implicit only; Count, $W_{combined}$ and normalized PMI scores.

The third part of Tables 5.5 and 5.6 presents the values for each triple of two examples of measures, $W_{combined}$ and normalized PMI, as well as the raw counts.

5.2.2.4 Evaluation of the association measures

Out of context evaluation

We evaluated the association measures to see if they could discriminate between the two groups, those judged positively or negatively according to human annotations. For this purpose, we performed a MannWhitney-U statistical test, reported in Table 5.7.

According to this test, all of our measures proved to be discriminative, with the

Verb pair	**HAR**		***Lecsie* scores**		
	All	**Impl.**	**Count**	W_{comb}	**norm. PMI**
Cause		**avg. Kappa: 0.82**			
work/succeed	0.47	0.38	194	0.52	0.23
read/know	0.51	0.33	565	0.52	0.23
think/feel	0.51	0.35	881	0.30	0.04
run/look	0.32	0.17	556	0.44	0.07
push/fall	0.50	0.41	116	0.23	0.02
Contrast		**avg. Kappa: 0.75**			
divorce/marry	0.50	0.10	238	0.61	0.44
believe/understand	0.38	0.00	477	0.41	0.10
acknowledge/defend	0.60	0.26	162	0.43	0.28
lose/fight	0.54	0.11	358	0.45	0.10
work/play	0.28	0.00	414	0.31	0.02
Temporal		**avg. Kappa: 0.90**			
ask/smile	0.90	0.85	875	0.70	0.29
fire/wound	0.87	0.89	2150	0.67	0.41
travel/visit	0.74	0.53	353	0.73	0.42
arrive/greet	0.95	1.00	1005	0.72	0.45
disappear/find	0.88	0.84	536	0.69	0.23

Table 5.6: Annotation results for English: Average kappa values for each subtask; Human association ratio (HAR) values for all instances and implicit only; Count, $W_{combined}$ and normalized PMI scores.

exception of overall raw counts and intrasentential counts for which p>0.05. Concerning intersentential counts, the sparsity of this measure makes it a bad candidate in terms of measuring association strength, even if it turned out to be discriminative for our task.

In context evaluation

In order to assess the validity of the association measures defined previously with respect to human intuition, we computed the Pearson correlation between the HAR values and scores for each association measure. This was performed in two ways: first, considering all annotated contexts, and secondly considering only the subset of implicit contexts. The latter is important to quantify the actual impact of the

Measure	p-value
specificity	2.5e-11
U_{do}	2.9e-11
normalized_PMI	1.28e-10
weighted_PMI	1.96e-10
PMI	1.86e-10
$W_{combined}$	4.93e-10
local_PMI	4.95e-08
raw intersentential count	0.000904
raw intrasentential count	0.0721
raw count	0.116

Table 5.7: MannWhitney-U test results for Contrast annotation set: p-value for each association measure, sorted in increasing order. The values in the last part of the table are non-significant at the 0.05 level.

method, since explicit marking is already used as the basis of verb association in the same corpus. Implicit contexts, however, never appeared in the computation of the verb pair associations.

Table 5.8 shows a high correlation between human association and the measures $W_{combined}$ and specificity, in particular, much higher than simple counts.

We also observed results on each relation separately, although one should be careful drawing conclusions from these results since the correlations are then computed on 5 points only. These results (not shown here) show a lot of variation between relations. The U_{do} measure, designed for causal relations, does indeed produce good results for these relations, but does not generalize well to our other chosen relations.

5.2.3 Ranking the triples

From these experiments we conclude that the best three measures are: normalized PMI, specificity, and $W_{combined}$. The last two assign their maximal value to several pairs, so we used them in a lexicographical ordering to sort all associated pairs, using normalized PMI to break ties.

As an example, high-scoring triples according to these measures are shown in Table 5.9.

Association measures	**Correlation values**			
	Lecsie-fr		*Lecsie-en*	
	All	Impl.	All	Impl.
$W_{combined}$	0.72	0.74	0.79	0.70
specificity	0.75	0.76	0.77	0.70
normalized_PMI	0.75	0.81	0.70	0.72
U_{do}	0.38	0.50	0.70	0.57
PMI	0.71	0.76	0.65	0.50
discounted_PMI	0.72	0.76	0.65	0.50
raw count	0.17	0.24	0.49	0.58

Table 5.8: Pearson correlation between HAR and association measures in *Lecsie* (all contexts, or implicit only), in decreasing order for English data. All correlations are significant at the 0.05 level.

In observing these triples, we remark that all the pairs of predicates indeed seem to be intuitively related. The specific relation with which they are most strongly associated in *Lecsie* is however not necessarily the one which would have naturally come to mind if one had been asked to chose a relation among the six we propose. For instance, the pair *Hate/Love* might have been associated with Contrast instead of Alternation; the pair *Record/Forecast* might have been associated with Temporal instead of Cause.

Of course, as we have seen with our out of context evaluation, it can be very difficult to make judgments about the best relation to associate to a pair of predicates when they are taken out of context, and the typical contexts we might think of can be very different from what is actually seen in a corpus. However, these observations might also indicate that our set of relations, or the way we chose to group the discourse connectives to get at these relations, may not be optimal.

The next section addresses the question of finding better groups of connectives through data-driven methods.

5.3 Automatically defining optimal sets of markers

The previous experiments have aimed to evaluate the potential of our measures to assess the significance of association between verb pairs and our manually defined relations. In this section we reconsider these relations and propose to use our measures as a way to automatically define data-driven groups of connectives.

Verb 1	Verb 2	Relation	Norm. PMI
Concern	Dissatisfy	Alternation	1.0
Put up	Shut up	Alternation	0.94
Damage	Destroy	Alternation	0.93
Capsize	Drown	Temporal	0.82
Stir	Thicken	Temporal	0.82
Hate	Love	Alternation	0.77
Not drink	Smoke	Alternation	0.77
Desire	Publish	Cause	0.74
Record	Forecast	Cause	0.73
Acquit	Convict	Contrast	0.66

Table 5.9: Examples of high-scoring triples in *Lecsie-en* with corresponding normalized PMI values.

5.3.1 Manually defined classifications of connectives

In Section 3.2.1.3, we justified our groupings of discourse connectives into relation classes, based on manually constructed lexicons of connectives designed especially for NLP applications, and in particular discourse parsing: the *Lexconn* resource for French (Roze et al., 2012), and the lexicon of connectives used for PDTB annotations (Prasad et al., 2007).

Apart from such application-oriented resources, the literature contains many different classifications of discourse connectives, drawing upon a wide range of evidence including textual cohesion (Halliday, 1976), hypotactic conjunctions (Martin, 1992), cognitive plausibility (Sanders et al., 1992), substitutability (Knott et al., 2001), and psycholinguistic experiments (Louwerse, 2001).

Due to the different theoretical aspects at the basis of each classification and their motivations, there is a general lack of consensus on the characterization of discourse connectives. Additionally, the usual limitations of manually constructed resources apply in this domain as well: the process is very labor-intensive and the resulting resources are often incomplete.

5.3.2 Approaches for data-driven classification

Due to these observations, some research has been directed at constructing such classifications of connectives automatically. The idea is to use non-biased evidence from natural instances of connective usage in large corpora to induce empirically-grounded classes.

Hutchinson (2004) aimed to automatically acquire the meaning of discourse connectives with regard to three aspects often found in hand-coded taxonomies: polarity, veridicality and type. The latter aspect is the most relevant to our interests, as it concerns the type of relation expressed by a connective.

In this approach, classes were first manually defined for each aspect. Concerning types, three distinct classes are defined: *additive*, *temporal* and *causal*. Instances of connective usage were then extracted from a large corpus based on string patterns. Features were then extracted to describe each instance, in terms of lexical co-occurrence in the clauses linked by the connective, as well as other linguistic infromtion. A portion of these instances was then manually annotated to produce training data. Finaly, classification models were trained on this data to obtain a classification which was compared to a golden standard compiled from previous manual classifications (Knott et al., 2001; Louwerse, 2001).

Although this approach obtained highly accurate results with respect to this gold standard, it doesn't seem to appropriately solve the problems we mentioned. Indeed, the classes to which connectives are assigned have been decided manually, and are thus necessarily biased. Additionally, since the evaluation of the results is based on previous manual classifications, they can only prove that they manage to recreate a similar classification, or, perhaps, to validate existing ones. Finally, the reliance on manually annotated instances to learn the models implies the necessity of important manual work prior to the applicability of this method, as well as a bias relative to the annotators.

A more suitable approach to alleviate these problems is that of Alonso et al. (2002), who propose to use a clustering method to automatically group instances of connective usage extracted from a large corpus. It should be noted however that the aim is not to classify connectives directly, but to classify instances of their usage in context.

The main goal of clustering is to identify partitions in an unstructured set of objects described by certain features. This identification relies only on these features, and no annotated data is required. Instead of manually identifying classes to which the instances need to be assigned, only the number of clusters needs to be defined. Groups of instances are then created based on their similarity with respect to the features.

Alonso et al. (2002) rely on two sets of features. The first set is derived from a hand-coded lexicon of connectives with syntactic, discourse segmental, and rhetorical information, including "rhetorical content", which consists in relations such as reinforcement, concession, consequence or enablement. The second set of features is based on shallow text processing of the instances, and contains features relative to the position of the connective in the segment, the words surrounding the connective, the presence of a negation, etc.

The analysis of the results demonstrates that the clusters contain mostly instances with similar syntactic behavior of the connective. Various rhetorical contents can be found across clusters, and there is not clear-cut distinction between subordinating and coordinating connectives, contrary to what is found in manual and supervised classifications.

5.3.3 Empirically-grounded characterization of connectives

We propose a different approach, with a slightly different goal. Starting from the consideration that our manually defined groups of connectives might not be optimal, we aim to automatically derive empirically-grounded clusters of connectives based on the significance of association between connectives and pairs of predicates in context.

5.3.3.1 Process

In order to arrive at such clusters, we start from the initial data we collected from the corpus: triples composed of two predicates and a marker, and their occurrence count. The association measure of each triple is then computed (here we consider only $W_{combined}$, as our evaluation showed it to be the best measure – other measures could be used for comparison). We should emphasize that contrary to our previous experiments, the association measure concerns the connective only, no previous grouping by relation is performed. This step results in a list of triples **{verb1, verb2, connective}**) with a corresponding association score.

The next step consists in grouping our instances by pairs of predicates to obtain their representation in terms of their association scores with the different connectives. We thus produce a matrix of dimensions {number of verb pairs} x {104 connectives}. When a pair never appears with a certain connective, the score is set to zero in the corresponding position of the matrix. In effect, each verb pair is thus represented by a set of 104 feature values.

Finally, we apply a dimensionality reduction in the feature space, that is the space of the connectives.

5.3.3.2 Dimensionality reduction and its interpretation

The idea of dimensionality reduction is to decompose a multivariate dataset in a set of successive orthogonal components that explain a maximum amount of the variance. We use a Non Negative Matrix Factorization (NMF) algorithm as implemented in the scikit-learn toolkit (Pedregosa et al., 2011). NMF has the advantage of producing interpretable models.

NMF assumes that the data and the components are non-negative. It produces a factorization of the input matrix to reveal the meaningful "latent features" that are hidden in the data. Each feature in the input matrix can be viewed as being built up from these latent features, and NMF allows us to identify clusters of input features, in our case connectives, that share the same latent features. It is important to remember that each connective can be associated with multiple latent features and thus appear in multiple clusters.

Let L be our original matrix of dimension $m \times n$, where m is the number of verb pairs and n is the number of features (connectives). NMF aims at finding a rank k approximation of the form: $L_{(m \times n)} \approx W_{(m \times k)} H_{(k \times n)}$, where W and H are nonnegative matrices of dimensions $m \times k$ and $k \times n$, respectively.

W is the basis matrix, whose columns are the basis components. H is the mixture coefficient matrix, whose columns contain the contribution of each basis component to the corresponding column of L. The decomposition is obtained by optimizing for the squared Frobenius norm:

$$\operatorname*{argmin}_{W,H} ||L - WH||^2 = \sum_{i,j} L_{ij} - WH_{ij}.$$

The matrix W thus contains a new representation of our pairs of predicates, with k feature values describing each pair. The rows of the matrix H provide weights for the input connectives relative to the k groups. We can therefore conceive these weights as the strength of association between connectives and groups.

5.3.3.3 Results

The number of dimensions k to which the feature space is factorized needs to be predefined. In order to mirror the granularity of our initial groups and the usual descriptions of human analysts, we set this number to six dimensions, for easier manual analysis. Depending on the intended application, this number might not be optimal however: machine learning methods might perform better with a higher number of features. An optimization targeted at the intended application would therefore potentially be necessary.

In Figure 5.2, we show the strength of association of each connective in each group.

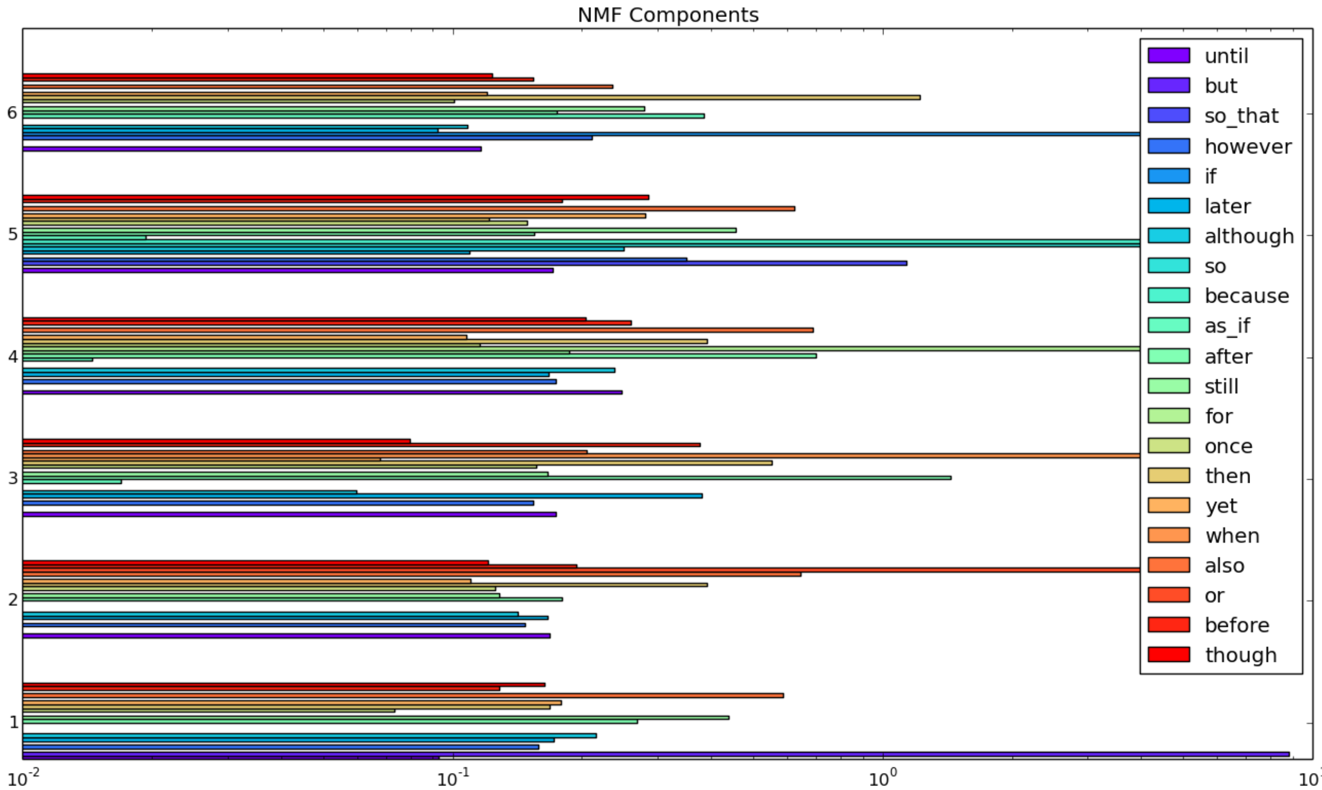


Figure 5.2: Weight of each connective for each group.

In Figure 5.3, we show the reversed representation, that is, for each connective we show its weight in the different groups.

In Table 5.10 we show the most prominent markers for each new dimension. Many markers appear with much lower weight in the new space, mainly because their usage is too infrequent.

