KATHOLIEKE UNIVERSITEIT
LEUVEN

Arenberg Doctoral School of Science, Engineering & Technology
Faculty of Engineering
Department of Computer Science

Automatic Detection and Classification of Argumentation in a Legal Case

Raquel MOCHALES

Dissertation presented in partial
fulfillment of the requirements for
the degree of Doctor
in Engineering

July 2011

Automatic Detection and Classification of Argumentation in a Legal Case

Raquel MOCHALES

Jury:
Prof. dr. ir. Paul Sas, chair
Prof. dr. Marie-Francine Moens, promotor
Prof. dr. Danny De Schreye, co-promotor
Prof. dr. ir. Maurice Bruynooghe
Dr. Steven Bethard
Prof. dr. Henry Prakken
(Utrecht University, The Netherlands)
Prof. dr. Patrick Saint Dizier
(IRIT-CNRS)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering

July 2011

D/2011/7515/98
ISBN 978-94-6018-396-6

Preface

First, I would like to thank my promotor Prof. Marie-Francine Moens for offering me the opportunity to pursue my Ph.D and for her guidance throughout the research. I would also like to include my gratitude to my co-promotor Prof. Danny De Schreye, for his kind interest in my work and progress.

I am very grateful to all the members of the PhD examination committee, Prof. Maurice Bruynooghe, Dr. Steve Berthard, Prof. Henry Prakken and Prof. Patrick Saint Dizier, for offering their expertise. Their comments and suggestions have improved the text significantly.

Furthermore, I am deeply indebted to my colleagues at LIIR that have provided the environment for sharing experiences, challenges and frustrations. Working together with you has been a pleasure and your friendship much appreciated. Special thanks to my office mates Phi, the Coca-Cola master, and Wim, who owes, among other things, his coffee addiction to me. Thank you for making our office a fun and interesting place to work in. Outside of the research group but still at K.U. Leuven, I would like to thank Wannes and Yves for their companionship during coffee breaks. Special thanks to Aagje Ieven with whom I explored the challenges of argumentation annotation.

During these years in Belgium I have met many good friends with whom I have shared many things in my daily life in Leuven. Thank you each of you, and forgive me if I forgot your name: Adriaan, Alex, Angela, Angels, Antonia, Christof, Dave, Dirk, Eric, Eva, Giorgos, Gorka, Inge, Leen, Leti, Marilu, Mauro, Mia, Patricia, Peter, Robin, Silke, Stef, Stephanie, Tassos, Thijs and Xavier. To my friends in Spain, thanks for your support from abroad: Anna B., Anna T., Aurora, Irene, Josephine, Jorg, Katy, Maria, Noe and Neus.

Finalment, vull agraïr a la meva familia pel seu suport i afecte. Moltes gràcies per creure i confiar en mi. Mama, Carol, Salvador i Helena gràcies per venir fins a Leuven. Thomas, bedankt voor je liefde en voor je geduld.

Abstract

The huge amount of documents available in the legal domain calls for computational tools supporting efficient and intelligent search and filtering of information. Over the last several years, machine-learning oriented research in information retrieval and document classification has spawned a number of systems capable of handling structural content management, helping users to automatically or semi-automatically identify relevant structured portions of legal texts, such as paragraphs, chapters or intertextual references.

This PhD thesis explores a novel idea to identify relevant portions of legal texts by using argumentation analysis. Many legal texts are argumentative, such as the case files exchanged by the parties in a case, a court's decision, scholarly publications and discussions and opinions in legal blogs. Therefore, argumentation can be used as a means to structure the texts contents to search and filter their information. However, there has been little research done on the automatic detection of argumentation or its structure.

This thesis presents a general way to automatically detect the arguments of a legal text and how they interact. This allows one to obtain a structured representation of the information of the text which later on can be used as a novel means to search or filter documents.

To this end this thesis introduces and discusses the development of the first corpus of legal texts fully annotated by their argumentation, including the four stages of the corpus creation process (design, collection, annotation and analysis). It also presents different approaches to obtain an automatic method to detect argumentation in legal cases. All the approaches are based on state-of-the-art information retrieval and natural language processing methods.

Beknopte samenvatting

Het grote aanbod van beschikbare documenten in de juridische wereld creëert een nood aan computationele hulpmiddelen die het efficiënt en intelligent zoeken in en filteren van informatie ondersteunen. Gedurende de laatste jaren heeft onderzoek in het terugvinden van gegevens en classificatie van documenten een aantal systemen geproduceerd die het beheer van gestructureerde informatie vergemakkelijken, en gebruikers helpen om automatisch of semi-automatisch relevante delen in wetteksten te identificeren, zoals paragrafen, hoofdstukken of intertekstuele referenties.

Deze doctoraatsverhandeling onderzoekt een nieuw idee om relevante delen van juridische teksten te identificeren via argumentatie-analyse. Vele wetteksten zijn argumentatief, zoals de dossiers uitgewisseld tussen de verschillende partijen van een rechtszaak, beslissingen van het hof, juridische publicaties, discussies en meningen in juridische weblogs. Zodoende kan argumentatie gebruikt worden als een middel om de inhoud van dergelijke teksten te structureren, zodat op hun informatie gezocht en gefilterd kan worden. Desondanks is er slechts een kleine hoeveelheid onderzoek geleverd naar de automatische detectie van argumentatie of de structuur hiervan.

Deze thesis stelt een generische methode voor om automatisch argumenten en hun onderlinge interactie in een wettekst te detecteren. Dit laat toe om een gestructureerde representatie van de informatie in de tekst op te stellen, dewelke later kan gebruikt worden als een nieuw middel om in de inhoud van documenten te zoeken of erop te filteren.

Om dit doel te bereiken, presenteren we de eerste collectie wetteksten die volledig geannoteerd zijn op basis van hun argumentatie, inclusief de vier fases van hoe deze collectie tot stand is gekomen (ontwerp, verzameling, annotatie en analyse). We presenteren ook twee verschillende methodes om automatisch argumentatie in wetteksten te detecteren. Al onze aanpakken zijn gebaseerd op grensverleggende informatieontginnings- en taalverwerkingsmethodes.

Abbreviations

AIF	Argumentation Interchange Format
ASCE	Assurance and Safety Case Environment
Aux	Auxiliar
AZ	Argumentative Zoning
BNF	Backus-Naur Form
CFG	Context Free Grammar
COE	Council of Europe
CRF	Conditional Random Fields
Det	Determinant
ECHR	European Court of Human Rights
HUDOC	Human Rights Documentation
IE	Information Extraction
MAP	Maximum A Posteriori probability
MaxEnt	Maximum Entropy Model
MLN	Markov Logic Network
MNB	Multinomial Naïve Bayes
NB	Naïve Bayes
NLP	Natural Language Processing
NOM	Nominal
NP	Noun Phrase
PCFG	Probabilistic Context Free Grammar

PSG	Phrase Structure Grammar
Q&A	Question Answering
RST	Rhetorical Structure Theory
SVM	Support Vector Machine
TC	Tree correctness
TFIDF	Term Frequency Inverse Document Frequency
VP	Verb Phrase

List of Symbols

(w, b)	Hyperplane
α	Krippendorff's alpha
β	A terminal symbol
Δ	Set of formulae
η	A particular parse of a structure T
γ	Margin
ϕ	Mapping function
σ	A sequence of terminals and non-terminals
Θ	A parameter vector
ε	The empty string
ϱ	A classifier
ϱ'	Another classifier
$\vdash$	Classical consequence relation
A	Argument (single or complex)
A, B, C	Non-terminal symbols of a grammar
C	A conclusion of an argumentative structure
c_{arg}	Number of detected arguments incomplete
D	The final decision of an argumentative structure
d_{arg}	Number of detected arguments

$Data$	Data which bears on the hypothesis h
F	Formula first-order logic
f	Target function
$F1$	F measure
F_n	The number of sentences that are incorrectly classified as negative
F_p	The number of sentences that are incorrectly classified as positive
f_{arg}	Final decision detected/not detected
F_{space}	Feature space
g	Ranges over the constituents of η
H	Entropy
h	A hypothesis
is_{arg}	Number of detected arguments with incorrect structure
K	Kappa coefficient
N	Set of non-terminals of a grammar
n	Nexus
n_{arg}	Total number of arguments
N_{auto}	Number of automatically classified sentences
N_{corr}	Number of correctly classified sentences
NT	A non-terminal symbol
P	A premise of an argumentative structure
$P(A)$	Agreement
$P(E)$	Agreement by chance
P_i	Group of premises
P_{ag}	A premise detected by the presence of an against rhetorical marker
P_{art}	A premise detected by the presence of an article reference
P_{sup}	A premise detected by the presence of a support rhetorical marker

P_{verbP}	A premise detected by the presence of a verb with function "*premise*"
$prop_i$	A sentence of the corpus
R	Set of rules/productions of a grammar
r_a	Contrast rhetorical marker
r_c	Conclusion rhetorical marker
r_i	Rule of a grammar
R_o	Input space
r_s	Support rhetorical marker
r_{art}	Article reference
S	Start symbol of a grammar
s	A sentence neither premise or conclusion of an argumentative structure
s_i	A simple argument
T	Argumentative structure of a legal case
t	Structure
T_a	Type of argument
T_n	The number of sentences that are correctly classified as negative
T_p	The number of sentences that are correctly classified as positive
TC	Tree correctness
Ter	Set of terminals of a grammar
v_c	Verb conclusion
v_p	Verb premise
w	A real number
X	Input object
x	A vector value of X
Y	Class label
y	A vlaue of Y

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Legal experts are exposed to huge masses of information when working on a legal case. This makes it hard for them to find the needle in the haystack as it is not possible to skim even a portion of the potentially relevant material. Moreover, the tradition of using extensive paper collections is undergoing a big change, as more and more legal cases are available in electronic form, offering resources to create automatic searching tools.

Legal experts have to keep abreast of developments in the legal system in general (e.g. laws, facts, exceptions). However, more practical requirements emerge when legal experts who are working on one case start recollecting information from other cases which might be relevant for their allegations. Their information needs change, the experts need understanding of the facts and laws but they might also benefit from a better understanding of the court's reasoning behind its decisions.

Current searches in legal case archives[1] are based on at least one of the following elements: citations (e.g. articles of the law, docket number[2]), party name, court, legal topic, date. None of these searches allow legal experts to obtain a collection of relevant cases with similar use of a type of reasoning (e.g. the court justifies its acceptance or rejection of a claim based on the margin of

[1] Online legal cases search tools: www.findlaw.com, www.law.cornell.edu/supct, www.justis.com

[2] A docket number is a number assigned to a case by a court clerk. Each new case is assigned a different number, which it carries for as long as it takes to be resolved. This number provides a key to finding out what is happening (or what has happened) in a particular case.

appreciation factor). This type of search would allow for example: (a) to study which characteristics of the facts were used in favour of or against a reasoning, (b) to study how previous case facts affected the case reasoning to foresee the play-role of the current ones, and (c) to detect the role playing reasoning patterns in judicial procedures.

To search a legal case by its reasoning one first needs to identify its argumentative structure, which is formed by all the arguments presented in the legal case together with different relations between those arguments. The detection of arguments and their relations is a difficult task with issues related to many fields, e.g. logics, dialectics. Furthermore, an argumentative structure can become quite complex if the reasoning behind it is extensive. Mingled arguments, connections between non-consecutive arguments or ambiguous sentences can complicate the understanding of the structure. Figure 1.1 presents a general view of the argumentative structure for a legal case that can be read in Appendix A. Closer views of the different parts of the structure are represented by Figure 1.2, Figure 1.3, Figure 1.4 and Figure 1.5.

Even if the detection of argumentation structures is a complex issue, this thesis argues that this task can be performed automatically using natural language processing and information extraction techniques. To this aim, this chapter starts introducing the concepts "argumentation", "natural language processing" and "information extraction". Then, it lists the main contributions of this work and outlines the structure of the rest of the document.

1.1 Argumentation in a legal case

Argumentation is a verbal activity normally conducted in an ordinary language (such as English). A speaker or writer, engaged in argumentation, uses certain words and sentences to state, question, or deny something, or to respond to statements, questions or denials. In this process arguments are constructed and handled [65].

Handling arguments may involve comparing arguments or evaluating them in some respects. To achieve that the arguments have to be broken down. The practice of breaking down arguments is called argument analysis. Argument analysis has different parts. The first part is only concerned with understanding the reasoning process of the author. It aims to pick out the ultimate premises, the final conclusion, and any intermediate steps. Once such an argument has been taken apart and its structure has been laid out so one can clearly see it, one can ask whether the argument is valid, whether the premises are

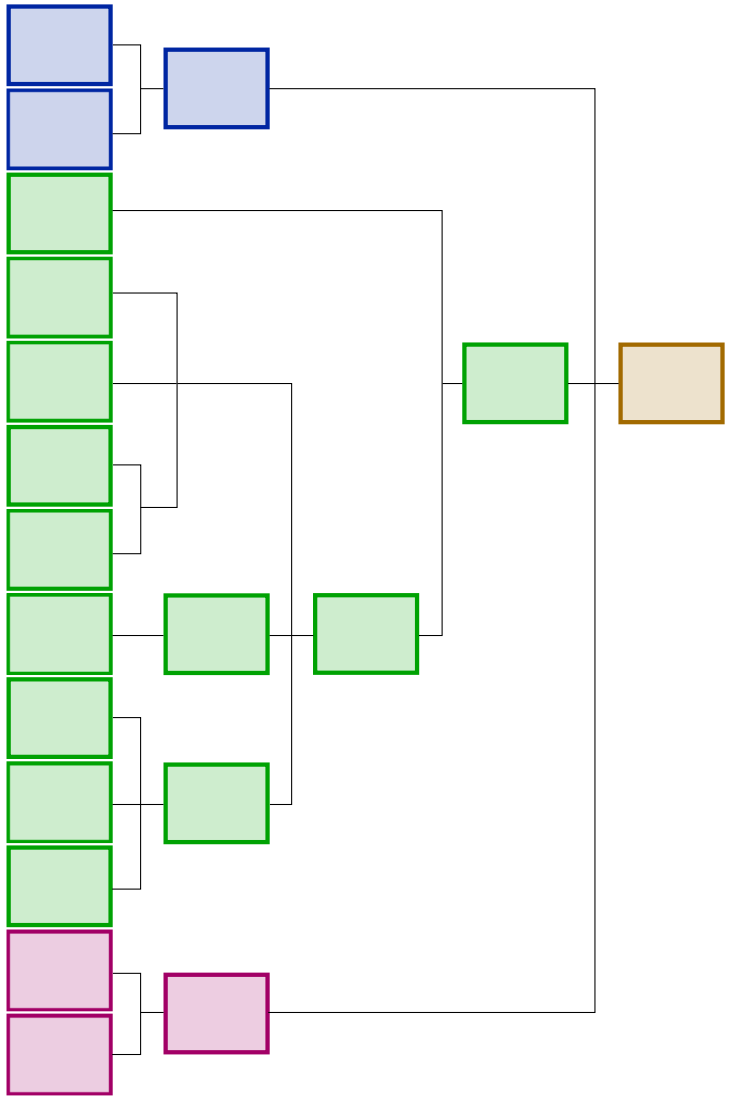

Figure 1.1: Reasoning structure of the legal case in Appendix A. Each block is a sentence of the legal case. There are 3 arguments (blue, green and red) that justify the final decision (brown). The contents of each argument and the final decision can be seen in detail in Figures 1.2, 1.3, 1.4 and 1.5

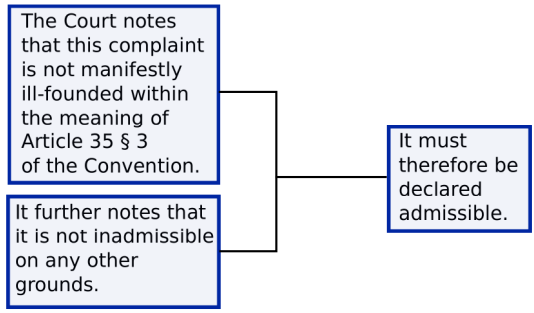

Figure 1.2: Closer view 1st Argument

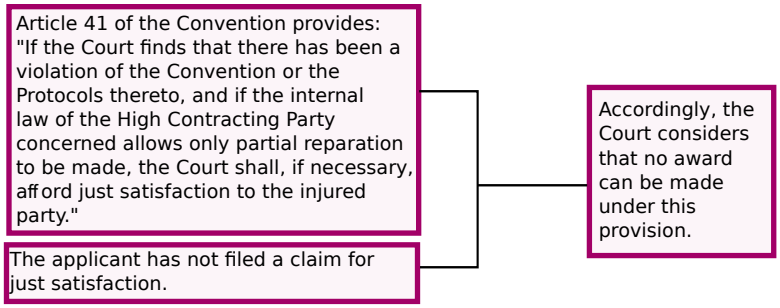

Figure 1.3: Closer view 2nd Argument

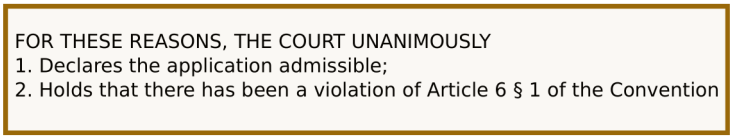

Figure 1.4: Closer view Final Decision

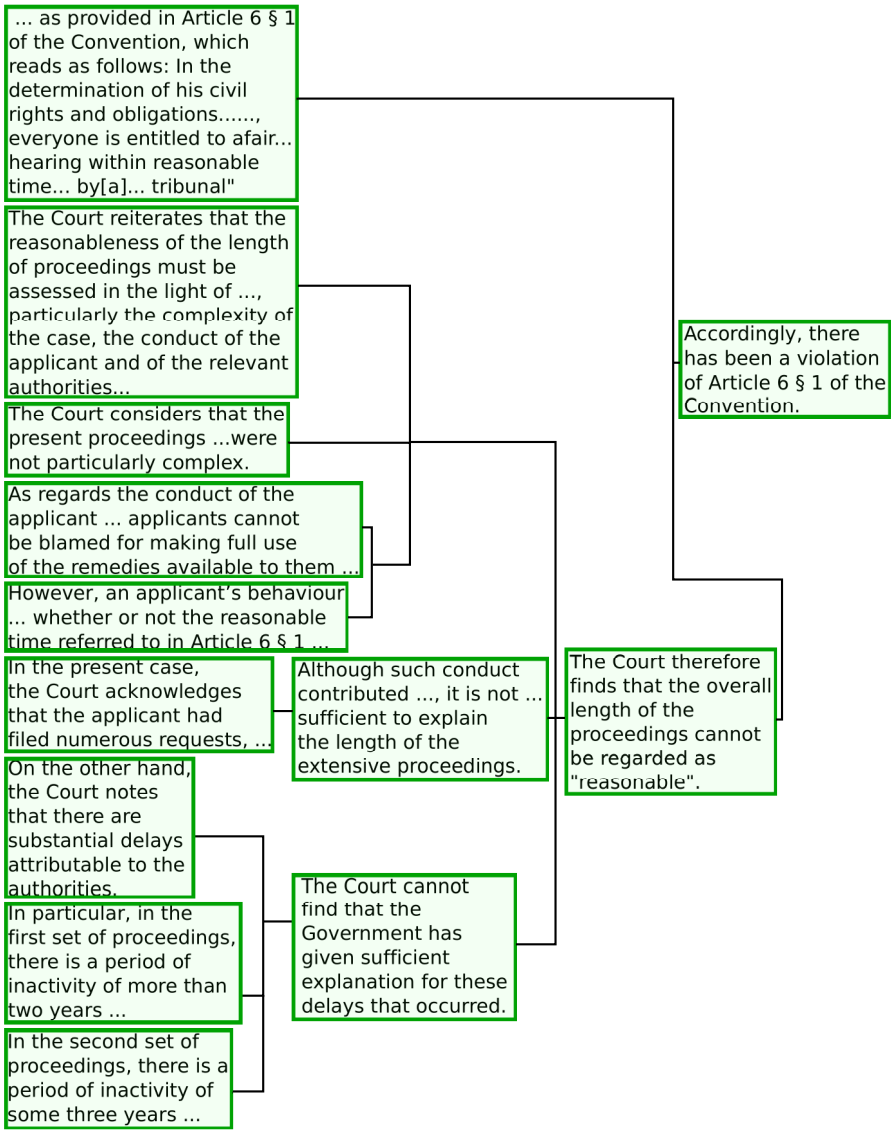

Figure 1.5: Closer view 3rd Argument

true and whether they really provide (probable or necessary) support for the intermediate or final conclusions.

This thesis focuses on the first part of argument analysis, i.e. the detection of all arguments in a text, their inner structure and their interactions. The detected arguments might or might not be valid or complete, i.e. the truth of the detected premises might not entail the truth of the detected conclusion or there might be missing premises to complete the reasoning behind the detected premises and conclusion. Note that a semantic analysis of the detected arguments is not the aim of this thesis.

Argumentation in a legal case is written in natural language. In general, various problems can arise when detecting arguments which have been written in natural language propositions in a text document. First, readers have to deal with lack of clarity. Not all writers use the language in an unambiguous and precise manner without making the document wordy and difficult to read. Second, in natural language one can say the same thing in completely different ways, it is up to the reader to find out the degree of relation between statements.

The language use in legal cases is intended to be correct and clear as its nature aims to deliver clear legal information without causing confusion. In legal language every chosen word is extremely important, ensuring correctness. Therefore, one can assume that in general legal arguments are clear and correctly presented. However, the need of detail in the presentation of legal information increases the complexity of the legal language, while the tendency to use arcane words and professional expressions affect the general comprehension. Therefore, legal arguments might be extensive, given the degree of detail needed, and difficult to comprehend. In Chapter 4 a detailed study on the linguistic characteristics of argumentative language in legal cases is presented.

1.2 Natural language processing

The aim of natural language processing (NLP) is the automatic analysis, transformation and generation of natural language texts using computer algorithms. NLP is a general methodology which has come back into fashion recently, and which is now applied in several tasks in theoretical linguistics, e.g. lexicography, syntax and lexical semantics [34].

As shown in the following chapters, this thesis mainly focuses on empirical or corpus-based natural language processing. In general, the focus is on using the automatic analysis of natural language to solve a specific information need

of an end-user. The general idea is that a linguist's or system developer's introspection alone cannot predict the unexpected turns of real language use. Rather than dealing with invented or artificially simplified examples, a large sample of naturally occurring language should be used instead. Empirical linguists aim to describe as much of the data as possible, but accept the fact that it is not normally the case that 100% of the data can be accounted for. It is generally accepted that large corpora are a reliable source of frequency-based data and provide the basis of a much more systematic approach to the analysis of language, as corpus methodology is open to verification of results [32].

Natural language processing is only a small part of the full process of solving an information need. An idealized description of such a process is:

- An end-user has a specific information need that can possibly be satisfied by the use of automated NLP from natural language texts.
- Someone familiar with NLP analyses the information need and specifies: (a) formal task definition, (b) task corpus description.
- A NLP expert designs and implements a computer algorithm to perform the defined task on the given corpus. The computer algorithm normally is not perfect but performs up to a certain accuracy.
- The computer algorithm is run on the entire task corpus. This produces automatic outputs for all input texts.

This thesis performs a study of the full process for a specific information need. It pays special attention to step 2: the formalization of the task and its corpus, and of step 3: the design of computer algorithms for the automatic analysis of natural text. However, on occasions it also addresses the other steps involved in the process.

1.3 Information extraction

Natural language processing is a broad discipline that comprises many different tasks. This thesis is only interested in one of these sub-tasks, information extraction (IE). Although this term is commonly used to refer to a number of related tasks, it does not have a commonly agreed definition. In this thesis information extraction is defined as the extraction of a predefined structure in natural language using computer algorithms, where elements in the structure have a mapping to individual words or phrases in the text.

Automatic analysis of text has been a goal of artificial intelligence researchers from the very beginning of computer science. The first extensively studied information extraction task was syntactic sentence parsing, that tried to discover the structure of a sentence according to a predefined grammar. Research on this topic started at the end of 1950's with the study of formal language theory, generative syntax and automatic parsing algorithms. These early parsing algorithms (e.g. [19]) used pattern matching and keyword search combined with simple heuristics for reasoning. By the end of 1960's more formal logical systems were developed: Colmerauer ([10]) defined a total precedence context free grammar and used the logic programming language Prolog to implement a deterministic sentence parser. An important disadvantage of this type of methods is that for an ambiguous sentence multiple parses are found without any indications of which parse is more likely. An important shift in methods used for information extraction occurred with the introduction of large corpora, such as the Penn Treebank ([38]), the Penn Discourse Treebank ([40]) and TimeBank ([50]). These corpora made it possible to use stochastical methods, which had already been succesfully applied to other problems, such as speech analysis [24]. The beginning of the 20th century saw a wide application of machine learning methods, such as support vector machines ([4]) and maximum entropy models ([2]). The new corpora also allowed a comparison of different information extraction algorithms on identical test-corpora.

Another well-studied IE trend is discourse analysis, which attempts to find patterns in language that are not explainable at the sentence level. There is substantial work in both, spoken and written, discourse analysis. However, most discourse theories depend on large amounts of world knowledge and have not been tested empirically on large text collections. These theories, therefore, do not adapt to most real texts, which are irregular and arbitrary. IE tries to develop techniques to automatically acquire discourse knowledge from real texts, so it can be added to discourse theories. In the 90's, there has been a strong interest in methods to automatically or semi-automatically extract domain-specific knowledge, see for example [23, 27, 55].

This thesis relates to extracting patterns of argumentative discourse in the legal domain. The second part of this thesis (Part II) presents experiments performed with a number of different information extraction tasks, moving from sentence extraction to parsing, to extract and analyse the patterns of an argumentative discourse.

1.4 Contributions

The aim of this work is to research if it is possible to automatically detect the argumentative structure of a legal case. In order to attain this aim, the different characteristics of legal argumentation, argumentative structures and argumentation annotation are studied. Moreover, some general NLP and IE techniques are applied to develop methods that automatically detect argumentation in legal cases. Below, a summary of the main contributions of this thesis is presented. More details on the contributions can be found at the end of each chapter.

- This thesis proposes a framework for argumentation annotation in legal cases, that allows to define the argumentation elements and their relations in natural argumentation.
- This thesis presents a study on the main errors humans incur when dealing with argumentation annotation, which has already been used to improve state-of-the-art argumentation theories.
- This thesis presents the first practical approaches to argumentation detection in legal cases, that are based on state-of-the-art NLP and IE techniques. This includes feature choosing, implementation of two automatic systems, and the training and testing of each system.
- This thesis presents the first tool that automatically extracts the argumentative structure of a legal case. The structure is shown as an argumentative tree.

1.5 Outline thesis

This thesis aims to contribute towards the indexation of documents by their argumentation in the framework of a document retrieval environment for legal cases. The practical topic of this thesis is whether it is possible to automatically detect the argumentation which will be the means to allow the indexation of legal cases. The rest of this thesis is structured as follows.

Chapter 2 starts with a description of basic concepts and techniques used throughout this work. It studies the main characteristics of legal cases, which influence many of the decisions during this work. It also gives an extensive introduction to different IE tasks, which will be used to develop all the approaches to automatic argumentation detection.

Then, Part I discusses argumentation annotation. It presents the state of the art of text annotation in Chapter 3. It discusses different corpora that could be used in this work, paying special attention to the Araucaria corpus. However, it concludes that a new corpus, specific for this task, is necessary. Therefore, Chapter 4 starts with a detailed definition of the characteristics the ideal corpus should present and justifies why the ECHR corpus was selected. Then, in Chapter 5, a detailed description of the annotation framework and the annotation process is given. Finally, Chapter 6 analyses the errors encountered during the design and development of the final corpus.

Part II discusses the approaches to the automatic detection of argumentation. It starts, in Chapter 7, with a description of the state of the art on automatic detection of argumentation. It focuses on two well-known theories related to discourse structure and argumentative information detection, which are the inspiration for most of this thesis work. Then it presents, in Chapter 8, the first approach, which is based on statistical classifiers. It shows the development of a method to distinguish, in a first stage, between argumentative and non-argumentative sentences, and later on, between a premise and a conclusion. Chapter 9 presents a second approach based on rule based parsing. It presents the development of a method to detect the argumentative structure of a legal case, cf. it detects the arguments of the legal case, the relations between them and their connection to the decision of the case. In Chapter 10, different approaches and issues for future research are presented. First, different ideas to modify the annotation process are presented and the implications of such modifications are studied. Second, the problems which a semantic approach should deal with are analysed. Third, a possible improvement of the rule-based parser based on the addition of statistical knowledge is studied with different methods. Fourth, the initial goal of document retrieval by argumentation is analysed given the results of the current research. Fifth, a study of other possible domains where argumentation detection could be feasible is presented.