Some of these dimensions isolate broad classes of semantic relations: for instance dimension 1 represents a contrastive or opposition type relation, dimension 2 an alternation type relation, dimension 3 a temporal one, dimension 5 a causal

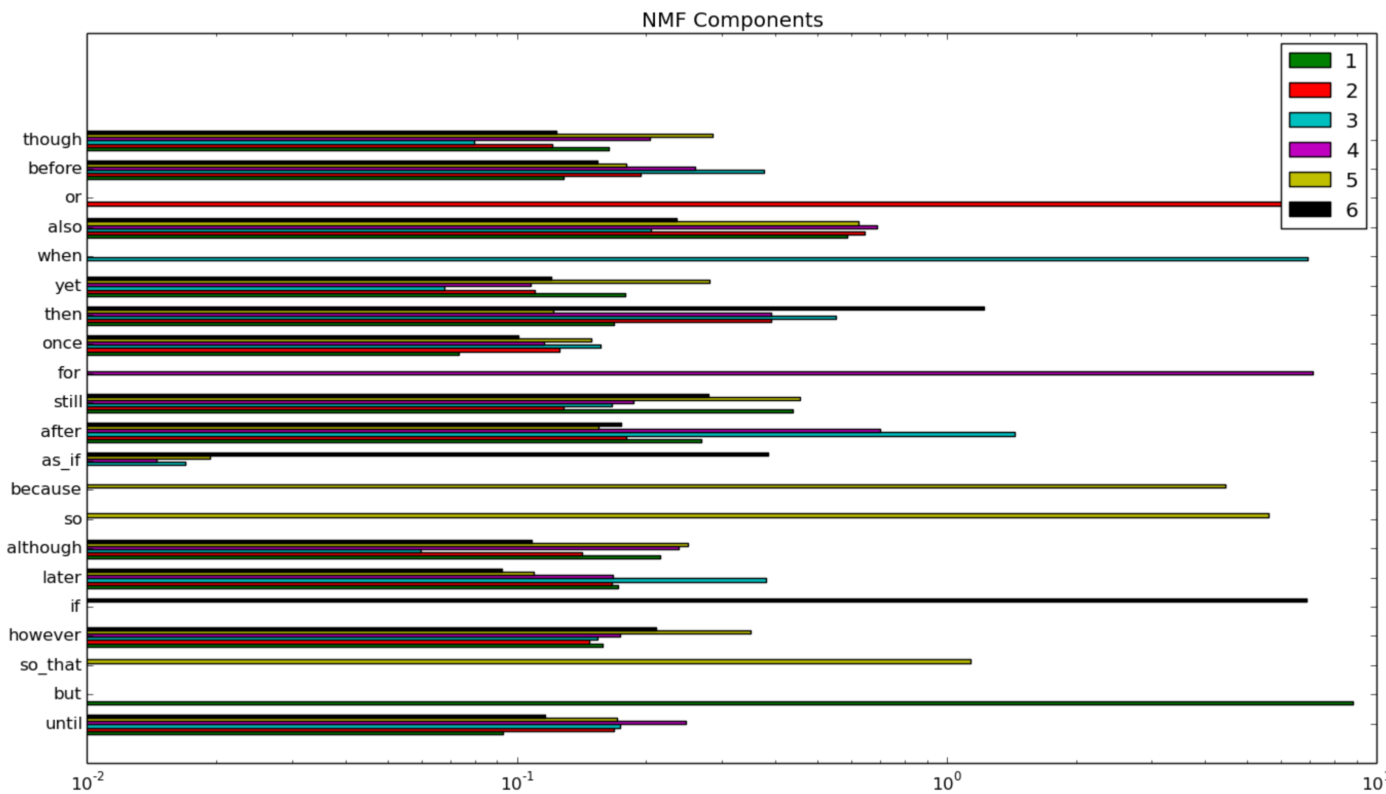


Figure 5.3: Weight of each group for each connective.

one (regrouping explanation, result and intentional causal relations like goal), dimension 6 a conditional or suppositional relation. Dimension 4 might also be a causal intentional family of relations, though the distinction between dimensions 4 and 5 is not so clear. Parallelism marked with *also* seems to pervade several of the semantic classes, but this is not so surprising for theories in which several relations may be at play between two discourse units: parallelism often combines with several other relations. We can see this as multiple markers in clauses are relatively common and natural when one of the markers is an indication of parallelism *also, too, as well—because he also, as a result she also, but he ... too, although she ...*

Dimension #	**Markers**
Dimension 1	but, (also, still)
Dimension 2	or, (also)
Dimension 3	when, after, (then)
Dimension 4	for, (also)
Dimension 5	because, so, so that, (still, also)
Dimension 6	if, then

Table 5.10: Most prominent markers for each dimension after NMF. Markers with significant weights, sometimes in several dimensions, are shown between parentheses.

as well (Asher, 1993).

In Section 7.1 we will discuss further the potential of using these representations in corpus-based applications, and specifically in discourse parsing.

5.4 Conclusion

This chapter has been concerned with defining measures of the association strength of the triples contained in *Lecsie*.

We have then presented our intrinsic evaluation of these associations in *Lecsie*, first by examining them manually out of context, and then, refining our evaluation approach, by manually annotating contexts in which these associations could be found. From this evaluation we have determined the association measures which are most correlated with human intuition.

Finally, we have looked into the potential of using our data to automatically derive optimal sets of connectives. Our manual analysis has shown that the groups we obtain seem to remain sensible to human intuition, and since they are obtained directly from the data, they might be better adapted to represent semantic associations in corpora.

In the next chapter we aim to use the scores obtained by applying our measures to construct vector-based representations of pairs of predicates, and to use these representations in the task of discourse parsing.

CHAPTER 6

SEMANTIC RELATIONS FOR DISCOURSE PARSING

Contents

In Chapter 2 we have shown that semantic relations are an essential part of the analysis of discourse structure and its interpretation, in which links are made between units of text at least in part because of the semantic information about the two main verbs of those clauses.

Discourse structure is important for many NLP tasks; it is responsible for the overall coherence of text, plays a role in information extraction tasks, like determining the temporal structure of eventualities in a text or automatically summarizing a text (Liu et al., 2007), and affects textual inference (Feng and Hirst, 2012), in which one has to infer certain relations between eventualities (Hashimoto et al., 2009; Tremper and Frank, 2013).

The task of inferring such discourse relations has two natural parts: first determining where a particular discourse unit attaches in the overall structure, and second labelling that attachment with a particular relational content.

Our resource provides vector-based representations of verb pairs based on their association with several semantic relations. This type of information has the potential to be valuable for discourse, by providing a semantically-informed representation of verb pairs found in discourse clauses, from which relations between events may be inferred.

In this chapter we aim to assess this potential by evaluating the impact of lexical features extracted from *Lecsie* on the tasks of attachment prediction and relation labelling.

We should emphasize that we are not trying to build a full discourse parser comparable to state-of-the-art approaches, but to demonstrate the potential of lexical information such as that collected in Lecsie, for discourse applications.

6.1 Attachment prediction

In this section we examine whether the information in *Lecsie* could help with the task of predicting attachment of discourse units. Predicting which discourse units are attached has proven to be a difficult task. As we discussed in Section 2.4.1, a common, simplifying assumption is that only adjacent units are related or that attachments are simply given in the discourse parsing task. For instance in the PDTB, most attachments are between adjacent units, with the exception of some Explicit instances. Most proposals for discourse parsing thus do not consider the problem of predicting which units are related but assume this as given and focus on predicting which relation links the units.

An exception is the study by Wellner and Pustejovsky (2007), who aim to identify the units related by connectives in the PDTB, or rather their lexical heads. However, this approach only focuses on explicit relations. The connective, which is considered given, serves as anchor to which each of the two units is connected. The lexical head of each unit is then search for among candidates in a fixed window of syntactic dependencies, using moslty syntactic features, except for some lexical features aiming to identify attribution relations. Each of the two units is searched for independently of the other, although inter-argument dependencies are considered in a second step.

Similarly, in their end-to-end discourse parser, Lin et al. (2010) aim to locate the first argument of PDTB explicit relations, using mostly information about the position of the connective in the sentence and its part-of-speech. However they only consider two possibilities: the first argument is either in the same sentence as the connective, or in the immediately previous sentence.

Such approaches are only applicable for explicit relations and depend heavily on the previous identification of the connective. Contrary to these approaches, we study the identification of related units without restricting to explicit relations. We posit that lexical information about related predicates, extracted from *Lecsie*, can help identify lexical cohesion between discourse units, and thus predict whether two units are related by a discourse relation, without having to rely on the presence of a connective. Indeed, intuitively, information about how strongly the events specified in two units are lexically related seems to be particularly relevant for predicting whether those units are related and should be attached in the structure. In order to assess the validity of this assumption, we present an evaluation based on the RST Discourse TreeBank (RST-DT, Carlson et al. (2007)).

6.1.1 Parsing in the RST framework

As detailed in Section 2.2, the RST tree representation is a hierarchical structure where EDUs are linked by rhetorical relations, forming larger units which are themselves recursively combined until the whole document constitutes a parse tree. Thus, an analogy can be made between such a hierarchical discourse structure and constituency based syntactic trees, where the constituents in the discourse tree are text spans from a document instead of words from a sentence for the syntactic tree. At each level of the tree, only adjacent units can be related, and more distant units can only be attached on higher levels in the structure, as part of larger complex units. Because of this analogy, previous work based on the RST framework largely explored the adaptation of common constituency based syntactic parsing techniques for discourse parsing (Soricut and Marcu, 2003; Subba and Di Eugenio, 2009; Sagae, 2009; Hernault et al., 2010; Feng and Hirst, 2012).

However, as discussed in Section 2.4.1, Li et al. (2014b) noted that several difficulties arise in the application of this idea, and proposed to use dependency parsing techniques instead. This idea was initially proposed by Baldridge et al. (2007), and applied on the Annodis corpus by Muller et al. (2012) (see Section 2.4.1). In order to apply this idea to the RST corpus, Li et al. (2014b) translate the RST trees into discourse dependency structures, represented as labeled directed graphs. In such a representation, each EDU is a node, and the nodes are linked by directed arcs labeled with dependency relations.

Through this transformation, a "head" EDU is distinguished in each CDU, representing the node to which units related to this CDU are attached. With this transformation, non-local dependencies can therefore be represented, and related EDUs are not necessarily adjacent: a unit can be attached to one occurring in a distant sentence in the discourse. Lexical cohesion can then be an insightful cue to identify the head unit to which another unit is connected.

Example 6.1 is a simplified extract from the RST corpus.

(6.1) [Some strongly feel the need]$_{\textbf{(A)}}$ [to be successful in their jobs]$_{\textbf{(B)}}$ [while others prioritize]$_{\textbf{(C)}}$ [spending time with their families.]$_{\textbf{(D)}}$

Figure 6.1 shows its constituency tree representation, and Figure 6.1.1 shows its conversion to a dependency structure representation.

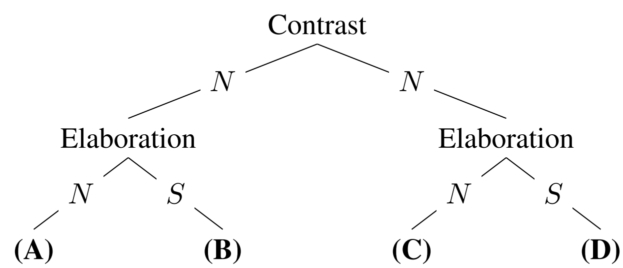


Figure 6.1: Constituency tree corresponding to Example 6.1.

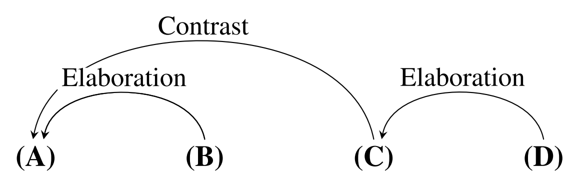


Figure 6.2: Discourse dependency structure corresponding to Example 6.1.

6.1.2 Studying attachment via *Lecsie*

In order to apply knowledge from *Lecsie* for attachment prediction, it is necessary to extract the relevant verbs from each EDU. This was done by applying the method described in Section 3.2 on the RST corpus, previously parsed using the Stanford Parser.

The previous Example 6.1 includes four units with the following relevant predicates: {*feel*, *be successful*, *prioritize*, *spend*}. These units form three pairs of attached units with specific RST relations: {**(A)**-**(B)**, *elaboration*}, {**(A)**-**(C)**, *contrast*}, {**(C)**-**(D)**, *elaboration*}. Two of these pairs are adjacent: **(A)**-**(B)** and **(C)**-**(D)**.

Some EDUs do not contain verbs however. This is partly due to the fact that some segments in RST are divided into two EDUs related by the *Same Unit* relation in order to avoid embedded units.

The RST corpus contains a total of 22174 EDUs in the training and test set, among which we identified 15600 containing at least one verb. The annotations contain 21111 pairs of related EDUs in total, and among these, 11006 pairs of EDUs containing at least one verb in each EDUs: 9775 in the training set and 1231 in the test set. The proportions of instances (pairs of related EDUs) containing two, one and no verb are reported in Table 6.1 for each set.

Dataset	**Total count**	**Two verbs**	**One verb**	**No verb**
Train set	18765	0.52	0.40	0.08
Test set	2346	0.52	0.38	0.10
All	21111	0.52	0.40	0.08

Table 6.1: Number of verbs found in the instances annotated in the RSTDT.

Since we are looking to evaluate the potential benefits of *Lecsie*'s lexical features, the following experiment is based on these instances only, as well as negative instances defined as all pairs of EDUs which are not annotated with a relation in each document of the corpus. Since documents can be quite long and long distance attachments are relatively rare, there is a considerable amount of possible pairs with no relation: 516862 in the training set and 80806 in the test set.

With this level of imbalance between positive and negative instances, classification algorithms would be considerably biased towards negative instances and result in negative predictions for most cases. A solution to this would be to artificially balance the training set by randomly selecting as many negative instances as the available positive instances, and training on this set only.

However we did not follow this solution for two reasons. First, it is deemed better to keep the natural distribution of the corpus to ensure the construction of a well adapted model. Secondly, it is important to consider that the information contained in *Lecsie* was extracted exclusively from adjacent units. Thus it might be better adapted for local context decisions. A valuable first step in predicting the discourse structure of a text is to decide whether each unit is attached to a contiguous one, or a more distant one.

Adjacency

For our experiment, we focused on predicting whether adjacent units are attached (that is, whether adjacent units are annotated with a relation). Considering only adjacent units, the training set is composed of a total of 8521 instances, with 63,6% positive (attached) instances, and the test set contains a total of 1051 instances, with 64,4% positive instances: both sets have similar distributions. The counts of positive and negative instances in each set are displayed in the first column of Table 6.2.

Dataset	**Adjacent instances**	**Ratio intra**	**Ratio explicit**	**Ratio $\in$ *Lecsie***
Training set positive	5423	0.87	0.28	0.75
Training set negative	3098	0.44	0.28	0.46
Total training set	8521	0.72	0.28	0.71
Test set positive	677	0.85	0.33	0.84
Test set negative	374	0.44	0.35	0.82
Total test set	1051	0.70	0.34	0.84

Table 6.2: Counts of instances in each set and ratio of instances covered in *Lecsie*, of intrasentential instances and of explicit instances.

Relative position of EDUs

For these adjacent instances, we also considered the proportion of intra- versus intersentential instances. The results for positive and negative instances in each set are shown in column **Ratio intra** of Table 6.2. The proportion of intrasentential instances is considerably higher for positive instances than for negative instances: over 85% of the positive instances in both sets, against 44% of the negative instances in both sets. Indeed, discourse relations appear most often intrasententially, as was also observed in the results of our extraction of explicit relations from Gigaword for the construction of *Lecsie*: 83% of extracted triples in *Lecsie* are intrasentential.

Relation types

During the extraction of verb pairs from instances of the RST corpus, we also applied our method to find out whether an explicit connective was linked in the dependency parsing to either of the two verbs, in which case the instance was considered as an explicit relation. We must note however that this method is quite imprecise, since contrary to *Lecsie*'s construction method, here we process each

EDU separately and cannot determine whether the explicit relation that could be expressed by the connective indeed links the two EDUs forming the instance. The resulting ratios seem inconclusive: positive and negative instances display similar values, as reported in column **Ratio explicit** of Table 6.2.

Coverage in *Lecsie*

Each instance is associated with the pair of verbs extracted from the two EDUs it is composed of. We can then study *Lecsie*'s coverage of these instances, that is the proportion of verb pairs appearing in *Lecsie*. The results for positive and negative instances in our datasets are reported in the column **Ratio** $\in$ ***Lecsie*** of Table 6.2. In the training set, 75% of verb pairs from the positive instances are found in *Lecsie*, meaning they appeared in explicitly related units in our construction corpus. In comparison, 46% of verb pairs from the negative instances appear in *Lecsie*. Such lower coverage indicates that *Lecsie*, as intended, captures proportionnally more pairs appearing in related EDUs in unseen texts. It is probable that the verb pairs from negative instances which do appear in *Lecsie* have lower association score than verb pairs from positive instances, and thus machine learning techniques should allow to make distinctions between attached and non-attached instances. This is studied in the next section.

6.1.3 Classification

Features

For each pair of units, we extracted the association values corresponding to the verb pair with any relation in *Lecsie*, for the three most significant measures: $W_{combined}$, normalized PMI and specificity. In cases where the verb pair is not found in *Lecsie*, all values are set to zero. Thus each pair is represented by a feature vector of 18 values corresponding to the six relations considered in *Lecsie*, for each of the three measures, as exemplified in 6.3. Note that the relation labels do not matter in this task, the values serve the purpose of measuring the strength of association of the verb pair in different dimensions (contexts).

Hate/Love	**Cause**	**Temporal**	**Contrast**	**Continuation**	**Alternation**	**Elaboration**
$W_{combined}$	0.34	0.22	0.57	0.31	1.0	0.0
norm. PMI	0.32	0.22	0.46	0.22	0.77	0.0
specificity	0.04	0.02	0.15	0.25	0.59	0.0

Table 6.3: 18 features for the pair *Hate/Love*.

Datasets and experiment protocol

The RST Discourse TreeBank comes with preassigned training and test sets for classification experiments. The contents of these sets in terms of adjacent instances, as reported in Table 6.2 in the previous section, are the following: 8521 total instances in the training set, with 63.6% positive instances, and 1051 instances in the test set, with 64.4% positive instances. Each instance is represented by a label, 1 if positive (if the units are attached) and 0 if negative, and a set of 18 features as described previously.

For our classification experiments, we use the implementations of different classification algorithms available in the scikit-learn toolkit (Pedregosa et al., 2011): Maximum Entropy, Random Forest and Nearest Neighbors. First, we perform a cross-validation on the training set, to determine which algorithms result in the best predictions. Then, a model is built on the whole training set and applied on the test set. The results are compared to the baselines described below.

Baselines

To evaluate our classification results, we compare them to results obtained with several baselines.

- The first baseline, namely **Baseline Stratified**, assigns random labels to each instance, while preserving the distribution of the labels in the training set.
- The second baseline, **Baseline Always attached** assigns a positive label to each instance: it is equivalent to considering that all adjacent units are attached.

These two baselines do not take any feature into account to assign a label to each instance.

We also implemented two baselines using lexical features, also based on the verb pairs appearing in each instance.

- In the first one, further referred to as **Baseline Two verbs**, each verb is used as a distinct feature, producing two categorical features;
- in the second one, **Baseline Verbpair**, the verbs are combined as one categorical feature.

For each categorical feature, one-hot encoding is used to map each category (distinct verb or pair of verbs) to a numerical value.

The two latter sets of baselines features are used to train the same classifiers as with features extracted from *Lecsie*. We thus produce three models for each type of classifier.

Results

In Table 6.4 we present our results for cross-validation on the training set, and in Table 6.5 we present our results on the test set.

	Baseline Stratified		**Baseline Attached**		**MaxEnt**		**Random Forest**		**Nearest Neighbors**	
	f1	acc	f1	acc	f1	acc	f1	acc	f1	acc
***Lecsie* feats**	0.60	0.49	0.78	0.64	0.82	0.76	0.82	0.75	0.77	0.66
Baseline Verbpair					0.76	0.62	0.19	0.41	0.10	0.39
Baseline Two verbs					0.74	0.61	0.59	0.52	0.53	0.49
Intra only	0.80	0.67	1.00	1.00	0.98	0.97	0.97	0.94	0.93	0.86

Table 6.4: Cross-validation attachment prediction results.

	Baseline Stratified		**Baseline Attached**		**MaxEnt**		**Random Forest**		**Nearest Neighbors**	
	f1	acc	f1	acc	f1	acc	f1	acc	f1	acc
***Lecsie* feats**	0.63	0.53	0.78	0.64	0.81	0.75	0.81	0.75	0.78	0.71
Baseline Verbpair					0.78	0.64	0.16	0.40	0.09	0.38
Baseline Two verbs					0.76	0.65	0.59	0.54	0.56	0.51
Intra only	0.80	0.67	1.00	1.00	0.99	0.98	0.98	0.95	0.93	0.88

Table 6.5: Attachment prediction results on the test set.

The best results are obtained with the Maximum Entropy and Random Forest models, which yield similar results.

Baselines

The baseline of always attaching adjacent EDUs obtains high scores since it is often an accurate assumption. We can see that is in fact always correct when considering only intrasentential instances (as seen in the line **Intra only**). Overall, compared to **Baseline Stratified**, **Baseline Always Attached** yields an improvement of 15 points for F-score, and 11 points for accuracy. The best results using our baseline lexical features are comparable to those obtained with **Baseline Always Attached**. This shows that simply considering verb pairs and representing them via one-hot encoding is not a satisfying approach.

Results with *Lecsie* features

With our features extracted from *Lecsie*, we obtain an improvement of 20 points over the **Baseline stratified** results, both in terms of F-score and accuracy, and a 3-point improvement of the F-score, and almost 10 points for accuracy, compared to the other baselines. From these results, we can conclude that *Lecsie*'s features are good indicators of attachment for pairs of adjacent units.

Using shared arguments

We also experimented with using the information about shared arguments which we collected thanks to the anaphora resolution tool of the Stanford Parser. However, this experiment was inconclusive, yielding no significant difference compared to not using this information. This is most probably due to the sparsity and lack of precision of this type of information. Using a better anaphora resolution method would be necessary in order to exploit predicate-argument mappings, which intuitively should help better identify the pairs of predicates, if the information were correct and precise enough.

Comparison to related results

We are not aware of any comparable experiment in the literature, but a related study is that of Li et al. (2014b), who propose a discourse parser based on exploiting dependency parsing techniques, as we detailed in Section 6.1.1. Among other results they present *unlabeled accuracy* scores, representing the correct identification of heads of EDUs, while our study focuses on predicting whether of not two EDUs are related. Using complex features and state-of-the-art dependency parsing techniques, they obtain an accuracy of 0.74 for their best model. Although it is not possible to make true comparisons, we obtain similar values to theirs, and can hypothesize that using our lexical features within their system might be a valuable addition.

Perspectives

Based on these positive results, we intend to include our features in state-of-the-art models for attachment prediction. Another possible perspective in terms of attachment prediction is to build similar models for more distant units, in order to form a cascade prediction model.

In the next section, we turn to the evaluation of our lexical features in the task of labelling the attachments between discourse units with discourse relations.

6.2 Labelling relations

In this section we turn to the task of labelling discourse relations, and especially implicit relations. Based on two discourse-annotated corpora, Annodis for French and the PDTB for English, we first assess the coverage of our resource and then evaluate its impact by using the information it contains to produce lexical features which we use to construct a model for supervized discourse relation prediction.

6.2.1 Predicting discourse relations in French

Our experiment on discourse relation prediction for French is based on the Annodis corpus (Afantenos et al., 2012). This resource is a corpus of French texts enriched with a manual annotation of discourse structure. It is the first discourse annotated French corpus, although we should note that a French version of the PDTB, based on the 1 million word French TreeBank corpus (Abeillé et al., 2003), is currently in preparation (Danlos et al., 2012).

6.2.1.1 Annodis data

In Annodis, each text is divided in EDUs which are linked by discourse relations, forming complex discourse units which in turn may be linked by discourse relations. As a simplification we only consider EDUs, since the question of what is a main verb of a complex unit is difficult to answer.

The corpus is relatively small, with about 2000 annotated relations between EDUs. For our purpose, evaluating the impact of *Lecsie*'s contents on relation prediction, only annotations between EDUs containing at least one verb each can be considered. In total, only 427 instances were found to contain one verb in each EDU. Indeed, Annodis contains many EDUs formed on short segments which do not contain a verb (incises, detached segments, ...).

Relation types

Unlike the PDTB, it is not specified in the annotation whether the relation is explicit or implicit. Thus, using the same lexicon of connectives as was used for the construction of *Lecsie*, Lexconn (Roze et al., 2012), we relied on our extraction method to detect whether a connective is linked to the identified verb. An instance is considered explicit if a connective is extracted from either of the EDUs it is composed of. This implies the same caveat as with the RST corpus: if an EDU contains a marker of an explicit relation, it does not necessarily indicate that the annotated relation is explicit, as the connective could be used to express a relation with another EDU or even a complex unit. This approximation may result in an overestimation of the number of explicit instances, but ensures that all implicit instances are indeed implicit (assuming a good enough coverage of the marker resource).

6.2.1.2 Coverage in *Lecsie*

Table 6.6 presents results for coverage, for the main relations in the annotated corpus.

Annodis relation	**All instances**				**Implicit instances**			
	Total	Pairs	Triples	Triples best relation	Total	Pairs	Triples	Triples best relation
Narration	73	0.71	0.43	0.34	52	0.52	0.49	0.25
Cause	67	0.81	0.71	0.50	53	0.54	**0.51**	0.40
Contrast	41	0.78	0.75	0.71	15	0.32	0.28	0.32
Elaboration	96	0.68	0.39	0.00	95	0.67	0.68	0.00
Continuation	92	0.62	0.29	0.21	87	0.59	**0.57**	0.20
Background	24	0.74	0.50	0.11	21	0.67	**0.71**	0.11
Other	16	0.63	0.44	0.00	16	0.63	0.63	0.00
Overall	427	**0.69**	0.41	**0.27**	357	**0.57**	**0.53**	**0.18**

Table 6.6: Coverage of all instances in Annodis.

The table includes: the proportion of verb pairs found in Annodis EDUs that appear in *Lecsie*, the proportion of triples from Annodis that appear in *Lecsie* (with the correct relation), and the restriction of these proportions to implicit contexts in Annodis. Except for a few exceptions due to lemmatisation errors, all verbs in Annodis are in *Lecsie* in at least one pair, and we can see that the pairs in *Lecsie* cover most of the pairs appearing in Annodis (almost 70% globally and between 60 and 80% depending on the relation), and a little less of implicit cases (around 57% on average).

We note that a high proportion of the implicit cases contains verb pairs that have been collected in a marked context, even for rarely marked relations like elaboration or continuation – contexts with these relations are the majority in Annodis. Furthermore more than half of these contexts are associated with the right relation in *Lecsie*. Thus the hypothesis of the partial redundancy of connectors appears useful when isolating verbal associations relevant for discourse from a large corpus. We also looked at semantic neighbors of the verbs in *Lecsie* but this did not increase coverage significantly.

6.2.1.3 Relation prediction

A good test of the predictive power of the semantic information we gathered is also to include the association measures as additional features to a predictive model, to improve classically low results on implicit discourse relations. The only available discursive corpus in French, Annodis, is small, and as shown above only about 400 instances have a verb in both related EDUs. We trained and tested a maximum entropy model with and without the association measures as features, on top of features presented in Muller et al. (2012), who trained a relation model on the same corpus. We did a 10-fold cross-validation on the 400 instance subset as evaluation, and did not find a significant difference between the two set-ups (F1 score was in the range .40–.42, similar to the cited paper), which is unsurprising given the size of the subset.

6.2.2 Predicting discourse relations in English

6.2.2.1 PDTB data

As the largest and most widely-used English language discourse annotated corpus, the PDTB is our resource of choice for the evaluation of *Lecsie*'s potential for discourse relation prediction. To perform our study, we applied our extraction method described in 3.2 to the PDTB annotated instances, in order to produce triples similar to those in *Lecsie*, composed of the main relevant verb (if any) from each of the two discourse units in an instance, and the annotated relation.

However, there are some aspects of its constitution that need to be taken into account as they differ from the corpus *Lecsie* was built on (Gigaword) and our extraction method.

Considered instances

The PDTB contains 40600 annotated instances from the Wall Street Journal corpus. These instances link two arguments of variable extents. They can be single clauses or full sentences, similarly to the extraction method used to build *Lecsie*. The PDTB also includes arguments comprising multiple sentences, which makes the extraction of a single verb per argument problematic: in this case we chose to extract the main verb from the last sentence in the first argument, and the first verb from the second argument, as this seems to be the closest result to what would be obtained on non-annotated text and is thus most coherent with *Lecsie*'s contents. Finally, in some exceptional cases, PDTB arguments can be non-clausal elements. For instance these elements can be nominalizations, which cannot be processed with our method since this aspect has not been implemented yet, or anaphoric expressions, which are also problematic since the treatment of these arguments would necessitate an anaphora resolution mechanism to be applied on the whole text or paragraph they are included in.

As a result, when applying our method to extract pairs of verbs from PDTB annotated instances, some instances had to be excluded as their arguments did not contain one verb each. As shown in Table 6.7, 83% of all PDTB instances (33766 instances) were found to contain at least one verb in each argument. The results are similar accross all types of relations which can be used to link the arguments.

Relation type	**Total count**	**Two verbs**	**One verb**	**No verb**
Explicit	18458	0.85	0.14	0.01
Implicit	16061	0.81	0.17	0.02
AltLex	626	0.80	0.19	0.01
EntRel	5202	0.78	0.20	0.02
NoRel	253	0.72	0.24	0.05
All instances	40600	0.83	0.16	0.01

Table 6.7: Number of verbs found in the instances annotated in the PDTB, per relation type.

Relation types

The different types are the following:

1. Explicit, where a discourse connective linking the arguments is identified,
2. Implicit, where no discourse connective is used in the text,
3. Alternative lexicalization (AltLex), where a “non-connective expression” expressing the discourse relation is identified,
4. Entity relation (EntRel), where the second argument provides further description of an entity in the first argument, but no discourse relation is expressed,
5. No relation (NoRel), where neither a discourse relation nor an entity relation is expressed.

Annotations of the first three types always include one or several discourse relations linking the two arguments, while the two others (EntRel and NoRel) never do, according to their definitions. Since the focus of this study is on discourse relations, further on we only consider the first three types. Annotations of these types including two verbs amount to 29513 instances. When mentioned, the two other types, used in 4253 instances with two verbs, are grouped together as "Other".

Relation classes

The PDTB uses a three-level hierarchy of discourse relations: classes, types and subtypes. Depending on the context, annotations do not necessarily contain descriptions for all levels but can be limited to a class specification, or a class and type specification, leaving the subtype unspecified. For more simplicity and in order to show coherent numbers, the quantitative descriptions of the annotations consider only the class level of relations, which is always specified.

For each annotation of the Explicit or Implicit types, at least one connective has to be identified (found in the text for explicit instances, and inserted by the annotator for implicit instances), but there can be two connectives. At least one discourse relation has to be associated with each connective, but a second relation can be given if the meaning is found to be ambiguous. Thus, for Explicit and Implicit types, up to four discourse relations can be identified for each instances, while at least one and maximum two discourse relations appear in AltLex instances.

When several relations are identified in an instance, we produce the corresponding number of triples, associating the verb pair found in the instance with each of the relations. Thus, there are more triples than instances containing discourse relations (that is, instances of types Explicit, Implicit or Altlex): overall, the PDTB contains 29513 instances of such types with two identified verbs, and 30854 triples.

Table 6.8 shows the distribution between all three types of annotations for each relation class, as well as for all instances.

Comparison and Temporal relations are clearly more often explicitly expressed, while Expansion and Contingency are more balanced between implicit and explicit expressions, although slighlty more often implicit. Overall, 54% of the instances are explicit. We should note that Prasad et al. (2008) presented a result of 52% implicit instances: the difference could be due to fact that we exclude instances which do not contain one verb in each argument. Generally speaking, the corpus is quite balanced between implicit and explicit instances, with alternative lexicalizations being very rare.

It is also interesting to consider the reversed distribution of relation classes per relation type. These results are presented in Table 6.9. This demonstrates that

Relation class	Total count	Implicit	Explicit	AltLex
Expansion	12915	**0.56**	0.42	0.01
Contingency	6841	**0.50**	**0.47**	0.03
Comparison	6854	0.30	**0.69**	0.00
Temporal	4244	0.20	**0.78**	0.02
All instances with relation	29513	0.45	**0.54**	0.02

Table 6.8: Implicit, explicit and alternative lexicalization distribution of the instances/triples annotated in the PDTB, per relation class.

explicit triples are mostly annotated with Comparison and Expansion relations, while implicit triples majorly express Expansion relations, and finally Contingency relations are the most common for alternative lexicalizations. Overall, Expansion relations are the most commonly annotated.

Relation type	Totalcount	Temporal	Contingency	Comparison	Expansion
Explicit	16745	0.20	0.19	**0.28**	**0.33**
Implicit	13601	0.06	0.25	0.15	**0.53**
AltLex	508	0.16	**0.45**	0.06	0.33
All triples	30854	0.14	0.22	0.22	**0.42**

Table 6.9: Relation class distribution of the triples annotated in the PDTB, per relation type.

Relative position of arguments

Another aspect to consider is whether the arguments are adjacent or not. In the PDTB, only explicit instances can link arguments which are not adjacent. Since we are mainly interested in implicit instances and these can only be adjacent, and since *Lecsie* was built on adjacent clauses only, adjacency of arguments is not considered in this study.

The PDTB contains both intra- and intersentential instances, however except for explicit relations, annotated intrasentential instances are only between clauses delimited by a semi-colon or colon. The PDTB Annotation Manual mentions that implicit relations can hold between intrasentential clauses which are not delimited in this way, but they are not annotated.

For this reason, except for explicit relations, most annotations are in fact intersentential, which constitutes an important difference from the results contained in *Lecsie* which were obtained from intrasentential instances for 83% of the triples. The distributions between intra- and intersential instances in the PDTB for each relation type and relation class, as well as for all instances, are presented in Table 6.10. As a reminder, the distributions found in *Lecsie* were presented in Table 4.1 on page 112.

In explicit relations from the PDTB, with 61% of intrasentential instances, the distribution is closer to *Lecsie*'s contents. It also differs depending on the relation class: Contingency and Temporal relations are significantly more often intrasentential, as is the case for most relations in *Lecsie*, while Expansion and Comparison relations are more balanced between the two. In *Lecsie*, the most noteworthy difference in distributions is for Continuation and Elaboration relations (which correspond to Expansion relations), with 61 and 76% intrasentential instances respectively, and for Contrast relations (which correspond to Comparison relations in the PDTB) with 80%, while all other relations in *Lecsie* are at least 90% intrasentential. Although less clearly marked in *Lecsie*, similar tendencies are thus observed in both distributions.

6.2.2.2 Coverage in *Lecsie*

As discussed in Section 3.2.1.3, we established a correspondence table with *Lecsie* relations based on the first two levels of the hierarchy used in the PDTB, reproduced in Table 6.11 for easier readability. Note that for annotations where only the relation class Expansion is specified, the relation is mapped to Elaboration in *Lecsie*.

Using this mapping, we studied *Lecsie*'s coverage of PDTB verb pairs and triples for all relation types, as well as for implicit relations only. Indeed, we are interested in seeing if lexical information contained in *Lecsie* can be extrapolated to implicit relations, even if it was extracted from explicit relations only. The coverage results are presented in Table 6.12.

Lecsie's pairs of semantically associated verbs capture a very high number of the related pairs in the PDTB, for all instances as well as for implicitly related arguments only (81%).

When considering triples, that is examining whether the relation that is annotated in the PDTB is also associated with the verb pair in *Lecsie*, we obtain about 50% coverage. However these results are brought down by the fact that Elaboration, which is the most common implicit relation, is rarely explicitly marked, and

Relation type	Relation class	Total count	Intrasentential	Intersentential
Explicit	Expansion	5473	0.45	**0.55**
	Contingency	3197	**0.91**	0.09
	Comparison	4744	0.46	**0.54**
	Temporal	3331	**0.87**	0.13
	All instances	15851	**0.61**	0.39
Implicit	Expansion	7276	0.03	0.97
	Contingency	3414	0.03	0.97
	Comparison	2079	0.03	0.97
	Temporal	832	0.02	0.98
	All instances	13159	0.03	**0.97**
AltLex	Expansion	166	0.01	0.99
	Contingency	230	0.00	1.00
	Comparison	31	0.00	1.00
	Temporal	81	0.00	1.00
	All instances	503	0.00	**1.00**
All types with relation	Expansion	12915	0.21	0.79
	Contingency	6841	0.44	0.56
	Comparison	6854	0.33	0.67
	Temporal	4244	0.69	0.31
	All instances	29513	0.34	**0.66**
Other	All instances	4253	0.01	**0.99**
All types	All instances	33766	0.30	**0.70**

Table 6.10: Intra- and intersentential distribution of the instances annotated in the PDTB, per relation type and relation class.

is thus the least common in *Lecsie*. Still, more than 30% of implicit Elaboration triples are captured. Contrast and Temporal relations are very well captured in *Lecsie*, for all types of relations as well as for implicit relations only (about 60% of implicit triples).