This thesis concludes with a summary of the work that was performed and the lessons that were learned in the process in Chapter 11.

Chapter 2

Foundations

This chapter introduces the basic concepts and techniques used throughout this thesis. It starts with an introduction to the fundamentals of argumentation that circumscribe this thesis, see section 2.1. Then, in section 2.2, it introduces some important facts of legal argumentation, specifically in legal cases. Finally, section 2.3 describes a number of information extraction tasks, which will help to situate the foundations for the solving methodology of this thesis.

2.1 Argumentation fundamentals

This section presents some of the most widely accepted properties of argumentation to situate the problems tackled later on the creation of argumentation corpora (Part I) and the automatic detection of argumentation (Part II).

Argumentation is a vast topic which influences many fields, such as logic, philosophy, law or linguistics. This thesis focuses on a specific type of argumentation, i.e. the one found in legal cases. Therefore, this section studies different aspects of argumentation as found in legal cases. Nevertheless, first some basic concepts which will be repeatedly mentioned in this thesis are defined. Some definitions have been extracted from [3] (pp.2-3) and the others from the Standford Encyclopedia of Philosophy[1].

[1] http://plato.stanford.edu/

Definition 2.1. Argumentation *Argumentation is the process by which arguments are constructed and handled. Handling arguments may involve comparing arguments, evaluating them in some respects, and judging a constellation of arguments and counterarguments to consider whether any of them are warranted according to some principled criterion.*

Definition 2.2. Argument *An argument is a set of one or more declarative sentences or "assumptions", along with another declarative sentence that can be obtained by one or more reasoning steps (i.e. steps of deduction). The assumptions used are named the support, or equivalently, the* ***premises****, of an argument. The resulting declarative sentence is named the* ***conclusion*** *or the claim of the argument. The support of an argument provides the reason, or justification, for the claim of the argument.*

Definition 2.3. Premise *A premise is a declarative sentence that is a reason for, or objection against, some claim. A premise is a statement presumed true within the context of an argument toward a conclusion.*

Definition 2.4. Conclusion. *A conclusion is a declarative sentence that is supposed to be supported by the premises. In the context of ordinary argumentation, the rational acceptability of a disputed conclusion depends on both the truth of the premises and the soundness of the reasoning from the premises to the conclusion.*

Definition 2.5. Enthymeme. *Logicians use the term enthymeme for an argument with a hidden premise. Most arguments encountered in everyday contexts are enthymematic. Premises are left tacitly understood as being obvious or self-evident, or not conducive to succinct discourse. For example, in the argument "Socrates is mortal, since all men are." it is evident that a tacitly understood claim is that "Socrates is a man". The fully expressed reasoning is thus: "Since all men are mortal and Socrates is a man, it follows that Socrates is mortal.".*

Definition 2.6. Defeasible reasoning. *Reasoning is defeasible when the corresponding argument is rationally compelling but not deductively valid. The truth of the premises of a good defeasible argument provides support for the conclusion, even though it is possible for the premises to be true and the conclusion false. In other words, the relationship of support between premises and conclusion is a tentative one, potentially defeated by additional information.*

Definition 2.7. Natural argumentation. *During this thesis the term natural argumentation will be used when referring to the argumentation that can be found in real-life documents. This argumentation is expressed using natural language and has not been pre-processed to fit into a specific argumentation pattern.*

Most arguments found in natural argumentation are enthymemes or have unstated conclusions. However, any argument requires that enough information is provided so that a person can recognize the argument as being an argument. For example, someone might say the following: *When exposed to the agent "A", chimpanzees show a massive increase in aggression. Humans are very similar to chimpanzees. If exposed to "A", humans should show a massive increase in aggression*, that is clearly an argument with two premises followed by a conclusion. However, one might say: *A conviction requires that one is confident beyond a reasonable doubt about the defendant's guilt. But, the discussion shows low confidence about the defendant's guilt. So, it is obvious what one should do.* In this case, the person is most likely concluding that the jury should not convict the defendant so that would be the unstated conclusion. The implicity of the conclusion does not prevent the recognition of the argument.

2.1.1 Argumentation nature

Once the main concepts of argumentation have been defined it is possible to study two aspects of argumentation that influence most decisions taken in this thesis with respect to argumentation in legal cases; first, the distinction between the (predominantly verbal) process of an argument and the (predominantly textual) product of the argument, see [54]; second, the defeasible aspect of argumentation, see Standford Encyclopedia of Philosophy[2].

Process vs. Product

Arguments can be seen as processes or products of argumentation. An argument as a process is an act of inquiry characterized by fact finding, information gathering, and consideration of alternative points of view. As a product, an argument is someone's contribution to the "conversation" at any one moment, e.g. a written proposition paper or a speech act.

The aim of this thesis is to identify and analyse arguments found in a text of written discourse. Therefore, each argument is seen as a product. The argument is already there, and the analyst goes only by what is given in the text. What is given is a set of statements, one being a conclusion and the others playing the role of premises offering support for (or against the view represented by) that conclusion. However, the detection task quickly becomes one of argument as process. First, to identify the argument one has to identify the conclusion as a specific proposition that doubt is being expressed about, or

[2]http://plato.stanford.edu/

at least that is open to doubt. This identification assumes a dialectical aspect in which there are two sides to the argument. A proponent must put forward reasons to support the conclusion while a respondent must express doubt about the truth or acceptability of the conclusion. Therefore when identifying an argument, the view of argumentation as process is being implicitly appealed to.

Defeasibility

Most natural argumentation is defeasible. Humans make assumptions based in what is not said, therefore inferences based on this convention are defeasible as the unstated can be explicitly abrogated or suspended. This thesis works with natural argumentation in legal cases. Given the nature of current legal systems, legal argumentation generally presents a defeasible implication, i.e. arguments not always fully justify their conclusion. This enables lawyers to express what they know to be the case, rather than pretending they can make an exhaustive list of all possible exceptions.

Defeasible argumentation allows to define arguments as chains of reasons leading to a conclusion with consideration of potential counterarguments at each step. Therefore, one must expect that most arguments in natural legal argumentation present complex structures with different sub-arguments leading to a final conclusion.

When detecting defeasible arguments one deals with arguments rationally compelling but not deductively valid, i.e. the relation of support between premises and conclusion is tentative. The truth of the premises provide support for the conclusion, even though it is possible for the premises to be true and the conclusion false. Therefore, argument detection in legal cases must deal with a degree of uncertainty when detecting which premises support which conclusion.

2.1.2 Argumentation theory

The current state of the art in the study of argumentation does not afford a universally accepted theory, but it is based in the co-existence of a variety of approaches. These approaches differ considerably in conceptualization, scope, and degree of theoretical refinement [63]. This section first presents a brief summary of the evolution of argumentation theory. Then, it explores two state-of-the-art theories which were chosen as the theoretical background to develop the annotation framework use in this thesis (see Section 5.2). These argumentation formalizations were chosen to maximize the user-friendliness

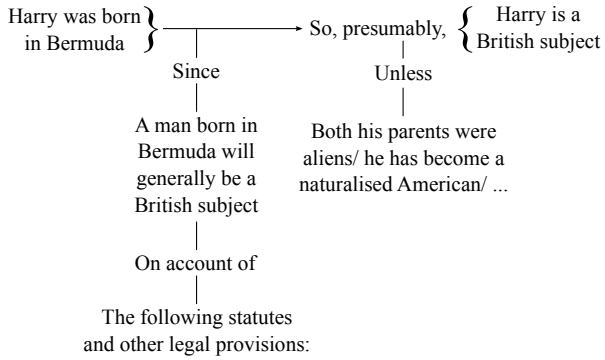

Figure 2.1: Example of Toulmin's layout of an argument found in [61]

of argumentation annotation while obtaining good results in the automatic detection of argumentation.

History

Argumentation has been studied for many years. The early voices of the history of argumentation are Plato and Aristotle. Until the 1950s, the approach of argumentation was based on rhetoric and logic, for example the work presented in [1]. In the 1960s and 1970s, two main theorists presented their view of argumentation, Perelman and Toulmin. Perelman was interested in finding a description of techniques of argumentation used by people to obtain the approval of others for their opinions, see [48]. He identified and defined many distinctive kinds of arguments used to convince a respondent on a provisional basis. Toulmin, on the other hand, was more interested on explaining how argumentation occurs in the natural process of an everyday argument, see [61]. He defined the layout of an argument as indicated in Figure 2.1, where he identified six elements of an argument: the claim, grounds, warrant, backing, qualifier and rebuttal.

Pragma-dialectics

At the beginning of this section natural argumentation has been defined as *the argumentation that can be found in real-life documents and is expressed using natural language*. The analysis of natural argumentation requires a pragmatic approach. Contextual information and background knowledge have an important role in natural argumentation, therefore, its analysis should take

into account those factors. Natural argumentation analysis need to show how argumentation works in practice in a "natural environment".

Pragma-dialectics, or pragma-dialectical theory, developed by Frans H. van Eemeren and Rob Grootendorst, is an argumentation theory that is used to analyse and evaluate argumentation in actual practice. The main concepts of pragma-dialectics as found in [63, 65, 64] are summarized as follows.

Unlike strictly logical approaches (which focus on the study of argument as a product), or purely communication approaches (which emphasize argument as a process), pragma-dialectics was developed to study the entirety of an argumentation as a discourse activity. Pragma-dialectical theory assumes argumentation is always part of an explicit or implicit dialogue in which one party attempts to convince the other party of the acceptability of his or her standpoint. In a fully explicit dialogue, the antagonist expresses his doubts and criticisms unequivocally, and all these doubts and criticisms must be answered by the protagonist by advancing more argumentation. In an implicit dialogue where the antagonist is silent, the protagonist can only anticipate the antagonist's doubts or criticism; he will only advance more argumentation if he assumes that doubts or criticism are to be expected.

Argumentation is seen as a complex whole made up of statements put forward to deal with real or anticipated critical reactions from an antagonist. Thus, argumentation for or against a standpoint can be simple, as in *"single argumentation"*, which consists of only one explicit reason for or against the standpoint, or can have a more intricate argumentation structure, as in *"complex argumentation"*, depending on the way in which the defence of the standpoint has been organized in view of (anticipated) doubts or criticism.

In complex argumentation, several reasons are put forward for or against the same standpoint. These reasons can be alternative defences of the standpoint that are unrelated, as in *"multiple argumentation"*, but they can also be interdependent, so that there is a *"parallel chain"* of mutually reinforcing reasons, as in *"coordinate argumentation"*, or a *"serial chain"* of reasons that support each other, as in *"subordinate argumentation"*. Figures 2.2, 2.3, 2.4 show examples of these types of argumentation found in [63].

A problem in the analysis of complex argumentation is that the literal presentation often makes insufficiently clear whether the argumentation is multiple, coordinatively compound, subordinatively compound, or some combination of these possibilities. In these cases all kinds of contextual and other pragmatic factors need to be taken into account in the analysis.

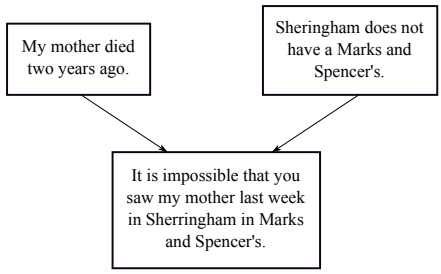

Figure 2.2: Multiple argumentation

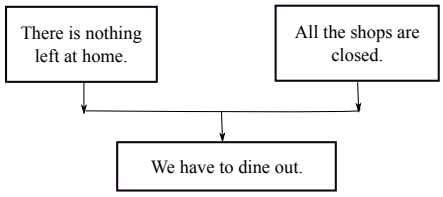

Figure 2.3: Coordinate argumentation

Argumentation schemes

It has been previously argued that argumentation found in legal cases is mostly defeasible argumentation, see Section 2.1.1. Different formalisms of defeasible argumentation have been studied, see [47, 49], one of the most relevant is, perhaps, argumentation schemes.

An argumentation scheme is a form of argument that can hold provisionally on a balance of considerations under conditions of uncertainty, but that might be defeated by the asking of critical questions that pinpoint weaknesses.

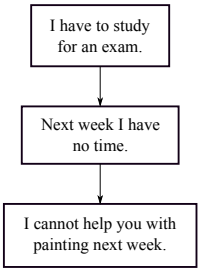

Figure 2.4: Subordinate argumentation

Major Premise: Source "a" is in a position to know about things in a certain subject domain "S" containing proposition "A".
Minor Premise: "a" asserts that "A" (in Domain "S") is true (false).
Conclusion: "A" is true (false).
Critical Question 1: Is "a" in a position to know whether "A" is true (false)?
Critical Question 2: Is "a" an honest (trustworthy, reliable) source?
Critical Question 3: Did "a" assert that "A" is true (false)?

Figure 2.5: Example of argumentation scheme: Argument from Position to Know found in [69]

Argumentation schemes pertain to the kind of relationship between the explicit premise and the standpoint that is established in the argumentation in order to promote a transfer of acceptability from the explicit premise to the standpoint. Therefore, argumentation schemes are more or less conventionalized ways of achieving this transfer.

The main concepts of argumentation schemes theory are summarized in [69], where most of the argumentation schemes used in natural argumentation are described. The list is not complete but identifies many of the most common forms of defeasible argumentation, see Appendix B. For example, the following instance of a defeasible argument is an instantiation of a particular scheme, the *Argument from waste*:

"*A PhD student, Susan, has spent more than five years trying to finish her thesis, but there are problems. Her advisers keep leaving town, and delays are continued. She contemplates going to law school, where you can get a degree in a definite period. But then she thinks: Well, I have put so much work into this thing. It would be a pity to give up now.*"

Each argumentation scheme is defined by a set of premises, a conclusion and a set of critical questions. For example, Figure 2.5 presents the *Argument from Position to Know*, which is based on the assumption by one party that another party has information that the first party needs.

The rationale behind argumentation schemes is as follows. A proponent puts forward an argument in a dialogue that meets the requirements indicated in the argument scheme. Then, the argument carries some weight as a presumption but it is defeasible by questioning. Therefore, the respondent can ask any one of the critical questions. Once the question has been asked the presumptive weight the argument had before is withdrawn. But if the proponent gives an acceptable answer to the question, the weight is restored.

The nature of both natural argumentation and argumentation schemes allows

that any argument presented in natural argumentation can be matched with a corresponding argumentation scheme. Moreover, argumentation schemes represent a mechanism to aid the process of analysis and reconstruction from both a product and process viewpoint. An analyst, by identifying claims and trying to link them with schemes (argument as a product), is guided towards critical questions by which to judge the strength of the claims, their relation, and the resulting argument (argument as a process). The schemes also highlight the type of reasoning being employed, allowing a more refined taxonomy of specific forms of support.

Argumentation schemes are therefore argumentation patterns easy to find in natural argumentation, such as the one found in legal cases. Moreover, the patterns are not too complex for an automatic tool to work with them, the basic elements are premises and conclusions. From the legal expert point of view argumentation schemes are easy to understand, to find in the text and to structure the sentences accordingly.

2.2 Legal fundamentals

The previous section has analysed the background of this thesis from the point of view of argumentation. This section presents the background from the point of view of law. In law, argumentation plays an important role specifically when someone presents a legal claim and wishes others to accept this claim. A lawyer who pleads a case in court must justify his or her case with arguments. The judge who makes a decision is expected to support it with arguments. In fact, in many legal systems the judge is obligated to justify his or her decision. When a legislator introduces a bill in Parliament, he or she is expected to support this proposal with arguments. In fact, everybody who advances a legal standpoint and wishes this standpoint to be accepted by others, will have to present justifying arguments [41].

Case law, also known as decisional law or judicial precedent, is that body of reported judicial opinions in countries that have common law legal systems, e.g. Australia, USA, UK or Canada, and also in special courts, e.g. the European Court of Human Rights (ECHR) or the Court of Wales. In the common law tradition, courts decide the law applicable to a case by interpreting statutes and applying precedents which record how and why prior cases have been decided. Unlike most civil law systems, common law systems follow the doctrine of *stare decisis*, by which most courts are bound by their own previous decisions in similar cases, and all lower courts should make decisions consistent with previous decisions of higher courts. Legal cases are the main decision tool

[PREMISE *It is indisputable that there is no common ground on the question.*] [PREMISE *Although most of the Contracting States do not expressly prohibit homosexuals from adopting where single persons may adopt, it is not possible to find in the legal and social orders of the Contracting States uniform principles on these social issues on which opinions within a democratic society may reasonably differ widely.*] [CONCLUSION *The Court considers it quite natural that the national authorities, whose duty it is in a democratic society also to consider, within the limits of their jurisdiction, the interests of society as a whole, should enjoy a wide margin of appreciation when they are asked to make rulings on such matters*]

Figure 2.6: Example of a judicial argument extracted from an ECHR legal case

of case law. Therefore, any legal decision presented in a legal case has long-term consequences and repercussion in following cases. Anyone involved in a legal claim must be sure his or her standpoints are clearly presented and understood by any legal processor, thus the decision of the case depends on them. Therefore, to ensure the case is correctly decided, defendants, plaintiffs and legal processors must be sure to study and present all the possible arguments. Furthermore, any argument presented in a legal process should be as unambiguous as possible and, therefore, follow a clear structure. In conclusion, in theory, legal argumentation should have a low rate of ambiguity and present none or few implicit information, making it a perfect test case for any argumentation study [41].

2.2.1 Legal case

A legal case is a dispute between opposing parties resolved by a court, or by some equivalent legal process. A legal case normally begins when a plaintiff files a document called a complaint with a court, informing the court of the wrong that the plaintiff has allegedly suffered because of the defendant, and requesting a remedy [41].

Legal cases are premised on the idea that a dispute will be fairly resolved when a legal procedure exists by which the dispute can be brought to a factfinder not otherwise involved in the case, who can evaluate evidence to determine the truth with respect to claims of guilt, innocence, liability, or lack of fault. This factfinder is normally a court, commission or committee, such as the Supreme Court of the United States or the High Court of Australia [41]. An example of argument from a judicial decision can be found in Figure 2.6.

All courts keep an accurate track of their cases and most of them maintain online databases with reports of the cases[3]. Moreover, given the importance of the previous cases in case law, different companies have developed specialized search tools, e.g. FindLaw[4] or Justis[5], that retrieve legal cases from different courts depending on different characteristics, e.g. the nationality of the applicant or defendant, relevant articles mentioned in the case or the type of court deciding on the case.

2.2.2 Legal case structure

A legal case has a fixed structure that depends on the Court processing the case. At the beginning of this thesis some random legal cases from different courts were selected and analysed by a legal expert. The results of this little analysis showed that the relevant parts for the argumentative process can be easily detected in all types of legal cases. For example, the *Supreme Court of the United States*, that files its reports as: *Syllabus*, *Opinion of the Court* and *Opinion of the judge X of the Court*, comprises the argumentative process of the Court in the second section and the argumentative process of each dissenting judge in the third section. On the other hand, the *Supreme Court of Canada* has a bit more complex structure. It starts deciding the admissibility of the case together with dissenting opinions of the judges on that point. Then, different citations, such as statutes, regulations or judges involved, are enumerated. This is followed by three possible sections: *Introduction*, *The facts* and *The judicial history of the case*. From this point on the structure is dependent on the type of case and the involved complaints. Some examples of possible sections are: *Restitution for Ultra Vires Taxes*, *Application of the Doctrine of Protest and Compulsion*, *Application of Limitations Law* or *Duty to Accommodate and Collective Labour Relations*. The report ends with the section *Conclusion*, which introduces the final decision of the court. Even if this structure is more complex, it is possible to see that the argumentation of the case will be encountered on the admissibility section, when dealing with the admissibility decision, and between the judicial history of the case and the conclusion, when dealing with the final court's decision [42].

Therefore, one can assume that all legal cases have similar structures that move from non-argumentative sections to argumentative sections finishing in conclusive sections. Furthermore, all courts use mostly the same keywords and verbs to denote conclusions or premises, e.g. "*accordingly*", "*however*", "*to hold*" or "*to reach*". However, there are some writing style differences, such

[3] See: http://csc.lexum.umontreal.ca/en/ or http://www.supremecourtus.gov
[4] http://www.findlaw.com
[5] http://www.justis.com

as the use of "*we think*" instead of "*the Court thinks*", i.e. direct vs. indirect speech, or referring to the applicant by name or with the term "*applicant*". There is however no consensus on the titles applied to each section or the number of sections. Note that this assumption cannot be fully confirmed till a complete study on legal case structures has been done, however the current initial observations seem to justify the assumption.

2.3 Information extraction fundamentals

In the previous chapter (Chapter 1) information extraction (IE) has been described as the task of extracting a predefined structure in natural language using computer algorithms. In this section this description is made more concrete by describing some popular information extraction tasks which are fundamental for this thesis and will be used in Part II to achieve the automatic detection of argumentation in legal cases. Note that these descriptions do not explore the more complex issues behind the IE tasks but are intended as a basic introduction for readers from a legal domain.

2.3.1 Classification

Classification is the act of distributing things into classes or categories of the same type. In machine learning, classification is the assignment of class labels (Y) to input objects (X).

This section provides an introduction to some classical classification methods, i.e. statistical classifiers. Statistical classifiers are among the most popular classifiers used in the machine learning community. These classifiers are derived from generative probability models which provide a principled way to the study of statistical classification in complex domains such as natural language and visual processing.

There are two types of classifiers depending on their training, discriminative and generative classifiers. Training classifiers involves estimating $f : X \rightarrow Y$, or $P(Y|X)$. Discriminative classifiers assume some functional form for $P(Y|X)$ and estimate parameters of $P(Y|X)$ directly from the training data. On the other hand, generative classifiers assume some functional form for $P(X, Y) = P(X|Y)P(X)$, estimate parameters of $P(X|Y), P(X)$ from the training data by using Bayes rule to calculate $P(Y|X = x_i)$.

Now three different state-of-the-art statistical classifiers are described. First, a generative classifier, the *multinomial naïve Bayes classifier*, is described, and

then two discriminative classifiers, the *maximum entropy classifier* and *support vector machines*.

Multinomial naïve Bayes classifier

A naïve Bayes classifier is an example of a generative classifier which learns a model of the joint probability, $P(X, Y)$ and makes its predictions by using Bayes rule to calculate $P(Y|X)$ [39].

Definition 2.8. Bayes Rule. *Given a hypothesis h and data $Data$ which bears on the hypothesis, then:*

$$P(h|Data) = \frac{P(Data|h)P(h)}{P(Data)}$$

where:

- $P(h)$*: independent probability of* h*: prior probability*
- $P(Data)$*: independent probability of* $Data$
- $P(Data|h)$*: conditional probability of* $Data$ *given* h*: likelihood*
- $P(h|Data)$*: conditional probability of* h *given* $Data$*: posterior probability*

Consider a target function $f : X \to Y$, where each instance x is described by a vector of n attributes $a_1, \ldots, a_n$. Applying Bayes rule, one can represent $P(Y = y_i|X)$ as:

$$P(Y = y_i|X = x_k) = \frac{P(Y = y_i)P(X = x_k|Y = y_i)}{\sum_j P(Y = y_j)P(X = x_k|Y = y_j)}$$

where y_i denotes the ith possible value for Y, x_k denotes the kth possible vector value for X, and where the summation in the denominator is over all legal values of the random variable Y.

To learn $P(Y|X)$ one can use training data to estimate $P(X|Y)$ and $P(Y)$, and then use these estimates together with Bayes rule to determine $P(Y|X = x_k)$ for any new instance x_k. To obtain reliable estimates with feasible complexity Naïve Bayes classifiers make a conditional independence assumption that dramatically reduces the numbers of parameters to be estimated when modelling $P(X|Y)$.

Definition 2.9. Conditional Independence *Given random variables X, Y and Z, we say X is conditionally independent of Y given Z, if and only if the probability distribution governing X is independent of the value of Y given Z; so $\forall i, j, k\ P(X = x_i|Y = y_j, Z = z_k) = P(X = x_i|Z = z_k)$.*

In conclusion, the Naïve Bayes algorithm is a classification algorithm based on Bayes rule, that assumes the attributes $X_1, \ldots, X_n$ are all conditionally independent of one another, given Y. Therefore, $P(X|Y)$ can be represented as follows:

$$P(X_1, \ldots, X_n|Y) = \prod_{i=1}^{n} P(X_i|Y);$$

so one can represent the probability that Y will take on its k_{th} possible value as:

$$P(Y = y_k|X) = P(Y = y_k|X_1, ..., X_n) =$$

$$= \frac{P(Y = y_k)P(X_1, ..., X_n|Y = y_k)}{\sum_j P(Y = y_j)P(X_1, ..., X_n|Y = y_j)} =$$

$$= \frac{P(Y = y_k)\prod_i P(X_i|Y = y_k)}{\sum_j P(Y = y_j)\prod_i P(X_i|Y = y_j)}.$$

If one is interested only in the most probable value of Y, then one has the Naïve Bayes classification rule:

$$Y \leftarrow argmax_{y_k} \frac{P(Y = y_k)\prod_i P(X_i|Y = y_k)}{\sum_j P(Y = y_j)\prod_i P(X_i|Y = y_j)}$$

which can be simplified as:

$$Y \leftarrow argmax_{y_k} P(Y = y_k)\prod_i P(X_i|Y = y_k)$$

In a variation of this model, which is called multinomial naïve Bayes classifier [39] (MNB) and which is often used in text categorization tasks, the number

of occurrences of each feature is captured in a feature vector. Therefore, the feature vectors capture word frequency information, not just presence or absence. This model normally performs better than other NB variations.

Maximum entropy classifier

The maximum entropy classifier (MaxEnt) is an example of a discriminative classifier, which models the posterior probability $P(Y|X)$ directly while learning a direct map from inputs X to the class labels Y [2]. The features are described by binary variables called feature functions.

This classifier adheres to the maximum entropy principle. This principle states that, when one makes inferences based on incomplete information, one should draw them from that probability distribution that has the maximum entropy permitted by the information one has. Therefore, when nothing is known, the distribution should be as uniform as possible, that is, have maximal entropy.

$$Entropy = H(P) = -\Sigma_x P(x) \log P(x)$$

Therefore, the goal is to choose the one distribution P^*, among all the distributions that satisfy the constraints, that maximizes $H(P)$.

$$P^* = argmax_{p \in P} H(p)$$

Entropy maximization with no testable information takes place under a single constraint: the sum of the probabilities must be one. Under this constraint, the maximum entropy discrete probability distribution is the uniform distribution,

$$P_i = \frac{1}{n} \text{ for all } i \in \{ 1, \ldots, n \}.$$

Natural language often deals with incomplete patterns in training sets given the variety of natural language patterns that signal similar content. Hence, this type of classifier is often used in information extraction from natural language texts.

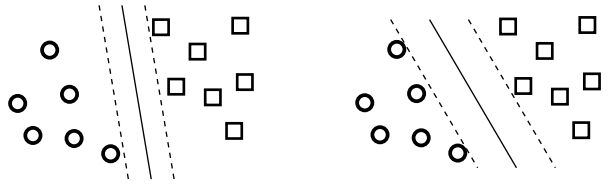

Figure 2.7: Support Vector Machine: idealized example

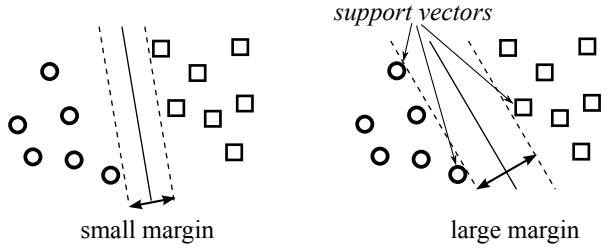

Figure 2.8: Support Vector Machine: support vectors

Support vector machines

Another well-known discriminative classifier is the support vector machine, see [57]. Support vector machines (SVM) are a group of supervised learning methods that can be applied to classification or regression. Support vector machines represent an extension to nonlinear models of the generalized portrait algorithm developed by Vladimir Vapnik.

A Support Vector Machine (SVM) performs classification by constructing an N-dimensional hyperplane that optimally separates the data into two categories [25]. In the idealized example of Figure 2.7, the cases with one category are in the lower left corner and the cases with the other category are in the upper right corner; the cases are completely separated.

Figure 2.8 illustrates the SVM analysis to find a 2-dimensional hyperplane (i.e. a line) that separates the cases based on their target categories. There are an infinite number of possible lines; the figure shows two candidate lines. The question is which line is better, and how does one define the optimal line. The dashed lines drawn parallel to the separating line mark the distance between the dividing line and the closest vectors to the line. The distance between the dashed lines is called the margin. The vectors (points) that constrain the width of the margin are the support vectors. The best line is the line oriented so that the margin between the support vectors is maximized.