Finally, when considering *Lecsie*'s best guess as to the semantic relation between discourse arguments, that is the highest scored relation (in terms of normalized PMI) among the possible relations associated with the verb pair, we obtain 23% for all triples, and 19% for implicit triples, which is reasonably high. The results are notably better for Temporal relations, at about 30%.

PDTB		*Lecsie*
Level 1	Level 2	
Temporal	Asynchronous Synchronous	Temporal
Contingency	Cause Pragm. cause Condition Pragm. condition	Cause
Comparison	Contrast Pragm. contrast Concession	Contrast
Expansion	Instantiation Restatement	Elaboration
	Alternative	Alternation
	Exception	Contrast
	Conjonction List	Continuation

Table 6.11: Correspondence table between PDTB relations levels 1 and 2 and *Lecsie* relations.

However this leaves an important number of triples where the highest scored relation in *Lecsie* is different from the annotated relation in the PDTB. Thus it is worthwhile to examine the distribution of highest scored *Lecsie* relations for each annotated implicit relation in the PDTB. In order to obtain these results, for each verb pair annotated with an implicit relation in the PDTB, the best relation in *Lecsie* among all relations existing with this verb pair is identified. This way we can obtain, for each PDTB relation, the number of triples containing each *Lecsie* relation as its best relation, and thus compute the distribution of best relation in *Lecsie* for each PDTB relation of implicit type.

Table 6.13 presents these distributions. When looking at all implicit instances which are covered in *Lecsie*, we find that the most common best relation in *Lecsie* is Continuation. For Elaboration relations, which are the most common annotation for implicit instances in the PDTB, but are also the least represented in *Lecsie* triples, Continuation is the most common best relation in *Lecsie*. Alternation relations, which are very uncommon in the PDTB, making the results less significant than for other relations, the most common best relation in *Lecsie* is Contrast. For

PDTB relation	**All instances**				**Implicit instances**			
	Total	Pairs	Triples	Triples best relation	Total	Pairs	Triples	Triples best relation
Continuation	8146	0.80	0.53	0.24	3290	0.81	0.55	0.23
Contrast	6869	0.81	0.62	0.25	2081	0.81	**0.61**	0.22
Cause	6841	0.81	0.54	0.18	3414	0.81	0.53	0.17
Alternation	448	0.83	0.39	0.14	137	0.85	0.40	0.11
Elaboration	4306	0.81	0.31	0.11	3847	0.80	**0.31**	0.11
Temporal	4244	0.84	0.63	0.31	832	0.81	**0.59**	0.28
No relation	4253	0.84	-	-	-	-	-	-
All instances with relation	29513	**0.81**	**0.54**	**0.23**	13159	**0.81**	**0.49**	**0.19**
All instances	33766	0.81	-	-	-	-	-	-

Table 6.12: Coverage of all instances in the PDTB.

the other four relations, we find that the most common best relation in *Lecsie* is identical to the annotated relation in the PDTB. This result is very promising, as it shows that, for these relations at least, our hypothesis that discourse markers are at least partially redundant holds.

Best relation in *Lecsie* **Annotated relation in PDTB**	Total	Elaboration	Cause	Continuation	Contrast	Temporal	Alternation
Elaboration	3090	0.13	0.17	**0.25**	0.19	0.16	0.1
Cause	2760	0.13	**0.21**	0.18	0.2	0.17	0.11
Continuation	2668	0.12	0.15	**0.28**	0.17	0.18	0.1
Contrast	1681	0.12	0.15	0.22	**0.27**	0.14	0.1
Temporal	670	0.08	0.11	0.23	0.12	**0.35**	0.1
Alternation	116	0.09	0.21	0.14	**0.27**	0.16	0.13
All implicit instances	10616	0.12	0.17	**0.23**	0.2	0.18	0.1

Table 6.13: Distribution of highest scored relations in *Lecsie* for the implicit triples in the PDTB covered in *Lecsie*, per PDTB relation.

6.2.2.3 Relation classification

As we have seen in Section 2.4.2, automatic annotations of these implicit instances through machine learning approaches have had limited success in this area, as relatively little manually annotated data is currently available. *Lecsie* provides automatically detectable features in non-annotated data for finding typical contexts (pairs of discourse units) in which various discourse relations occur and for determining potential arguments for discourse relations. Sporleder and Lascarides (2008); Braud and Denis (2013) suppose such contexts display regular lexical associations, in particular with verbs in those discourse units. While a manually compiled list of the possible associations between two verbs and the semantic relations they suggest is infeasible, *Lecsie* provides an automatic method for compiling such a list.

Features

As in the attachment prediction experiment (Section 6.1.3), the feature vector associated with each instance is composed of the 18 values corresponding to the three association values for the six relations.

Datasets and experiment protocol

To create the training and test sets, we followed the recommendations given in the PDTB annotation manual (Prasad et al., 2007), also followed by most relation prediction studies based on this corpus in the literature: the training set is composed of sections 2 to 20, and the test set is composed of sections 21 and 22. We performed a binary classification for each relation class (level 1 in the hierarchy), meaning that a model for each relation is built to predict whether each instance should be assigned this relation (positive label) or any other relation, or none (negative label). For each relation, the training set was down-sampled to contain equal numbers of positive and negative instances. All instances from the test set were used, to maintain a natural distribution. This is also the strategy followed in most experiments in the literature. The resulting counts of instances in each set, as well as the ratio of positive instances in the test set for each relation are displayed in Table 6.14.

Relation	Training instances	Test instances	Positive ratio in test set
Expansion	8924	1408	0.47
Contingency	4926	1408	0.21
Comparison	2960	1408	0.19
Temporal	1038	1408	0.13

Table 6.14: Instance counts in training and test sets, and distribution in test set, for each relation. Each training set contains 50% positive instances.

As in the attachment prediction experiment presented in Section 6.1.3, we perform a cross-validation on the training set for each relation, and then build a model for each relation on the whole training set and apply it on the test set.

Baselines

For comparison, we use the same baselines as in the attachment prediction experiment, except for **Baseline Always attached** which is undefined in this context. They are defined and referred to as previously: **Baseline Stratified**, **Baseline Two verbs** and **Baseline Verbpair**.

Results

The cross-validation results for each relation and each algorithm are presented in Table 6.15.

	Baseline stratified		**Nearest Neighbors**		**Naive Bayes**		**MaxEnt**		**Random Forest**	
	f1	acc	f1	acc	f1	acc	f1	acc	f1	acc
Comparison	0.50	0.50	0.55	0.52	0.48	0.55	0.53	0.55	**0.57**	0.55
Baseline Verbpair			0.53	0.51	0.50	0.51	0.17	0.52	0.36	0.52
Baseline Two verbs			0.54	0.50	0.52	0.54	0.52	0.53	0.52	0.54
Expansion	0.50	0.50	0.55	0.53	0.40	0.54	0.52	0.54	**0.56**	0.54
Baseline Verbpair			0.54	0.51	0.53	0.51	0.21	0.53	0.20	0.53
Baseline Two verbs			0.55	0.52	0.53	0.52	0.52	0.53	0.53	0.54
Contingency	0.50	0.50	0.56	0.52	**0.64**	0.55	0.57	0.58	0.53	0.55
Baseline Verbpair			0.52	0.50	0.12	0.52	0.55	0.54	0.53	0.52
Baseline Two verbs			0.51	0.51	0.32	0.53	0.57	0.55	0.53	0.54
Temporal	0.50	0.50	0.53	0.56	**0.64**	0.61	0.60	0.60	0.58	0.60
Baseline Verbpair			0.51	0.51	0.06	0.51	0.59	0.51	0.51	0.51
Baseline Two verbs			0.51	0.52	0.40	0.55	0.60	0.52	0.55	0.56

Table 6.15: Crossvalidation results.

The results for our lexical baselines should be observed as a conjunction of F-scores and accuracy scores: we obtain either good F-scores but low accuracy scores (much lower than our random baseline accuracy scores), or high accuracy scores but very low F-scores. This is due to either low recall or low precision scores for the positive class, respectively.

These results show that *Lecsie* features alone can achieve an improvement of between 6 and 14 points in F-scores compared to our baseline assigning random predictions with the same distribution as the training set. Depending on the relation, the best algorithms are Random Forest and Naive Bayes, although Maximum Entropy models also produce good predictions.

The final prediction results obtained on the test set are presented in Table 6.16.

With *Lecsie*'s features, we obtain an improvement of 5 points on the F-score for all relations except Expansion which gains 2 points compared to the random

	Baseline stratified		Nearest Neighbors		Naive Bayes		MaxEnt		Random Forest		State of the art
	f1	acc	f1	acc	f1	acc	f1	acc	f1	acc	f1
Comparison	0.28	0.51	0.30	0.49	0.32	0.65	0.32	0.61	**0.33**	0.53	0.31
Baseline Verbpair			0.32	0.22	0.33	0.22	0.07	0.78	0.07	0.79	
Baseline Two verbs			0.32	0.32	0.33	0.31	0.29	0.51	0.31	0.53	
Expansion	0.51	0.51	0.52	0.51	0.36	0.53	0.48	0.52	**0.53**	0.53	0.69
Baseline Verbpair			0.63	0.48	0.63	0.48	0.16	0.54	0.16	0.54	
Baseline Two verbs			0.61	0.49	0.60	0.49	0.47	0.51	0.47	0.52	
Contingency	0.29	0.50	0.31	0.46	0.35	0.38	**0.35**	0.51	0.34	0.54	0.51
Baseline Verbpair			0.35	0.24	0.06	0.77	0.35	0.25	0.35	0.25	
Baseline Two verbs			0.35	0.32	0.23	0.71	0.32	0.47	0.34	0.49	
Temporal	0.24	0.52	0.25	0.53	0.28	0.53	**0.29**	0.60	0.27	0.55	0.24
Baseline Verbpair			0.22	0.14	0.22	0.13	0.01	0.86	0.22	0.14	
Baseline Two verbs			0.23	0.28	0.20	0.66	0.24	0.49	0.24	0.50	

Table 6.16: Classification results on the test set.

baseline. Indeed, the baseline score is quite high for this relation since it is the most frequent in the training set. Overall, our results with respect to simple baselines show that information from *Lecsie* has potential to improve performance on relation classification.

Additionally, we show state-of-the-art F-scores for each relation in the last column of Table 6.16, as reported by Braud and Denis (2015) who compared different representations of word pairs from the literature as well as their own, in addition to other features. Since we only aimed at evaluating the potential of our features alone, and did not implement any other features, our results are not directly comparable as we can expect that the addition of other classical features such as syntactical information to be beneficial. However, we can observe that we manage to obtain better results for Comparison and Temporal instances with our features alone, while our results are lower for Expansion and Contingency relations.

6.3 Conclusion

This chapter has been concerned with the extrinsic evaluation of our resource through its application in the task of discourse parsing. We have focused on two subtasks in which semantic information about the predicates found in discourse clauses would intuitively be particulary valuable.

First, we considered the possibility of using *Lecsie* to improve the prediction of related clauses. We based our study on the RST corpus, and showed that our lexical features alone yield significant improvement in comparison to several baselines. Then, we examined the task of labelling discourse relations linking clauses in the absence of discourse connectives. Again we observed significant improvements. These results open the way to further investigations, where *Lecsie*'s semantic information could be used as additional lexical features in state-of-the-art discourse parsing models.

In the next chapter, we discuss other potential applications of our resource and report some preliminary results.

CHAPTER 7

FURTHER APPLICATIONS OF *Lecsie*

Contents

In the previous chapter, we showed how the semantic relations between verbal predicates provided by *Lecsie* can be exploited in the task of discourse parsing. There are of course other promising applications for such semantic knowledge.

In this chapter, we discuss some other tracks we considered. In Section 7.1, we go back to the idea of finding better dimensions to represent pairs of predicates, and discuss the possibilities of using these dimensions in the task of discourse parsing, instead of the ones we obtained by manually grouping the connectives which helped extract these pairs. In Section 7.2, we show the potential of *Lecsie* to expand FrameNet, a notoriously sparse lexical knowledge base. In Section 7.3, we discuss other applications where *Lecsie* would potentialy be useful.

7.1 Using more optimal representations in discourse parsing

In Chapter 6, we studied the potential of *Lecsie* in the task of discourse parsing. More specifically, we extracted the association scores between a pair of predicates and each of our relations, to construct a vector-based representation of this pair. This representation can then be used as features representing the semantic association between the main predicates of two discourse clauses.

However, as we discussed in Section 5.3, more optimal representations can be obtained by applying dimension reduction methods in the space of connectives. Indeed, rather than using manually-defined sets of connectives representing specific relations, and measuring associations with these relations, it seems more sensible to exploit the knowledge provided by our data to induce the desired dense representation. As we have seen, a manual analysis of these dimensions reveals that the resulting dimensions still hold meaning by human intuition.

We posit that such data-driven representations would be better adapted for applications on large corpora, and propose to evaluate this assertion on the task of discourse parsing. Indeed our previous experiments can be used as a baseline to estimate the benefit of these new dimensions in comparison with manually defined ones.

In preliminary experiments using the six dimensions presented in Section 5.3, we unfortunately did not obtain conclusive results. A possible reason for this is that we did not optimize the number of dimensions which would be useful in our model. Since the initial model also had six dimensions, it might mean that more dimensions would be necessary to give a better description of the semantic information expressed by the association of two predicates. It would also be worth to investigate other clustering methods, aside from the Non Negative Matrix Factorization algorithm which we applied. Indeed, while this method is particularly interesting when we are interested in analyzing the resulting dimensions in comparison with the input dimensions (the markers), other approaches are less transparent but might give better results when applied to other tasks.

7.2 Automatic expansion of FrameNet

As introduced in Section 1.1.3.3, FrameNet is a lexical database which provides a set of “Frames”, descriptions of types of events and their participants. A promising perspective for *Lecsie* is to help enrich that structure, by providing candidate relations between verbal frames, based on associations between their verbs.

7.2.1 FrameNet structure

Frames group lexical units of different syntactic categories that evoke the event type of the frame. Lexical units of the same frame share semantic arguments. For example, the KILLING frame has lexical units such as *assassin*, *assassinate*, *blood-bath*, *fatal*, *murderer*, *kill*, *suicide* that share semantic arguments such as *Killer*, *Instrument*, *Cause*, *Victim*. FrameNet (release 1.5) contains 1195 frames, and 12989 lexical units.

FrameNet frames are not isolated units but are related to each other in various ways. FrameNet 1.5 contains 1774 frame-to-frame relations. It employs eight different types of relations between frames, which fall into three groups (Fillmore and Baker, 2010):

- Generalization relations (**inherits from**, **is perspective on**, **uses**);
- event structure relations (**is subframe of**, **precedes**);
- "systematic" relations (**is causative of**, **is inchoative of**).

Inheritance is the strongest relation between frames. If a frame (the child frame) **inherits from** another frame (the parent frame) then all frame elements (semantic roles) of the parent frame occur as frame elements of the child frame, possibly under a different name. The semantic type of a child frame is a subtype of the parent frame it inherits from, and the same condition holds for all role filler types. For example, the CUTTING frame inherits from the INTENTIONALLY AFFECT frame, with the *Item* element of CUTTING bound to the *Patient* element of INTENTIONALLY AFFECT and all other element names kept equal. And the COMMERCE BUY frame inherits from GETTING, with the roles *Buyer*, *Goods*, and *Seller* bound to *Recipient*, *Theme*, and *Source*, respectively.

The **is subframe of** relation holds between a pair of frames if the first frame represents a subevent of the (complex) event represented by the second frame. The COMMERCIAL TRANSACTION frame has the subframes COMMERCE GOODS TRANSFER and COMMERCE MONEY TRANSFER. Apparently, there is no clear temporal precedence between the events denoted by these two subframes. This is different for the subframes of the GIVING SCENARIO frame: PRE-GIVING, GIVING and POST-GIVING. They represent a temporal succession of states and events, and are linked by the **precedes** relation.

The **is inchoative of** and **is causative of** relations, finally, hold between pairs of frames of which the first denotes respectively the inchoative and the causative of the event denoted by the second frame. For instance, the KILLING frame is causative of the DEATH frame, the CAUSE EXPANSION frame is causative of the EXPANSION frame, and BECOMING DETACHED is inchoative of BEING DETACHED.

7.2.1.1 Applications of FrameNet in NLP tasks

Besides providing an organisation of Frames, these relations can contribute to the discourse interpretation process, both for establishing discourse relations and for resolving indirect anaphora. For instance, Burchardt et al. (2005) aimed to connect predicate-argument structures across sentences in a network of frame-to-frame relations.

Other approaches have aimed to identify reference to eventualities in discourse based on the world knowledge represented in frames and their relations. Bos and Nissim (2008) proposed a method for augmenting a DRT-based formalism with FrameNet roles, aiming both to produce richer semantic representations and to improve semantic role labelling.

Irmer (2011) proposed to use the hierarchical structure of FrameNet to help establishing discourse relations. Indeed, according to Asher and Lascarides (2003), the occurence of an event followed by a state is for instance a strong indicator for the presence of a Background relation between the discourse segments containing the eventualities. This can be used as a cue indicating a Background relation between events belonging to frames which inherit from the Event and State frames. Similar cues can be obtained for other relations, for instance a precedence relation between frames can indicate a Narration relation between events belonging to these frames.

FrameNet has also been used to improve performance in other NLP tasks. For instance, Shen and Lapata (2007) show the improvement that semantic roles obtained from FrameNet can bring on the performance of a Question Answering (QA) system. Similarly, several other studies (Bar-Haim et al., 2005; Garoufi, 2007) indicate that frame semantics plays a central role in Recognizing Textual Entailment (RTE).

7.2.1.2 Network sparseness

The main issue of FrameNet, which limits its usability in such applications, is its lack of coverage, as reviewed respectively by Shen and Lapata (2007) for the QA task and Burchardt et al. (2005) for the RTE task. Additionally to the inherent limitation of any manually constructed resource, frames are based on more complex information than the word senses considered in WordNet for instance, and requires a much more demanding construction methodology (Subirats and Petruck, 2003; Burchardt et al., 2006).

The FrameNet workflow proceeds in a top-down fashion – one frame is created at a time. First, a frame and its lexical units are created and then, representative corpus instances are annotated. As the process is labor-intensive, FrameNet remains incomplete in several respects. Three types of coverage problems can be distinguished:

- Missing lexical units: not all relevant words which can evoke a frame are listed;
- Missing frames: not all prototypical situations are described by an existing frame;
- Missing frame relations: not all relevant relations between frames are identified, and some frames appear without any relation to other frames;

Therefore, there is nowadays a pressing need to design automatic methods for the expansion of FrameNet. Quite a lot of work has been dedicated to address the first aspect described above, aiming to automatically assign new lexical units to correct frames. This was indeed the object of a SemEval-2007 task (Baker et al., 2007). Most work dedicated to this task uses WordNet to extend the FrameNet lexicon (Johansson and Nugues, 2007; Pennacchiotti et al., 2008), although other resources have been used, such as the paraphrase database PPDB (Rastogi and Van Durme, 2014).

Here we focus on the third aspect: the sparseness of the network. Not much work has been dedicated to this issue. An exception is Ovchinnikova et al. (2010), who aimed to discover missing frame relations. For this purpose they proposed a methodology for detecting clusters of frames which are likely to be semantically related, based on frame relatedness measures investigated by Pennacchiotti and Wirth (2009).

We propose to use the knowledge about semantic relations collected in *Lecsie* to infer new connections between frames which are not currently related in FrameNet.

7.2.2 *Lecsie* and FrameNet

In accordance with our focus on semantic relations between predicates, we only consider frames evoked by verbs. There are 605 frames containing at least one verbal lexical unit, and 739 relations involving frames with at least one verbal lexical unit in each.

As an example, we display in Table 7.1 the verbal lexical units included in the SLEEP and WAKING UP frames, which are related by the **precedes** relation.

SLEEP	WAKING UP
catnap	awake
doze	come back around
drowse	come to
hibernate	get up
kip	revive
nap	wake up
sleep	wake
slumber	

Table 7.1: Verbal lexical units in the SLEEP and WAKING UP frames.

In Table 7.2 we display the $W_{combined}$ scores found in *Lecsie* for some pairs of verbs belonging to each of these two frames.

Verb 1	**Verb 2**	contrast	cause	continuation	temporal	alternation	elaboration
Sleep	Wake up	0.37	0.39	0.10	0.74	0.19	0
Sleep	Rouse	0	0.36	0	0.46	0	0
Nap	Wake up	0.42	0.20	0	0.41	0.20	0

Table 7.2: Examples of $W_{combined}$ scores found in *Lecsie* for pairs of verbs belonging to the SLEEP and WAKING UP frames respectively.

As a starting point, we aimed to evaluate the existing relations in FrameNet with the knowledge found in *Lecsie*. For this purpose we did the following preliminary experiment: for each frame relation in FrameNet between two frames f_1 and f_2, we considered all pairs of distinct verbs $(v_1 \in f_1, v_2 \in f_2)$ and computed the ratio of such pairs that are in *Lecsie*, named “Frame Association Score” (FAS):

$$FAS(f_1, f_2) = \frac{\sum_{(v1 \in f_1, v_2 \in f_2)} (v_1, v_2) \in \textit{Lecsie}}{\sum_{(v1 \in f_1, v_2 \in f_2)} v_1 \neq v_2}$$

We also computed a “weighted” version, where the contribution of a pair in *Lecsie* to the score is the value of the measure we found best correlated with human judgments ($W_{combined}$).

Typically the ratio should be high compared to non-related frames: as a control experiment, we selected frame pairs randomly and did the same computation.

Table (7.3) sums up the results when averaged on relations between verbal frames. There are less than 739 because some relations are between frames with only one (identical) verb.

	Global evaluation for *Lecsie*	**Control experiment**
nb of frame relations	735	738
average FAS	0.52	0.42
weighted FAS	0.22	0.16

Table 7.3: FrameNet evaluation

We can see there is a significant (and important) difference between the two sets, both weighted and unweighted. This shows that the semantic information found in *Lecsie* can be used to infer whether two frames are related or not.

7.2.3 Inducing frame relations

Of course when considering random frame pairs in a sparse resource as Framenet, we might have selected satisfactory associations, which is exactly what we want to use *Lecsie* for. A manual exploration of FrameNet quickly shows that some intuitive, obvious relations are missing from the network. For instance, there is no **is causative of** relation between the CAUSE TO WAKE frame and the WAKING UP frame. Table 7.4 displays the verbal lexical units included in these two frames.

CAUSE TO WAKE	WAKING UP
awaken	awake
get up	come back around
rouse	come to
wake up	get up
wake	revive
	wake up
	wake

Table 7.4: Verbal lexical units in the CAUSE TO WAKE and WAKING UP frames.

In Table 7.5 we display the $W_{combined}$ scores found in *Lecsie* for some pairs of verbs belonging to each of these two frames.

Verb 1	**Verb 2**	contrast	cause	continuation	temporal	alternation	elaboration
Get up	Wake up	0.14	0.23	0	0.35	0	0
Awaken	Wake up	0.35	0	0.44	0.17	0	0
Rouse	Wake	0	0.41	0	0.38	0	0.61

Table 7.5: Examples of $W_{combined}$ scores found in *Lecsie* for pairs of verbs belonging to the CAUSE TO WAKE and WAKING UP frames respectively.

Table (7.6) shows the ten best random frame pairs, according to the measure $W_{combined}$; we pointed to the pairs that seem to be genuinely related, although this should be the subject of further investigations.

A promising perspective in this regard would be to use machine learning techniques to build a model of related frames based on *Lecsie*'s vector-based representations of verb pairs, and apply this model to unrelated frames in order to predict whether they should be related or not.

Frame 1	**Frame 2**	**FAS**	
Commerce_buy	Renting_out	0.72	*
Respond_to_proposal	Taking_time	0.56	*
Meet_specifications	Usefulness	0.54	*
Expensiveness	Estimating	0.53	
Taking	Passing	0.48	*
Grasp	Being_named	0.47	
Meet_with_response	Exchange	0.46	?
Agree_or_refuse_to_act	Meet_with_response	0.46	*
Assemble	Have_as_requirement	0.45	
Usefulness	Exporting	0.45	*

Table 7.6: Top scored random relations

Additionally to the lexical relatedness between predicates represented in *Lecsie*, analogies between the lexical relations proposed in *Lecsie* and those targeted in FrameNet can be conceived. For instance, a strong association between a pair of predicates and a temporal relation might indicate that two frames respectively evoked by these two predicates are related by a precedence relation. A causal relation between predicates in *Lecsie* might indicate a causal relation between their corresponding frames.

7.3 Other perspectives

In this chapter and the previous one we have discussed a few extrinsic uses of the semantic knowledge we managed to induce with our method, namely in the field of discourse parsing and for the automatic expansion of FrameNet. It is however quite clear that many other applications could benefit from such information.

7.3.1 Recognizing Textual Entailment

The task of Recognizing Textual Entailment (Dagan et al., 2006) refers to identifying whether the meaning of a piece of text can be inferred from that of another one. As we have discussed previously, FrameNet and other lexical knowledge bases can be used in systems dedicated to this task, which is a motivation to expand these resources. However, *Lecsie* and more generally semantic knowledge about relation between pairs of predicates, can be used directly in this task.

Indeed, systems based on semantic relatedness between verbs have been very successful, such as those of Wang et al. (2009), Majumdar and Bhattacharyya (2010) and Kouylekov et al. (2010), who used in various ways information on specific relations between verbs derived from VerbOcean (Chklovski and Pantel, 2004) to construct features aiming to represent the semantic relatedness of the verbs found in each piece of text.

The lexical features offered by *Lecsie* are very likely to bring out crucial semantic information implicitly conveyed in pairs of sentences, and to help RTE systems in determining inference relations between sentences.

7.3.2 Temporal relations between events

Another application in which *Lecsie* could be beneficial is that of understanding the temporal structure of a text. This application was the object of a SemEval task, namely TempEval (Verhagen et al. (2007), and follow-ups). It is divided in three subtasks:

- Identification of events;
- Identification of time expressions;
- Identification of temporal relations.

This task is based on data from the TimeBank corpus (Pustejovsky et al., 2003), which contains new artices tagged for events, time expressions, and relations between the events and times. There are six main relations and their inverses in

Timebank: **before**, **ibefore**, **includes**, **begins**, **ends** and **simultaneous**. TempEval used a simplified set temporal relation labels: **before**, **overlap** and **after**. Example 7.1 illustrates the type of events to be related.

(7.1) Iraq's Saddam Hussein promised to *withdraw* from Iranian territory and *release* captured soldiers.

Most systems dedicated to this task are based on machine learning approaches and treat the assignment of temporal relations as a classification task. A majority of the systems rely mostly on morphosyntactic features, although lexical semantic knowledge has been used in a few proposals, mainly for event processing. For instance, WordNet has been used to derive features from *event* and *state* hyponyms (Saurí et al., 2005; Bethard and Martin, 2006; Grover et al., 2010).

An interesting perspective of research would be to evaluate the benefits of using lexical semantic features derived from *Lecsie* as additional features in current systems. Although our resource does not distinguish between different temporal relations or the ordering of events, and thus could not offer sufficient information on its own, we posit that such semantic knowledge about events would be useful and relevant to this task.

For instance, we display in Table 7.7 the w_combined scores found in *Lecsie* for the pair of predicates *release* and *withdraw* extracted from Example 7.1, which show a clear preference for temporal relations between the two predicates.

Verb 1	Verb 2	contrast	cause	continuation	temporal	alternation	elaboration
Release	Withdraw	0.14	0.08	0.20	0.43	0.07	0

Table 7.7: w_combined scores found in *Lecsie* for the pair of predicates *release* and *withdraw*.

The tasks we have described in this chapter are just some examples of the usefulness of our resource. We aim to pursue the investigation of such exciting avenues of research in future work.

CHAPTER 8

CONCLUSION

Contents

The work presented in this thesis arose from the observation that natural language understanding often relies on common-sense reasoning, for which knowledge about semantic relations, especially between verbal predicates, may be required.

In order to be potentially useful in applications, semantic processing models need to have the capability to support these common-sense inferences, which may be valid only in the typical case and can be overwritten if more specific contradicting information is added. As we discussed in Chapter 1, standard predicate-logic deduction just adds information, extending the knowledge base in a monotonic way, and has no mechanism for knowledge revision. Alternative logic frameworks, in particular default logic (Asher and Lascarides, 2003), have been proposed to support non-monotonic inference, especially in the analysis of discourse structure, which we discussed in Chapter 2. These frameworks however need to have access to such common-sense knowledge.

The advent of large text corpora has open the door to the collection of implicit semantic information with wide coverage, which is not feasible through manual efforts. Distributional approaches offer the possibility of obtaining such information in a fully unsupervised way, requiring no annotation or other preparatory manual work. However, standard distributional models offer only a rough approximation to lexical meaning, and a rather vague notion of semantic relatedness which includes all kinds of lexical relations, such as synonymy, topical relatedness, or antonymy (Budanitsky and Hirst, 2006). These methods therefore do not seem to have the ability to offer the type of semantic information that is crucial to common-sense inference.

In this thesis we aimed to address the challenge of using a distibutional method to automatically extract the necessary semantic information for common-sense inference, by targeting a specific set of semantic relations. In the next section we recall our approach and summarize our main contributions. In Section 8.2 we discuss the next steps to improve our approach, and the promising avenues of research that lie ahead.

8.1 Contributions

In this thesis we have proposed a method to collect typical associations between pairs of predicates from large corpora, with the potential to model defeasible common-sense knowledge.

Our main contributions are the following:

- We designed a method to extract event predicates linked by semantic relations from a large corpus;
- We designed measures representing the significance of association between a pair of predicates and a relation;
- We built a resource based on this method, consisting of triples: pairs of event predicates associated with a relation; each triple is assigned significance scores based on our measures;
- We used the resource to construct vector-based representations of pairs of predicates, and showed the potential of using these representations as features to build models for external applications, and specifically discourse parsing.

Our extraction method, presented in Chapter 3, uses syntactic patterns linking predicates and discourse connectives to extract these related predicates from corpus contexts. Indeed, we assume that there is at least some redundancy in explicitly expressed discourse relations, and that these relations are not only lexicalized by the connective, but also by the combination of predicates used in each related clause.

We analyzed the possible syntactic dependencies between a connective and these predicates, which vary depending on the syntactic category of the connective, among other factors. We also considered intersentential instances, where we assume that the first clause is located in the sentence immediately before the one enclosing the connective.

Using only unambiguous discourse connectives allowed us to make a direct correspondence between each connective and the discourse relation it conveys. We then grouped the collected information into triples composed of a pair of predicate and a relation, each associated to its count of occurrences in the source corpus. These triples form the basis of our resource, *Lecsie*.

In Chapter 4, we explored the resulting data in terms of the semantic information which it encloses, and the regularities which can be observed about the usage of semantic relations, in particular in the presence of negations and different coreference patterns.

In order to represent the typicality of the associations we extracted, we designed several association measures, which we presented in Chapter 5. We then performed intrinsic evaluations based on manual annotations, which allowed us to find the most significant measures. We also investigated a different way of representing the semantic relations between two predicates through automatic induction of optimal sets of connectives.

In Chapter 6, we aimed to perform an extrinsic evaluation of our resource by using it in the application of discourse parsing. Current approaches to discourse parsing are mostly based on supervized models, which extract features from each instance and build models aiming to make predictions, using manually annotated instances. One of the main difficulties consists in designing the most appropriate features to represent the data. Features representing lexical semantic information are crucial to such models, and many attempts have been made to find appropriate representations of pairs of words found in related clauses, for instance with word classes obtained from external resources (Lapata and Lascarides, 2004; Pitler et al., 2009), cluster representations (Rutherford and Xue, 2014), or word embeddings (Braud and Denis, 2015). These representations have the drawback of not being specifically designed for the targeted task. Since *Lecsie* specifically contains association scores between pairs of predicates and relations, it is clearly well-adapted to derive appropriate lexical semantic features. We therefore evaluated its potential in the subtasks of predicting which clauses are related, and of labelling the relation between clauses. Using only features from *Lecsie*, we obtained significant improvements for both tasks in comparison to several baselines, including ones using other representations of the pairs of predicates.

8.2 Perspectives

8.2.1 Refining our extraction method for *Lecsie*

Although we have demonstrated the benefits of the semantic knowedge we extract, our method could clearly be refined, both in terms of the scope of the predicates we consider and the precision of our extraction.

For instance, the current version of our extraction method considers only verbs and predicative adjectives as possible event predicates. A more complete method should also account for noun predicates, and in particular deverbal nouns, such as *building*, corresponding to the verb *build*. However this class of nouns is known for a specific lexical ambiguity phenomenon: they may denote the event or the result of the action of the corresponding verbs. This is illustrated in Examples 8.1 and 8.2.

(8.1) The building of the bridge lasted three years.

(8.2) This building is huge.

Therefore, taking such nouns into account would require a disambiguation analysis which might be quite complex. Despite this difficulty, it remains a promising track for the improvement and densification of *Lecsie*. Other non-deverbal event nouns such as *party* or *conflict*, which refer to an event but cannot be identified by their morphology, would also be interesting to consider.

Another aspect in which our method could be improved concerns the extraction of predicates occuring in different sentences. As reported by Prasad et al. (2007), 77% of explicit intersentential instances relate clauses in immediately adjacent sentences. This justifies our simplifying assumption that this is always the case. A deeper analysis might however allow us to identify the first clause with better precision.

We also consider that in such cases, where we cannot make use of dependency links to identify the first predicate, it is always the main verb of the previous sentence. An exception to this is in our method for French, and specifically for the *narration* relation, where we consider that it is the last predicate of the previous sentence, instead of the main one. A manual study of the different cases existing for each relation might help determine the best scheme to use for each relation, although our hypothesis seems to be justified in most cases.

As we mentioned in Section 3.2, our method also extracts information on tense and modality, but this information is stored in a secondary knowledge base and unused in the current version. Indeed, it would induce considerable sparsity issues if it were to be included in the main resource. This knowledge could however prove to be useful in specific tasks, such as temporal ordering prediction, and it would be interesting to find a way to link the information found in the secondary resource to the main one, for example by analyzing regularities in terms of tense or modality for each relation.

Another very important refinement concerns coreference resolution, which we use to derive coreference patterns. As we have discussed in Section 6.1, our preliminary attempts to use the information based on coreference patterns have not yielded significant results in the task of predicting attachment between discourse clauses. This result is counter-intuitive, since we expect that such information would be very relevant as a refinement of the associations between predicates. We can suppose that the lack of improvement is due to the imprecision and unreliability of current coreference resolution systems. This is however an active area of research, and we hope that further improvements of these systems will allow us to obtain better results in our applications.

8.2.2 Applications of *Lecsie*

In Chapter 6 we presented our study of the application of *Lecsie* in the task of discourse parsing, where we obtained very positive results. As we discussed in that chapter, the next step in this direction would be to implement our features into state-of-the-art discourse parsing models. Once we obtain more reliable coreference patterns, we also hope to derive valuable features from this information. In Section 7.1 we described some preliminary experiments using potentially better adapted representations derived from *Lecsie* by applying dimension reduction methods in the space of connectives. Our results were unfortunately not conclusive, but this study invites for more investigation as to the best dimensions, and number of dimensions, to use for our vector-based representation in machine learning approaches such as those used in discourse parsing.

In Chapter 7, we have discussed some promising perspectives for using *Lecsie* in other NLP tasks, such as the expansion of FrameNet, for which we presented some preliminary encouraging results, the RTE task or the ordering of temporal events.