In general, the hyperplane is a nonlinear decision boundary in the input space. Assume $X \in R_0 \subseteq \Re^n$ is the input vectors, $y \in \{-1, +1\}$ is the labels, and $\phi : R_0 \rightarrow F_{space}$ is the mapping from input space to feature space. Then, the SVM learning algorithm finds a hyperplane (w, b) where the margin γ is the maximized quantity of: $\gamma = \min_i y_i\{\langle w, \phi(X_i)\rangle - b\}$. The decision function is then: $f(X) = sgn(\langle w, \phi(X)\rangle - b)$.

2.3.2 Parsing

Parsing is the process by which a sequence of characters or values are split into smaller parts according to a set of rules. One of the most common parsers is the *syntactic parser*, which is a process that recognizes the grammatical structure of sentences, for instance, which groups of words go together (as "phrases") and which words are the subject or object of a verb. Another type of parsing is document parsing, where a document is parsed for example by its different sections or by its argumentation.

This section presents in detail the two main concepts of parsing, the grammar and the parser. It starts describing the concept of *grammar*, which is (basically) the set of rules that will be followed during the parsing process. Then, it focuses on the *parser*, where it describes the different types of parsers depending on how they derive the input from the start symbol of the grammar.

Grammars

In order to parse any textual data, one must first agree on the grammar to be used. The choice of grammar is affected by both linguistic and computational concerns. There are some basic principles of sentence organization: linear order, hierarchical structure (constituency) or subcategorization, and grammatical relations.

A sentence has different meanings based on the linear order of its words. For example, *"John loves Mary"* has a different meaning than *"Mary loves John"*. Languages vary as to what extent this is true, but linear order is still a guiding principle for organizing words into meaningful sentences. However, one cannot only use a linear order to determine sentence organization. One cannot simply say *"The verb is the second word in the sentence"*, as this is not always true, for example *"Some professors eat at really fancy restaurants"*. However, it is possible to separate each sentence in "meaningful units", e.g. *"Some professors", "really fancy restaurants", "really fancy", "at really fancy restaurants"*. These meaningful groupings are named *constituents* of a sentence.

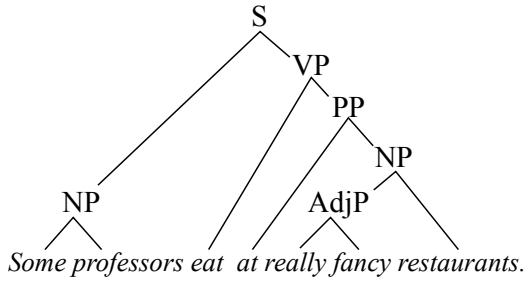

Figure 2.9: Syntactic tree

Note that constituents appear within other constituents. It is then possible to represent this in a bracket form or in a syntactic tree, see 2.9.

There has also to be noted that not all constituents seem of the same type or have the same function in the sentence. Two aspects are referred: their lexical and phrasal categories and their grammatical relations. The constituents *Some professors* and *really fancy restaurants* are both noun phrases even if the first is formed by two nouns and the second by an adverb, an adjective and a noun. However, *at really fancy restaurants* is a prepositional phrase. Furthermore, the constituent *Some professors* functions in this sentence as the subject, while in another sentence it could be the object. Therefore, both the categories and the grammatical relations are important in the grammar definition.

A well-known type of grammars based on all these factors are Context-Free Grammars (CFG), an idea which dates back to Wilhem Wundt (1890), but was not formalized until Chomsky in 1956, and, independently, by Backus in 1959. This type of grammar is also referred to as Phrase Structure Grammar (PSG) or Backus-Naur Form (BNF).

A formal definition of a context-free grammar is as follows:

CFG = < Ter, N, S, R >

- Ter is set of terminals (lexical symbols)
- N is set of nonterminals (phrasal symbols).
- S is start symbol (one of the nonterminals)
- R is a set of rules/productions of the form NT $\rightarrow \sigma$, where NT is a nonterminal and σ is a sequence of terminals and nonterminals.

An example context-free grammar, with S as the start symbol, is as follows:

CFG = < Ter, N, S, R >

- Ter = that, this, a, the, man, book, flight, meal, include, read, does
- N = S, NP, NOM, VP, Det, Noun, Verb, Aux
- R = {

  S → NP VP
  S → Aux NP VP
  S → VP
  NP → Det NOM
  NOM → Noun
  NOM → Noun NOM
  VP → Verb
  VP → Verb NP
  Det → that | this | a | the
  Noun → book | flight | meal | man
  Verb → book | include | read
  Aux → does

  }

A context-free grammar is said to be in Chomsky normal form if all of its production rules are of one of the following forms:

A → BC

A → β;

where A, B and C are nonterminal symbols and β is a terminal symbol. Also, neither B nor C may be the start symbol. Every grammar in Chomsky normal form is context-free, and conversely, every context-free grammar can be transformed into an equivalent one which is in Chomsky normal form.

Probabilistic context-free grammars (PCFG)

A PCFG is a context-free grammar in which every rule is assigned a probability [7]. Probabilities are interpreted as the probability of expanding a constituent, e.g. S, NOM, using a particular rule, as opposed to any of the other rules that could be used to expand this kind of constituent. The probabilities of all the rules for a constituent should sum to one. Once the probability of individual rules is obtained, the probability of the entire parse is calculated by taking the product of the probabilities for each of the rules used therein. Therefore, the probability of a particular parse is calculated as:

$$p(t, \eta) = \prod_{g} p(r(g)); \tag{2.1}$$

where:

- t is the entire structure
- η is a particular parse of t
- g ranges over the constituents of η
- $r(g)$ is the rule one uses to expand g

A PCFG can be obtained by extending a CFG with statistical knowledge. Given a CFG = < Ter, N, S, R > where:

- Ter is set of terminals (lexical symbols)
- N is set of nonterminals (phrasal symbols).
- S is start symbol (one of the nonterminals)
- R is a set of rules/productions of the form NT $\rightarrow \sigma$, where NT is a nonterminal and σ is a sequence of terminals and nonterminals.

For convenience we suppose rules R being placed in an arbitrary order, $r_1, r_2, \ldots, r_n$. A PCFG = <Ter, N, S, R, Θ > also includes a parameter vector $\Theta \in \Re^n$ which assigns a probability to each rule in R. Given an input x, e.g. a sentence from a corpus, and a tree t spanning the input, one assumes a function $f(x, t)$ which tracks the counts of the rules in (x, t). In concrete, the i-th component of $f(x, t)$ is the number of times rule r_i is seen

in (x, t). The PCFG is then trained. Assume there is a training set of trees $\{t_1, t_2, \ldots, t_n\}$.The the maximum-likelihood estimate for the probability of a rule $r : NT \to \sigma$ is $P(NT \to \sigma) = \sum_j c(t_j, NT \to \sigma) / \sum_j c(t_i, NT)$, where $c(t_j, NT)$ is the number of times nonterminal NT is seen in t.

Another way to obtain a PCFG is from scratch using the knowledge from a tree-bank. These PCFGs are called *Tree-bank grammars*. Tree-bank grammars are expected to lead to inaccurate parsings due to the nearly impossibility to have a large enough tree-bank to contain all possible structures. Thus it is highly possible that a new *input structure* requires rules not in the derived grammar. However, even if this happens, if the rules that are not in the tree-bank are really rare, then missing them has little effect on parsing and just creates a parse slightly incorrect. This extra incorrect parses are not statistically significant when added to the higher number of incorrect parses due to grammar ambiguity. Tree-bank grammars for syntactic parsing were evaluated in [7] and proved to be reasonably effective, with results yielding a 75% average precision and recall of the labellings.

Parsers

The task of a parser is essentially to determine if and how the input can be derived from the start symbol of the grammar. There are two possible ways to accomplish this task:

1. **Top-down parsing:** it attempts to find left-most derivations of an input-stream by searching for parse trees using a top-down expansion of the given formal grammar rules. The tokens are consumed from left to right. Inclusive choice is used to accommodate ambiguity by expanding all alternative right-hand-sides of grammar rules. An example of top-down parser is the *LL parser*. A *LL parser* parses the input from *Left to right*, and constructs a *Leftmost* derivation of the input. A leftmost derivation deterministically determines the leftmost nonterminal as the next nonterminal to rewrite.

2. **Bottom-up parsing:** The idea behind this type of parsing is that a parser can start with the input and attempt to rewrite it to the start symbol. The parser attempts to locate the most basic elements, then the elements containing these, and so on. An example of bottom-up parser is the *LR parser*. A *LR parser* reads input from *Left to right*, as it would appear in a visual display, and produces a *Rightmost* derivation. A rightmost derivation deterministically determines the rightmost nonterminal as the next nonterminal to rewrite.

An important type of parsing in NLP is shallow parsing [17]. Shallow (or partial) parsing is the task of recovering only a limited amount of syntactic information from natural language sentences. It has proved to be a useful technology for written and spoken language domains and has many applications in question answering or text mining applications. Shallow parsers use rule sets. However, they need thousands of rules and the rule sets tend to be largely "soft". Therefore, building a shallow parser is a labour intensive task. Shallow parsers are usually automatically built, using machine learning techniques. However, applying machine learning to shallow parsing is not straight forward, because (a) the amount of data to be processed will push systems to the limit, (b) labelled training material is frequently noisy and exists in small quantity, and (c) real world sentences tend to be long so learners which do not operate in (near) linear time are simply unfit for the task.

Most modern parsers are at least partly probabilistic; that is, they rely on a corpus of training data which has already been annotated (parsed by hand), also known as treebank. This approach allows the system to gather information about the frequency with which various constructions occur in specific contexts. Treebanks can be created completely manually, where linguists annotate each sentence with its syntactic structure, or semi-automatically, where a parser assigns some syntactic structure which linguists then check and, if necessary, correct. Treebank creation is a labour intensive project that can take teams of graduate linguists several years.

2.4 Summary

This chapter has, first, defined the type of argumentation that aims to be detected in this thesis, i.e. natural argumentation, defining the main elements of any argumentation, which will be used recurrently in the document. Then, some characteristics of natural legal argumentation have been described. These characteristics will be used to motivate some decisions in future chapters. Then, a study of the special characteristics of legal argumentation, specifically in case law, was introduced. It has been argued that legal cases are in theory one of the best resources to study argumentation. Moreover, a definition of the general structure of a legal case has been given, which will be used in the definition of the annotation framework in Section 5.2. Finally, a set of basic techniques used in information extraction have been introduced. These techniques are the foundations for the solving methodologies on automatic argumentation detection presented in Part II and some of the future approaches that could solve automatic argumentation detection (Chapter 10).

Part I

Study on argumentation annotation

Outline part I

This part of the thesis is focused on the linguistic resources necessary for the automatic detection of argumentation in legal cases. We argue that the existing resources at the start of this thesis were not adequate for our aim. Therefore, it presents a complete anylisis of the characteristics required in the desired corpus. Then, it motivates the chosen resources used to gather the corpus during this thesis and how they have altered the state of the art on argumentation analysis. Moreover, the annotation methodology is studied and the main problems encountered during the annotation process are analysed.

First, it presents, in Chapter 3, the state of the art before this thesis started, from the existing resources to the used methodologies. The resources were very limited and forced us to create our own. Therefore, in Chapter 4 and Chapter 5 a motivation of the selection of documents and annotation methodology is presented. To end this part of the thesis Chapter 6 presents a detailed discussion of all the problems found during the corpus gathering, annotation and use.

The work in this part of the thesis has been partially published in the following articles.

- MOCHALES PALAU, Raquel and MOENS, Marie-Francine (2011) Argumentation Mining. Artificial Intelligence and Law. 19(1), (pp.1-22).
- MOCHALES PALAU, Raquel and IEVEN, Aagje (2009) Creating an Argumentation Corpus: Do Theories Apply to Real Arguments? A Case Study on the Legal Argumentation of the ECHR. In Proceedings of the Twelfth International Conference on Artificial Intelligence and Law (ICAIL). pp. (21-30). New York: ACM.
- REED, Chris, MOCHALES PALAU, Raquel, ROWE, Glenn and MOENS, Marie-Francine (2008) Language Resources for Studying Argument. In Proceedings of Language Resources and Evaluation Conference (LREC). (pp. 91-100) Marrakech (Morocco).
- MOCHALES PALAU, Raquel and MOENS, Marie-Francine (2008) Study on the Structure of Argumentation in Case Law. In Proceedings of JURIX 2008: The 21st International Conference on Legal Knowledge and Information Systems. (pp. 11-20) IOS Press.

Chapter 3

State of the art

Even if the study of argumentation is a well-known topic, the study of argumentation annotation is a quite recent research topic. There are few collections of texts where each sentence is annotated with information about the argumentative role that the sentence plays in the paper. This is due to the limited attempts to automatically process argumentation (either by symbolic or machine learning techniques) which would be the main promoters of this type of collection.

The first attempt to gather a collection of this kind was done in 1999, where a collection of research articles were annotated by their argumentative role, see [60]. The roles were defined based on the sentence functionalities in the research article, e.g. aim, general background, description. This research was replicated in 2004 over a collection of legal documents, the gathered collection contained 188 judgements from the House of Lords [16].

Other more recent attempts have been more focused on dialogical argumentation, specifically on the analysis of meeting discussions, see [45, 46]. This research has provided a corpus of meeting discussions and a coding scheme for their annotation. The coding scheme proposed is aimed at reconstructing the linked sequence of argumentative actions that are common in meeting discussions and that highlight the main effort of participants to solve conflicts of opinion in order to end up with some agreed decisions.

None of the previous collections was adequate for our research, as their annotation schemes were not based on signalling arguments and their relations.

We found one last attempt which provided a collection that adjusted to our needs. This collection was used in the first attempts to argumentation detection. Section 3.1 presents the main characteristics of this collection.

The Question-Answering (Q&A) field has also shown an interest in argumentation annotation. Most of the work has been focused on how to automatically create knowledge repositories. This type of studies are generally domain specific. For example, in [14] a quite detailed analysis of the textual structure of procedural texts, identifying their main basic components as well as their global structure, was developed to automatically construct a textual database of advices and warnings. Note that advices and warnings are specific subtypes of argumentation and procedural texts a type of documents with specific patterns of discourse. These types of studies present research similarities with our work, such as the choosing of an adequate annotation framework depending on the type of argumentation found in the document collection. However, we are not aware of any relevant studies in the legal domain.

Another main obstacle in the field of argumentation annotation is the lack of a shared, agreed notation or "interchange format" for argumentation and arguments. In the last years a number of different argument mark-up languages have been proposed in the context of tools developed for argument visualisation and construction. For example, the Assurance and Safety Case Environment (ASCE), a graphical and narrative authoring tool for developing and managing assurance cases, safety cases and other complex project documentation. ASCE relies on an ontology for arguments about safety based on claims, arguments and evidence [13].

The analysis and study of human argument has also prompted the development of specialised argument mark-up languages and tools. A particularly relevant development in this direction is AML [11], a proposal for an XML mark-up of argumentation started in 1999. The annotation was thought to help the reader or for further processing of the documents. The author proposed a set of markers derived from manual corpus annotation and describes a way to assign them using surface cues and limited syntax for scoping. A subset of the tags collection can be seen in Table 3.1.

These various attempts at providing argument mark-up languages have a major limitation, each particular language is designed for use with a specific tool rather than for facilitating inter-operability of arguments among a variety of tools [9]. As a consequence, the semantics of arguments specified using these languages is tightly coupled with particular schemes to be interpreted in a specific tool and according to a specific underlying theory. In order to address these limitations, a group of researchers, including us, have attended different workshops whose aim was to sketch an Argumentation Interchange Format

Type	Tag	Description
Reasoning and Rhetoric	<JUSTIF>	justification
	<EVID>	evidence
	<BACKGROUND>	background
	<CONTRAST>	contrast
	<CONDITION>	condition or premise
	<FOCUS>	use of focus (e.g. importance of an element in the reasoning)
Modelling	<FRAME_OF_REF>	model
	<SITUATION>	description of current situation
	<BACKGROUND>	background of the situation
	<ACTOR>	person or institution saying or doing something
Narrative	<INTRO>	introduction
	<COMMENT>	author's comment
	<REQUEST>	requesting information
Evaluative	<FALLACY>	fallacy
	<AGREE>	pro

Table 3.1: Subset of AML tags collection extracted from [11]

(AIF) [51, 9]. This format aims to: (a) facilitate the development of systems using a shared formalism; (b) to facilitate data interchange among tools. Even if AIF is one of the most important bases for future research in argumentation annotation, it is still in an early development step. Therefore, the annotation framework used in this thesis was not based on AIF, but was limited to a specific task, i.e. argumentation detection in legal cases. However, the analysis of the framework and document collection presented in this thesis (see Chapters 4 and 5) exposed the need for new characteristics in the AIF.

3.1 The Araucaria corpus

The Araucaria corpus is an annotated document collection constructed by Chris Reed at the University of Dundee, where different legal and non-legal arguments have been annotated and analysed in the scope of the Araucaria project [52]. This is the only collection which contains legal documents and has a coding scheme adequate to the aim of this thesis.

The Araucaria collection comprises two distinct sets of data: a structured set in English collected and analysed according to a specific methodology as a part

of a project at the University of Dundee, and an unstructured multi-lingual set of user-contributed analyses. The data was collected over a six week period in 2003, during which time a weekly regime of data collection scheduled a regular harvest of arguments from the following sources: 19 newspapers (from the UK, US, India, Australia, South Africa, Germany, China, Russia and Israel, in their English editions where appropriate), 4 parliamentary records (in the UK, US and India), 5 court reports (from the UK, US and Canada), 6 magazines (UK, US and India), and 14 further online discussion boards and "cause" sources such as Human Rights Watch and GlobalWarming.org. These sources were selected because they offered (a) long-term online archive of material; (b) free access to archive material; (c) reasonable likelihood of argumentation. Each week, the first argument encountered in each source was identified and analysed by hand [43].

The analysis of the Araucaria collection employed a rigorous process based on scheme-theory analysis using Walton's classification [69]. The annotators were experts on argumentation and had a good knowledge of scheme-based argumentation characteristics and limitations.

The Araucaria collection is an ongoing dataset, it is normal to find the collection altered by the addition of new texts or re-annotation of existing texts. Therefore, during this thesis a stable subset of the Araucaria collection was used as corpus. In this corpus there are 1899 sentences that contain an argument and 827 sentences without arguments, which were used for the experiments. In addition, 1072 new sentences containing no argumentative information were extracted from the same sources as the ones used for the Araucaria project and added to the corpus. The sentences were classified by their text type: newspapers, parliamentary records, legal judgements, weekly magazines, discussion fora, cause information sources and speeches. A sub-corpus for each text type was also built by picking sentences as to have a maximum balanced set of positive and negative examples. Table 3.2 presents some examples of argumentative and non-argumentative sentences of the Araucaria corpus. Note that some of the sentences do not seem clear examples of an argument. This is mainly because the sentences are out of context and it is difficult to think in which context they would behave as argumentative or not. However, they were annotated as such by the legal experts involved in the Araucaria project.

Text type	Argumentative	Non-argumentative
Discussion fora	"On this occasion, however, I shall not vote for any individual or party but will spoil my paper."	"I have been voting since 1964 and at one time worked for my chosen party."
Legal judgments	"He is aware of the risks involved, and he should bear the risks."	"Let there be any misunderstanding one point should be clarified at the outset."
Newspapers	"Labor no longer needs the Liberals in the Upper House."	"The independents were a valuable sounding board for Labor's reform plans."
Parliamentary records	"I have accordingly disallowed the notice of question of privilege."	"Copies of the comments of the Ministers have already been made available to Dr. Raghuvansh Prasad Singh."
Weekly magazines	"But for anyone who visits Rajasthan's Baran district, the apathy of the district administration and the failure of the Public Distribution System (pds) is clear to see."	"This time in Rajasthan."

Table 3.2: Examples of argumentative (sentence is part of an argument) and non-argumentative (sentence is not part of an argument) sentences and their source text type from the Araucaria corpus from [43]

3.2 Discussion

The Araucaria corpus allowed an initial study of argumentation in different fields and a first overview of the main characteristics of legal argumentation. However, it also showed limitations for our aim, mostly due to its gathering process and annotation methodology.

First, the Araucaria project was started as a means to facilitate the human analysis of argumentation, where humans are able to assume missing information due to their world knowledge. Therefore, it decided to focus on analysing small pieces of text out of context as long as these pieces were

complete arguments. Taking this into account, the gathering process consisted of extracting complete argumentative pieces, losing structural, linguistic and reasoning links between the selected piece and the rest of its source. Furthermore, the source of the piece was not saved, only a reference to its type and date. When the aim is the automatic detection of argumentation based mainly on structural, linguistic and reasoning features the lack of information about the source and the links between the piece and the source is a major limitation. Computers do not posses world knowledge and even if it is possible to introduce some knowledge on a system, it is difficult with the current state-of-the-art technology to obtain results similar to those of humans when working with a variety of topics.

Second, the Araucaria project did not present a set of guidelines for their annotators. The annotators had a good background on argumentation and they were instructed to annotate the text following a concrete theory, which restricted their annotation choices. However, there were no linguistic or structural restrictions. For example, there was not a rule that indicated if linguistic connectors were to be introduced as part of the annotated information. Figure 3.1 shows two possible analysis for the following argument:

"*We can create all the legal issues, laws and strict gun control we want. But if someone wants to kill, they could kill someone with a pen, pencil, knife, baseball bat or even a sling shot. So are we going to make it illegal for everything we can think of that might kill.*"

It is possible to observe that the first analysis of the argument does not include rhetorical markers, such as *"so"*, or punctuation marks, such as *"."*. The second analysis is more complete and includes all the information. Even if these differences between the analyses have no impact on the argument understanding, it is obvious that the first analysis contains less information about the linguistic structure of the argument than the second analysis.

A strict annotation framework has a great impact when dealing with a computational system that must learn patterns from the annotations. The smallest change on the annotation scheme, even if it is not directly related to the main annotation elements, can alter the patterns that are being learned by the system and create non-realistic outcomes. In the previous example (Figure 3.1), the loss of information about the linguistic structure of the argument could alter learned patterns. For example, *"a premise cannot start with a rhetorical marker so"*, if the information on the second analysis is taken into account, should be updated to *"a premise CAN start with a rhetorical marker so"*.

In conclusion, there are not many resources where argumentation is annotated and the available resources, such as Araucaria, do not contain enough complete

and reliable content to properly support the type of automatic argumentation detection desired in this thesis.

3.3 Summary

This chapter has discussed the state of the art on argumentation corpora in the period before the start of this thesis. It has described the main developments in the field during this research. It has shown that there exist limited resources which contain annotations of argumentation. Furthermore, most of these resources present annotation frameworks which do not adapt to the requirements of the current task.

Later on, it has described a specific resource, a data set from the Araucaria project. This resource has been analysed in detail because its annotations are the closest to this thesis requirements. A part of the Araucaria annotations have been used in the early experiments on automatic argumentation detection in Chapter 8. However, for the purposes of this thesis, there are fundamental problems with this data set such as its inability to retain context information or the non-restrictive linguistic guidelines. This partly motivates some decisions taken when gathering the corpus (see Chapter 4) and defining the annotation framework (see Chapter 5).

Analysis 1
Conclusion: Identifier: C : We should not impose gun control
Scheme: Argument from the Constitution of Negative Normative Facts
Support:

- Identifier: B
  - are we going to make it illegal for everything we can think of that might kill.
  - *Scheme: Argument from the Constitution of Negative Normative Facts*
- Identifier: A
  - We can create all the legal issues, laws and strict gun control we want. But if someone wants to kill, they could kill someone with a pen, pencil, knife, baseball bat or even a sling shot
  - *Scheme: Argument from the Constitution of Negative Normative Facts*
- Identifier: E (enthymeme)
  - If gun control alone does not ensure that people are not able to kill and the additional steps needed in order to do so are not going to be taken, we should not impose gun control.
  - *Scheme: Argument from the Constitution of Negative Normative Facts*

Analysis 2
Conclusion: Identifier: C : We should not impose gun control
Scheme: Argument from the Constitution of Negative Normative Facts
Support:

- Identifier: B
  - **So** are we going to make it illegal for everything we can think of that might kill.
  - *Scheme: Argument from the Constitution of Negative Normative Facts*
- Identifier: A
  - We can create all the legal issues, laws and strict gun control we want. But if someone wants to kill, they could kill someone with a pen, pencil, knife, baseball bat or even a sling shot**.**
  - *Scheme: Argument from the Constitution of Negative Normative Facts*
- Identifier: E (enthymeme)
  - If gun control alone does not ensure that people are not able to kill and the additional steps needed in order to do so are not going to be taken, we should not impose gun control.
  - *Scheme: Argument from the Constitution of Negative Normative Facts*

Figure 3.1: Two different annotations for an argument in the Araucaria corpus. In *Analysis 2* the *Identifier:B* contains '**so**' and *Identifier:A* contains '**.**', while *Analysis 1* contains neither feature.

Chapter 4

Gathering an argumentation corpus

The gathering of a corpus adequate for automatic detection of argumentation in legal cases has some basic requirements. First, the source or sources from where the documents are gathered should be reliable and stable and the gathered documents should be representative of the legal cases domain. The selected set of documents to form the corpus were extracted from a unique source, the European Court of Human Rights (ECHR), a human rights documentation (HUDOC)[1] database of judgements and decisions, which is available on CD-ROM and online. Figure 4.1 shows an example of ECHR argumentation. Section 4.1 justifies the reliability and adequacy of this source and its documents.

Second, in this thesis the automatic detection of argumentation aims to be accomplished mainly using the structural, linguistic and reasoning information encountered in each document of the corpus. Therefore, to preserve all this information each document of the corpus should contain a full texts, not just pieces.

Third, the gathered documents should present similar structures and linguistic and reasoning patterns without losing the general overview of the domain. In this way, it is possible to extract rules of behaviour that represent the domain. Section 4.2 explore the different characteristics of the gathered documents to

[1] http://www.echr.coe.int/ECHR/EN/Header/Case-Law/HUDOC/HUDOC+database/

prove their generality and representativity.

Fourth, the gathered documents should provide a consistent and complete set for the annotation framework. Each argument encountered in the documents should have a representation in the framework. This issue is discussed in Chapter 5.

{ (*PREMISE*: The Court recalls that the rule of exhaustion of domestic remedies referred to in Article x of the Convention art. x obliges those seeking to bring their case against the State before an international judicial or arbitral organ to use first the remedies provided by the national legal system.
CONCLUSION: Consequently, States are dispensed from answering before an international body for their acts before they have had an opportunity to put matters right through their own legal systems.)

(*PREMISE*: The Court considers that, even if it were accepted that the applicant made no complaint to the public prosecutor of ill-treatment in police custody, the injuries he had sustained must have been clearly visible during their meeting.
PREMISE: However, the prosecutor chose to make no enquiry as to the nature, extent and cause of these injuries, despite the fact that in Turkish law he was under a duty to investigate see paragraph above.
PREMISE: It must be recalled that this omission on the part of the prosecutor took place after Mr Aksoy had been detained in police custody for at least fourteen days without access to legal or medical assistance or support.
PREMISE: During this time he had sustained severe injuries requiring hospital treatment see paragraph above.
CONCLUSION: These circumstances alone would have given him cause to feel vulnerable, powerless and apprehensive of the representatives of the State.)

CONCLUSION: The Court therefore concludes that there existed special circumstances which absolved the applicant from his obligation to exhaust domestic remedies. }

Figure 4.1: A fragment of the argumentation of an ECHR legal case. The fragment contains two sub-arguments.

4.1 The source

The ECHR is the court watching over the application of and compliance with the European Convention on Human Rights, a treaty to which 47 Council of Europe (COE) member states, including the Russian Federation and Turkey, are parties. As such it has jurisdiction over all these states in matters regarding

human rights. It is the highest court on these matters and its judgements must be implemented in all member states. For this reason alone, automated annotation of a corpus of ECHR could have a wide field of application. The topics of the decisions cover all human rights articles, e.g. freedom of expression, suspicious deaths, torture, privacy or fair trial.

Both states and individuals can file complaints, and thus act as plaintiffs or "applicants" in the Court's jargon, but only states can be tried as defendants, and are referred to as "the government". Previously there was also a Commission which decided on admissibility of cases before the Court could decide on the merits, but since 1998 the Court has taken over the previous commission's duties. However, some of the cases before that date have also been selected, to have a better overview of the problems. This, of course, leads to a more difficult automatic classification, but makes the corpus more representative.