The work presented in this thesis leaves much room for further investigations, and we are convinced that the type of knowledge we collected has much potential, in addition to what we managed to demonstrate.

Bibliography

Anne Abeillé, Lionel Clément, and François Toussenel. Building a treebank for french. In *Treebanks*, pages 165–187. Springer, 2003. (Cited pages 91 and 171)

Stergos Afantenos, Nicholas Asher, Farah Benamara, Myriam Bras, Cecile Fabre, Mai Ho-Dac, Anne Le Draoulec, Philippe Muller, Marie-Paul Pery-Woodley, Laurent Prevot, Josette Rebeyrolles, Ludovic Tanguy, Marianne Vergez-Couret, and Laure Vieu. An empirical resource for discovering cognitive principles of discourse organisation: the ANNODIS corpus. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may 2012. European Language Resources Association (ELRA). ISBN 978-2-9517408-7-7. (Cited pages 56, 62, 64, 113 and 171)

Laura Alonso, Irene Castellón, Karina Gibert, and Lluís Padró. An empirical approach to discourse markers by clustering. In *Topics in Artificial Intelligence*, pages 173–183. Springer, 2002. (Cited page 154)

Nicholas Asher. *Reference to abstract objects in discourse*, volume 50. Springer Science & Business Media, 1993. (Cited pages 43 and 159)

Nicholas Asher. *Lexical meaning in context: A web of words*. Cambridge University Press, 2011. (Cited page 16)

Nicholas Asher and Alex Lascarides. *Logics of Conversation*. Studies in Natural Language Processing. Cambridge University Press, Cambridge, UK, 2003. (Cited pages 36, 42, 48, 91, 188 and 195)

Nicholas Asher and James Pustejovsky. The metaphysics of words in context. *Journal of Logic, Language and Information*, 2000. (Cited page 42)

Nicholas Asher and Laure Vieu. Subordinating and coordinating discourse relations. *Lingua*, 115(4):591–610, 2005. (Cited pages 52 and 55)

Fatemeh Torabi Asr and Vera Demberg. Uniform information density at the level of discourse relations: Negation markers and discourse connective omission. *IWCS 2015*, page 118, 2015. (Cited pages 86 and 122)

Collin Baker, Michael Ellsworth, and Katrin Erk. Semeval'07 task 19: frame semantic structure extraction. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 99–104. Association for Computational Linguistics, 2007. (Cited page 189)

Collin F Baker and Josef Ruppenhofer. Framenet's frames vs. levin's verb classes. In *Proceedings of the 28th annual meeting of the Berkeley Linguistics Society*, pages 27–38, 2002. (Cited page 19)

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The berkeley framenet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*, pages 86–90, Montreal, Quebec, Canada, August 1998. Association for Computational Linguistics. doi: 10.3115/980845.980860. URL http://www.aclweb.org/anthology/P98-1013. (Cited pages 23, 24 and 227)

Jason Baldridge, Nicholas Asher, and Julie Hunter. Annotation for and robust parsing of discourse structure on unrestricted texts. *Zeitschrift für Sprachwissenschaft*, 26(2):213–239, 2007. (Cited pages 62 and 163)

Roy Bar-Haim, Idan Szpektor, and Oren Glickman. Definition and analysis of intermediate entailment levels. In *Proceedings of the ACL Workshop on Empirical Modeling of Semantic Equivalence and Entailment*, pages 55–60. Association for Computational Linguistics, 2005. (Cited page 188)

Maja Bärenfänger, Daniela Goecke, Mirco Hilbert, Harald Lüngen, and Maik Stührenberg. Anaphora as an indicator of elaboration: A corpus study. *JLCL*, 23(2):49–73, 2008. (Cited page 88)

Marco Baroni and Alessandro Lenci. How we blessed distributional semantic evaluation. In *Proceedings of the GEMS 2011 Workshop on GEometrical Models of Natural Language Semantics*, pages 1–10. Association for Computational Linguistics, 2011. (Cited page 31)

Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. The wacky wide web: a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation*, 43(3):209–226, 2009. (Cited pages 90 and 111)

Regina Barzilay and Michael Elhadad. Using lexical chains for text summarization. *Advances in automatic text summarization*, pages 111–121, 1999. (Cited page 26)

Regina Barzilay and Noemie Elhadad. Inferring strategies for sentence ordering in multidocument news summarization. *Journal of Artificial Intelligence Research*, pages 35–55, 2002. (Cited page 42)

Regina Barzilay and Mirella Lapata. Modeling local coherence: An entity-based approach. *Computational Linguistics*, 34(1):1–34, 2008. (Cited page 42)

Matthew Berland and Eugene Charniak. Finding parts in very large corpora. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 57–64. Association for Computational Linguistics, 1999. (Cited page 32)

Steven Bethard and James H Martin. Identification of event mentions and their semantic class. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 146–154. Association for Computational Linguistics, 2006. (Cited page 194)

Or Biran and Kathleen McKeown. Aggregated word pair features for implicit discourse relation disambiguation. In *Proceedings of the Conference*, page 69, 2013. (Cited pages 74 and 79)

Sasha Blair-Goldensohn, Kathleen McKeown, and Owen Rambow. Building and refining rhetorical-semantic relation models. 2007. (Cited pages 70 and 72)

Johan Bos. Towards wide-coverage semantic interpretation. In *Proceedings of Sixth International Workshop on Computational Semantics IWCS*, volume 6, pages 42–53, 2005. (Cited page 25)

Johan Bos. Wide-coverage semantic analysis with boxer. In *Proceedings of the 2008 Conference on Semantics in Text Processing*, pages 277–286. Association for Computational Linguistics, 2008a. (Cited page 25)

Johan Bos. Introduction to the shared task on comparing semantic representations. In *Proceedings of the 2008 Conference on Semantics in Text Processing*, pages 257–261. Association for Computational Linguistics, 2008b. (Cited page 25)

Johan Bos and Katja Markert. When logical inference helps determining textual entailment (and when it doesn't). In *Proceedings of the Second PASCAL RTE Challenge*, 2006. (Cited page 26)

Johan Bos and Malvina Nissim. Combining discourse representation theory with framenet. *Frames, Corpora, and Knowledge Representation*, pages 169–183, 2008. (Cited page 188)

Gerlof Bouma. Normalized (pointwise) mutual information in collocation extraction. *Proceedings of GSCL*, pages 31–40, 2009. (Cited page 139)

Chloé Braud and Pascal Denis. Identification automatique des relations discursives "implicites" à partir de données annotées et de corpus bruts. In *TALN - 20ème conférence du Traitement Automatique du Langage Naturel 2013*, volume 1, pages 104–117, Sables d'Olonne, France, June 2013. URL `http://hal.inria.fr/hal-00830983`. (Cited page 181)

Chloé Braud and Pascal Denis. Comparing word representations for implicit discourse relation classification. In *Empirical Methods in Natural Language Processing (EMNLP 2015)*, 2015. (Cited pages 74, 79, 183 and 197)

Peter F Brown, Peter V Desouza, Robert L Mercer, Vincent J Della Pietra, and Jenifer C Lai. Class-based n-gram models of natural language. *Computational linguistics*, 18(4):467–479, 1992. (Cited page 74)

Alexander Budanitsky and Graeme Hirst. Evaluating wordnet-based measures of lexical semantic relatedness. *Computational Linguistics*, 32(1):13–47, 2006. (Cited pages 28 and 196)

A. Burchardt, A. Frank, and M. Pinkal. Building text meaning representations from contextually related frames – a case study. In *Proceedings of IWCS-6*, 2005. (Cited page 188)

Aljoscha Burchardt, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Padó, and Manfred Pinkal. The salsa corpus: a german corpus resource for lexical semantics. In *Proceedings of LREC*, volume 6, 2006. (Cited page 188)

Marie Candito, Benoît Crabbé, and Pascal Denis. Statistical french dependency parsing: Treebank conversion and first results. In *LREC*, 2010. (Cited page 90)

Sharon A Caraballo. Automatic construction of a hypernym-labeled noun hierarchy from text. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 120–126. Association for Computational Linguistics, 1999. (Cited pages 33 and 34)

L. Carlson, D. Marcu, and M.E. Okurowski. Rst discourse treebank ldc2002t07. Web download, 2007. (Cited pages 51, 56 and 163)

Lynn Carlson and Daniel Marcu. Discourse tagging reference manual. *ISI Technical Report ISI-TR-545*, 54, 2001. (Cited pages 56 and 87)

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. *Building a discourse-tagged corpus in the framework of rhetorical structure theory*. Springer, 2003. (Cited page 64)

Rudolf Carnap. Meaning and necessity; a study in semantics and modal logic. 1947. (Cited page 12)

Nathanael Chambers and Dan Jurafsky. Unsupervised Learning of Narrative Event Chains. In *Proceedings of ACL-08: HLT*, pages 789–797, Columbus, Ohio, June 2008. Association for Computational Linguistics, Morristown, NJ, USA. URL `http://www.aclweb.org/anthology/P/P08/P08-1090`. (Cited pages 36, 37 and 78)

Gennaro Chierchia. A Semantics for Unaccusatives and its Syntactic Consequences', MS., Cornell University.(19980).'Reference to Kinds Across Languages'. *Natural Language Semantics*, 6:339–405, 1989. (Cited page 46)

Timothy Chklovski and Patrick Pantel. Verbocean: Mining the web for fine-grained semantic verb relations. In Dekang Lin and Dekai Wu, editors, *Proceedings of EMNLP 2004*, pages 33–40, Barcelona, Spain, July 2004. Association for Computational Linguistics. (Cited pages 24, 34, 35, 38, 78, 82 and 193)

Kenneth W Church and William A Gale. A comparison of the enhanced good-turing and deleted estimation methods for estimating probabilities of english bigrams. *Computer Speech & Language*, 5(1):19–54, 1991. (Cited page 139)

Philipp Cimiano, Aleksander Pivk, Lars Schmidt-Thieme, and Steffen Staab. Learning taxonomic relations from heterogeneous sources of evidence. *Ontology Learning from Text: Methods, evaluation and applications*, 2005. (Cited page 33)

Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46, 1960. (Cited page 146)

Allan M Collins and M Ross Quillian. Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior*, 8(2):240–247, 1969. (Cited page 16)

Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. Natural language processing (almost) from scratch. *The Journal of Machine Learning Research*, 12:2493–2537, 2011. (Cited page 75)

Terry Copeck, Alistair Kennedy, Martin Scaiano, Diana Inkpen, and Stan Szpakowicz. Summarizing with roget's and with framenet. In *The Second Text Analysis Conference*, 2009. (Cited page 26)

Simon H Corston-Oliver. Identifying the linguistic correlates of rhetorical relations. In *Proceedings of the ACL Workshop on Discourse Relations and Discourse Markers*, pages 8–14. Citeseer, 1998. (Cited page 88)

Rosalind A Crawley, Rosemary J Stevenson, and David Kleinman. The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research*, 19(4):245–264, 1990. (Cited page 129)

D Alan Cruse. *Lexical semantics*. Cambridge University Press, 1986. (Cited page 17)

Ido Dagan, Oren Glickman, and Bernardo Magnini. The pascal recognising textual entailment challenge. In *Machine learning challenges. evaluating predictive uncertainty, visual object classification, and recognising tectual entailment*, pages 177–190. Springer, 2006. (Cited pages 26 and 193)

Laurence Danlos. "discourse verbs" and discourse periphrastic links. In *Proceedings of KONVENS 2006 (Konferenz zur Verarbeitung natürlicher Sprache) Universität Konstanz*, page 160. (Cited page 46)

Laurence Danlos. Event coreference in causal discourses. *The Language of Word Meaning*, page 216, 2001. (Cited page 46)

Laurence Danlos. D-stag: a formalism for discourse analysis based on sdrt and using synchronous tag. In *Formal Grammar*, pages 64–84. Springer, 2011. (Cited page 96)

Laurence Danlos and Owen Rambow. Discourse relations and propositional attitudes. In *CID 2011-Fourth International workshop on Constraints in Discourse*, 2011. (Cited page 85)

Laurence Danlos, Diégo Antolinos-Basso, Chloé Braud, and Charlotte Roze. Vers le fdtb: French discourse tree bank. In *TALN 2012: 19ème conférence sur le Traitement Automatique des Langues Naturelles*, volume 2, pages 471–478. ATALA/AFCP, 2012. (Cited page 171)

P. Denis and B. Sagot. Coupling an annotated corpus and a lexicon for state-of-the-art pos tagging. *Language Resources and Evaluation*, (46):721–736, 2012. (Cited page 91)

Quang Do, Yee Seng Chan, and Dan Roth. Minimally supervised event causality identification. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 294–303, Edinburgh, Scotland, UK., July 2011. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/D11-1027`. (Cited pages 37, 137, 140, 144 and 148)

Bonnie J Dorr and Doug Jones. Role of word sense disambiguation in lexical acquisition: Predicting semantics from syntactic cues. In *Proceedings of the 16th conference on Computational linguistics-Volume 1*, pages 322–327. Association for Computational Linguistics, 1996. (Cited page 19)

DR Dowty. *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's Ptq*, volume 7. Springer Science & Business Media, 1979. (Cited pages 14 and 46)

Eladio Duque. Signaling causal coherence relations. *Discourse Studies*, 16(1): 25–46, 2014. (Cited pages 46, 47 and 88)

Stefan Evert. *The statistics of word cooccurrences*. PhD thesis, Stuttgart University, 2005. (Cited page 138)

Stefan Evert and Brigitte Krenn. Using small random samples for the manual evaluation of statistical association measures. *Computer Speech & Language*, 19(4):450–466, 2005. (Cited page 143)

Robert M Fano and David Hawkins. Transmission of information: A statistical theory of communications. *American Journal of Physics*, 29(11):793–794, 1961. (Cited page 138)

C. Fellbaum. *Wordnet, an Electronic Lexical Database for English*. Cambridge: MIT Press, 1998. (Cited pages 3, 19, 21, 34, 71 and 227)

Vanessa Wei Feng and Graeme Hirst. Text-level discourse parsing with rich linguistic features. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 60–68, Jeju Island, Korea, July 2012. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/P12-1007`. (Cited pages 65, 161 and 163)

Charles Fillmore. Frame semantics. *Linguistics in the morning calm*, pages 111–137, 1982. (Cited pages 18 and 23)

Charles J Fillmore. *Pragmatics and the description of discourse*. (Cited page 43)

Charles J Fillmore and Collin Baker. A frames approach to semantic analysis. *The Oxford handbook of linguistic analysis*, pages 313–339, 2010. (Cited page 187)

John R Firth. A synopsis of linguistic theory, 1930-1955. 1957. (Cited pages 3 and 27)

Seeger Fisher and Brian Roark. The utility of parse-derived features for automatic discourse segmentation. In *ANNUAL MEETING-ASSOCIATION FOR COMPUTATIONAL LINGUISTICS*, volume 45, page 488, 2007. (Cited page 66)

Jerry Fodor and Ernest Lepore. The red herring and the pet fish: why concepts still can't be prototypes. *Cognition*, 58(2):253–270, 1996. (Cited page 13)

Katherine Forbes, Eleni Miltsakaki, Rashmi Prasad, Anoop Sarkar, Aravind Joshi, and Bonnie Webber. D-ltag system: Discourse parsing with a lexicalized tree-adjoining grammar. *Journal of Logic, Language and Information*, 12(3):261–279, 2003. (Cited pages 48 and 49)

Robert Gaizauskas and Yorick Wilks. Information extraction: Beyond document retrieval. *Journal of documentation*, 54(1):70–105, 1998. (Cited page 42)

Konstantina Garoufi. *Towards a better understanding of applied textual entailment*. PhD thesis, Master Thesis, 2007. (Cited page 188)

Maayan Geffet and Ido Dagan. The distributional inclusion hypotheses and lexical entailment. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 107–114. Association for Computational Linguistics, 2005. (Cited pages 30 and 31)

Roxana Girju, Adriana Badulescu, and Dan Moldovan. Automatic discovery of part-whole relations. *Computational Linguistics*, 32(1):83–135, 2006. (Cited page 33)

Joseph Evans Grimes. *The thread of discourse*, volume 207. Walter de Gruyter, 1975. (Cited page 43)

Barbara J Grosz and Candace L Sidner. Attention, intentions, and the structure of discourse. *Computational linguistics*, 12(3):175–204, 1986. (Cited pages 48, 49 and 54)

Barbara J Grosz, Scott Weinstein, and Aravind K Joshi. Centering: A framework for modeling the local coherence of discourse. *Computational linguistics*, 21(2): 203–225, 1995. (Cited page 36)

Claire Grover, Richard Tobin, Beatrice Alex, and Kate Byrne. Edinburgh-ltg: Tempeval-2 system description. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 333–336. Association for Computational Linguistics, 2010. (Cited page 194)

MAK Halliday. &r. hasan. cohesion in english, 1976. (Cited pages 43, 47 and 153)

Zellig Sabbettai Harris. Mathematical structures of language. 1968. (Cited page 27)

Chikara Hashimoto, Kentaro Torisawa, Kow Kuroda, Stijn De Saeger, Masaki Murata, and Jun'ichi Kazama. Large-scale verb entailment acquisition from the Web. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 1172–1181, Singapore, August 2009. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/D/D09/D09-1122`. (Cited page 161)

Marti A Hearst. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th conference on Computational linguistics-Volume 2*, pages 539–545. Association for Computational Linguistics, 1992. (Cited pages 31, 32, 33 and 34)

Hugo Hernault, Helmut Prendinger, David A. duVerle, and Mitsuru Ishizuka. HILDA: A Discourse Parser Using Support Vector Machine Classification. *Dialogue and Discourse*, 1(3):1–33, 2010. (Cited pages 65 and 163)

Ryuichiro Higashinaka and Hideki Isozaki. Automatically acquiring causal expression patterns from relation-annotated corpora to improve question answering for why-questions. *ACM Transactions on Asian Language Information Processing (TALIP)*, 7(2):6, 2008. (Cited page 42)