An important factor in choosing this court was that the ECHR, in the years since its instalment several decades ago, has developed its own patterns of reasoning, using specific structures and types of argumentation. Because of the direct applicability of its judgements in all COE member states, many of these patterns have been taken over by judges and courts at the national level. These patterns took some time to develop but occur in most documents from after 1985, which is the large majority.

In conclusion: (a) the ECHR covers the whole of Europe in an important field, i.e. human rights, (b) the European Court of Justice (ECJ) has been known to use some of the same patterns of reasoning and has taken over some others from the ECHR, (c) there is also an ongoing process to make the European Union a party to the Convention which will make this corpus all the more significant, (d) the Inter-American and African regional human rights treaties have installed courts very similar in structure to the ECHR, using some of the same phrasing and reasoning patterns as the ECHR uses, and (e) several constitutional courts use discursive expressions similar to those used by the ECHR. Therefore, the ECHR is a reliable and stable source, and its documents have general reasoning patterns which are representative of the legal domain. The next sections show these reasoning patterns and prove that the documents also posses clear linguistic and structural characteristics, which can be exploited for the automatic detection of argumentation.

4.2 The legal cases

The corpus contains both admissibility decisions and judgments on the merits, from both the ECHR commission and the ECHR court, so that it allows to work with all types of ECHR cases. In this section the main characteristics of these documents are presented. Therefore, an analysis of the structural differences, reasoning patterns and linguistic features that influence the annotation process are shown.

<table>
<tr><th>Option A</th><th>Option B</th><th>Option C</th><th>Option D</th><th>Option E</th><th>Option F</th></tr>
<tr><td rowspan="2">The Court's assesment of the evidence and establishment of the facts</td><td rowspan="2">General Approach</td><td rowspan="2">Preliminary observation</td><td rowspan="2">-</td><td rowspan="2">Scope of the case</td><td>The Court's assessment of the facts</td></tr>
<tr><td>The Government's preliminary objection</td></tr>
<tr><td colspan="5" rowspan="2">Alleged violation of article X</td><td>The merits</td></tr>
<tr><td>Alleged violation of article X</td></tr>
<tr><td colspan="6">Application of article X</td></tr>
</table>

Table 4.1: Structures of *The Law* section in the ECHR

4.2.1 Document structure

The structure of admissibility reports and legal cases reports is similar in its basis and it pre-defines where the argumentation of the decision can be found. A typical admissibility report consists of a document with the following sections, where sections 3 and 4 can be omitted in some cases, while sections 1, 2 and 5 are always present:

1. *Introduction*: Different elements of the case are presented, e.g. application number, plaintiff, defendant and the responsible ECHR members. Furthermore, some details on previous submissions and the related article of the "Convention for the Protection of Human Rights and Fundamental Freedoms" are given.

2. *The Facts*: This section presents a summary of the main facts of the case. All the facts are presented in past tense in chronological order.

3. *Proceedings before the Commission:* This section contains previous decisions of this court on the current case and previous applicant steps in front of this court. There are no justifying reasons for each court decision and all the steps are presented in chronological order.

4. *Complaints*: The applicant's complaints are presented in this section. More than one complaint can be presented in front of the court.

5. *The Law*: (sometimes called: *Reasons for the Decision*) For each complaint presented in the previous section the reasons exposed by the applicant are presented, together with all the court's deductive procedure to arrive to a final admissibility conclusion. Sometimes in one admissibility report some complaints will be declared admissible while others inadmissible.

On the other hand, a typical report of a legal case consists of a document with the following sections, where sections 3, 4 and 6 can be omitted, while sections 1, 2 and 5 are always present:

1. *Introduction*: As in the admissibility reports different elements of the case are presented, e.g. the application number, plaintiff, defendant and the responsible members of the ECHR. Also some details on previous submissions and the related article are given. However, in this case a complete list of all the persons present in the court is given, together with their role, e.g. agent, adviser or counsel. Furthermore, the procedure, i.e. a chronological summary of the main steps followed previously in this case, is also presented.

2. *The Facts*: (sometimes called: *As to the Facts*) This section presents the circumstances of the case in chronological order. Then some relevant articles to the case, i.e. domestic law, administrative law remedies, civil proceedings and others, are mentioned.

3. *Proceedings before the Commission*: As in the admissibility reports this section contains previous decisions of this court on the current case and the previous steps of the applicant in front of this court.

4. *Final Submission to the Court*: This section presents the complaints that are analysed in this specific process.

5. *The Law*: (sometimes called: *As to the Law*) As in the admissibility reports, for each complaint presented in the previous section the reasons exposed by the applicant are presented, together with all the court's deductive procedures to arrive to a final conclusion. Sometimes in one case some complaints will be dismissed while others accepted. Table 4.1 shows some possible structures of this section.

6. *Partly Dissenting Opinion of Judge X*: In this section a judge, or a group of judges, can clarify his or her opinion and discrepancies from the final decision. There can be more than one dissenting opinion, in this case they are presented one after the other separated in different sections.

Apart from these typical sections, the *registar* of the case can introduce some notes during the writing of the case. These notes can appear in any section but they are always marked with a specific and identifiable title.

The fix structure of the documents allows to narrow the search of argumentation in sections *The Law* or *Partly Dissenting Opinion of Judge X*. Any other argumentation found in other sections would be reported argumentation, see Section 4.2.4 for more details about reported argumentation. Moreover, in Section 2.2.1 the typical structure of a legal case has been presented and it is possible to observe that the structure of the ECHR documents follows the same patterns. Therefore, a ECHR document can be considered a representative form of legal case.

4.2.2 Reasoning structure

The ECHR uses certain specific patterns of argumentation that can be reconstructed using argumentation schemes theory, see Section 2.1.2 for the theory explanation and Appendix B for the definitions of the various schemes.

First, an applicant always presents a complaint about the violation of a specific article of law from the Convention. This prompts the court to use the scheme of *Argument from an Established Rule* to prove or disprove the validity of the complaint. Normally this is the "top" argument of a chain of arguments that extends from the beginning till the end of the legal case.

Second, given that the ECHR is a court which relies on its previous decisions in so-called precedent cases for the interpretation of the Convention, it is also quite common to see the court use the scheme of *Argument from Precedent* to justify a decision. This occurs often in combination with the *Argument from an Established Rule* as the precedent establishes the rule.

The Court recalls that it falls to each Contracting State, with its responsibility for "the life of (its) nation", to determine whether that life is threatened by a "public emergency" and, if so, how far it is necessary to go in attempting to overcome the emergency. By reason of their direct and continuous contact with the pressing needs of the moment, the national authorities are in principle better placed than the international judge to decide both on the presence of such an emergency and on the nature and scope of the derogations necessary to avert it. Accordingly, in this matter a wide margin of appreciation should be left to the national authorities.

Figure 4.2: Example of an argument from position to know in the ECHR

Third, when applying an article it is not always immediately clear which is its scope and whether it applies to the facts currently under review. Therefore, the court presents a recurrent use of the scheme *Argument of Verbal Classification*, again often in combination with the *Argument from an Established Rule*.

Fourth, the *Argument from Position to Know* is regularly used by the ECHR not in cases of witness testimony or expert opinion, or even of the jury's authority, but in a pattern of argumentation that is specific and typical to the ECHR and which has attracted wide attention and sometimes following, for example, from other international human rights courts, from federal constitutional courts, and from those who study them - like commentators of the US Supreme Court. This pattern of argumentation is called the "margin of appreciation doctrine" and embodies the principle that there can be a certain moral or cultural relativism in the application of human rights. The pattern then always includes the Court's assertion that "local authorities are, in principle, better placed" to assess what is acceptable to the society in question and what is the best way to implement certain human rights in the particular circumstances of the country and culture, see Figure 4.2.

In conclusion, the ECHR documents can be analysed using argumentation theories, in this case argumentation schemes. Therefore, these documents are adequate to perform formal annotations of argumentation.

4.2.3 Linguistic structure

The argumentative structure of the ECHR's arguments has its ending in the court's final decision. Therefore, it can be said that the whole argumentation of the case branches from the final decision. This final decision of the court can be

easily identified by its linguistic characteristics: "*For these reasons, the Court*".[2] Following this phrase the different parts of the decision are presented. There are three possible linguistic formats, where the text between square brackets is optional:

- *holds* (*by X votes to Y/unanimously*)
- *dismisses* (*by X votes to Y/unanimously*)
- (*by a majority/unanimously*) *declares* (*the application*) (*in*)*admissible*

Conclusions	(*A*) *the Court* (*B*) *"Verb-Conclusion"* (*C*) (*A*) *the Court "Verb-Aux-Conclusion"* (*NOT*) (*B*) *"Verb-Conclusion"* (*A*) *the Court has* (*B*) *no difficulty in accepting* *The Court* (*B*) *proposes to proceed* *The Court has found it established* (*A*) *there has been a violation of* *It* (*B*) *follows that* *It must be* (*B*) *declared admissible / inadmissible* *It is* (*B*) *inappropriate to* *There has* (*B*) *been a breach of* *This aspect / part of the case / application "Verb-Aux-Conclusion"* (*B*) *be rejected / accepted as*
Premises	(*A2*) *the Court* (*B*) *"Verb-Premise"* (*C*) (*A2*) *the Court "Verb-aux-Premise"* (*NOT*) (*B*) *"Verb-Premise"* (*C*) *The Court sees no reason to "Verb-Premise"*

Table 4.2: Typical expressions in the ECHR corpus

The different interrelated arguments that justify the final decision also have some linguistic characteristics that can help identifying parts of their hidden structure. Table 4.2 shows some examples of phrasal linguistic structures that are common in ECHR argumentation. On these structures *"the Court"* can be replaced by *"it"* or *"the Commission"* and any text between square brackets is optional. Other linguistic traits that are relevant when dealing with argumentation are linguistic keywords. These are some words or groups of words that help to linguistically connect the different parts of the

[2]In the admissibility reports *the Court* is named *the Commission*

argumentation, see Table 4.3. Note that these expressions are not exclusive for argumentative statements, they are just expressions which introduce an argumentative statement in some ECHR documents.

A	*In the present case, / For all these reasons, / Accordingly, / In the light of the foregoing / In the light of the parties' submissions, / Having reached this conclusion, / In these circumstances, / In short, / In conclusion, / Like the xxxx, / In contrast, however,/ Consequently / Like the Delegate of the xxxx, / In such a situation*
A2	*Although / In the present case, / In such a situation / In contrast, however, / Indeed, / On the other hand, / In particular / Thus, / in this respect / Since, / However, / Furthermore, / As regards / Moreover, / However, again, / In addition, / Like the Delegate of the xxx,*
B	*, like the xxxx and the xxxx, / , like the xxxx, / therefore / , firstly / accordingly / clearly / also / further*
C	*, in the light of the parties' submissions / , in the light of all the material before it*

Table 4.3: Typical keywords and keyphrases in the ECHR corpus

Verb-Conclusion	find / conclude / decide / accept / reject / dismiss
Verb-Premise	consider / note / recall / agree / disagree / reiterate / acknowledge / is of the opinion / point out / emphasise / stress / is of the view / is satisfied
Verb-Aux-Conclusion	must / cannot / can / do
Verb-Aux-Premise	would

Table 4.4: Typical verbs in the ECHR corpus

The statements in legal cases are a complex linguistic set. If one observes the statements in Figure 4.2 it is possible to see some of the linguistic cues, such as *"The Court recalls"* or *"Accordingly,"*, in the first and third statement, but the second statement lacks any linguistic clue which might highlight its functionality. Moreover, one can observe that the first and second statements are rather long and complex, with various degrees of subordination and coordination. The syntactical analysis of both statements suffers from both

ambiguity and structural complexity. A more close study of the words used in each statement does not show a high degree of domain specific words or archaic expressions. However, the statements are not easily comprehensible. This type of long complex statements are the norm in legal cases. Figure 4.3 and Figure 4.4 present some examples of premises and conclusions as encountered in the ECHR cases.

The ECHR documents share enough linguistic characteristics to allow the extraction of some feature patterns for the automatic analysis of argumentation. Therefore, these documents are an adequate resource to achieve the aim of this thesis.

There is no doubt that these acts, in addition to giving rise to a violation of Article 3, constituted particularly grave and unjustified interferences with the applicants' rights to respect for their private and family life and home, and to the peaceful enjoyment of their possessions (see SelÃğuk and Asker, cited above, Âğ 86).

The Court notes at the outset that it appears from the material before it that the documents relating to the evaluation referred to by the Government were not appended to their letter of 17 May 1999 to the Commission, as was pointed out to them by the Secretary to the Commission on 31 May 1999.

It is furthermore established that the burden of proving the existence of available and sufficient domestic remedies lies upon the State invoking the rule (cf. Eur. Court. H.R., Deweer judgment of 27 February 1980, Series A no. 35, p. 15, para. 26, and No. 9013/80, Dec. 11.12.82, D.R. 30 p. 96, at p. 102).

The Commission recalls that Article 26 (Art. 26) of the Convention only requires the exhaustion of such remedies which relate to the breaches of the Convention alleged and at the same time can provide effective and sufficient redress.

Figure 4.3: Examples of premises as encountered in ECHR cases

4.2.4 Reported arguments vs. non-reported arguments

An important fact when analysing the ECHR argumentation linguistic characteristics is the difference between reported and non-reported arguments. In both, admissibility reports and case reports, arguments can come from

It considers that the applicant's complaints raise serious issues of fact and law under the Convention, the determination of which should depend on an examination of the merits.

It follows that the application cannot be dismissed as manifestly ill-founded. No other ground for declaring it inadmissible has been established.

In these circumstances, the Commission finds that the application cannot be declared inadmissible for non-exhaustion of domestic remedies.

The applicants' complaints in this regard are therefore arguable for the purposes of Article 13

The Court concludes that there has been a breach of Article 13 of the Convention.

Figure 4.4: Examples of conclusions as encountered in ECHR cases

the applicant (or plaintiff), the defendant or the court. The arguments from the applicant and the defendant are reported arguments, as they have been previously filed to the court. These reported arguments are written in the past and can be found in the sections *The Law* and *Partly Dissenting Opinion* together with the rest of argumentation from the court. These arguments are seen as facts by some lawyers given that they reflect past decisions and cannot be attacked in the current process. On the other hand, arguments analysed by the court are the ones that defend and attack the current decision. Therefore, they are part of the current process and they are written in present tense. These arguments are named non-reported arguments and are found in the section *The Law*. Non-reported arguments can also be found in the section *Partly Dissenting Opinion* as part of the argumentative opinion of a single dissenting court member. In conclusion, the arguments from the court, i.e. non-reported arguments, are the core of the argumentative process from the ECHR legal cases and the most important arguments to detect.

4.2.5 Argument from precedent in the ECHR

At some point during the annotation process, defined in Chapter 5, the annotators detected a specific reasoning pattern used recurrently in the ECHR argumentation, a related form of *Argument from Precedent*. This form of *Argument from Precedent* is always presented in a single sentence.

An example of this specific form can be seen in Figure 4.5, where it is obvious that a previous case (*"Eur. Court H.R., De Jong, Baljet and Van Den Brink judgment of 22 May 1984, Series A no. 77, p. 18, para. 36, 11.5.89, D.R. 61, pp. 250, 262"*) is used as a precedent to be able to assert the beginning of the sentence (*"It is furthermore established that the burden of proving the existence of available and sufficient domestic remedies lies upon the State invoking the rule"*). The connection between conclusion and premise is represented by *"cf."*. Other connectors found in a similar way in the ECHR corpus are *"see", "see eg.", and "see also"*.

It was decided that this type of argument was of great importance for the understanding of the reasoning structure of a legal case. Therefore, the ECHR corpus was modified by an automatic process to separate these sentences in two parts, premise and conclusion. The connectors were saved in the second part of the sentence, i.e. the premise.

It is furthermore established that the burden of proving the existence of available and sufficient domestic remedies lies upon the State invoking the rule (cf. Eur. Court H.R., De Jong, Baljet and Van Den Brink judgment of 22 May 1984, Series A no. 77, p. 18, para. 36, 11.5.89, D.R. 61, pp. 250, 262).

Premise: (cf. Eur. Court H.R., De Jong, Baljet and Van Den Brink judgment of 22 May 1984, Series A no. 77, p. 18, para. 36, 11.5.89, D.R. 61, pp. 250, 262).

Conclusion: It is furthermore established that the burden of proving the existence of available and sufficient domestic remedies lies upon the State invoking the rule.

Figure 4.5: Argument from Precedent using "cf." format

4.3 ECHR corpus in numbers

To form the ECHR corpus 54 documents composed of 25 legal cases and 29 admissibility reports were randomly selected (see Appendix C). These

documents contain between 40 and 300 sentences each, having an average size of 145 sentences per document. Each sentence contains an average of 49 tokens, that makes a final total average of more than 6.900 tokens per document. We would like to remark that an average of almost 50 tokens per sentence represents a corpus with quite long sentences, and this is in accordance with general knowledge about legal texts that characterize these texts as containing long and complex sentences. There are a total of 12.904 sentences, 10.133 non-argumentative and 2.771 argumentative, 2.355 premises and 416 conclusions. The presence of non-argumentative sentences is considerably higher due to the fact that even in legal texts the presentation of facts, explanations and summaries is more frequent than the argumentation itself. It is also clear that the amount of premises is much higher than the conclusions. This is due to the fact that for arriving to a conclusion first it is needed to explain the reasons for it, therefore there will be some explanations, in favour or against, while only one conclusion will be reached.

4.4 Summary

This chapter has first presented the requirements of an argumentation corpus for automatic argumentation detection in a legal case. Then, a corpus has been selected to fulfil all the requirements. First, it has been motivated the source choice, presenting its main characteristics and working procedures, arguing that the source is reliable and contains legal cases that can be considered standard in the case law domain. Then, the main characteristics of the corpus documents have been analysed, from their discourse structure to their main reasoning patterns. The important linguistic cues that will be later used as features in the experiments in Part II have been described, followed by two discussions over specific characteristics of the ECHR argumentation that influence the corpus structure and annotation. To conclude the chapter, a description of the numerical characteristics of the corpus has been reported.

Chapter 5

Annotating the argumentation corpus

In the previous chapter the documents that form the corpus have been selected according to the task requirements. In this chapter the annotation of the selected documents is discussed. Section 5.1 studies which elements should be annotated to achieve an annotated corpus suitable as training resource for automatic argumentation detection. Then, in Section 5.2 an annotation framework to achieve high performance annotations is designed. Section 5.3 details the process the selected annotators followed during the annotation of the corpus.

5.1 Annotation requirements

After the initial experiments with the Araucaria corpus discussed in Chapter 3 it was established that contextual information is an important requirement to perform automatic argumentation detection, see Chapter 8. The nature of the intended IE methods also requires to preserve as much structural and linguistic information of the source documents. Moreover, the selected documents characteristics also influence which elements should be annotated in each case. For example, the specific features of the sentences used to express the final decision of a legal case make it recommendable to treat these sentences as a separated argumentative class in the argumentative process.

Taking into account these requirements, a suitable annotation should annotate all the sentences in each document and for each sentence it should be necessary to determine if it is part of the argumentative process or not. If the sentence is part of the argumentative process then, it has to be decided if it is part of the final decision or the set of arguments to achieve the final decision. If it is part of an argument its function in the argument has to be determined. Moreover, if two or more arguments are related it should be indicated.

Given the nature of the selected documents and the annotation requirements it is expected the resulting annotations will be tree-like structures with a final decision as root and different arguments hanging from it. Each argument might contain other sub-arguments. At this point, it is unclear how an argument and its relations should be represented in the annotation. The next section discusses a framework to define this and other annotation issues.

5.2 Annotation framework

In general, designing an annotation framework is not a straightforward process. The annotation framework has to be (a) predictive and informative, so that it will prove useful for the end task and (b) intuitive, or at least learnable, such that it can be applied consistently both by different annotators and over time [59]. If an annotation framework is simple and intuitive and the task well-described, it will result in high consistency, but there is a danger that the information contained in it might not be informative enough for the given task. On the other hand, if the annotations are too informative their definition is necessarily vague, leaving a lot of leeway for subjective interpretation. In this case, it is likely that different annotators will disagree in their judgements. The process of finding a workable annotation framework is thus a tight rope act between the conflicting requirements of informativeness and consistency.

In the previous section, the requirements of the current annotation framework have been outlined. Taking into account those requirements the desired framework has to deal with three aspects:

- Detection and classification of arguments (5.2.1)
- Classification of elements of an argument (5.2.2)
- Relations between arguments (5.2.3)

The ambiguous nature of argumentation gave rise to the option to adopt a multiple annotation framework. This type of framework allows an annotator

to give more than one output annotation for an input. However, allowing for multiple annotations led to so many sentences with multiple annotated outputs that the informativeness of the framework contained in multiple annotation was unacceptable. Therefore, this option was dismissed in favour of a single annotation framework.

After reading the guidelines, the annotators marked up the first two training cases, followed by a discussion, then the other two training cases, followed by another discussion. In these discussions, disagreements in the annotators' judgements were settled and unclear passages in the instructions changed.

5.2.1 Detection and classification of arguments

In Chapter 2 the concept of defeasible argumentation has been introduced together with the theory of argumentation schemes. Argumentation schemes theory is part of the theoretical background of this framework. This theory was selected given its current impact on argumentation theory and annotator-friendly characteristics. Argumentation schemes present non-abstract definitions of argumentative patterns which seem to easily transform in categories of an informative and intuitive framework. However, only extrinsic evaluation of other frameworks, e.g. based on logical argumentation, can prove or disprove if this is valid.

The first approach to an annotation framework contained 25 categories (see Appendix B) defined directly by the argumentative schemes classification given in [69]. Three annotators worked independently on the definition phase. The end result of this phase was a written collection of guidelines detailing criteria for the choosing of each category. As is typical for high-level, information-rich classification tasks, the annotation framework had to be changed repeatedly during this time.

Settling on an exhaustive list of schemes on which annotators agreed on proved very difficult. After studying some legal cases it was decided that the first 25 categories will not manage to capture all the arguments presented in a legal case. Therefore, some subtypes, which were mentioned in a previous work of [71] but not added to the final list of this annotation, see [69], were included. The following schemes were added: (a) the *Argument from Falsification of a Hypothesis* in addition to the *Argument from Evidence to a Hypothesis*, (b) the *Argument from Popular Practice* in addition to the *Argument from Popularity*, (c) the *Negative Ethotic Argument* in addition to the (positive) *Ethotic Argument*, (d) the *Argument from an Exceptional Case* in addition to the *Argument from an Established Rule* and (e) the *Arguments from Ignorance*.

Therefore, at this stage, the framework contained 30 different categories or argumentation schemes.

Another decision was to leave out the list all of the *Slippery Slope Argument* instances (causal, precedent, verbal and full), because, as Walton already noted in [69], they are all combinations of the *Argument from Gradualism* with other types of argument (*Argument from Verbal Classification*, *Argument from Precedent*, *Argument from Popularity*, etc.). This was done to prevent conflicts between annotations, where one annotator might choose a *Slippery Slope Argument* while the other might choose a combination of different arguments. Therefore, the framework was reduced to 26 different categories or argumentation schemes.

Note that in later studies Walton approached the schemes in the specific context of legal argumentation [70]. There, Walton asserts that the *Argument from Position to Know* includes often as subtypes the *Argument from Witness Testimony* and the *Argument from Expert Opinion*, among others. These types of arguments are common in legal argumentation and it seems they should also be added to our list. However, it was decided not to add them but to include guidelines for the annotators specifying these as subtypes of the argument from position to know. This was done to keep the classification scheme general enough for application to ordinary language argumentation, outside the legal field.

Walton also stated in [70] that the argument from precedent is a subtype of the argument from analogy. However, the *Argument from Precedent* was not removed from the list of arguments. The representations of *Argument from Analogy* and *Argument from Precedent* encountered in the corpus were clearly distinctive, so the sub-classification did not present a problem to the annotators.

After all these considerations, a second attempt at an annotation framework was obtained. It consisted of 26 categories (Table 5.1). A pilot study with the three, by now task-trained annotators, was ran and proved that the new framework is very reliable. It achieved a respectable stability when re-annotating parts of the corpus. Nevertheless, fundamental problems with the type of annotation were noticed. It was observed that even if within the mind of the annotators, private understandings of the categories were rather consistent, some categories were more difficult to differentiate. At that point in time, it was believed these categories could be wrapped up in a bigger common category, but there was not enough evidence to assume this was really the case, or if it would improve the current results. An example of such "wrapping" would be to group *Argumentation from Sign* and *Argument from Evidence to Hypothesis* both as forms of abductive reasoning. These

observations were later confirmed by a study presented by Walton [71], where the supposed subcategories were established as such and the need for a more structural categorization in argumentation schemes was confirmed. The use of this hierarchical classification could be beneficial for the annotators. They could choose a "general" category of argument and then move into more specific subcategories allowing a study of annotators agreement per level. However, this thesis has not studied the development and impact that these changes would have in the current results.

Another problem encountered during the design of the framework was the observation of a low agreement between annotators when dealing with reported arguments. Some annotators will count these reported arguments as argumentation, others do not. This is due to their different background conceptions about the ontology of the law. The annotators which identified reported arguments were either incapable of distinguishing between the two types of argumentation, i.e. reported and non-reported, or understood the two types of argumentation to be all one unique argumentation. On the other hand, the annotators who could not identify reported arguments, understood argumentation to be only the current argumentation by the Court itself, i.e. non-reported. This problem was solved by adding a new element in our framework that indicates when the argument is reported. This allows to calculate the agreement without the reported arguments, while keeping their analysis and detection as part of the collection.

Pilot studies with the annotators with the final framework showed that they were much more comfortable and accurate when applying this framework to real texts. This is the framework that will be used for the extensive annotation of our corpus in Section 5.3 and for the prototypical implementation reported in Part II.

Note that the selected argumentation schemes are the ones we expect to find in the ECHR legal cases. However, the experimental study cannot assure the completeness or correctness of this list due to the limited number of involved cases. Still, it gives a good indication of which schemes are generally used in the ECHR legal cases. Table 5.2 presents one example of the most common schemes as found in the ECHR corpus.

5.2.2 Classification of elements of an argument

In the previous section the categories of argument that are included in the framework have been defined using argumentation schemes. Each detected argument has, at this point, been assigned a category and a reported/non-reported status. The next step is to break the argument in its elements, e.g. premise, conclusion, etc. Even if the annotators had a clear understanding of the argumentation categories, it was noticed they had difficulties when associating textual units (i.e. sentences) with the different elements of the argument category, i.e. its premises, conclusion and critical questions.

The treatment of critical questions in the annotation framework was one of the main discussion points given its importance on argumentation schemes theory. Critical questions are important tools for the evaluation of argumentation when working with argumentation schemes. In the ECHR corpus they figure as premises, either implicit or explicit, in the argumentation of both parties, either in anticipation of critical questions from the counter-party (this is usually the case for the applicant's argumentation) or in answer to the critical questions from the counter-party (which is usually the case for the government). In the Court's argumentation, critical questions sometimes occur in the evaluation of the so-called reported argument from the parties, and they also figure as implicit and/or explicit premises in the Court's own argumentation. Therefore, the annotation of the critical questions could present a problem when working only over non-reported arguments. However, the choice to separate reported arguments from the non-reported arguments facilitates the identification of the critical questions that have been answered in the premises and those that have not been answered, which in turn helps identify the arguments that can be criticized.

The classification of a textual unit as a premise or a conclusion is a difficult task. In the first analysis, it was observed that the annotators had difficulties to distinguish a premise from a conclusion. Some of the errors came from a lack of attention in the rhetorical markers used in the sentence. This was easy to prevent after the addition of some linguistic guidelines. However, the complex argumentation structure of the ECHR legal cases created other errors which are presented in Chapter 6.

5.2.3 Relations between arguments

In the two previous sections the categories to identify arguments and their elements in the framework have been defined. At this point, each detected argument is represented by the following information:

- Group of Premises: $\{P_i, i = [0 \dots n]\}$
- Conclusion: C
- Type of argument: $T_a = [1 \dots m]$
- Reported/Non-reported

Each argument is restricted to these elements to maintain simplicity with an adequate level of information. Given the early nature of this thesis topic it was decided that the definition of argument should not include internal relations between premises or between premises and conclusion, the argument is seen as a set of propositions that work "together". The only information about how they work together is related to the function they maintain, i.e. premise or conclusion.

The full argumentation structure, i.e. the relations between arguments, can be extracted from this information and a pragma-dialects approach, see Chapter 2. If the premises of two arguments maintain a coordinative or multiple relation, both will have the same conclusion. If two arguments maintain a subordinate relation, the conclusion of one will be a premise of the other. Note that without the semantic information of the premises it cannot be distinguished between coordinative and multiple relations. Appendix D shows the annotation from an ECHR legal case fragment.

5.3 Annotation process

The detection of arguments is a difficult and tedious task. Legal experts in the corporate world are required to identify arguments but not to annotate their elements. Argument annotation, such as the one required in this thesis, is not generally part of an educational training in law. However, legal experts seem to be generally better suited for the task than other annotators given the complexity of the legal language. Non-legal experts might not understand all the information presented in a legal case or the implicit references to legal background. Therefore, all the annotators involved in this thesis were legal experts. However, only extrinsic evaluation of other annotations can prove or disprove if legal experts are the most adequate annotators for this task.

For the annotation process before the reliability study three annotators were selected: two annotators and a "judge" to solve disagreements between the annotators. The two annotators were two lawyers, a European and non-European. They had previous knowledge on argumentation detection, achieved

during work at different legal firms. However, these experiences did not include argumentation analysis or good knowledge of argumentation theories. The "judge" was a legal expert on ECHR cases with extensive knowledge of argumentation theories and previous experience on argumentation annotation.

After the reliability studies had confirmed that the annotation can in principle be done reliably by trained annotators. Then two new annotators, with specific training on the annotation framework and the type of argumentation done in the ECHR corpus, were selected. The "judge" from the previous round maintained her role.

The full corpus was annotated by the new set of annotators. These annotations are used to compute the final agreement between annotators, see Chapter 6. Then, the "judge" chose an annotation between the two provided by the annotators. This annotation is used as system training material in the second part of this thesis.

All the annotators worked individually during five weeks, 5 days per week, on 8 hours shifts. Given the length and complexity of legal documents and time restrictions the annotators were asked to follow a strict procedure of annotation. Each document was annotated under the following procedure:

- Lecture for general knowledge
- Identification final decision
- Identification of single arguments
  - Identification conclusion
  - Identification premises
  - Identification type
- Recursive till all connected to final decision:
  - Connection between single arguments
  - Integration in super-argument

This procedure was established in order to fight the complexity of ECHR argumentation annotation and the possible loss of attention due to long working hours. Some of the documents contained more than 10 pages of argumentation, all of it connected to the final decision, creating really complex argumentation structures. Most legal experts have never been asked to find full argumentation structures, their general work is to find pieces that are relevant for a concrete statement. Therefore, it was not a surprise that our

pilot studies showed that their annotations were incomplete. For example, some arguments were unannotated because of the large distance between them and any other argumentative fragments, or some arguments were never connected to the final decision, which can mean two things: (a) there was an unannotated argument connecting the argument with the final decision or (b) the conclusion of the argument was directly related to the final decision but the relation was unannotated, i.e. the argument was not annotated as a premise of the final decision. The use of the above procedure produced much more complete and accurate annotations in shorter time. Some statistics on the annotations are found in Table 5.1.

5.4 Summary

This chapter has presented the annotation framework. It has discussed the main issues involved in the framework definition, dealing with the annotators limitations and the task complexity. The framework background has been related to the argumentation theories presented in Chapter 2, which motivate most of the decisions. It has also been introduced the first aspects of the evaluation of the framework, such as pilot studies with corpus subsets to determine the annotators reliability. The rest of the framework evaluation is considered in the next chapter with the analysis of the main annotation problems.

There are typically three outcomes of this process. First, the exhaustive analysis of a whole text or corpus is a more empirically sound procedure for discovering linguistic phenomena, compared to choosing examples; annotation of the electronic text forces the analyst to test and refine the system of categorization to account for all cases. Second, it is possible to extract statistics relating to frequency, distribution and co-occurrence of forms from the annotated text. Third, an annotated corpus is obtained, available for studies aiming to replicate or further develop the research, and usable for other areas of literary or linguistic research.

Argumentation Scheme	% Occurrences in dataset	% Errors annotation
Argument from Sign	7.9	7.2
Argument from Example	5.2	2.4
Argument from Verbal Classification	4.7	5.1
Argument from Commitment	0.6	0
Circumstantial Argument Against the Person	0	0
Argument from Position to Know	2.1	2.3
Argument from Expert Opinion	2.7	0
Argument from Evidence to a Hypothesis	0.3	0
Argument from Falsification of a Hypothesis	4.2	0
Argument from Correlation to Cause	1.1	1.7
Argument from Cause to Effect	4.7	10.6
Argument from Consequences	8.2	5.4
Argument from Analogy	21.6	25.4
Argument from Waste	2.3	0
Argument from Popularity	1.2	0
Argument from Popular Practice	2.4	2.2
Ethotic Argument	0	0
Negative Ethotic Argument	0.3	0
Argument from Bias	0	0
Argument from an Established Rule	19.6	28.0
Argument from an Exceptional Case	0.4	1.7
Argument from Precedent	6.1	1.7
Argument from Gradualism	4.4	6.3
Argument from Vagueness of a Verbal Classification	0	0
Argument from Arbitrariness of a Verbal Classification	0	0
Argument from Ignorance	0	0

Table 5.1: Argumentation Schemes found on our framework. Percentage of occurrences in the text of the given scheme type. Percentage errors on the annotation related to arguments of that scheme type, i.e. arguments scheme x with annotation error divided by total of arguments scheme x in corpus.

Argument from Sign	The Commission observes that whereas in his original application the applicant stated that his house was burnt down on January 1994, it appears that the incident under investigation by the public prosecutor of Kulp district and the Kulp District Administrative Board occurred on November 1993. In these circumstances , the question arises whether the complaints insofar as they relate to specific acts carried out on November 1993 have been introduced out of time, given that the application was introduced on June 1993
Argument from Analogy	The Court recalls its decision in the case of Brogan and Others v. the United Kingdom (...) that a period of detention without judicial control of four days and six hours fell outside the strict constraints as to time permitted by Article 6 (...). It clearly follows that the period of fourteen or more days during which Mr Aksoy was detained without being brought before a judge or other judicial officer did not satisfy the requirement of "promptness".
Argument from an Established Rule	The Court is sensitive to the subsidiary nature of its role and must be cautious in taking on the role of a first-instance tribunal of fact, where this is not rendered unavoidable by the circumstances of a particular case (...). Where domestic proceedings have taken place, it is not the Court's task to substitute its own assessment of the facts for that of the domestic authorities and, as a general rule, it is for those authorities to assess the evidence before them (...). Though the Court is not bound by the latter's findings, in normal circumstances it requires cogent elements to lead it to depart from the findings of fact they have reached (...)
Argument from Precedent	The Commission observes that Article 26 of the Convention "should be applied with some degree of flexibility and without excessive formalism; it is sufficient that the complaints intended to be made subsequently before the Convention organs should have been raised at least in substance and in compliance with the formal requirements and time-limits laid down in domestic law" ((...)Castells judgement of 23 April 1992 (...)). (...) The Commission considers that the applicant did invoke before the Greek courts, at least in substance, the complaints relating to Article 10 (Art. 10) of the Convention which he now puts to the Commission. He may therefore be said to have exhausted domestic remedies. The Commission concludes that the applicant has complied with the requirements of Article 26 (Art. 26) of the Convention.

Table 5.2: Examples of the most common argumentation schemes encountered on the ECHR corpus

Chapter 6

Evaluation of the annotation

The previous chapter has defined the framework and methodology of the annotation. This chapter focuses on evaluating the efficiency of the framework and the analysis of the main problems encountered during the annotation process. The aim is also to rise open issues on argumentation annotation which should be dealt with in the future.

6.1 Evaluation of the framework

The evaluation of the annotation framework is particularly interested in the framework stability and reproducibility. Stability refers to the extent to which one annotator will produce the same classifications at different times [29]. Stability is important, because in unstable annotation frameworks the definition of the categories is not even consistent within one annotator's private understandings, and as a result, such frameworks are very unreliable. High stability shows at the very least that there must be some consistent definition of semantics in the standard, even if one does not know yet if this definition can be communicated to others. Reproducibility refers to the extent to which different annotators will produce the same classifications, which measures the consistency of shared understandings (or meaning) held by more than one annotator [29]. Reproducibility is an important part of the evaluation of a framework. It is commonly assumed that a proof of the reproducibility of a framework implies its stability, as consistent shared understandings require

consistent private understandings, therefore an unstable annotation can never be reproducible.

Reproducibility can be seen as of little importance when one is interested in user satisfaction. If there are two or more intuitively "good" but different gold standards, two judges might disagree over which one to choose, resulting in a low reproducibility. However, both of these gold standards might have satisfied the user. Therefore, there is an accepted theoretical priority of stability over reproducibility, but at the end of the day, only extrinsic evaluation can prove or disprove if this argument is valid.

The annotation task is a mutually exclusive categorical assignment. There are different ways to evaluate agreement between humans for such task [6], using either majority opinion or percentage agreement as measurement. This evaluation is opposed to using majority opinion: the average does not reflect anybody's understanding of the categories. We want to treat all the annotator's opinions as a valid judgement. None of these is by definition wrong or right, we are dealing with a difficult "high-level" task, where a certain level of subjective disagreements can be expected.

The evaluation of the annotations due to the proposition categories, i.e. categories: premise, conclusion and non-argumentative, was done using two metrics, Kappa coefficient (Section 6.1.1) and Krippendorff's alpha (Section 6.1.2). The correctness of the tree was analysed between two annotators element by element independently (Section 6.1.3).

6.1.1 Kappa coefficient

The Kappa coefficient, K, measures agreement, $P(A)$, respect agreement by chance, $P(E)$ [56].

$$K = \frac{P(A) - P(E)}{1 - P(E)}$$

Chance agreement is defined as the level of agreement which would be reached by random annotation using the same distribution of categories as the real annotators. Kappa is stricter than percentage agreement $P(A)$: its value is always lower or equal to percentage agreement. No matter how many items or annotators, or how the categories are distributed, $K = 0$ when there is no agreement other than what would be expected by chance, and $K = 1$ when agreement is perfect. If two annotators agree less than expected by chance, Kappa can also be negative.

Kappa has the following advantages. First, it allows for comparisons between arbitrary numbers of annotators and items. Second, it factors out random agreement. Third, as a side effect of taking random agreement into account, Kappa treats agreement in a rare category as more surprising, and rewards such agreement more than an agreement in a frequent category, i.e. it treats less frequent categories as more important.

There are different scales of how to interpret Kappa values. Krippendorff [29] starts from the assumption that there are two independently annotated variables which show a clear correlation. If the agreement of an annotation of one of these is so high that it reaches a value of $K = 0.8$ or above on a reasonably-sized dataset, then the correlation between these two variables can be shown with a statistical significance of $p \leq 0.05$. That is, the annotation contains enough signals to be found among the noise of disagreement. If agreement is in a range of $0.67 \leq K \leq 0.8$, the correlation can be shown with a (marginal) statistical significance of $p = 0.06$, which allows for tentative conclusions to be drawn. Krippendorff's strict scale considers annotations with $K < 0.67$ as unreliable. More forgiving scales take into account that most practical annotation frameworks only mark one dependent variable and assume that $K = 0.6$ is still a reasonable agreement. However, is some cases, where the task complexity is high, a low Kappa agreement, even as slow as the threshold below chance (e.g. $K = 0.4$), can nevertheless exclude chance and it is considered reliable.

The pilot studies with the first framework showed an agreement between annotators of $K = 0.57$, when analysing the agreement on the identification of individual propositions categories, i.e. identification between premise, conclusion and non-argumentative. This K, in normal Kappa standards, is a quite low agreement. The final framework managed a better annotation agreement of $K = 0.75$, which it is argued to be reliable enough given the task complexity.

It was also analysed which type of arguments where more prone to present annotation errors. Table 5.1 shows for each argument category of the framework its percentage weight on the annotation errors rate. *Argument from Analogy* and *Argument from an Established Rule* are the two argument categories which are more prone to present annotation errors. .

6.1.2 Krippendorff's alpha

Krippendorff's alpha (α) is a reliability coefficient widely applicable wherever two or more methods of processing data are applied to the same set of units of analysis [30].

The α's general form is: $\alpha = 1 - \frac{D_o}{D_e}$ where D_o is the observed disagreement and D_e is the disagreement one would expect when the coding of units is attributable to chance rather than to the properties of these units.

When observers agree perfectly, observed disagreement $D_o = 0$ and $\alpha = 1$, which indicated perfect reliability. When observers agree as if chance had produced the results, $D_o = D_e$ and $\alpha = 0$, which indicates the absence of reliability. The α would measures 0 if observers failed to observe and made up their data randomly. When $\alpha = 0$, data are totally uninformative of anything outside the process of generating them. Therefore, for reliability considerations, α's range is:

$$1 \geq \alpha \geq 0 \begin{cases} - \text{ Systematic disagreement} \\ \pm \text{ Sampling errors} \end{cases}$$

The α coefficient applies to any number of annotators and categories. It works with large and small sample sizes alike, not requiring a minimum.

The pilot studies with the first framework showed an agreement between annotators of $\alpha = 0.49$, when analysing the agreement on the identification of individual propositions categories, i.e. identification between premise, conclusion and non-argumentative. This α is a low agreement. The final framework obtained a better annotation agreement of $\alpha = 0.642$, a comparable result to the agreement by Kappa.

6.1.3 Tree-structure agreement

The agreement on the tree-structure of the cases was computed between two annotators as follows. First, for each case it was evaluated if the annotators agree on the final decision. The results showed a perfect agreement on the root of the tree. Moreover, there was no confusion between non-argumentative and the propositions of a final decision. Note that given the characteristics of the ECHR cases final decisions have clear markers. Figure 6.1 shows some examples of typical structures for final decisions. Second, the agreement between numbers of arguments hanging from the final decision (tree root) and the elements of each of the arguments was studied. In 85% of the cases the annotators agreed on the number of arguments hanging from the root. Up to 70% of these arguments were completed, while in 30% of the cases some argumentative propositions were not annotated as part of the argument. The inner structure of the arguments also caused errors when the arguments presented complex structures, specifically on structures with more than 2 levels of depth or large number of sub-arguments, which represented 75% of the inner structural errors.

For these reasons, the Commission, by a majority,
DECLARES ADMISSIBLE, without prejudging the merits of the case, the applicant's complaint that his conviction for having disrupted public peace amounts to a violation of his rights set forth in the Convention;
DECLARES INADMISSIBLE the remainder of the application.

FOR THESE REASONS, THE COURT
1. Holds by seventeen votes to four that there has been no violation of Article 5 para. 1 of the Convention (art. 5-1);
2. Holds by seventeen votes to four that Article 5 para. 5 of the Convention (art. 5-5) is not applicable;
3. Holds unanimously that there has been a violation of Article 6 paras. 1 and 3 (c) of the Convention taken together (art. 6-1+6-3-c);
4. Holds by nineteen votes to two that the finding of a violation constitutes adequate satisfaction for the non-pecuniary damage suffered by the applicant;
5. Holds unanimously
(a) that the respondent State is to pay the applicant, within three months, in respect of costs and expenses, Âč10,000 (ten thousand pounds sterling) less 25,510 (twenty-five thousand, five hundred and ten) French francs to be converted into pounds sterling at the rate applicable on the date of delivery of the present judgment;
(b) that simple interest at an annual rate of 8% shall be payable from the expiry of the above-mentioned three months until settlement;
6. Dismisses, unanimously, the remainder of the claim for just satisfaction in respect of costs and expenses.

For these reasons, the Commission by a majority
DECLARES THE APPLICATION ADMISSIBLE, without prejudging the merits.

Figure 6.1: Examples of final decisions of ECHR case

6.2 Analysing the annotation problems

The skill of distinguishing argument from non-argument is sophisticated and requires training: it is a typical learning outcome of an undergraduate critical thinking course. The analysis of an argument, including the categorisation of its text by an argumentation scheme, is more challenging yet, and faces the additional problem that multiple analyses may be possible.

This chapter presents the results of the analysis of the annotation process presented in Chapter 5. This helps us to study different aspects of human

argument perception. In the following sections the different disagreements between annotators are analysed.

6.2.1 Argumentation or fact?

The first disagreement between annotators is related to the distinction between argumentation and fact. The cause of this disagreement is, so we believe, due to different conceptions of law held by the respective annotators. [31] suggests that the positivist conception of the law[1], still taught in many law schools, is challenged by theories of legal reasoning which highlight the argumentative aspects of legal decision making. The hypothesis used in this thesis is that, in the annotation study, one is possibly confronted with something of a reverse mechanism, where someone holding a positivist conception of law does not identify as much argumentation in legal decisions. It is difficult to fully confirm this presumption, but two elements nevertheless seem to support it. First, during this study it was observed that the annotator who identified the least arguments had enjoyed a strongly positivism-oriented legal training, whereas another of the annotators, who had a firm background in both legal theory and argumentation theory, identified more argumentative sentences than either all the other lawyers together did. They detected till 18% more argumentative information. Second, the majority of the argumentative sentences that remained undetected were classified by the other annotators as *Arguments from Precedent or Analogy*. The second most commonly non-detected type of argument was the *Argument from an Established Rule*. It is common for legal positivists to count rules of law, in whichever way established, be it by statute or precedent, as "facts" in their ontology. Possibly, then, the non-identification of instances of *Argument from an Established Rule* and *Argument from Precedent* points to such a positivist ontology in the annotator's (implicit) conception of law. In the annotations of the annotators with a positivist conception of law, around 6% of the instances of these arguments were not detected.

6.2.2 Limits of the argument

The second disagreement between annotators relates to the limits of an argument. This disagreement presented itself in the form of premises going

[1] From: *http://plato.stanford.edu/entries/legal-positivism/*. Legal positivism is the thesis that the existence and content of law depends on social facts and not on its merits. The positivist thesis does not say that law's merits are unintelligible, unimportant, or peripheral to the philosophy of law. It says that they do not determine whether laws or legal systems exist.

undetected by certain annotators. This occurs most often in *Argument from an Established Rule*. It is believed this relates to the fact that the general overarching argument of a judicial decision is an instance of *Argument from an Established Rule*, with premises to the final conclusion scattered all over the text of the document. These scattered premises are most difficult to detect. It is believed there are two reasons for this: (a) the larger the distance between a premise and its conclusion, the harder it is to acknowledge their relation and (b) the larger the distance between premise and conclusion, the higher the chance of human error, e.g. oblivion or lack of concentration. Even if there is no real solution to human errors, it is expected that automatic detection and visualization of argumentation will be of great service in making the structure of those complex arguments, such as *Argument from an Established Rule*, more easily accessible to readers. The experiments during this thesis presented an improvement of around 25% when the annotators were asked to represent the cases in tree-structures.

6.2.3 Identifying the structure of the argumentation

The third disagreement between annotators is related to the identification of argumentation structures. This disagreement arises from the structure attributed to the detected argumentative sentences. There are three main causes for this disagreement, which we discuss in the following subsections.

Distinguishing premises from conclusions

In the earliest annotations, disagreements commonly arise concerning the nature of argumentative sentences, that is on the question of whether an argumentative sentence was a premise or a conclusion. The reason is, once more, that the reasoning of the Court usually takes the form of one or more large, complex and multilayered instances of *Argument from an Established Rule*, leading to one or several simple conclusions concerning the violation of an article of law. Therefore, there are only few "pure" conclusions, with a large number of subordinative, coordinative and multiple arguments leading to it, the premises of which can in turn be the conclusions of other arguments. In this way a reasonably large percentage of the premises are also conclusions to foregoing arguments, which explains the high ratio of disagreement between the first two annotators as they lacked familiarity with the ECHR's characteristics. The only way to solve this problem is to better train the annotators and make them aware of the specific overall argumentation structure of the documents to be annotated, previous to starting the annotation process. This strategy

was successfully carried out with the second set of annotators. In the final annotations only 26.4% of the middle conclusions were annotated as premises.

Identifying the structure of complex argument

The second cause of disagreement on argument structure presented itself, in almost two thirds of the cases, with the premises or conclusions of instances of *Argument from an Established Rule*: what was identified as a premise by one annotator was identified as a conclusion by the other and vice versa. Also, almost half of the cases in which super-arguments, sub-arguments or co-arguments were left undetected, or in which subordinative arguments were mistakenly identified as coordinative arguments, an *Argument from an Established Rule* or an Argument from Gradualism was concerned - two types of argument which are likely to have complex structures. Again, as most of the complex overarching arguments in the ECHR's decisions are instances of *Argument from an Established Rule*, scattered sometimes over several pages of text - it is no surprise that the structure of these particular arguments was most difficult to detect, for the similar reasons as cited in the Section 6.2.2: the larger the distance between different elements of an argument, (a) the more difficult it is to acknowledge their relation and (b) the larger the chance of human error. There is no real solution for this problem, but visualization will make argument structures more easily accessible. As in previous cases, the annotators presented an improvement of nearly 10% when representing the cases as tree-structures. This improvement was lower than the one achieved when detecting the limits of the argument as this task was more fine grained and involved many decisions.

Subordinative or coordinative?

The third and last cause of disagreement due to argument structure is also related to complex arguments, but this time to the overarching complex argument. As stated before, many sub-, super- and co-arguments were left undetected, and lots of coordinative and subordinative premises met the same fate. Most importantly, subordinative arguments were sometimes wrongly annotated as coordinative, and the other way around, whereas the difference should not be too difficult to understand. In both cases there is a chain of premises (say, more than two) working together and leading to a conclusion but in the case of subordinative premises the chain is serial and one could place the premises in a logical order that cannot be altered, or the argument would lose meaning. In the case of coordinative premises, the order of the premises would not necessarily make much difference as they work parallel to

each other. Also in a subordinative structure there can be different types of argument, though this is not necessarily the case. When the premises are coordinative there is only one argument and therefore only one possible argument scheme. There are many reasons for this type of error, e.g. the argument presents an incorrect or unclear discourse structure (the ECHR judges stated their standpoints not enough clearly) or on the other hand it could be that the annotators needed a more exhaustive training in this type of complex arguments. There was around 6% of subordinative arguments annotated, fully or partially, as coordinative and around 4% of coordinative arguments were included incorrectly in a subordination.

6.2.4 Identifying the type of argument

The fourth disagreement between annotators is non-corpus related, but argumentation theory related. The annotators had difficulties identifying the correct argument type, i.e. argumentation scheme, that was most suitable to the argument reasoning. The annotators presented an agreement of $K = 0.85$ when detecting the type of argument. The two main reasons for this disagreement are discussed in detail in the next subsections.

The argument from gradualism. Type or structure?

The *Argument from Gradualism* proved particularly difficult to detect. Annotators frequently disagreed on what was an *Argument from Gradualism*. In almost all cases of conflict, one of the annotators saw gradualism while another would see different types of argument, and thus different arguments, in a subordinative structure. This raised the following question: is gradualism not simply another term for a subordinative structure of argumentation, as Walton admitted to be the case for the *Slippery Slope* argument schemes in [69]? The chosen framework in which chains of premises are either broken up into chains of separate, subordinate arguments or, if listed as one argument, presumed to be coordinative (see Chapter 5) only complicated this issue. In this framework there is no place for subordinative structures of *Argument from Gradualism*. In the next paragraphs it is proven that an *Argument from Gradualism* can be coordinative.

Walton describes the *Argument from Gradualism* as consisting of a chain of premises, but does not explicitly say whether it is subordinative or coordinative in structure. Yet he admits in [69] that this type of argument can also be understood as a strategy or tactic of argumentation. In other words it can be understood to describe how chains of subordinative argument can best

be built in certain cases, by gradually augmenting and shifting the weight of probability. The visual representation given there, gives the same impression of subordinative argument as the premises are listed serially, one supporting the other. However, he also says that sometimes gradualism is used as a distinctive type of argumentation - the example given in Figure 6.2 argues for a gradual introduction of taxes as this will be more easily acceptable for the population. Questions arise as to whether this is not rather a combination of arguments from *Argument from Consequences* and *Argument from Popular Acceptance*, but these questions aside, the key ingredients to a "real" *Argument from Gradualism* seem to be: (a) the fact that one of the premises does not lead to the conclusion by itself, the other premises are also needed for the conclusion to gain probability and (b) an element of augmentation, of gradually increasing the weight of probability.

> A government knows that it needs to get an 18 percent value-added tax (VAT), sometimes also called a goods and services tax (GST) in order to deal with the deficit. However, the public would never vote for, or approve such a large tax, in one single step. Therefore, the government adopts the strategy of introducing a 3 percent VAT, and then increasing it every few years, when politically appropriate, until the 18 percent level is reached.

Figure 6.2: An instance of *Argument from Gradualism* as found in [69]

In fact, in the corpus, the annotators found few examples of arguments which seemed to be non-subordinative, hence assumed to be coordinative, which could not be classified as any other type of argument and which carried an element of gradualism, and there was disagreement on almost all of these examples between annotators. Therefore, the question of whether the *Argument from Gradualism* is really a type of argument and not rather a strategy or structure remains, evidently, in need of further theoretical elaboration.

Classification of schemes

This is a disagreement between two types of argumentation which in fact belong to the same general class according to [71]. For example, conflict occurred between an *Argument from Correlation* and an *Argument from Cause to Effect*, both classified as causal reasoning, between an *Argument from Sign* and an *Argument from Evidence to Hypothesis*, both classified as abductive reasoning, and between *Argument from (lack of) Analogy* and *Argument from an Established Rule (or exception to it)*, both arguments applying rules to cases according to [71]. This seems to imply that it is easier for humans to detect

more general classes of argumentation than it is for them to identify the more specific types.

It must also be noted that in cases of "negative" argumentation schemes, where the argumentation scheme was used negatively, for example to say that an analogy or established rule did not apply, annotators would wrongly pick a negative argumentation scheme on the list belonging to the same class (e.g. *Argument from Exceptional Case*) instead of the correct argumentation scheme which was used negatively (e.g. *Argument from Analogy*), but of which the "negative" was not in the list of choices. Another example would be a lawyer arguing that the witness is not in a position to know about a fact, and therefore anything he might say must be discounted in advance as representing a limited and biased point of view. In this case, what could be understood as a negative instance of *Argument from Position to Know* could be easily confused with *Argument from Bias*. See the argument presented in Figure 6.3, where a negative instance of *Argument from Analogy* was classified as an *Argument from Exceptional Case*.

> The Court recalls its decision in the case of Brogan and Others v. the United Kingdom (...) that a period of detention without judicial control of four days and six hours fell outside the strict constraints as to time permitted by Article 6 (...). In the present conditions, however, it follows that the period of fourteen or more days during which Mr Aksoy was detained without being brought before a judge or other judicial officer does satisfy the requirement of "promptness".

Figure 6.3: A negative instance of *Argument of Analogy* annotated as *Argument from Exceptional Case*

It is inferred then that a classification of argumentation schemes into broader classes, as done in [71], would be more adequate to use when creating corpora to let annotators select the general class of argumentation first and only clear out the specific details later. On this basis, "pathways of classification", in the form of a real taxonomic hierarchy, could be drawn up, leading annotators to the correct detailed argumentation scheme by means of answering simple taxonomic questions.

6.2.5 Common argumentation schemes in legal cases

One could assume that if an argumentation scheme is common in the domain, i.e. occurs frequently, then the annotators must have a good understanding of it and they should have fewer problems annotating its instances. However,

we have observed that some of the most common schemes in legal cases, such as *Argument from Analogy*, *Argument from Precedent* and *Argument from an Established Rule*, which represent a 47.3% of the total number of arguments in the ECHR corpus, are the reason behind our problems. First, the use of *Argument from Analogy* and *Argument from Precedent* in the legal domain, where one can take a positivism conception of law, is the reason behind the problem in Section 6.2.1. Second, the incorrect detection of the negative use of *Argument from Analogy* is one of the main factors behind the errors in Section 6.2.4. Finally, in the legal domain, the *Argument from an Established Rule* is a complex argument with many sub-arguments presenting why the established rule (e.g. article from the human rights declaration) applies or not in the current case. Therefore, this type of argument in the legal domain will always extend itself over many pages of the legal case and it is one of the main causes of the errors in Section 6.2.3.

6.3 Other remarks

During the late phases of this thesis (see Part II), where automatic detection of argumentation was studied, an important problem was observed in the annotation corpus. This problem has an important impact in all the results of the experimental approaches. All the models owe an important part of their errors to a subtype of argument, *Argument from Precedent*.

These problems are not due to the incapability to detect premises or conclusions, or the ambiguity of the argumentation structure. The reason behind the problems is the decision of considering the specific form of *Argument from precedent* represented by a single sentence, where premise and conclusion are connected using words such as *"cf." or "see"*, as an argument, where the first part of the sentence is the conclusion and the second part the premise (see Figure 4.5).