Donald Hindle. Noun classification from predicate-argument structures. In *Proceedings of the 28th annual meeting on Association for Computational Linguistics*, pages 268–275. Association for Computational Linguistics, 1990. (Cited page 29)

Julia Hirschberg and Diane Litman. Empirical studies on the disambiguation of cue phrases. *Computational linguistics*, 19(3):501–530, 1993. (Cited page 45)

Jerry R Hobbs. Coherence and coreference. *Cognitive science*, 3(1):67–90, 1979. (Cited pages 43 and 52)

Jerry R Hobbs. *On the coherence and structure of discourse*. CSLI, 1985. (Cited page 36)

Jerry R Hobbs. *Literature and cognition*. Number 21. Center for the Study of Language (CSLI), 1990. (Cited page 48)

Julie Hunter, Nicholas Asher, Brian Reese, and Pascal Denis. Evidentiality and intensionality: Two uses of reportative constructions in discourse. In *Proceedings of the Workshop on Constraints in Discourse*, pages 99–106. Citeseer, 2006. (Cited page 85)

Ben Hutchinson. Acquiring the meaning of discourse markers. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, page 684. Association for Computational Linguistics, 2004. (Cited page 154)

Sabine Schulte Im Walde and Maximilian Köper. Pattern-based distinction of paradigmatic relations for german nouns, verbs, adjectives. In *Language Processing and Knowledge in the Web*, pages 184–198. Springer, 2013. (Cited page 39)

Takashi Inui, Kentaro Inui, and Yuji Matsumoto. Acquiring causal knowledge from text using the connective marker tame. *ACM Transactions on Asian Language Information Processing (TALIP)*, 4(4):435–474, 2005. (Cited pages 35 and 78)

Matthias Irmer. *Bridging inferences: Constraining and resolving underspecification in discourse interpretation*, volume 11. Walter de Gruyter, 2011. (Cited page 188)

Ray Jackendoff. Semantic and cognition. 1983. (Cited pages 2 and 14)

Richard Johansson and Pierre Nugues. Using wordnet to extend framenet coverage. In *Building Frame Semantics Resources for Scandinavian and Baltic Languages*, pages 27–30. Department of Computer Science, Lund University, 2007. (Cited page 189)

Karen Sparck Jones and Julia R Galliers. *Evaluating natural language processing systems: An analysis and review*, volume 1083. Springer Science & Business Media, 1995. (Cited page 7)

Shafiq Joty, Giuseppe Carenini, and Raymond T Ng. Codra: A novel discriminative framework for rhetorical analysis. *Computational Linguistics, page*, 2015. (Cited page 65)

Daniel Jurafsky and James H Martin. Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition, 2009. (Cited page 42)

Hans Kamp and Barbara Partee. Prototype theory and compositionality. *Cognition*, 57(2):129–191, 1995. (Cited page 13)

Hans Kamp and Uwe Reyle. *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and discourse representation theory*. Number 42. Springer Science & Business Media, 1993. (Cited page 49)

Jerrold J Katz and Jerry A Fodor. The structure of a semantic theory. *Language*, 39(2):170–210, 1963. (Cited pages 2, 13, 14 and 227)

Andrew Kehler, Laura Kertz, Hannah Rohde, and Jeffrey L Elman. Coherence and coreference revisited. *Journal of Semantics*, 25(1):1–44, 2008. (Cited pages 46, 128 and 129)

Judith Klavans and Min-Yen Kan. Role of verbs in document analysis. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, pages 680–686. Association for Computational Linguistics, 1998. (Cited page 26)

Alistair Knott, Jon Oberlander, Mick O'Donnell, and Chris Mellish. Beyond elaboration: The interaction of relations and focus in coherent text. *Text representation: linguistic and psycholinguistic aspects*, pages 181–196, 2001. (Cited pages 60, 153 and 154)

Lili Kotlerman, Ido Dagan, Idan Szpektor, and Maayan Zhitomirsky-Geffet. Directional distributional similarity for lexical inference. *Natural Language Engineering*, 16(04):359–389, 2010. (Cited page 31)

Milen Kouylekov, Yashar Mehdad, Matteo Negri, Elena Cabrio, et al. Fbk participation in rte6: Main and kbp validation task. *Proceedings of the Sixth Recognizing Textual Entailment Challenge*, 2010. (Cited page 193)

Klaus Krippendorff. Content analysis: an introduction to its methodology. 1980. (Cited pages 147 and 148)

Mirella Lapata and Alex Lascarides. Inferring sentence-internal temporal relations. In *HLT-NAACL*, pages 153–160, 2004. (Cited pages 71, 79 and 197)

Alex Lascarides and Nicholas Asher. Temporal interpretation, discourse relations and commonsense entailment. *Linguistics and philosophy*, 16(5):437–493, 1993. (Cited page 20)

Beth Levin. English verb classe and alternations: A preliminary investigation. 1993. (Cited pages 18, 22 and 71)

Beth Levin and Malka Rappaport. *Unaccusativity: At the syntax-lexical semantics interface*, volume 26. 1995. (Cited page 46)

David Lewis. *General semantics*. Springer, 1972. (Cited page 14)

Jiwei Li, Rumeng Li, and Eduard Hovy. Recursive deep models for discourse parsing. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2061–2069, 2014a. (Cited page 75)

Sujian Li, Liang Wang, Ziqiang Cao, and Wenjie Li. Text-level discourse dependency parsing. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 25–35, Baltimore, Maryland, June 2014b. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/P14-1003`. (Cited pages 63, 67, 163 and 170)

Dekang Lin. Automatic retrieval and clustering of similar words. In *Proceedings of the 36th ACL and 17th COLING joint conference*, volume 2, pages 768–774, Montreal, 1998. (Cited pages 29 and 37)

Dekang Lin and Patrick Pantel. Discovery of inference rules for question-answering. *Natural Language Engineering*, 7(04):343–360, 2001. (Cited page 35)

Dekang Lin and Patrick Pantel. Concept discovery from text. In *Proceedings of Coling 2002*, pages 1–7. Association for Computational Linguistics, 2002. (Cited pages 138 and 139)

Dekang Lin, Shaojun Zhao, Lijuan Qin, and Ming Zhou. Identifying synonyms among distributionally similar words. In *IJCAI*, volume 3, pages 1492–1493, 2003. (Cited pages 31 and 33)

Ziheng Lin, Min-Yen Kan, and Hwee Tou Ng. Recognizing implicit discourse relations in the penn discourse treebank. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 343–351. Association for Computational Linguistics, 2009. (Cited page 73)

Ziheng Lin, Hwee Tou Ng, and Min-Yen Kan. A PDTB-styled end-to-end discourse parser. *Natural Language Engineering*, 20(2):151–184, 2010. (Cited pages 66 and 162)

Maofu Liu, Wenjie Li, Mingli Wu, and Qin Lu. Extractive summarization based on event term clustering. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 185–188, Prague, Czech Republic, June 2007. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/P07-2047`. (Cited page 161)

Annie Louis and Ani Nenkova. Automatically evaluating content selection in summarization without human models. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 306–314. Association for Computational Linguistics, 2009. (Cited page 42)

Annie Louis, Aravind Joshi, Rashmi Prasad, and Ani Nenkova. Using entity features to classify implicit discourse relations. In *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 59–62. Association for Computational Linguistics, 2010. (Cited page 74)

Max Louwerse. An analytic and cognitive parametrization of coherence relations. *Cognitive Linguistics*, 12(3):291–316, 2001. (Cited pages 153 and 154)

John Lyons. *Introduction to theoretical linguistics*. Cambridge university press, 1968. (Cited pages 3, 16 and 17)

Debarghya Majumdar and Pushpak Bhattacharyya. Lexical based text entailment system for main task of rte6. 2010. (Cited page 193)

William C. Mann and Sandra A. Thompson. Rhetorical Structure Theory: Towards a Functional Theory of Text Organization. *Text*, 8(3):243–281, 1988. (Cited pages 43, 48, 49, 51, 52, 53, 57, 87 and 227)

Christopher D Manning and Hinrich Schütze. *Foundations of statistical natural language processing*. MIT press, 1999. (Cited page 138)

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, 2014. URL `http://www.aclweb.org/anthology/P/P14/P14-5010`. (Cited page 91)

Daniel Marcu. Building up rhetorical structure trees. In *Proceedings of the National Conference on Artificial Intelligence*, pages 1069–1074, 1996. (Cited pages 53, 54 and 227)

Daniel Marcu. From discourse structures to text summaries. In *Proceedings of the ACL*, volume 97, pages 82–88. Citeseer, 1997. (Cited page 42)

Daniel Marcu. The rhetorical parsing of unrestricted texts: A surface-based approach. *Computational linguistics*, 26(3):395–448, 2000. (Cited page 67)

Daniel Marcu and Abdessamad Echihabi. An Unsupervised Approach to Recognizing Discourse Relations. In *Proceedings of ACL*, pages 368–375, 2002. (Cited pages 68, 69, 70, 71, 72, 73, 79 and 80)

James R Martin. *English text: System and structure*. John Benjamins Publishing, 1992. (Cited page 153)

Shachar Mirkin, Ido Dagan, and Maayan Geffet. Integrating pattern-based and distributional similarity methods for lexical entailment acquisition. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 579–586. Association for Computational Linguistics, 2006. (Cited pages 34 and 37)

Seyed Abolghasem Mirroshandel, Alexis Nasr, and Benoît Sagot. Enforcing subcategorization constraints in a parser using sub-parses recombining. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 239–247, Atlanta, Georgia, June 2013. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/N13-1024`. (Cited page 139)

María Begona Villada Moirón. *Data-driven identification of fixed expressions and their modifiability*. PhD thesis, GRODIL, Secretary Department of General Linguistics, 2005. (Cited page 138)

Richard Montague. Universal grammar. *Theoria*, 36(3):373–398, 1970. (Cited page 10)

Richard Moot. Wide-coverage french syntax and semantics using grail. In *TALN 2010*, 2010. (Cited page 25)

Philippe Muller, Stergos Afantenos, Pascal Denis, and Nicholas Asher. Constrained decoding for text-level discourse parsing. In *Proceedings of COLING 2012*, pages 1883–1900, Mumbai, India, December 2012. The COLING 2012 Organizing Committee. URL `http://www.aclweb.org/anthology/C12-1115`. (Cited pages 65, 66, 163 and 173)

Joakim Nivre, Johan Hall, Jens Nilsson, Atanas Chanev, Gülsen Eryigit, Sandra Kübler, Svetoslav Marinov, and Erwin Marsi. Maltparser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(2):95–135, 2007. (Cited page 91)

Adrian Novischi and Dan Moldovan. Question answering with lexical chains propagating verb arguments. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 897–904. Association for Computational Linguistics, 2006. (Cited page 26)

Ekaterina Ovchinnikova, Laure Vieu, Alessandro Oltramari, Stefano Borgo, and Theodore Alexandrov. Data-driven and ontological analysis of framenet for natural language reasoning. In *LREC*, 2010. (Cited page 189)

Patrick Pantel and Dekang Lin. Discovering word senses from text. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 613–619. ACM, 2002. (Cited pages 29 and 34)

Patrick Pantel and Deepak Ravichandran. Automatically labeling semantic classes. In *HLT-NAACL*, volume 4, pages 321–328, 2004. (Cited pages 33 and 34)

Patrick Pantel, Deepak Ravichandran, and Eduard Hovy. Towards terascale knowledge acquisition. In *Proceedings of the 20th international conference on Computational Linguistics*, page 771. Association for Computational Linguistics, 2004. (Cited pages 32 and 34)

Joonsuk Park and Claire Cardie. Improving implicit discourse relation recognition through feature set optimization. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 108–112. Association for Computational Linguistics, 2012. (Cited page 73)

R. Parker, D. Graff, J. Kong, K. Chen, and K. Maeda. English gigaword fifth edition ldc2011t07. Web download, 2011. (Cited pages 72, 90 and 111)

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. (Cited pages 156 and 168)

Viktor Pekar. Discovery of event entailment knowledge from text corpora. *Computer Speech & Language*, 22(1):1–16, 2008. (Cited pages 36, 37 and 78)

Marco Pennacchiotti and Patrick Pantel. Entity extraction via ensemble semantics. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 238–247. Association for Computational Linguistics, 2009. (Cited pages 33 and 34)

Marco Pennacchiotti and Michael Wirth. Measuring frame relatedness. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 657–665. Association for Computational Linguistics, 2009. (Cited page 189)

Marco Pennacchiotti, Diego De Cao, Roberto Basili, Danilo Croce, and Michael Roth. Automatic induction of framenet lexical units. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 457–465. Association for Computational Linguistics, 2008. (Cited page 189)

Emily Pitler and Ani Nenkova. Using syntax to disambiguate explicit discourse connectives in text. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, ACLShort '09, pages 13–16, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics. URL `http://dl.acm.org/citation.cfm?id=1667583.1667589`. (Cited pages 45, 67 and 81)

Emily Pitler, Mridhula Raghupathy, Hena Mehta, Ani Nenkova, Alan Lee, and Aravind K Joshi. Easily identifiable discourse relations. 2008. (Cited page 68)

Emily Pitler, Annie Louis, and Ani Nenkova. Automatic sense prediction for implicit discourse relations in text. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, pages 683–691. Association for Computational Linguistics, 2009. (Cited pages 70, 73, 74, 79 and 197)

Livia Polanyi. A formal model of the structure of discourse. *Journal of pragmatics*, 12(5):601–638, 1988. (Cited pages 52, 53 and 54)

Livia Polanyi and Remko Scha. A syntactic approach to discourse semantics. In *Proceedings of the 10th international conference on Computational linguistics*, pages 413–419. Association for Computational Linguistics, 1984. (Cited page 48)

Rashmi Prasad, Eleni Miltsakaki, Nikhil Dinesh, Alan Lee, Aravind Joshi, Livio Robaldo, and Bonnie L Webber. The penn discourse treebank 2.0 annotation manual. 2007. (Cited pages 59, 87, 93, 153, 181 and 198)

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie L. Webber. The Penn Discourse TreeBank 2.0. In *Proceedings of LREC 2008*, 2008. (Cited pages 37, 49, 56, 58, 64, 66, 81, 83, 114 and 175)

James Pustejovsky. The generative lexicon. 1995. (Cited pages 15 and 47)

James Pustejovsky. Generativity and explanation in semantics: A reply to fodor and lepore. *Linguistic Inquiry*, 29(2):289–311, 1998. (Cited page 15)

James Pustejovsky, Patrick Hanks, Roser Sauri, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro, et al. The timebank corpus. In *Corpus linguistics*, volume 2003, page 40, 2003. (Cited page 193)

Pushpendre Rastogi and Benjamin Van Durme. Augmenting framenet via ppdb. *ACL 2014*, page 1, 2014. (Cited page 189)

Deepak Ravichandran and Eduard Hovy. Learning surface text patterns for a question answering system. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 41–47. Association for Computational Linguistics, 2002. (Cited page 32)

Philip Stuart Resnik. Selection and information: a class-based approach to lexical relationships. *IRCS Technical Reports Series*, page 200, 1993. (Cited page 36)

Michael Roth and Sabine Schulte im Walde. Combining word patterns and discourse markers for paradigmatic relation classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, volume 2, pages 524–530, 2014. (Cited pages 38 and 78)

Charlotte Roze, Laurence Danlos, and Philippe Muller. Lexconn: A french lexicon of discourse connectives. *Discours*, (10), 2012. (Cited pages 91, 153 and 171)

Attapol T Rutherford and Nianwen Xue. Discovering implicit discourse relations through brown cluster pair representation and coreference patterns. *EACL 2014*, page 645, 2014. (Cited pages 74, 75, 79 and 197)

Kenji Sagae. Analysis of discourse structure with syntactic dependencies and data-driven shift-reduce parsing. In *Proceedings of the 11th International Conference on Parsing Technologies*, pages 81–84. Association for Computational Linguistics, 2009. (Cited page 163)

Benoît Sagot. The lefff, a freely available and large-coverage morphological and syntactic lexicon for french. In *7th international conference on Language Resources and Evaluation (LREC 2010)*, 2010. (Cited pages 101, 223, 226 and 232)

Ted JM Sanders, Wilbert PM Spooren, and Leo GM Noordman. Toward a taxonomy of coherence relations. *Discourse processes*, 15(1):1–35, 1992. (Cited pages 48, 50 and 153)

Roser Saurí and James Pustejovsky. Are you sure that this happened? assessing the factuality degree of events in text. *Computational Linguistics*, 38(2):261–299, 2012. (Cited pages 84, 85 and 86)

Roser Saurí, Robert Knippen, Marc Verhagen, and James Pustejovsky. Evita: a robust event recognizer for qa systems. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 700–707. Association for Computational Linguistics, 2005. (Cited page 194)

Roger C Schank and Robert P Abelson. Scripts, plans, goals and understanding: An inquiry into human knowledge structures. 1977. (Cited pages 18 and 36)

Holger Schauer and Udo Hahn. Anaphoric cues for coherence relations. In *Proceedings of the Conference on Recent Advances in Natural Language Processing*, pages 228–234. Citeseer, 2001. (Cited page 88)

Karin Kipper Schuler. Verbnet: a broad-coverage, comprehensive verb lexicon. 2005. (Cited pages 22, 105 and 227)

Amy Sheldon. The role of parallel function in the acquisition of relative clauses in english. *Journal of verbal learning and verbal behavior*, 13(3):272–281, 1974. (Cited page 129)

Dan Shen and Mirella Lapata. Using semantic roles to improve question answering. In *EMNLP-CoNLL*, pages 12–21, 2007. (Cited page 188)

Ron Smyth. Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research*, 23(3):197–229, 1994. (Cited page 129)

Radu Soricut and Daniel Marcu. Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 149–156. Association for Computational Linguistics, 2003. (Cited pages 65 and 163)

C Sporleder and A Lascarides. Exploiting linguistic cues to classify rhetorical relations. 2005. (Cited pages 69, 79 and 80)

Caroline Sporleder and Alex Lascarides. Using Automatically Labelled Examples to Classify Rhetorical Relations: An Assessment. *Natural Language Engineering*, 14(3):369–416, July 2008. (Cited pages 69, 70, 80 and 181)