It is observed that these arguments, even if correctly annotated by humans and automatically detected, pose a bigger problem to the developed systems. The conclusions of these arguments, such as "*It is furthermore established that the burden of proving the existence of available and sufficient domestic remedies lies upon the State invoking the rule*", are clearly introduced as premises. This is correct and normal, *Argument from precedent* is used here as a sub-argument of a more general reasoning.

Given that an argumentative structure where *Argument from Precedent* is used as a sub-argument is not an exception but is quite common, the decision to "save" these instances as arguments instead of as single premises causes

major problems when identifying the general features or rules that characterize premises and conclusion. For example, in "*However, an applicant's behaviour constitutes an objective fact which cannot be attributed to the respondent State and which must be taken into account for the purpose of determining whether or not the reasonable time referred to in Article 6 § 1 has been exceeded (see Erkner and Hofbauer v. Austria, no. 9616/81, Commission decision of 23 April 1987, A 117, § 68)*", the conclusion "*However, an applicant's behaviour constitutes an objective fact which cannot be attributed to the respondent State and which must be taken into account for the purpose of determining whether or not the reasonable time referred to in Article 6 § 1 has been exceeded*" is clearly marked as a premise of a super-argument, which might confuse the classifier as it could infer that "*However*" is a marker of conclusion.

6.4 Summary

This chapter has evaluated the framework used during the annotations finding that the annotators reached an agreement of $K = 0.75$ and $\alpha = 0.642$. This agreement would normally be considered good enough only to draw tentative conclusions. However, it has been argued that given the task complexity the agreement is more than reliable. It has also been described the main problems encountered during the annotation process, analysing the possible reasons behind them. However, we believe a more extensive study should be performed to obtain more concrete justifications. This chapter has concluded with the analysis of a problem created by the decisions taken during the corpus development. This problem will have an impact in any experiment trained on the ECHR corpus.

Part II

Automatic Argumentation Detection

Outline part II

This part of the thesis is focused on the automatic detection of argumentation. All the experiments presented in this part use the linguistic resources presented in Part I as training and test corpora. The study on human argumentation annotation (Chapter 5) has shown that argumentation detection is a difficult task with a human agreement K=0.75. In this part of the thesis different approaches to obtain an automatic method for argumentation detection is studied. It starts with simple models with traditional features for NLP and move into more specific features for argumentation while increasing the complexity of the models.

It first present, in Chapter 8, a simple classification task between argumentation and non-argumentation, moving into more specific classifications between premises and conclusion. This approach is limited by the complex semantics of legal cases and does not allow the detection of argument structures. Therefore, it moves to the study of a more structural model, a rule-base parser. Chapter 9 presents an argumentative grammar for legal cases and a parser that allows to detect the full argumentative structure of a legal case.

The work in this part of the thesis has been partially published in the following articles. Parts of this research have not been previously published.

- MOCHALES PALAU, Raquel and MOENS, Marie-Francine (2011) Argumentation Mining. Artificial Intelligence and Law. 19(1), (pp.1-22).
- WYNER, Adam, MOCHALES-PALAU, Raquel, MOENS, Marie-Francine, MILWARD, David (2010) Approaches to Text Mining Arguments from Legal Cases. In E. Francesconi, S. Montemagni, W. Peters and D. Tiscornia (Eds.), Semantic Processing of Legal Texts: Where the Language of Law Meets the Law of Language (LNAI 6036) (pp. 60-79). Berlin: Springer.
- MOCHALES PALAU, Raquel and MOENS, Marie-Francine (2009) Argumentation Mining: The Detection, Classification and Structure of Arguments in Text. In Proceedings of the Twelfth International Conference on Artificial Intelligence and Law. (pp. 98-109) New York: ACM. (Best Student's Paper Award)
- MOCHALES PALAU, Raquel and MOENS, Marie-Francine (2008) Automatic Argumentation Detection and Its Role in the Semantic Web. In J. BREUKER, P. CASANOVAS, M. KLEIN and E. FRANSCESCONI (Eds.), Law, Ontologies and the Semantic Web (Frontiers in Artificial Intelligence and Applications). (pp. 115-129) Amsterdam: IOS Press.

- MOCHALES PALAU, Raquel and MOENS, Marie-Francine (2008) Study on the Structure of Argumentation in Case Law. In Proceedings of JURIX 2008: The 21st International Conference on Legal Knowledge and Information Systems. (pp. 11-20) IOS Press.
- MOENS, Marie-Francine, BOIY, Erik, MOCHALES PALAU, Raquel and REED, Chris. (2007) Automatic Detection of Arguments in Legal Texts. In Proceedings of the Eleventh International Conference on Artificial Intelligence and Law (pp. 225-230). New York: ACM.
- MOCHALES PALAU, Raquel and MOENS, Marie-Francine (2007) ACILA - Automatic Detection of Arguments in Legal Cases. In Proceedings of Workshop on Semantic Web Technology for Law. (pp. 5-9) Palo Alto, CA: Stanford University.
- MOCHALES PALAU, Raquel and MOENS, Marie-Francine (2007) Study on Sentence Relations in the Automatic Detection of Argumentation in Legal cases. In Proceedings of JURIX 2007: The 20th Anniversary International Conference on Legal Knowledge and Information Systems. (pp. 89-98) IOS Press.

A condensed form of the third and sixth articles was also presented at the 2007 and 2009 BNAIC conferences:

- MOCHALES PALAU, Raquel and MOENS, Marie-Francine (2009) Argumentation Mining: The Detection, Classification and Structuring of Arguments in Text. In Proceedings of the 21st Benelux Conference on Artificial Intelligence.
- MOCHALES PALAU, Raquel and MOENS, Marie-Francine (2007) Automatic Detection of Arguments in Legal Texts. In Proceedings of the 19th Belgian-Dutch Conference on Artificial Intelligence. Utrecht, the Netherlands.

Chapter 7

State of the art

Argumentation detection is a new task, but there is much work in computational and theoretical linguistics which is closely related. First, there is research on text *zoning*, i.e. methods which break documents into segments. Most approaches try to segment documents into topic-related zones [44, 20]. The general notion behind work like this is that there is a connection between the discovery of aboutness or discourse topics and textual organization. There are a few approaches that do not focus on topic-related zones but on rhetorical-related zones. The approach presented in [59] attempts to determine argumentative zones, where the interest is in detecting rhetorically coherent segments on scientific articles. This approach is presented in more detail in Section 7.1.

Another group of work related to argumentation detection are discourse theories for rhetorical structure. Discourse structure is concerned with two aspects of the organization of sentences: (a) the fact that the sentences in one topical or rhetorical segment of the text are in relation to each other and (b) that different segments also have an inter-segmental ordering of intentional relations. This is referred to as *micro vs. macro structure* [62]. The macro structure can also be referred to as discourse-level structure or large scale text structure. In well-written texts, the function of micro segments with respect to the macro segment, as well as the function of a macro segment with respect to the text as a whole, is signalled by cues. For example, at micro level one finds connectives between clauses such as "*but*", "*thus*" or enumeration markers such as "*first*" or "*second*", while on a macro level one finds phrases such as "*later we will show that...*". Rhetorical Structure Theory (RST) is presented in Section 7.2.

7.1 Argumentative zoning

Argumentative Zoning (AZ) attempts to divide scientific articles by their *argumentative zones* [59].

Scientific articles show a considerable level of variation in their writing style. It is possible to encounter articles overtly argumentative, arguing against another author's views; articles presenting empirical work, such as a linguistic survey; other articles describe practical work, such as an implementation for a given problem. In interdisciplinary fields, articles might combine research methodologies from more than one discipline, providing a wide range vocabulary. The linguistic expressions occurring in the articles mirror this variety. Scientific articles describe the author's work from his own viewpoint. This bias is present in the argumentation, the articles are written to convince the reader of the validity of a given research. Therefore, the texts typically contain explicit markup of this rhetorical information. It is not trivial to identify which kind of document structure underlies scientific articles.

During this research, annotators were asked to annotate all sentences in the document depending on the rhetorical status of the sentence with respect to the communicative function of the whole paper. An example of the desired result can be seen in Figure 7.1. The sentential-rhetorical speech act of single, important sentences, defines the argumentative Zones of the document. The zones, each associated with one or more sentences, are the following:

- BKG: General scientific background
- OTH: Neutral descriptions of other people's work
- OWN: Neutral descriptions of the own, new work
- AIM: Statements of the particular aim of the current paper
- TXT: Statements of textual organization of the current paper (e.g. *"in chapter 1, we introduce..."*)
- CTR: Contrastive or comparative statements about other work; explicit mention of weaknesses of other work
- BAS: Statements that own work is based on other work

The authors relied on well-known probabilistic classifiers and very simple features to identify and classify sections in scientific documents. The features included location of a sentence within a document and within subsections

and paragraphs; sentence length; whether the sentence contains a word from the title; whether the sentence contains significant terms spotted by the *TFIDF* (Term Frequency Inverse Document Frequency) metric, see equation 7.1; whether the sentence contains a citation; linguistic features of the first finite verb; cue phrases; and the presence of certain named entity types, to divide scientific documents in the different zones. For features examples check Tables 7.1 and 7.2.

$$
\begin{aligned}
TFIDF_w = \quad & TF_w * log(\tfrac{N}{DF_w}) \\
\\
TFIDF_w : \quad & \text{TFIDF weight for diagnostic unit w} \\
TF_w : \quad & \text{term frequency of w in document} \\
DF_w : \quad & \text{number of documents containing diagnostic unit w} \\
& \text{or number of occurrences of w in document collection} \\
N : \quad & \text{number of documents in collection}
\end{aligned}
\tag{7.1}
$$

The authors present reliability studies which were performed on a difficult test bed, a corpus of 80 conference papers in computational linguistics that showed great variation with respect to subdomain, writing style, register and linguistic expression. The seven categories (argumentative zones) were defined not to be specific to the domain, but to the text type, based on the typical argumentation to be found in scientific articles.

Type	Example
GAP INTRODUCTION	to our knowledge
PREVIOUS CONTEXT	elsewhere, we have
OUR AIM	main contribution of this paper
FUTURE	avenue for improvement
TEXT STRUCTURE	then we describe
AFFECT	hopefully
CONTINUATION	following the argument in
SOLUTION	insight
SIMILARITY	similar to
IN ORDER TO	in order to
COMPARISON	when compared to our
POSITIVE ADJECTIVE	appealing
CONTRAST	however
NEGATIVE ADJECTIVE	unsatisfactory

Table 7.1: Some formulaic expression types used in Argumentative Zoning [59]

Type	Example
AFFECT	we hope to improve our results
NEED	this approach, however, lacks
ARGUMENTATION	we argue against a model of
PRESENTATION	we present here a method for
AWARENESS	we are not aware of attempts
PROBLEM	this approach fails
BETTER SOLUTION	our system outperforms
RESEARCH	we collected our data from
SIMILAR	our approach resembles that of
COMPARISON	we tested our system against
CONTRAST	our approach differs from
USE	we employ Suzuki's method
FUTURE INTEREST	we intend to improve
INTEREST	we are concerned with

Table 7.2: Some types of actions used in Argumentative Zoning [59]

Argumentative Zoning was applied later on legal documents by Ben Hachey and Claire Grover, see [16]. The authors trained a classifier on 141 House of Lords judgements and tested it on 47 judgements, where a judgement contained 105 sentences on average. Different classification algorithms were used: decision tree learning algorithms, naïve Bayes classifier, support vector machines and maximum entropy modelling. Among the best results, the maximum entropy classifier achieved a precision of 51% and a recall of 17%.

Argumentative Zoning presents two main differences with respect to the aims of this thesis. First, the argumentative information captured by Argumentative Zoning is divided in statements of various natures, e.g. contrast, aims, descriptions. There is no intention of finding which argumentative function, e.g. premise, conclusion, these statements undertake in the text. Second, even if Argumentative Zoning detects all claims of the document and for each claim it detects the statements to support or contradict it, Argumentative Zoning does

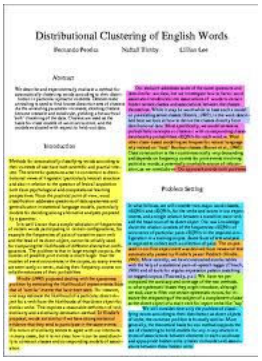

Figure 7.1: Argumentative Zoning document output example [59]

not show that those statements relate to that specific claim. Argumentative Zoning just detects the elements of the document arguments but not how those elements function or relate between themselves. The aim of this thesis goes further than simple detection of argumentative information, it also aims to structure it.

7.2 Rhetorical Structure Theory

Rhetorical Structure Theory (RST) is a descriptive theory of an important aspect of organization of natural text, see [33]. It is based on the notion that text structure serves a communicative role. RST is a method to describe natural text by characterizing its structure in terms of relations that hold between parts of the text, also known as text spans. The relations between the text parts are described in functional terms. RST identifies hierarchic structure in text, for any text size.

The main claims of RST are that discourse is characterized by strong hierarchical relations and by the predominance of structural patterns of nucleus/satellite type. However, there are relations which do not have a particular span of text which is more central to the author's purposes, e.g. a neutral contrast relation. These relations are called *multinuclear*.

RST provides schemas to define possible structural constituency arrangements of text. The schemas are patterns formed by a small number of text spans and relations between them. A schema application links a number of consecutive spans, and creates a complex span which can in turn be linked by a higher level schema application. This enables tree-structures to be built. RST recognizes

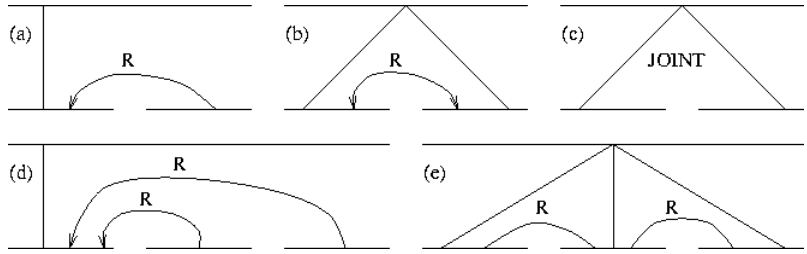

Figure 7.2: RST five kinds of schemas: (a) circumstance, (b) contrast, (c) joint, (d) motivation and enablement and (e) sequence and sequence. [33]

five kinds of schemas defined in terms of the relations that can co-occur between two text parts. Figure 7.2 shows one example of each type of schema. Schemas for relations not mentioned in Figure 7.2 follow the simple pattern represented by (a). The schema names are the same as the relation names. The straight lines identify the nuclear span(s) and the curves are the relations holding between them. The schemas do not constrain the order of nucleus or satellites in the part of text where the schema is applied.

The relations among the parts of text are described whether or not they are grammatically or lexically signalled. The relations holding between any two adjacent spans of text are rhetorical relations. The set of relations is in principle open and the relations are typically asymmetric. The set includes the following relations between others: CIRCUMSTANCE, SOLUTIONHOOD, ELABORATION, BACKGROUND, ENABLEMENT, MOTIVATION, EVIDENCE, JUSTIFICATION, CAUSE (volitional and non-volitional), RESULT (volitional and non-volitional), PURPOSE, ANTITHESIS, CONCESSION, CONDITION, INTERPRETATION, EVALUATION, RESTATEMENT, SUMMARY, SEQUENCE and CONTRAST.

The definitions of the rhetorical relations are kept general on purpose, cf. "*JUSTIFY:* a JUSTIFY satellite is intended to increase the reader's readiness to accept the writer's right to present the nuclear material" [33]. The definitions do not rely on morphological or syntactic signals, but on functional and semantic judgments alone.

A relation definition consists of four elements: constraints on the nucleus, constraints on the satellite, constraints on the combination of nucleus and satellite and the effect. Each element specifies judgments the text analysts must take into account when building the RST structure of the text. See Figure 7.3 for an example of RST relation definition.

An RST analyst effectively provides a plausible reason the writer might have

Relation name: EVIDENCE Constraints on Nucleus: R might not believe N to a degree satisfactory to W Constraints on Satellite: The reader believes S or will find it credible Constraints on the Nucleus + Satellite combination: R's comprehending S increases R's belief of N The effect: R's belief of N is increased Locus of the effect: N

Figure 7.3: EVIDENCE relation definition [33]

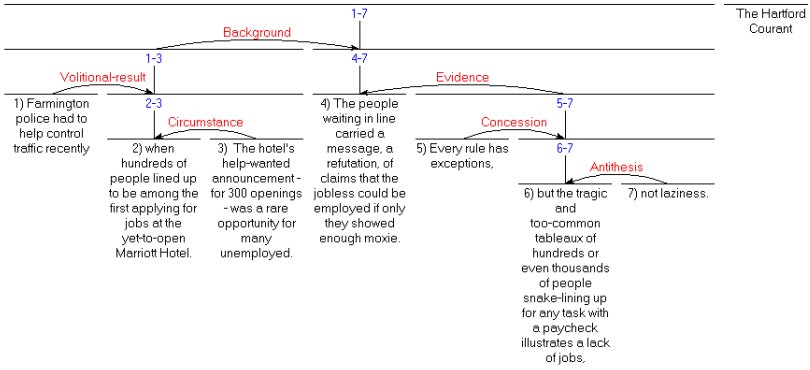

Figure 7.4: RST example - It's not laziness [33]

had for including each part of the whole text, see Figure 7.4. There is nothing in RST which would force the analyst to find some structural role for every element of the text. Even so, for carefully written texts, virtually every text has an RST analysis that provides a structural place for every element of the text. Depending on the text complexity the analyst may sometimes provide more than one analysis. Ambiguity of relations and structure are considered normal in RST [33]. This vagueness poses a problem for computational applications as it leads to multiple RST analyses for a given piece of text.

Another issue is the annotation unit. Unit size is arbitrary but a good unit size should provide units with independent functionality integrity. It is still entirely unclear, even if it has long been debated, what the formal linguistic criteria defining such units should be. Consider the unit "*not laziness*" in Figure 7.4. This unit carries a lot of information in the argument and thus has been determined as "clause-like". However, syntactically, the unit is only a single NP in a VP ellipsis construction. Therefore, a general syntactic criterion which defines this phrase as a clause, but excludes other NPs must be found.

RST has been extensively and successfully applied to the generation of texts describing ship movements and air traffic control procedures [22]. Moreover, Marcu used heuristics based on punctuation and cue phrases to recognize fully hierarchical RST structures in popular science text, see [36, 35]. A small number of people used RST as a writing guide, however all such use has been entirely informal.

RST has the ability to express interrelationships among argument elements in meaningful and useful ways. For example, the EVIDENCE relation can be used to represent a simple argument where the satellite is the premise and the nucleus is the conclusion. The same applies to the VOLITIONAL-CAUSE relation. However, the current state of the art on automatic rhetorical structure analysis in propositions as complex as the ones encountered in the ECHR corpus is limited. Therefore, in this thesis the concept of a complex text span formed by a satellite and nucleus is inspiration for the work presented in Chapter 9, but the RST categories involved on the complex text span and its elements relations are not explored.

7.3 Summary

In this chapter the state of the art in argumentation detection has been discussed. It has been shown that, even if argumentation detection is a new task, the research in text zoning and discourse analysis offers a background for this thesis. The approach used in text zoning is a simple but efficient model which is explored in Chapter 8, where the automatic detection of argumentation is seen as a classification problem between argumentation and non-argumentation. The rhetorical structure analysis is the inspiration for Chapter 9, where a complex model instead of providing a structural analysis of a text by its rhetorical relations looks for argumentative relations.

Chapter 8

Statistical classification

In this chapter a first approach to automatic argumentation detection is addressed. This approach is based on the work in text zoning presented in Chapter 7, where statistical classification methods were used to classify a text in rhetorically, rather than topically, coherent segments.

It is argued that argumentation detection can be achieved by a binary classification of all text sentences as argumentative and non-argumentative. If each sentence of the text can be classified as being part of the argumentation or not, then all units classified as argumentative constitute all the arguments of the text. This approach cannot detect the delimiters of each argument or their relations. Therefore, it is known which information forms the argumentation but not how this information is split into the different arguments.

The classification problem is formalized as follows: given a training set $(x_1, y_1), \ldots, (x_m, y_m)$ a classifier ϱ is produced, such that $\varrho(x)$ can be evaluated for any possible value of x (not only those included in the training set) and such that the class attributed to any new observation, specifically $\hat{y} = \varrho(x)$, is as close as possible to the true class label y. In the classification at hand the set used for training consists of an x and y for each sentence of the corpus, where x denotes a vector of observed features for the sentence and y denotes the class label, e.g. argumentative or non-argumentative, attributed to that sentence.

Given the problem formalization, there are three aspects to analyse: the training set (i.e. corpus), the features to represent the sentences and the classifier. The first experiments are focused on the Araucaria corpus, a non-

specific legal corpus (see Chapter 3). It is studied how simple features, typical of NLP, can be used to identify argumentation in a general domain. The results are compared by the different type of document source, e.g. blog vs. legal case, which shows the differences between working with formal and non-formal texts. The focus then moves on the ECHR corpus, a legal corpus presented in detail in Chapter 4. This corpus presents specific characteristics of legal argumentation which are transformed in complex features for a deeper analysis of the text. All classifiers studied during the experiments are state-of-the-art models and their specifics have been defined in Section 2.3.

8.1 Evaluation metrics

To evaluate the output of the automatic argumentation detection system well-known evaluation metrics that count the numbers of correct classified sentences are used. A sentence is considered as being correctly classified, if the given label corresponds to the label in the manual annotation for that sentence.

For a given test set, in the context of classification tasks (cf. classification between two class labels: E1 and E2) the following four terms are used to compare the given labels with the label the items actually belong to:

- T_p is the number of sentences that are correctly classified as E1
- T_n is the number of sentences that are correctly classified as E2
- F_p is the number of sentences that are incorrectly classified as E1
- F_n is the number of sentences that are incorrectly classified as E2

Precision, recall and F1 measure are defined as follows:

$$precision = \frac{T_p}{T_p + F_p}, \quad recall = \frac{T_p}{T_p + F_n}, \quad F1 = 2 * \frac{precision * recall}{precision + recall}$$

Accuracy is computed as the number of correctly classified sentences divided by the number of sentences that were classified: $accuracy = \frac{N_{corr}}{N_{auto}}$, note that $N_{corr} = T_n + T_p$ and $N_{auto} = T_n + T_p + F_n + F_p$.

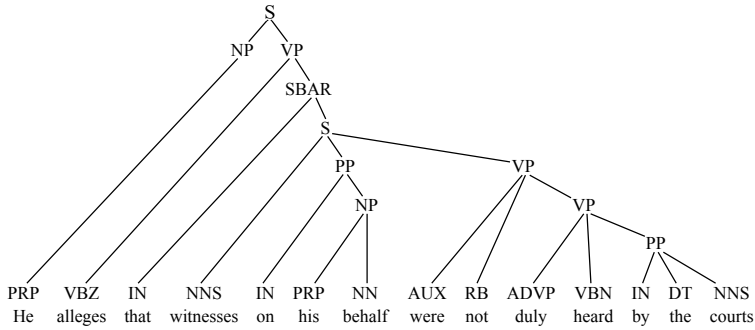

Figure 8.1: Syntactic dependency tree of a sentence in the ECHR corpus

8.2 Baseline

The first experiments discussed are focused on the classification of sentences as argumentative and non-argumentative. The methods used are state-of-the-art statistical classifiers. One is a generative model (naïve Bayes classifier) and the other a discriminative model (maximum entropy classifier), both presented in Section 2.3. The features used to represent the input text are presented in Section 8.2.1 and Section 8.2.2 presents the evaluation of the results.

The task at hand can be formalized as: given a training set $(x_1, y_1), \ldots, (x_m, y_m)$, where x_i denotes a vector of observed features for sentence of the corpus and y_i denotes a human annotated class label, cf. argumentative or non-argumentative. We train a classifier ϱ that for a new observation x_j determines its label $\hat{y_j} = \varrho(x_j) \approx y_j$.

8.2.1 Features

For the first experiments, generic features were defined. These features are typical of text extraction so they can easily be extracted from the texts. Every word in the Araucaria and ECHR corpora is tagged with its part-of-speech (POS) by an automatic tagger (QTag [1]) and a syntactic dependency tree is constructed for every sentence by an automatic parser (Charniak [8]), see Figure 8.1. These automatic annotations are, together with the word tokens, converted to a number of features used in the statistical classifiers. The sentence "*He alleges that witnesses on his behalf were not duly heard by the courts.*" is used as example to show the set of tokens given by each feature.

[1] www.english.bham.ac.be/staff.omason/vare/qtag.html

- **Unigrams** Each word in a sentence. In the example the set of unigrams would be {*he, alleges, that, witnesses, on, his, behalf, were, not, duly, heard, by, the, courts*}.
- **Bigrams** Each pair of successive words in a sentence. In the example the set of bigrams would be {*he alleges, alleges that, that witnesses, witnesses on, on his, his behalf, behalf were, were not, not duly, duly heard, heard by, by the, the courts*}.
- **Trigrams** Each three successive words in a sentence. In the example the set of trigrams would be {*he alleges that, alleges that witnesses, that witnesses on, witnesses on his, on his behalf, his behalf were, behalf were not, were not duly, not duly heard, duly heard by, heard by the, by the courts*}.
- **Adverbs** Detected with a POS tagger. In the example the set of adverbs would be {*duly*}.
- **Verbs** Detected with a POS tagger. Only the main verbs (excluding "*to be*", "*to do*" and "*to have*") are considered. In the example the set of verbs would be {*alleges, heard*}. Note that only the verbs detected as verbs by the parser are used as features. Parsing errors are thus carried into the features.
- **Modal auxiliary** Indicates if a modal auxiliary is present using a POS tagger. In the example the set of modal auxiliary would be empty.
- **Word couples** All possible combinations of two words in the sentence are considered. In the example the set of word couples would be {*he alleges, he that, he witnesses, he on, he his, he behalf, he were, he not, he duly, he heard, he by, he the, he courts, alleges that, alleges witnesses, alleges on, alleges his, alleges behalf, alleges were, alleges not, alleges duly, alleges heard, alleges by, alleges the, alleges courts, that witnesses, that on, that his, that behalf, that were, that not, that duly, that heard, that by, that the, that courts, witnesses on, witnesses his, witnesses behalf, witnesses were, witnesses not, witnesses duly, witnesses heard, witnesses by, witnesses the, witnesses courts, on his, on behalf, on were, on not, on duly, on heard, on by, on the, on courts, his behalf, his were, his not, his duly, his heard, his by, his the, his courts, behalf were, behalf not, behalf duly, behalf heard, behalf by, behalf the, behalf courts, were not, were duly, were heard, were by, were the, were courts, not duly, not heard, not by, not the, not courts, duly heard, duly by, duly the, duly courts, heard by, heard the, heard courts, by the, by courts*}.

- **Text statistics** Sentence length, average word length and number of punctuation marks.In the example the sentence length would be *14*, the average word length would be *4* and the number of punctuation marks would be *1*.

- **Punctuation** The sequence of punctuation marks present in the sentence (e.g. ":."). When a punctuation mark occurs more than once in a row, it is considered the same pattern (e.g. two or more successive commas both result in ",+"). In the example the set of punctuation marks would be empty.

- **Keywords** 286 words or word sequences obtained from a list of terms indicative for argumentation [28]. Examples from the list are "*but*", "*consequently*", and "*because of*". In the example the set of keywords would be empty.

- **Parse features** Information from the parse tree of each sentence was used as features. In the example the parse tree of Figure 8.1.

8.2.2 Evaluation

Table 8.1 presents some of the results obtained using the naïve Bayes classifier (MNB) and the maximum entropy classifier (MaxEnt) with some combinations of the simple shallow features presented in the previous subsection. It also presents results obtained with single individual features. These first experiments were done over the Araucaria corpus (see Section 3.1). To reduce variability 10-cross-validation was performed. In each of the 10 rounds of the cross-validation the corpus was partitioned into complementary subsets, the training set, where the analysis was performed and the classification model trained, and the test set, where the analysis was validated. After the 10 rounds the validation results were micro averaged.

One of the aims of these experiments was to observe if the classification between argumentative and non-argumentative sentences was possible. It was observed that even if the used features are simple, the experiments already yield acceptable results improving a basic baseline by taking every sentence as non-argumentative, which would lead results slightly above 60%. For example, using just the words of the text (unigram feature) the classification results achieve an accuracy slightly above 70%. Some features, e.g. keywords, present lower results that what might have been expected. However, note that unigrams include more word cues for the rhetoric than the keyword list and the keyword list does not perfectly fit argumentation discovery.

Analysing the results by the type of features, it is observed that in the category of lexical features a better score is obtained by considering bigrams or word couples than simple unigrams. Considering syntactic features, it is observed that verbs and adverbs contribute to the argument classification, but on their own, these features are insufficiently discriminative. Parse features, i.e. exploiting the depth of the parse tree and the number of subclauses, are weaker patterns for argument detection. Note that not all feature combinations were tested. Among the ones tested, it is observed that the combination of *word couples, verbs and text statistics* obtains the higher accuracy (73.75%). An example of sentence which is correctly classified by this combination, but wrongly classified using the features individually is: "But, I'm still convinced there will be no solution without a 24/7 approach to conflict resolution by the United States.".

Another important aim of these first experiments is to observe the influence that the different text sources have on the classification. Table 8.2 shows the results for the individual text types, where it is observed an indication that arguments in *newspapers* and arguments in *legal sources* are respectively the most easy and most difficult to detect. Explanations for the lower accuracy obtained for legal texts are: the small number of training examples and, at least in this test set, more ambiguous argumentation patterns. Although many training examples are available, *discussion fora* score also lower than *newspapers*. They contain more ambiguous and fewer well-formed texts compared to *newspapers*.

8.3 Improving the baseline: part I

The inference of the coherence structure of a text is a major NLP topic, that has been stimulating a lot of theoretical work since the 80's. It has been shown that coherence based approaches increase the accuracy in information extraction and textual summarization. Evidence shows that knowledge about rhetorical relations between different segments of texts contributes significantly to the increase of precision, and thus of the overall systems [21, 18]. It is argued that the rhetorical relations also contribute to an increase in precision of the argumentation classification system. As discussed in Section 2.1, there have been different studies that analysed the relations between the different parts of an argument, but here it is choosen to follow the approach of argumentation schemes, i.e. an argument is formed by a set of premises, critical questions (which it is seen as "subtype" of premise, see Chapter 5) and a conclusion. Due to the interest in full argumentation structure, the Arucaria corpus does not fulfil the desired requirements, therefore from now on the ECHR corpus will be used. Section 8.3.1 presents the new classification approach based on

argumentative coherence and Section 8.3.2 presents the new features involved to represent the input text preserving its coherence.

Features	Number features	MNB	MaxEnt
Unigrams	9238	73.06%	71.14%
Bigrams	38078	71.09%	70.62%
Trigrams	50929	64.24%	64.67%
Adverbs	435	55.74%	58.87%
Verbs	2043	60.19%	61.16%
Modal auxiliary	1	49.76%	57.35%
Word couples	402193	71.67%	72.90%
Keywords	108	53.32%	57.98%
Text statistics	3	58.48%	50.95%
Parse features	2	50.54%	50.26%
Unigrams + Text statistics	9247	73.12%	70.98%
Word couples + Text statistics	402196	73.70%	73.22%
Word couples + Verbs + Text statistics	404236	73.75%	72.59%
Word couples + Verbs + Text statistics + Keywords	404344	73.46%	72.72%

Table 8.1: Results in terms of accuracy for argumentative classification

The task at hand can be formalized as: given a training set $(x_1, y_1), \ldots, (x_m, y_m)$, x_i denotes a vector of observed features for a sequence of N sentences of the corpus and y_i denotes a human annotated class label. The N sentences can represent one of the following two meanings:

- $sentence_i$ & $N-1$ previous sentences $(sentence_{i-N}, \ldots, sentence_{i-1})$
- $sentence_i$ & $N-1$ next sentences $(sentence_{i+1}, \ldots, sentence_{i+N-1})$

The class labels are: non-argumentative, premise and conclusion. A classifier ϱ is trained so for a new observation x_j determines its label $\hat{y_j} = \varrho(x_j) \approx y_j$.

Text Source	Sentences	Number features	MNB	MaxEnt
Discussion fora	750	89613	71.73%	68.40%
Legal judgments	138	39681	65.94%	68.12%
Newspapers	702	119942	76.35%	73.22%
Parliamentary records	184	31207	72.83%	67.93%
Weekly magazines	176	33525	69.89%	69.32%

Table 8.2: Results in terms of accuracy for argumentative classification with features *Word couples + Verbs + Text statistics* depending on text source

Type	Precision	Recall	F1
Non-argumentative	83.68%	96.96%	89.83%
Premise	63.39%	29.34%	40.12%
Conclusion	72.97%	12.98%	22.04%

Table 8.3: Baseline over ECHR corpus with features *Word couples + Verbs + Text statistics*

8.3.1 N-gram model

An argument is always formed by premises and conclusions, which sometimes are left implicit, i.e. enthymemes (see Chapter 2). Someone could think that an argument could be presented as a single clause in its minimal representation. However, even for a human at least two argumentative parts are needed to have an appropriate certainty when distinguishing arguments from statements. An isolated argumentative sentence is hard to distinguish from a simple statement, for example the following sentence: "Councilman Smith voted in favour of the tax increase." does not look like an argumentative sentence by itself. However, placed on the right context: "Councilman Smith voted in favour of the tax increase. No one who voted in favour of the tax increase is a desirable candidate. Therefore, Councilman Smith is not a desirable candidate." it is completely clear that it is part of an argument. Furthermore, argumentation models also describe an argument as a group of non-overlapping elementary textual units with relationships between them, see [26, 3].

For these reasons it is assumed that the at hand model of the typical flow of argumentation can predict typical patterns in the given texts. It is possible to assume that a sentence is more likely to be of category CONCLUSION, for example, if the previous sentence was a PREMISE, than if the previous sentence was a NON-ARGUMENTATIVE sentence, even if it is not known anything about the features of the sentence to be classified yet.

N-gram models estimate a more accurate prior by taking the context of a sentence, in terms of surrounding categories, into account. N-gram models are typically used over letters in statistical language processing, but here they are applied to whole sentences instead. The prior can then be written as a conditional probability of the category y_i for the ith sentence in the document, $P(y_i|y_{i-1}, ..., y_{i-N})$, instead of the general $P(y_i)$. The index $N+1$ is called the *order* of the N-gram model. A system of order $N+1$ takes N items before the one to be classified into account, a bigram model ($N+1=2$) uses the formula $P(y_i|y_{i-1})$.

8.3.2 Features

As in the Araucaria corpus (see Section 8.2.1) every word in the ECHR corpus is tagged with its part-of-speech by an automatic tagger (QTag), and a syntactic dependency tree is constructed for every sentence by an automatic parser (Charniak [8]). However, to preserve the knowledge related to context the following features are added to the baseline features described in Section 8.2.1:

- **Unigrams in previous/next sentences** Each word in the N previous/next sentences. This simple feature can be considered as a baseline for more specific features. Punctuation marks (,;:.?!) are not included in the basic version. Unigrams have been proved a useful feature in other areas of text classification.
- **Bigrams in previous/next sentences** Each pair of successive words in the N previous/next sentences. As unigrams, they have also been proved useful in previous text classification research.
- **Word couples in previous/next sentences** All possible combinations of two words in each previous/next sentence are considered. This approach captures more context than bigrams, at the expense of increasing the feature vector size substantially.
- **Adverbs in previous/next sentences** Adverbs are detected with a POS tagger. The presence of words like "*Unfortunately*" could be representative for conclusions or other kinds of argumentative sentences.
- **Verbs in previous/next sentences** Verbs are also detected with the POS tagger. Only the main verbs (excluding "*to be*", "*to do*" and "*to have*") are considered. The presence of concrete verbs in the context around the sentence could be significant for the type of argumentative sentence.

- **Modal auxiliary in previous/next sentences** A binary feature that indicates if a modal auxiliary is present in each of the previous/next sentences. Modal auxiliary could be more present in argumentative sentences than in non-argumentative sentences.

- **Text statistics in previous/next sentences** Average previous/next sentence length (the sentences around an argumentative sentence form a more probable part of the same argument, so they should be longer, as it was proved in [17] that argumentative sentences tend to be longer than non-argumentative sentences). Average previous/next word length ("difficult" words might appear around the argumentative sentences). Average previous/next number of punctuation marks (the presence of argumentation may increase the amount of punctuation needed).

- **Punctuation in previous/next sentences** The possible patterns that could appear previously to an argumentative sentence are studied. For example the presence of a big amount of commas in the previous/next sentences could indicate that there was an explanation of facts before starting the argument.

- **Keywords in previous/next sentences** Keywords refer to 286 words or word sequences from a list of terms indicative for argumentation [28]. Examples from the list are "*but*", "*consequently*" and "*because of*". The presence of different keywords in the previous/next sentences could determine different argumentative patterns.

- **Negative/positive previous/next sentences** The presence of the word "*not*", in all its possible appearances, e.g. "*don't*" = "*do not*", "*won't*" = "*will not*", is studied. The presence or lack of this word in the previous/next sentences could give some kind of argumentative patterns for positive/negative behaviour around argumentation.

- **First/last words in previous/next sentences** The first or last words of a sentence should be a connector with the next/previous sentence. The connectors between the N previous/next sentences are collected in this feature to study the connections between argumentative and non-argumentative sentences.

- **Same words in previous/next sentences and current sentence** Sentences inside the same argument should talk about similar things, therefore they should contain similar words. This feature informs of the appearance or lack of the same words in the previous/next sentences and the current one. Words with a length smaller than four are not analysed, as it is considered that words like "*a*", "*in*" or "*the*" don't express real connections between sentences.

8.3.3 Evaluation

Experiments with N-gram models of order 2, 3 and 4 were performed. Table 8.3 shows the baseline results, i.e. N=1 (without associated context) and features *Word couple + Text statitics + Verbs*. The baseline classification has an average accuracy slightly below 82%. Tables 8.4, 8.5 and 8.6 show the results depending on the order of the N-gram. The results found on these experiments show a clear improvement on the classification of sentences as argumentative and non-argumentative, with precision higher than 80% and recall over 90%. However, the model does not allow a good classification between premises and conclusions. It is important to notice that even if these results are low, it is observed that there is not much difference when working with previous or next sentences, in fact in some cases the results are the same.

Note that the classification between premises and conclusions is one of the most complex tasks given the nature of legal cases where subordinated argumentative structures are a frequent occurrence. Subordinated structures present a final conclusion and a set of sub-arguments and premises to justify it. Each sub-argument has a conclusion. These sub-conclusions are sometimes not expressed as clearly as the final conclusions. There can be different reasons behind their ambiguous presentation. For example, the writer might have been focused on the "general" argument and thus uses rhetorical markers to emphasize the connection between the sub-conclusion and the final conclusion. This relation is obviously a relation of type premise. However, the sub-conclusion is nevertheless a conclusive proposition and should be detected as such.

A second set of experiments was performed with different combinations of features, some from previous sentences and some from next sentences, however they yield results similar to the ones presented in Tables 8.4, 8.5 and 8.6. The combination of: word couples of the previous sentences, word couples, verbs and text statistics for the current sentence, and bigrams for the next sentences, with an order N=4 presented the best results of these combinations. An accuracy of 89.13% was obtained. This was due to the increment of the precision in the detection of conclusions, which was 68.54%. However, the recall did not improve and therefore the $F1$ measure for conclusion detection remained in the same range than before around 24.26%. Therefore, it seems that there is no combination of N sentences which would achieve a reliable classification between premises and conclusions.

One possible reason behind this limitation could be the lack of training examples. It is difficult to find argumentative texts with a high number of conclusions, therefore it is hard to learn how these sentences are represented. However, we are more in favour of a second reason, the shallowness of the

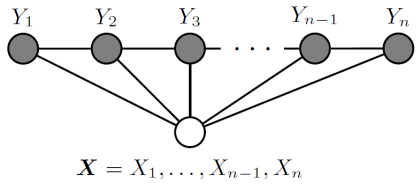

Figure 8.2: Graphical structure of a chain-structured CRFs for sequences [68]

features. It is argued that more specific features, which could capture the main differences between a premise and a conclusion, could help to improve the current results.

During the experiments with N-gram models it was also initially tested a Conditional Random Fields (CRF) approach. However, this approach performed similar to the other models. CRF is a discriminative undirected probabilistic graphical model [68], globally conditioned on X, the random variable representing observation sequences. Formally, $G = (V, E)$ is defined to be an undirected graph such that there is a node $v \in V$ corresponding to each of the random variables representing an element Y_v of Y. If each random variable Y_v obeys the Markov property with respect to G, then (Y, X) is a conditional random field. In theory the structure of graph G may be arbitrary, provided it represents the conditional independencies in the label sequences being modeled. However, when modeling sequences, the simplest and most common graph structure encountered is that in which the nodes corresponding to elements of Y form a simple first-order chain, as illustrated in Figure 8.2.

8.4 Improving the baseline: part II

Section 8.3 has improved the baseline model (see Section 8.2) applying a N-gram model. This was achieved by adding features of the context of a sentence (see Section 8.3.2) to the features of the sentence itself (see Section 8.2.1). However, even if this new model improves the classification between argumentative and non-argumentative sentences, the distinction between premise and conclusion presented low results. It was concluded that the specialization of the features to the argumentative domain would give a more accurate classification between premise and conclusion.

A possible way to obtain more specialized features is to study the behaviour of the current features depending on the different argumentative functions of

	Previous Sentences				Next Sentences			
	N=1	N=2	N=3	N=4	N=1	N=2	N=3	N=4
Unigrams	35.46	38.69	40.73	42.81	35.46	37.58	39.24	41.59
Bigrams	35.71	48.11	52.36	67.07	35.71	48.47	51.44	65.22
Word couples	38.31	67.63	71.58	74.17	38.31	66.29	70.34	73.1
Verbs	30.21	41.29	41.66	42.18	30.21	41.04	40.67	40.91
Adverbs	30.06	40.74	40.65	40.27	30.06	41.01	40.87	40.64
Modal auxiliary	31.20	40.65	40.62	40.59	31.20	40.64	40.48	40.35
Punctuation	31.16	40.57	40.44	40.36	31.16	39.09	39.09	39.08
Keywords	32.07	39.27	39.07	38.77	32.07	41.04	40.67	40.91
Text statistics	31.14	39.31	40.08	40.11	31.14	38.87	38.92	38.98
Negative	28.73	40.06	40.27	40.4	28.73	39.44	39.44	39.36
Same words	-	39.08	39.16	38.88	-	39.68	39.66	39.63
First words	-	41.33	41.34	41.33	-	41.33	41.33	41.34
Last words	-	40.86	40.86	40.87	-	39.66	39.54	38.86

Table 8.4: $F1$ measure from the classification of *Premises* in the ECHR depending on N-gram order. Baseline is 40.12%

a sentence. For example, Section 8.3.2 studied the verbs presented in the current, next and previous sentences, but it has not separated these verbs by their argumentative function. Verbs, such as "*conclude*" or "*decide*", have a higher chance of appearing in a conclusion, while verbs, such as "*present*" or "*note*", would occur normally in premises.

Given the previous good results on the classification between argumentative and non-argumentative, it was decided to perform classification in two steps. First, a classifier that uses shallow features, such as *Word couples*, *Text statistics* and *Verbs*, and the N-gram model. Then, the results obtained by this classifier are input in a second classifier, that uses the more specific features presented in Section 8.4.1.

The task at hand is formalized as: given a training set $(x_1, y_1), \ldots, (x_n, y_n)$, where x_i denotes a vector of observed features for each sentence of the corpus and y_i denotes a human annotated class label, cf. argumentative or non-argumentative. Train a classifier ϱ that for a new observation x_j determines its label $\hat{y_j} = \varrho(x_j) \approx y_j$. For each observation, add $\hat{y_j}$ to its feature vector x_j. Then, train a second classifier ϱ' that for the observation x_j determines its label $\hat{y'_j} = \varrho'(x_j) \approx y_j$, where y_j denotes premise, conclusion or non-argumentative.

	Previous Sentences				Next Sentences			
	N=1	N=2	N=3	N=4	N=1	N=2	N=3	N=4
Unigrams	18.15	21.03	20.25	20	18.15	19.83	20.77	19.63
Bigrams	20.31	24.45	24.7	23.53	20.31	25.35	26.64	22.94
Word couples	21.06	32.73	27.81	24.54	21.06	28.09	23.72	24.35
Verbs	18.75	21.86	21.22	21.27	18.75	22.18	21.54	21.54
Adverbs	20.06	22.49	22.13	22.22	20.06	22.54	22.49	22.18
Modal auxiliary	19.54	22.18	21.82	21.82	19.54	21.82	22.18	21.82
Punctuation	19.75	22.18	22.18	21.82	19.75	18.26	18.26	18.27
Keywords	20.34	21.05	21.18	21.18	20.34	22.18	21.54	21.54
Text statistics	19.32	21.18	21.54	21.5	19.32	18.07	18.03	18.37
Negative	18.15	21.82	21.82	21.82	18.15	18.26	18.26	18.26
Same words	-	18.3	18.33	17.92	-	21.54	21.18	21.22
First words	-	22.49	22.49	22.48	-	22.49	22.49	22.49
Last words	-	22.13	22.13	22.13	-	21.46	21.46	20.77

Table 8.5: F-1 measure from the classification of *Conclusion* in the ECHR depending on N-gram order. Baseline is 22.04%

8.4.1 Features

As in Section 8.3.2 the ECHR corpus is tagged with its part-of-speech by an automatic tagger and a syntactic dependency tree is constructed for every sentence by an automatic parser [67]. The features for the first classifier are: baseline features (*word couples*, *text statistics* and *verbs*) and the N-gram model features. The second classifier uses the following features:

- **Absolute location** Position of the sentence in the whole document. The document is divided in 7 segments, therefore feature value $\in$ [0,1,2,3,4,5,6].
- **Sentence length** A binary feature, which indicates that the sentence is longer than a threshold number of words (currently 12 words).
- **Tense of main verb** Tense of the verb from the main clause of the sentence; having as nominal values “Present”, “Past” or “NoVerb”.

- **History** The most probable argumentative category (among the 5 categories) of previous and next sentences.
- **Information 1st classifier** The sentence has been classified as argumentative or non-argumentative by a first classifier.
- **Rhetorical patterns** Type of rhetorical pattern occurring in current, previous and next sentences (e.g. "*however,*"); 5 types (Support, Against, Conclusion, Other or None) have been distinguished. These types were defined after the assumption that markers to introduce supportive propositions, contrasts or conclusions should appear frequently on argumentative propositions.
- **Article reference** A binary feature indicating whether the sentence contains a reference to an article of the law, detected with a POS tagger [67]. For example the proposition "*It follows that this part of the application is also manifestly ill-founded within the meaning of Article 27 para. 2 (Art. 27-2) of the Convention.*" contains a reference to article 27. The reference is expressed in two manners, "*Article 27 para. 2*" and "*Art. 27-2*". Both manners, together and individually, are detected with the POS tagger and a posterior automatic matching process.
- **Article** A binary feature indicating that the sentence includes the definition of an article detected again with the help of a POS tagger [67]. For example proposition "*The applicant invokes Article 6 paras. 1, 2, 3 (b) and (d) (Art. 6-1, 6-2, 6-3-b, 6-3-d) of the Convention, which provides certain guarantees to the defence in a criminal case, and Article 14 (Art. 14), which prohibits discrimination in the securement of Convention rights and freedoms.*" contains different article references but also some article definitions, such as "*which provides certain guarantees to the defence in a criminal case*" and "*which prohibits discrimination in the securement of Convention rights and freedoms*".
- **Argumentative patterns** Type of argumentative pattern occurring in sentence; 5 types of patterns have been distinguished in accordance with our 5 categories (Support, Against, Conclusion, Other or None). An example of a support pattern is "*see, mutatis mutandis,*", while "*having reached this conclusion*" and "*by a majority*" are examples of conclusive patterns.
- **Type of subject** The agent of the sentence is the applicant, the defendant, the court or other. The type of agent is detected with the POS tagger.

- **Type of main verb** Argumentative type of the main verb of the sentence; 4 types (premise, conclusion, final decision or none) have been distinguished, implemented as a list of corresponding verbs, which are detected in the text also with a POS tagger [67].

	Previous Sentences				Next Sentences			
	N=1	N=2	N=3	N=4	N=1	N=2	N=3	N=4
Unigrams	89.06	89.97	90.11	90.25	89.06	89.9	90.02	90.23
Bigrams	89.73	90.87	91.37	93.14	89.73	90.92	91.33	92.92
Word couples	89.85	93.27	93.81	94.14	89.85	93.04	93.57	94.06
Verbs	89.02	90.09	90.17	90.25	89.02	90.06	90.05	90.08
Adverbs	86.43	90.04	90.02	90.01	86.43	90.05	90.03	90.03
Modal auxiliary	80.93	89.99	89.99	89.99	80.93	89.99	89.98	89.96
Punctuation	87.96	89.98	89.96	89.96	87.96	89.8	89.8	89.81
Keywords	87.27	89.88	89.89	89.85	87.27	90.06	90.05	90.08
Text statistics	88.63	89.92	89.98	89.96	88.63	89.79	89.79	89.8
Negative	81.46	89.96	89.97	89.99	81.46	89.81	89.81	89.8
Same words	-	89.85	89.84	89.81	-	89.97	89.95	89.96
First words	-	90.02	90.02	90.02	-	90.02	90.02	90.03
Last words	-	90.02	90.02	90.03	-	89.94	89.94	89.89

Table 8.6: $F1$ measure from the classification of *Non-argumentative* in the ECHR depending on N-gram order. Baseline is 89.83%

Type	Precision	Recall	F1
Conclusion	77.49%	60.88%	74.07%
Premise	70.19%	66.16%	68.12%

Table 8.7: Classification results of the ECHR corpus between premises and conclusions. The classification is done in two steps: (i) classification between argumentative and non-argumentative with a Maximum Entropy classifier, (ii) classification between premise and conclusion with a Support Vector Machine using the results of the first classifier as an input feature.

8.4.2 Evaluation

Table 8.7 shows the best results for the classification between premise and conclusion attaining a 68.12% and 74.07% $F1$ measure respectively. This is an important improvement to the results in Section 8.3 where the conclusion classification had around 21% $F1$ measure. Figure 8.3 shows an example of the classifier output. The classification results achieve a general accuracy slightly above 83%.

1. The applicant, who was detained between 26 January and 30 March 1990, complains that this deprivation of liberty was contrary to Article 5 paras. 1, 3 and 4 (Art. 5-1, 5-3, 5-4) of the Convention. | -1
Article 5 para. 1 (Art. 5-1) of the Convention guarantees the right to liberty and security of person, subject to certain exceptions, such as the lawful detention of a person after conviction by a competent court, within the meaning of sub-paragraph (a) of the provision. | 0
Article 5 paras. 3 and 4 (Art. 5-3, 5-4) provide certain guarantees of judicial control of provisional release or detention on remand pending trial. | 0
The Commission notes that the applicant was detained after having been sentenced by the first instance court to 18 months' imprisonment. | 0
He was released after the Court of Appeal reviewed this sentence, reducing it to 15 months' imprisonment, convertible to a fine. | 0
The Commission finds that the applicant was deprived of his liberty "after conviction by a competent court" within the meaning of Article 5 para. 1 (a) (Art. 5-1-a) of the Convention. | 1
The Commission also finds no evidence in the case to suggest an infringement of paragraphs 3 and 4 of Article 5 (Art. 5-3, 5-4): | 1
The applicant was not detained on remand prior to his trial and the judicial control of the lawfulness of his subsequent detention after conviction was provided by the first instance court (cf. Eur. Court H.R., De Wilde, Ooms and Versyp judgement of 18 June 1971, Series A no. 12, p. 40, para. 76). | 0
It follows that this part of the application is manifestly ill-founded and must be rejected in accordance with Article 27 para. 2 (Art. 27-2) of the Convention. | 1

Figure 8.3: ECHR legal case fragment annotated output by a naïve Bayes classification model (0:Premise, 1:Conclusion, -1:Non-argumentative)

8.5 Summary

In this chapter it has been discussed how to approach the automatic detection of argumentation based on statistical classification methods. It has been shown that argumentation detection can be achieved by a binary classification

of all text sentences as argumentative and non-argumentative. Classifying each sentence of the text as part of the argumentation or not it is possible to constitute all the arguments of the text with the detected argumentative sentences.

A baseline for argumentative vs. non-argumentative classification has been established with the use of shallow features typical of NLP. The baseline yields 89.83% $F1$ measure in a legal corpus, see Table 8.3.

The baseline has been improved by two extensions of the basic classifier. First, it has been used the relevance of contextual information increasing the results closely to 94% $F1$ measure. Then, the focus has been on the classification between premises and conclusions, where it has been studied the importance of using specific features of argumentation. Results have then achieved 68.12% and 74.07% $F1$ measures respectively. Therefore, it has been proved the classification of argumentative sentences on a legal domain is feasible with results close to the human annotation level, which stand at an agreement of K=0.75 (see Chapter 6).

A major limitation of this approach is the impossibility of detecting the delimiters of each argument. It is known which information forms the argumentation, but not how this information is split into the different arguments. This segmentation problem could be solved by a semantic analysis of the different arguments of the text. Some minor semantic analysis using state-of-the-art clustering models was undertaken but the results were not favourable, all the different arguments of the legal case were closely semantically related, or they contained not enough semantic information to allow a topic differentiation. These results were in accordance with the a priori expectations given the linguistic characteristics of legal cases. As shown in Chapter 4, a legal case is formed by different arguments which justify a final decision. All the arguments are closely related as normally there is just one final point to be justified (e.g. guilty vs. non-guilty), and most arguments are formed as a response to previous arguments. Moreover, arguments are just formed by a few sentences which contain not enough information to detect a main topic. A different type of document or a more complex clustering model could achieve better results, however it was decided to leave this research line for future work.

Chapter 9

Rule-based parsing

In this chapter a second approach to automatic argumentation detection is presented. This approach is based on the work in document structure analysis presented in Chapter 7, and is especially inspired by the work of Daniel Marcu on automatic detection of discourse relations, see [37].

It is argued that argumentative discourse is just another type of discourse, therefore it should be possible to find a structural model to segment it. The problem of segmenting discourse into the elementary units appropriate for building up the structure of the discourse is an extremely difficult one. Each discourse theory must specify how "segments" should be identified in light of the questions the theory is set up to answer (see Chapter 7). Once the "segments" and the "relations" between them are analysed it is possible to define a grammar which can be used to parse new documents.

This chapter presents a description of a set of terminal and non-terminal symbols ("segments") and rules ("relations") to form a grammar that describes the argumentation structure of a legal case (Section 9.2). The specification of how the argumentative "segments" and the argumentative "relations" between them were identified is presented in Section 9.3. At the end, in Section 9.4, the grammar is tested over the ECHR corpus and evaluated.

9.1 Evaluation metrics

In order to separate the evaluation of parsing (tree-structure) and part of argument tagging (symbols) different metrics are used. Precision is the number of correct constituents (see Section 2.3.2) produced by the parser divided by the total number of constituents produced by the parser. Recall is the number of correct constituents produced by the parser divided by the total number of constituents in the gold standard parse tree. It is considered that a constituent is correct if there is a constituent in the gold standard parse tree which dominates the same sequence of terminal symbols.

Other metrics relate to argument-structure evaluation, i.e. evaluation based on argument-structure constituents, such as premise-conclusion relationships, which are important for correctly capturing the reasoning behind the structure. It also penalises certain attachment errors less harshly, specifically those related to "grouping" coordinate segments of an item (e.g. premises of an argument), as this is too sensitive to the annotation framework but does not modify the general reasoning. In general, the metric, which is named TC (*tree correctness*), is computed as follows:

- Check if final decision is correctly detected (+1)
- Check if all arguments are detected (cf. an argument is detected if the conclusion and at least one premise are detected) (+1)
- Check for each detected argument if it is complete (cf. all premises + conclusion detected) (-0.5 for each incomplete detected argument)
- Check for each detected argument if the structure is correct (cf. relations between subordinate elements detected, it is not considered as error the differences between internal structure of coordinate elements) (-0.1 for each incorrect structure of detected argument)

Therefore the normalized TC is as follows:

$$TC = f_{arg} + \frac{d_{arg} - (c_{arg} * \frac{1}{2}) - (is_{arg} * \frac{1}{10})}{n_{arg}} \tag{9.1}$$

where $f_{arg} = 1$ if final decision detected, $f_{arg} = 0$ if final decision not detected, d_{arg} = number of detected arguments, c_{arg} = number of detected arguments that are incomplete, is_{arg} = number of detected arguments with incorrect structure and n_{arg} = total number of arguments in the ground truth legal case.