Manfred Stede. Discourse processing. *Synthesis Lectures on Human Language Technologies*, 4(3):1–165, 2011. (Cited pages 44 and 65)

Rajen Subba and Barbara Di Eugenio. An effective discourse parser that uses rich linguistic information. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 566–574. Association for Computational Linguistics, 2009. (Cited pages 65 and 163)

Carlos Subirats and Miriam Petruck. Surprise: Spanish framenet. In *Proceedings of CIL*, volume 17, 2003. (Cited page 188)

Alexandros Tantos. Rhetorical relations in verbal eventual representations? evidence from psych verbs. 2004. (Cited page 46)

Marta Tatu and Dan Moldovan. A logic-based semantic approach to recognizing textual entailment. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 819–826. Association for Computational Linguistics, 2006. (Cited page 26)

Galina Tremper and Anette Frank. A discriminative analysis of fine-grained semantic relations including presupposition: Annotation and classification. *Dialogue & Discourse*, 4(2):282–322, 2013. (Cited pages 37, 38, 78 and 161)

Joseph Turian, Lev Ratinov, and Yoshua Bengio. Word representations: a simple and general method for semi-supervised learning. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 384–394. Association for Computational Linguistics, 2010. (Cited page 74)

T Van de Cruys and P Mertens. The application of singular value decomposition to dutch noun-adjective matrices. 2006. (Cited pages 28 and 227)

Tim Van de Cruys. Two multivariate generalizations of pointwise mutual information. In *Proceedings of the Workshop on Distributional Semantics and Compositionality*, pages 16–20. Association for Computational Linguistics, 2011. (Cited page 138)

K. Van Den Eynde and P. Mertens. Le dictionnaire de valence: Dicovalence, 2010. (Cited page 100)

Antoine Venant, Nicholas Asher, Philippe Muller, Pascal Denis, and Stergos Afantenos. Expressivity and comparison of models of discourse structure. In *Special Interest Group on Discourse and Dialogue*, 2013. (Cited page 62)

Marc Verhagen, Robert Gaizauskas, Frank Schilder, Mark Hepple, Graham Katz, and James Pustejovsky. Semeval-2007 task 15: Tempeval temporal relation identification. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 75–80. Association for Computational Linguistics, 2007. (Cited page 193)

Laure Vieu. Representing content semantics, ontology, and their interplay. 2009. (Cited page 11)

Rui Wang, Yi Zhang, and Guenter Neumann. A joint syntactic-semantic representation for recognizing textual relatedness. *Proceedings of the Text Analaysis Conference*, 2009. (Cited page 193)

Bonnie Webber. What excludes an alternative in coherence relations? In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)*, pages 276–287. Association for Computational Linguistics, 2013. (Cited pages 86 and 88)

Bonnie Webber, Matthew Stone, Aravind Joshi, and Alistair Knott. Anaphora and discourse structure. *Computational linguistics*, 29(4):545–587, 2003. (Cited page 56)

Bonnie Lynn Webber and Aravind K Joshi. Anchoring a lexicalized tree-adjoining grammar for discourse. In *In Coling/ACL Workshop on Discourse Relations and Discourse Markers*, 1998. (Cited pages 50 and 56)

Julie Weeds, David Weir, and Diana McCarthy. Characterising measures of lexical distributional similarity. In *Proceedings of the 20th international conference on Computational Linguistics*, page 1015. Association for Computational Linguistics, 2004. (Cited page 30)

Hila Weisman, Jonathan Berant, Idan Szpektor, and Ido Dagan. Learning verb inference rules from linguistically-motivated evidence. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 194–204. Association for Computational Linguistics, 2012. (Cited page 36)

Ben Wellner. *Sequence models and ranking methods for discourse parsing*. PhD thesis, Citeseer, 2009. (Cited page 47)

Ben Wellner and James Pustejovsky. Automatically identifying the arguments of discourse connectives. In *EMNLP-CoNLL*, pages 92–101, 2007. (Cited pages 66 and 162)

Ben Wellner, James Pustejovsky, Catherine Havasi, Anna Rumshisky, and Roser Saurí. Classification of discourse coherence relations: an exploratory study using multiple knowledge sources. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*, SigDIAL '06, pages 117–125, Stroudsburg, PA, USA, 2006. Association for Computational Linguistics. (Cited page 73)

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pages 347–354. Association for Computational Linguistics, 2005. (Cited page 73)

Terry Winograd. On primitives, prototypes, and other semantic anomalies. In *Proceedings of the 1978 workshop on Theoretical issues in natural language processing*, pages 25–32. Association for Computational Linguistics, 1978. (Cited pages 2 and 14)

Florian Wolf and Edward Gibson. Representing discourse coherence: A corpus-based study. *Computational Linguistics*, 31(2):249–287, 2005. (Cited page 53)

List of Tables

List of Figures

Appendices

APPENDIX A

TOOLS FOR SEMANTIC RELATION EXTRACTION

Verbal locution	
mener	campagne
reprendre	confiance
savoir	lourd,bézef,beaucoup,long,trop,lerche,gras,tout,rien
réclamer	justice
chercher	asile,noise,querelle,fortune,abri,malice,chicane,refuge
prêter	assistance
sortir	grandi
accorder	foi
lier	amitié,conversation,partie,connaissance
tirer	vengeance,avantage,parti,profit
jouer	serré
espérer	beaucoup,rien
peser	lourd
bailler	belle
refuser	rien,obéissance
rentrer	dedans
obtenir	compensation,justice,réparation,satisfaction
aller	loin
tenir	froid,compagnie,tête,conseil,parole,lieu,garnison,bon, registre,compte,promesse,boutique,chaud,frais,école
fausser	compagnie
opposer	diamétralement
livrer	passage
pouvoir	beaucoup
donner	carte blanche,cours,congé,suite,asile,instruction,confiance, protection,lourd,matière,corps,signe,priorité,raison,bézef, gain de cause,courage,lieu,audience,lecture,justice,effet, passage,envie,droit,quitus,abri,satisfaction,tort,vie,ordre, naissance,refuge,cher,sujet,accès
redonner	courage,vie
ouvrir	boutique
parler	haut,affaire,boutique,chiffons
vivre	maritalement
signifier	beaucoup,peu
rendre	visite,confiance,bien,justice,gloire,hommage,vie,espoir,service
se rendre	compte
perdre	confiance,patience,contenance,pied
interjeter	appel
mériter	beaucoup,confirmation
revenir	cher
prendre|avoir	confiance
causer	affaire,boutique,chiffons
avoir	trait,cours,réponse,honte,hâte,voix au chapitre,faim,froid, horreur,matière,plaisir,priorité,valeur,prise,maille à partir, raison,avantage,gain de cause,soif,soin,coutume,lieu, occasion,part,peine,consensus,recours,idée,conscience, intérêt,mission,tendance,vocation,envie,possibilité,égard, droit,préséance,affaire,cure,besoin,dessein,tort,chaud, ordre,marre,vent,espoir,peur,rapport,autorité,foi,accès
porter	rancune,assistance,confiance,chance,malheur,préjudice, ombrage,secours,intérêt,plainte,bonheur,atteinte,dommage

Verbal locution	
courir	sus
passer	maître,inaperçu,tout
élire	domicile
demeurer	sec,couvert,court
rester	sec,couvert,court
battre	froid
compter	beaucoup,peu,guère,énormément
tenter	fortune
établir	domicile
répondre	non,présent,affirmativement,négativement,mal,oui
vouer	obéissance
nécessiter	confirmation
trouver	asile,assistance,appui,faveur,écho,prise,créance, fortune,abri,malice,grâce,naissance,refuge,place
comprendre	rien
dire	bonsoir,bonjour,tu,vrai,tout,vous,adieu
miser	gros
poser	problème
valoir	rien
voir	clair,tout,malice
pleurer	misère
renouveler	connaissance
prendre	fin,congé,contact,haleine,asile,note,livraison,appui,feu, domicile,plaisir,corps,date,rang,en main,parti,conseil, patience,soin,souche,modèle,courage,garde,position, acte,pitié,avis,mal,part,jour,en compte,effet,assise, rendez-vous,conscience,forme,possession,exemple,vie, espoir,peur,pied,place,racine,goût
apprendre	lourd,bézef,long
jurer	amitié,fidélité,obéissance
payer	tribut
faire	étalage,référence,figure,attention,honneur,silence,gaffe, honte,suite,levier,sensation,fête,confiance,sécession, grève,carrière,tache,affront,diversion,état,plaisir,impression, corps,mystère,surface,signe,halte,semblant,date,faillite, écho,défaut,tête,merveille,cadeau,mouche,recette,erreur, office,pression,montre,fureur,appel,abstraction,front, escale,équipe,choix,concurrence,acte,pitié,voeu,mal,part, peine,bloc,allusion,enquête,obstacle,exception,outrage, face,envie,gloire,loi,crédit,foule,don,obstruction,illusion, affaire,allégeance,rage,fortune,campagne,partie,tort,mine, grâce,mention,peur,échec,barrage,autorité,irruption,foi, naufrage,connaissance,intrusion,maladresse,place,école
tomber	bien,mal,faible
coûter	tant,chaud,cher
nouer	connaissance
demander	aide,asile,assistance,conseil,réflexion,audience, justice,secours,confirmation,trop,abri,refuge,rien
conter	fleurette
inspirer	confiance
recevoir	livraison,notification,satisfaction

Table A.1: List of French verbal locutions extracted from the Lefff resource (Sagot, 2010).

Adjective	Count	Adjective	Count	Adjective	Count	Adjective	Count
able	160252	high	29344	critical	16694	small	11864
likely	104990	first	28962	dead	16610	ok	11836
available	82094	necessary	28524	popular	16398	glad	11698
good	78002	confident	28492	alive	15784	simple	11686
ready	76212	unlikely	28142	optimistic	15662	big	11368
important	74320	close	27240	serious	15616	unchanged	11310
clear	68566	unclear	26476	pleased	15472	positive	11296
sure	63306	wrong	25742	over	14958	crucial	11240
due	56412	strong	25716	capable	14652	expensive	11128
hard	55346	great	24216	nice	14280	comfortable	11128
difficult	53058	old	23202	effective	14126	excited	11090
happy	48344	prepared	22934	common	14090	fair	11054
willing	46510	free	22654	much	13884	eligible	11042
concerned	42412	afraid	22032	reluctant	13764	illegal	10776
possible	41722	impossible	21888	low	13614	busy	10706
easy	38210	worried	21162	fine	13518	welcome	10706
right	35562	bad	20750	subject	13324	new	10650
different	35376	tough	20040	dangerous	13252	long	10582
responsible	34816	safe	19872	present	13046	obvious	10406
interested	33242	same	19722	second	12942	essential	10272
true	32536	certain	18694	eager	12756	slow	10242
worth	32452	enough	18616	full	12598	lucky	10124
unable	32024	successful	18100	careful	12466	surprised	10060
aware	30076	proud	17712	active	12066	angry	9928
open	29554	convinced	16952	similar	11890	third	9802

Table A.2: List of the 100 most frequent predicative adjectives following copulas in the Gigaword corpus, and number of occurrences.

Appendix B

Searching for significant relations

Verb pair		Verb pair	
connecter	activer	dépasser	intervenir
agir	réfléchir	envoyer	vérifier
lancer	réaliser	démouler	attendre
arriver	suggérer	fournir	occuper
poster	lire	révéler	permettre
rejoindre	travailler	mourir	aimer
engager	convenir	installer	travailler
établir	ne pas souhaiter	mourir	confier
jeter	tuer	entraîner	ne pas entendre
éviter	ressembler	ne pas hésiter	impliquer
attester	recueillir	acheter	renseigner
écrire	réfléchir	prévoir	commettre
parcourir	suffire	cesser	inspirer
servir	refroidir	commencer	ne pas poser
servir	saupoudrer	séparer	exercer
accorder	constater	conduire	nommer
contacter	ne pas lire	ne pas disposer	retourner
démouler	refroidir	comprendre	disparaître
servir	laisser	adorer	corriger
tuer	vendre	servir	réchauffer
acheter	comparer	revenir	partir
poster	relire	retrouver	attendre
rejoindre	débuter	tuer	entretenir
hésiter	insister	ramener	transformer
ajouter	déclarer	tuer	ne pas vendre
envoyer	transmettre	disjoncter	déconnecter
composer	proposer	partir	oublier
acheter	essayer	tester	diminuer
partir	ne pas oublier	ne pas cesser	estimer
apparaître	dresser	abonner	combattre
acheter	lire	endormir	lire
atteindre	traverser	coller	affronter
rejoindre	traverser	situer	attacher
installer	vivre	croire	cuire
compter	ne pas permettre	ajouter	mélanger
rédiger	écraser	entrer	frapper
présenter	marcher	consommer	attendre
sortir	consacrer	ne pas valoir	former
acheter	réfléchir	nommer	tenter
ajouter	expliquer	ne pas hésiter	ne pas utiliser
écraser	réagir	entamer	inquiéter
lire	appuyer	ne pas poser	adopter
imaginer	affronter	publier	lire
consulter	étendre	découper	refroidir
servir	parsemer	acheter	tester
mourir	jeter	pleuvoir	appeler
retrouver	compromettre	dépendre	libérer
ne pas compter	entraîner	entrer	travailler
déguster	laisser	guérir	prévenir
appuyer	ne pas sentir	rêver	rajouter

Table B.1: List of French verb pairs for out of context evaluation: Narration.

Verb pair		Verb pair	
chercher	ne pas trouver	demander	refuser
attendre	supprimer	chercher	trouver
préférer	ne pas servir	tenter	rejoindre
aimer	ne pas arriver	penser	demander
répondre	demander	permettre	rendre
diminuer	augmenter	envoyer	posséder
ne pas connaître	aimer	ajouter	garder
valoir	ne pas croire	préférer	prévoir
pratiquer	apporter	regarder	oublier
demander	nuire	préciser	servir
construire	soumettre	regarder	permettre
appartenir	taper	aimer	jouer
consacrer	considérer	passer	ne pas passer
privilégier	fermer	aider	commencer
perdre	gagner	constituer	suivre
appartenir	décrire	penser	croire
considérer	rire	lancer	toucher
lever	paraître	indiquer	posséder
essayer	ne pas arriver	proposer	refuser
ne pas aimer	aimer	connaître	ne pas connaître
réaliser	ne pas venir	travailler	expliquer
vendre	attacher	essayer	ne pas marcher
ne pas agir	garantir	augmenter	baisser
choisir	ne pas risquer	engager	effectuer
porter	envoyer	ne pas changer	bousculer
posséder	sentir	ne pas connaître	penser
appeler	oublier	ne pas exister	exister
ne pas agir	montrer	entendre	parler
ajouter	atteindre	revenir	définir
bénéficier	entendre	penser	regarder
essayer	ne pas réussir	valoir	arrêter
écrire	parler	lire	relire
découvrir	éternuer	indiquer	porter
perdre	montrer	envoyer	aider
retrouver	paraître	ne pas agir	attirer
ne pas dévoiler	essayer	fermer	ouvrir
décliner	réserver	oublier	ne pas trouver
rappeler	perdre	ne pas agir	créer
augmenter	envoyer	attendre	venir
permettre	ne pas permettre	indiquer	rappeler
tenter	précéder	proposer	montrer
adorer	aimer	servir	engager
ne pas agir	proposer	réaliser	porter
exister	attribuer	comprendre	reprendre
ne pas parler	parler	oublier	ne pas agir
ne pas diffuser	partager	tenter	échouer
offrir	consister	accepter	refuser
fournir	ressembler	atteindre	profiter
former	descendre	trouver	ne pas trouver
comprendre	ne pas marcher	arriver	répondre

Table B.2: List of French verb pairs for out of context evaluation: Contrast.

Verb pair		Verb pair	
efforcer	fonder	gagner	porter
convenir	ne pas suivre	ne pas regarder	ne pas reconnaître
trouver	chercher	arranger	venir
rédiger	ne pas apparaître	contacter	rencontrer
continuer	rappeler	présenter	ajouter
découvrir	offrir	envoyer	attaquer
agir	exister	aimer	plaire
apparaître	considérer	essayer	apprendre
sentir	arrêter	préciser	tromper
participer	reconnaître	offrir	acheter
ajouter	partir	marquer	témoigner
participer	partir	imposer	écrire
garder	aimer	posséder	travailler
acheter	souhaiter	comprendre	rappeler
découvrir	ne pas connaître	vérifier	ne pas connaître
soumettre	relever	risquer	disposer
informer	orienter	créer	obtenir
déterminer	permettre	montrer	garder
bénéficier	découvrir	perdre	appeler
ne pas trouver	tenter	manquer	bénéficier
comprendre	lire	dérouler	prévoir
proposer	désirer	prononcer	justifier
apercevoir	regarder	partir	posséder
laisser	posséder	subsister	survenir
vérifier	trouver	découvrir	aimer
venir	passer	manquer	inscrire
lire	intéresser	lâcher	rapporter
réaliser	annoncer	titulariser	juger
promettre	élire	risquer	arriver
manquer	entendre	sentir	compter
connaître	ajouter	développer	utiliser
réserver	exiger	entendre	envoyer
passer	progresser	porter	oublier
offrir	regarder	refuser	lancer
supprimer	recommencer	arriver	continuer
ne pas connaître	entrer	répondre	appeler
disposer	offrir	mentir	respirer
montrer	venir	tenir	refuser
travailler	oublier	passer	réussir
conserver	obtenir	afficher	cliquer
contacter	rechercher	porter	disposer
précéder	dépasser	constituer	apprendre
essayer	décider	permettre	aider
jouer	connaître	augmenter	revenir
risquer	répondre	commencer	poser
lancer	penser	envoyer	comprendre
présenter	ajouter	créer	tomber
demander	obtenir	rappeler	tenir
apparaître	montrer	acheter	attendre
laisser	abandonner	assurer	ne pas tomber

Table B.3: List of French verb pairs for out of context evaluation: Cause.