Values for TC metric range from 0 to 2. Any parse tree with a correct root, cf. final decision, will have a $TC \geq 1$. For each argument detected TC will increase in the worse case by $\frac{2}{5n_{arg}}$ and in the best case by $\frac{1}{n_{arg}}$. It is considered that a parse tree where the final decision and all arguments have been detected, but all arguments are incomplete and wrongly structured, is the low threshold. The high threshold is 2, cf. parse tree with final decision and all argument detected, with all arguments complete and well-structured. Therefore, a parse tree is considered "correct" if:

$$1 + \frac{d_{arg} * \frac{4}{10}}{n_{arg}} \leq TC \leq 2 \tag{9.2}$$

Note that the incorrect detection of the structure of an argument can have an influence on the general precision and recall values. For example, a subordinate structure that has been detected as coordinate implies that a "middle" conclusion has been wrongly detected as premise. Therefore, there is a decrease on the precision and recall. On the other hand TC would be also double penalized. First, by the incompleteness of the sub-argument, cf. argument has no conclusion, and second, by the incorrectness of the super-argument, cf. argument has wrong structure.

9.2 Argumentative grammar

This sections presents a grammar to accomplish the parsing of a legal case by its argumentative structure. The chosen grammar is a Left to Right (LR) grammar (see Chapter 2). The LR argumentative grammar presented in Figure 9.1 was developed using theoretical knowledge of argumentation and the empiric observation of a subset of the ECHR corpus, 10 legal cases. This development process is analysed in Section 9.3, where definitions and motivations for all terminal and non-terminal symbols are provided.

The language of the grammar for real case law argumentation comprises all the possible argumentative structures of a case law document, in this case from the ECHR documents. However, the grammar could be generalized further to also accept texts from other courts. The grammar works on a sentence level. The symbols used for describing the grammar are shown in Table 9.1.

T	General argumentative structure of legal case
A	Argumentative structure that leads to a final decision of the factfinder $A = \{a_i, ..., a_j\}$, each a_i is an argument from the argumentative structure
D	The final decision of the factfinder $D = \{d_i, ..., d_j\}$, each d_i is a sentence of the final decision
P	One or more premises $P = \{p_i, ..., p_j\}$, each p_i is a sentence classified as premise
P_{verbP}	A premise with at least one instance of a verb related to function premise.
P_{art}	A premise with at least one reference to an article.
P_{sup}	A premise with at least one instance of a rhetorical marker related to the concept of support.
P_{ag}	A premise with at least one instance of a rhetorical marker related to the concept of contrast.
C	Sentence with a conclusive meaning
n	Sentence, clause or word that indicates one or more premises will follow
s	Sentence, clause or word neither classified as a conclusion nor as a premise ($s! = \{C\|P\}$)
r_c	Conclusive rhetorical marker (e.g. therefore, thus, ...)
r_s	Support rhetorical marker (e.g. moreover, furthermore, also, ...)
r_a	Contrast rhetorical marker (e.g. however, although, ...)
r_{art}	Article reference (e.g. terms of article, art. para. ...)
v_p	Verb related to a premise (e.g. note, recall, state,...)
v_c	Verb related to a conclusion (e.g. reject, dismiss, declare, ...)
f	The entity providing the argumentation (e.g. court, jury, commission, ...)

Table 9.1: Terminal and Non-terminal symbols from the ECHR grammar

$$T \Rightarrow A^{+}D \tag{9.3}$$

$$A \Rightarrow \{A^{+}C|A^{*}CnP^{+}|Cns|A^{*}sC|P^{+}\} \tag{9.4}$$

$$D \Rightarrow r_c f\{v_c s|.\}^{+} \tag{9.5}$$

$$P \Rightarrow \{P_{verbP}|P_{art}|PP_{sup}|PP_{ag}|sP_{sup}|sP_{ag}\} \tag{9.6}$$

$$P_{verbP} \Rightarrow sv_p s \tag{9.7}$$

$$P_{art} \Rightarrow sr_{art}s \tag{9.8}$$

$$P_{sup} \Rightarrow \{r_s\}\{s|P_{verbP}|P_{art}|P_{sup}|P_{ag}\} \tag{9.9}$$

$$P_{ag} \Rightarrow \{r_a\}\{s|P_{verbP}|P_{art}|P_{sup}|P_{ag}\} \tag{9.10}$$

$$C \Rightarrow \{\{r_c|r_s\}\{s|C|r_c^{*}P_{verbP}\}|s^{*}v_c s\} \tag{9.11}$$

Figure 9.1: Generalized argumentation grammar

9.3 Grammar development

General theoretical knowledge on argumentation and the analysis of ten random legal cases from the ECHR corpus is the basis to develop a grammar for real case law argumentation. The development task consists on defining which elements would be the terminal and non-terminal symbols of the grammar and constructing the rules which allow to move from one symbol to another.

As established in Section 2.2 argumentation of legal cases presents some specific characteristics. The selected 10 random legal cases from the ECHR corpus ("training" set) help to determine how these characteristics define possible formats for every different type of proposition in legal cases, e.g. *premise* or *conclusion*.

Lists of rhetorical markers were created from the observations in the "training

set". There was a list related to the concept of support (r_s), a list related to the concept of contrast (r_a), a list of related to the concept article (r_{arg}) and a list related to the concept of conclusion (r_c). Lists for verbs related to function premise (v_p) and function conclusion (v_c) were also created. The lists contain the following elements:

- r_c = *in conclusion*, *therefore*, *thus*, *in the light of*, *accordingly*, *consequently*, *in these circumstances*, *having reached this conclusion*, *in short*, *in such a situation*, *it follows that*, *thereafter*, *in view of the above*, *in sum*, *in view of [the gravity] of this conclusion*, *having reached this conclusion*, *altogether*, *overall*, *all in all*, *to conclude*, *hence*, *as a consequence*, *as a result*, *that implies*
- r_a = *however*, *although*, *on the other hand*, *in contrast*, *nevertheless*, *nonetheless*, *neither*, *nor*, *otherwise*, *this is contrary to*, *on the contrary*, *though*, *in spite of that*, *meanwhile*
- r_s = *in the present case*, *indeed*, *in particular*, *moreover*, *in addition*, *also*, *further*, *furthermore*, *in the instant case*, *in the case of*, *on the other hand*, *on the whole*, *in general*, *such*, *above all*, *see, mutatis*, *similarly*, *in other words*, *alternately*, *that is to say*, *for example*, *for instance*, *in particular*
- r_{art} = *article*, *para.*, *art.*, *under the terms of*, *paras.*, *cf.*
- v_c = *find*, *conclude*, *decide*, *accept*, *reject*, *dismiss*, *hold*, *declare*, *see no reason*, *see no cause*, *see no objection*, *be (dis)satisfied*, *be (un)convinced*, *establish*, *decide*, *deem*
- v_p = *consider*, *note*, *recall*, *agree*, *disagree*, *reiterate*, *acknowledge*, *is of the opinion*, *point out*, *emphasise*, *stress*, *is of the view*, *is satisfied*, *observe*, *recognise*, *endorse*, *point out*, *emphasize*, *state*, *reiterate*, *be persuaded*, *requires*

Note that the presence of verbs was studied in different verb tenses, even with the use of auxiliary verbs. Once these lists were created it was possible to move onto the definition of the most basic non-terminals of the grammar, P_{art}, P_{VerbP}, P_{ag} and P_{sup}, as follows:

- A premise with a reference to an article (P_{art}). For example, "*The relevant part of Article 10 (Art. 10) of the Convention provides as follows:*".
- A premise with an occurrence of a verb of premise (P_{verbP}). For example, "*The Commission notes that the applicant's conviction involved his writings.*".

- A premise with an occurrence of a rhetorical marker of concept contrast (P_{ag}). For example, "*However, this is far from clear at the present stage of the proceedings, and it is therefore impossible to reject the application on this ground.*".
- A premise with an occurrence of a rhetorical marker of concept support (P_{sup}). For example, "*In addition, the examination of the case has disclosed no circumstance which, according to the generally recognised principles of international law, might have absolved the applicant from raising these complaints during the cassation proceedings.*".

Then it is possible to determine some more complex formats for premise sets as follows. A premise (P) can be:

- One or more premises followed by a premise with a support reference (PP_{sup}). For example, "*The Commission notes the applicant's declaration that he had told the public prosecutor that he had been tortured. Moreover, when asked to sign a statement, he had answered that he could not sign because he could not move his hands.*".
- One or more premises followed by a premise with a contrast reference (PP_{ag}). For example, "*The Court finds it established that there was a causal link between the anxiety and distress suffered by the applicant and the breach found of the Convention. However, in the circumstances of the case, the Court considers that this finding constitutes adequate just satisfaction in respect of the damage claimed under this head.*".
- A sentence, clause or word neither classified as a conclusion nor as a premise followed by a premise with a support reference (sP_{sup}). For example, "*No evidence has been adduced before the Court to show that any other action was taken, despite the prosecutor's awareness of the applicant's injuries. Moreover, in the Court's view, in the circumstances of Mr Aksoy's case, such an attitude from a State official under a duty to investigate criminal offences was tantamount to undermining the effectiveness of any other remedies that may have existed.*".
- A sentence, clause or word neither classified as a conclusion nor as a premise followed by a premise with a contrast reference (sP_{ag}). For example, "*The Court has taken account of the unquestionably serious problem of terrorism in South-East Turkey and the difficulties faced by the State in taking effective measures against it. However, it is not persuaded that the exigencies of the situation necessitated the holding of the applicant on suspicion of involvement in terrorist offences for fourteen days or more*

in incommunicado detention without access to a judge or other judicial officer.".

A simple conclusion can be defined using the list of verbs and rhetorical markers related to the concept/function conclusion as follows. A conclusion (C) can be:

- A sentence, clause or word neither classified as a conclusion nor as a premise containing a conclusive verb and a conclusive verb ($r_c s^* v_c s$). For example, "*It follows that this part of the application cannot be dismissed as manifestly ill-founded within the meaning of Article 27 para. 2 (Art. 27-2) of the Convention.*".
- A conclusive rhetorical marker followed by a sentence, clause or word neither classified as a conclusion nor as a premise ($r_c s$). For example, "*In conclusion, there has been a violation of Article 13 of the Convention (art. 13).*".
- A conclusive rhetorical marker followed by a premise with a verb of premise ($r_c P_{verbP}$). For example, "*Having reached this conclusion it does not consider it necessary to examine the applicant's claim that there exists an administrative practice of withholding remedies in breach of the Convention.*".
- A support rhetorical marker followed by a conclusion ($r_s C$). For example, "*Further, the Commission finds that the applicants' choice of pursuing proceedings in the courts of Northern Ireland was not unreasonable or without basis in domestic law.*".

Once premises and conclusions have been defined it is necessary to define arguments. An argument is understood as the minimum unit of argumentation. There are *simple argument*, text spans that communicates information about not more than one argument each. But there can be also *complex arguments*, where a text span communicates information about more than one argument. Examples of arguments are given in Figure 9.2.

Inside each argument premises and conclusions are structured in different ways. In this work, it is understood that a premise or conclusion is represented by a single sentence. Therefore, each argument can be identified as a set of ordered sentences, where each sentence is identified as premise or conclusion. The formal definition of simple argument is as follows: given an argumentative discourse T formed by n argumentative segments, $\{s_1, \ldots, s_n\}$, s_i is a simple argument if s_i is formed by a set of ordered sentences $\{prop_1, \ldots, prop_m\}$ if there exists j such as $prop_j =$ conclusion and $\forall k \neq j, k \in [1, \ldots, m] : p_k$ is a

premise. A complex argument definition is as follows: given an argumentative discourse T formed by n argumentative segments, $\{s_1, \ldots, s_n\}$, s_i is a complex argument if s_i is formed by a set of ordered units $\{u_1, \ldots, u_m\}$, where there exists j such as $u_j =$ conclusion, there exists l such as $u_l =$ argument and $\forall k \neq j \neq l, k \in [1, \ldots, m] : u_k$ is either a premise or an argument. Note that the order of the sentences or units in the structure of the argument yield to distinctive argument structures.

Argument 1
It is not always an easy matter to trace the dividing line between procedural and substantive limitations of a given entitlement under domestic law. It may sometimes be no more than a question of legislative technique whether the limitation is expressed in terms of the right or its remedy. In the present case, the Court does not consider it necessary to settle the question of the precise nature of the defence of privilege for the purposes of Article para. art., since it is devoid of significance in the particular circumstances. If the Court were to treat the facts underlying the complaints declared admissible by the Commission as raising a substantive, rather than a procedural, complaint going to the right to respect for private life under Article 8 of the Convention (...) the same central issues of legitimate aim and proportionality as under Article 8 para. 2 would be posed.

Argument 2
There will always be a need for elucidation of doubtful points and for adaptation to changing circumstances. Indeed, in the United Kingdom, as in the other Convention States, the progressive development of the criminal law through judicial law-making is a well entrenched and necessary part of legal tradition.

Figure 9.2: Examples of argumentative discourse segments, i.e. arguments, in the ECHR

The possible structures are defined as follows. An argument (A) can be formed by:

- Other arguments and a conclusion (A^+C). See Figure 9.3.
- Other (none) arguments, a conclusion, a nexus and one or more premises (A^*CnP^+). A nexus can be different types of elements such as "*see eg.*", "*cf.*", "*as follows*", see Figure 9.4, or a full sentence, see Figure 9.5.
- A conclusion, a nexus and a sentence, clause or word neither classified as a conclusion nor as a premise (Cns). See Figure 9.6.
- Other (none) arguments, a sentence, clause or word neither classified as a conclusion nor as a premise and a conclusion (A^*sC). See Figure 9.7.

- One or more premises (P^+). See Figure 9.8.

Note that premises are joined in "arguments" (A) and then connected to a conclusion (C). This was done to try to distinguish between coordinative and multiple premises. For example, Figure 9.9 shows how two sets of premises justify a conclusion using multiple argumentation while Figure 9.10 shows how two sets of premises justify a conclusion using coordinative argumentation. Moreover, the option to define an argument (A) as a single premise, allows to connect independent premises that justify the final decision directly. The definition of (P) formed by a single (A) was not considered to diminish loop effects.

The last layer of the structure defines how the arguments interact between themselves and the final decision. The decision (D) can be formed by one or more sentences that contain: (i) conclusive rhetorical marker and reference to the entity providing the argument or (ii) the previous and conclusive verb. An argument (A), or more, and a decision (D) form the tree-structure (T) of the legal case. The relations forming the tree-structure are represented by the following rules:

- $T \Rightarrow A^+D$. For an example see Figure 9.12.
- $D \Rightarrow r_c f\{v_c s|.\}^+$. For example, *"DECLARES THE APPLICATION ADMISSIBLE, without prejudging the merits of the case."*.

9.4 Evaluation

The grammar was implemented using Java and JSCC[1], a parser development system. The grammar was tested over the documents from the ECHR corpus not included in the training set (10 documents) used to develop the grammar, i.e. 44 documents. The documents were pre-processed to adapt them to the grammar terminal symbols. A set of rules were described in Java as to transform the propositions of the documents in symbols such as s, v_c. A rule invocation order was used for propositions which could fit two or more grammar rules, e.g. a proposition with the following structure sv_psv_cs could be a P_{verbP} or a C. The preference order was given first to decisions, then conclusions and then premises, where the order was first P_{verbP} and then P_{art}. The order between r_{sup}, r_{ag} and r_c is first C, then P_{sup} and finally P_{ag}. All sentences that at this point were identified as s and were not directly situated next to

[1] http://jscc.jmksf.com/

The Commission considers that while the Inspectors' report was being studied by the Director of Public Prosecutions (DPP) and the Director of the Serious Fraud Office (SFO), the applicants could have considered themselves affected by a possible criminal prosecution and therefore, "charged", within the meaning of Article 6 para. 2 (Art. 6-2) of the Convention (cf. the Court's aforementioned Deweer judgment of 27 February 1990, p. 24, para. 46, and the Adolf judgment of 26 March 1982, p. 15, para. 30). However, once these officials had issued their public statements on 1 March 1990 that the applicants would not be prosecuted, there can, in the Commission's view, be no question of the applicants being deemed to have been "charged" any longer. In these circumstances the Commission concludes that the applicants' rights under Article 6 paras. 1 and 2 (Art. 6-1, 6-2) of the Convention in respect of the earlier risk of criminal prosecution were not infringed.

It follows that this aspect of the case is manifestly ill-founded, within the meaning of Article 27 para. 2 (Art. 27-2) of the Convention.

Figure 9.3: Example of the structure *AC* found in the ECHR corpus

The case-law of the Commission establishes that where no domestic remedy is available, the six months period runs from the act complained of

(see eg. No. 10530/83, Dec. 16.5.85, D.R. 42 p. 171)

Figure 9.4: Example of the structure *CnP* found in the ECHR corpus

a $Psup$, Pag, C or n were eliminated, as they would never be part of the argumentative structure, i.e. the grammar consider them non-argumentative. The resulting list of symbols was given to the JSCC system to parse. The sentences considered non-argumentative were written down in a file of non-argumentative information.

The main results for argumentation tagging can be seen in Table 9.2. It is observed that all final decisions (D) are encountered. Therefore, the limitations of the grammar are due to the structure of A, i.e. the justification given by the court. The results on argumentation parsing, i.e. tree-structures, show that around 80% of the automatic structures comply with $(1 + \frac{d*\frac{4}{10}}{n_{arg}}) \leq TC \leq 2$, making those structures correctly detected by TC standards. Section 9.4.1 presents a detailed study of the automatic annotation problems and a comparative between them and the problems encountered during human-annotation.

An important characteristic of the parser is that the tree-structures contain

The Commission does not deem it necessary to determine whether there exists an administrative practice on the part of Turkish authorities tolerating abuses of human rights of the kind alleged by the applicant, because it anyway finds that the applicant has done all that could be expected in the circumstances.

The Commission has formed this view for the following reasons:

The Commission notes the applicant's declaration that he had told the public prosecutor that he had been tortured.

It is not possible to establish in detail what happened during the applicant's meeting with the public prosecutor, but the Commission finds no reason to doubt that during their conversation there were elements which should have made the public prosecutor initiate an investigation or, at the very least, try to obtain further information from the applicant about his state of health or about the treatment to which he had been subjected.

Figure 9.5: Example of the structure $CnPP$ found in the ECHR corpus

The Commission finds that the criminal proceedings leading to the applicants' conviction were not conducted before an impartial District Court and Court of Appeal and that the proceedings were unfair and contrary to the presumption of innocence.

The Commission has formed this view for the following reasons:

(a) The interview was carried out in the absence of the first applicant, (b) the second applicant should have been assisted by a legal representative during the interview.

Figure 9.6: Example of the structure Cns found in the ECHR corpus

the argumentative information of the legal case, i.e. the reasoning process. The non-argumentative information was not kept in the tree (see Figure 9.12) even if it could be seen in a separate document known as facts. Therefore, these tree-structures can be seen as summaries of the legal cases, i.e. an argumentative summary of the legal case. A general study over the summaries showed that in average the compression range is a 65% of the original legal case. Moreover, the average summary only includes 10% of non-relevant information for the final decision, i.e. non-argumentative information detected as argumentation.

Moreover, a graphically representation of the information extracted as argumentative from the case is obtained (see Figure 9.12). Graphical

The Commission considers that while the Inspectors' report was being studied by the Director of Public Prosecutions (DPP) and the Director of the Serious Fraud Office (SFO), the applicants could have considered themselves affected by a possible criminal prosecution and therefore, "charged", within the meaning of Article 6 para. 2 (Art. 6-2) of the Convention (cf. the Court's aforementioned Deweer judgment of 27 February 1990, p. 24, para. 46, and the Adolf judgment of 26 March 1982, p. 15, para. 30).

Once these officials had issued their public statements on 1 March 1990 that the applicants would not be prosecuted, there can, in the Commission's view, be no question of the applicants being deemed to have been "charged" any longer.

In these circumstances the Commission concludes that the applicants' rights under Article 6 paras. 1 and 2 (Art. 6-1, 6-2) of the Convention in respect of the earlier risk of criminal prosecution were not infringed.

Figure 9.7: Example of the structure AsC found in the ECHR corpus

The Commission notes the applicant's declaration that he had told the public prosecutor that he had been tortured. Moreover, when asked to sign a statement, he had answered that he could not sign because he could not move his hands.

The Commission further notes that, after his detention, the applicant was in a vulnerable position, if he had, as he stated, been subjected to torture during his detention.

Figure 9.8: Example of the structure PP found in the ECHR corpus

representations of argumentation have been studied for a long time, and their importance in argumentation understanding is well proved. Different authors [53, 15, 66] have developed systems to visualize those structures. In contrast to these complex systems and due to the current inability of the grammar to take into account complex relations inside an argument (i.e. only premise-conclusion relations are studied) it was decided to use a simple tree-structure. However, once the grammar is extended to more complex relations it would be needed to accordingly adapt the graphical representation. Note that an extension of the grammar does not entail a change of type of grammar, in this case a CFG, but more a higher complexity of the grammar rules, e.g. more non-terminal symbols or longer right sides of the rules.

The Court notes, firstly, that the applicant was convicted by the Greek courts of disturbing, through his writings, the public peace and the peace of the citizens of Western Thrace. Without prejudice to its decision on the objection relating to non-exhaustion of domestic remedies, the Court considers that Mr Ahmet Sadik's widow and children have a legitimate moral interest in obtaining a ruling that his conviction infringed the right to freedom of expression which he relied on before the Convention institutions.

Furthermore, it notes that the applicant was sentenced to fifteen months' imprisonment, commutable to a fine of GRD 1,000 per day of detention, which sum he paid. Like the Delegate of the Commission, the Court considers that the applicant's heirs also have a definite pecuniary interest under Article 50 of the Convention (art. 50).

The Court accordingly finds that Mrs Isik Ahmet and her two children, Mr Levent Ahmet and Miss Funda Ahmet, have standing to continue the present proceedings in the applicant's stead.

Figure 9.9: Example of the structure AAC found in the ECHR corpus

The Court notes that the Convention forms an integral part of the Greek legal system, where it takes precedence over every contrary provision of the law (Article 28 para. 1 of the Constitution - see paragraph 19 above). It further notes that Article 10 of the Convention (art. 10) is directly applicable; Mr Ahmet Sadik could therefore have relied on that provision (art. 10) in the Greek courts and complained of a violation thereof in his case.

At no time, however, did the applicant rely on Article 10 of the Convention (art. 10), or on arguments to the same or like effect based on domestic law, in the courts dealing with his case.

Accordingly, domestic remedies were not exhausted in the instant case.

Figure 9.10: Example of the structure PPC found in the ECHR corpus

9.4.1 Annotation problems vs. detection problems

Section 6.2 has presented an analysis of the main problems during the human annotation process. Now, the focus is in the analysis of the behaviour of the automatic tool when dealing with those problems. Some of these problems are on the tagging of premise/conclusion and others on the correctness of the tree-structure.

Argument or Fact

Human annotators showed difficulties when detecting some types of argument (e.g. *Argument from Precedent*) due to their conceptions of law. These arguments were, by some annotators, seen as facts. The automatic tool uses features which should be clear enough to distinguish between argument and fact. Therefore, one should assume that if some arguments or argument elements are detected as facts, this is due to their ambiguous writing. In the automatic results, around 30% of the premises were classified as non-argumentative. Most of these premises were descriptions of principles or statements with no clear speaker. For example, *"An applicant does not need to exercise remedies which, although theoretically of a nature to constitute remedies, do not in reality offer any chance of redressing the alleged breach."* or *"The threats to which the applicant claimed to have been exposed after he had complained to the Commission, as well as his tragic death in circumstances which have so far not been fully clarified, are further elements which may at least support the view that the pursuance of remedies may not be devoid of serious risks."*. These sentences could have been less ambiguous if they were expressed as *"The Court notes that an applicant does not need to exercise remedies which, although theoretically of a nature to constitute remedies, do not in reality offer any chance of redressing the alleged breach."* or *"The threats to which the applicant claimed to have been exposed after he had complained to the Commission, as well as his tragic death in circumstances which have so far not been fully clarified, are further elements which may at least support the view of the Court that the pursuance of remedies may not be devoid of serious risks."* In the first sentence the introduction of "*The Court notes*" clarifies that the following is a description of a court's principle, while in the second sentence the addition of "*the Court*" after "*the view of*" explicitly states that this sentence is part of the view of the agent presenting the reasoning.

Limits of the argument

Human annotators showed difficulties when working with long arguments with scattered premises between different pages of the legal case. Premises far away from the conclusion were forgotten. The automatic tool does not suffer from this behaviour as it threats premises independent of the distance to a conclusion. The distance between premises does not affect their detection, however, it can affect the correctness of their argumentative structure. A detected premise will always be attached to a conclusion, but it might not be the correct conclusion. Therefore, the limits of the argument are sometimes incorrect, but in this case it is difficult to calculate when the automatic tool forgets a premise because of

the distance between the premise and its conclusion or the complexity of the argumentation structure.

Identifying the structure of argumentation

Human annotators had three problems related to this type of error: (i) the detection between premise and conclusion, (ii) the identification of complex argument structures, and (iii) the distinction between coordinate and subordinate structures.

First, the automatic tool presents some problems at tagging level when distinguishing between premises and conclusions. These are normally caused by the ambiguity of the judge's writing style. Around 2% of the premises were detected as conclusion, while 28% of the conclusions were detected as premises. Second, human annotators forgot sub-arguments when identifying complex structures, such as the structure of *Argument of an Established Rule*, the automatic tool does not normally have this type of problem, as it detects all arguments of the legal case in 80% of the cases, even if the arguments are not completely detected (i.e. some premises or even the conclusion is not detected). Third, human annotators had difficulties distinguishing subordinate and coordinate arguments. One possible reason for this is the ambiguity of the case's writing style, which presents middle conclusions of a subordinate structure as premises. The automatic tool owes most of its errors to this type of problem. The lack of markers between premises hinders their grouping in P or A. If they are grouped in different A then the relation is, or should be, multiple. The only rules that allow to group premises in P, coordinative relation, are through P_{sup} and P_{ag}. Therefore, only when there are explicit markers joining coordinative premises would the premises be detected as such. See Figure 9.11 where $PREMISE_1$, $PREMISE_2$, $ARGUMENT_2$ and $ARGUMENT_3$ maintain a coordinative relation, in this case the automatic tool would join these element with a coordinative relation. In fact the automatic tool would first join as coordinative $PREMISE_2$ and $ARGUMENT_2$, and then $PREMISE_1$ and $PREMISE_2 + ARGUMENT_2$, but the result would be the same. However, in the case of $PREMISE_4$ and $PREMISE_5$ the automatic tool would not find the coordinative relation between them as there is no explicit marker of the relation and $PREMISE_5$ would not be identified by itself as premise.

Simple subordinate structures, such as premises-conclusion-conclusion, are only completely correctly detected with a 60% accuracy and more complex structures have even lower results. The agreement between human-annotation

[$ARGUMENT_1$ [$PREMISE_1$] The Commission recalls that Article 26 (Art. 26) of the Convention only requires the exhaustion of such remedies which relate to the breaches of the Convention alleged and at the same time can provide effective and sufficient redress. [$PREMISE_2$] An applicant does not need to exercise remedies which, although theoretically of a nature to constitute remedies, do not in reality offer any chance of redressing the alleged breach (cf. No. 9248/81, Dec. 10.10.83, D.R. 34 p. 78). [$ARGUMENT_2$ [$CONCLUSION_1$] It is furthermore established that the burden of proving the existence of available and sufficient domestic remedies lies upon the State invoking the rule [$PREMISE_3$] (cf. Eur. Court. H.R., Deweer judgment of 27 February 1980, Series A no. 35, p. 15, para. 26, and No. 9013/80, Dec. 11.12.82, D.R. 30 p. 96, at p. 102).] [$ARGUMENT_3$ [$PREMISE_4$] The Commission notes that in the context of the section 9 powers the Secretary of State has a very wide discretion. [$PREMISE_5$] It appears that he has only exercised his discretion under section 9 in five cases and that there is no example of any enforcement of those directions being pursued in the courts. [$ARGUMENT_4$ [$PREMISE_6$] The Commission recalls that in the case of Temple v. the United Kingdom (No. 10530/83, Dec. 16.5.85, D.R. 42 p. 171) the Commission held that recourse to a purely discretionary power on the part of the Secretary of State did not constitute an effective domestic remedy. [$CONCLUSION_2$] The Commission finds that the suggested application for discretionary relief in the instant case cannot do so either.] [$CONCLUSION_3$] In these circumstances, the Commission finds that the application cannot be declared inadmissible for non-exhaustion of domestic remedies.]]

Figure 9.11: ECHR argument with coordinative and subordinative relations between elements

	Number of sentences	Precision	Recall	F1
Premises	1365	61%	68%	64.3%
Conclusions	509	60%	77%	67.4%
Non-argumentative	9610	87%	79%	82.8%
Final decision	43	100%	100%	100%

Table 9.2: Results from the argumentative parser over 44 documents from the ECHR corpus

and automatic-annotation presents a $K = 0.72$, with around a thousand agreements at premise detection and three hundred at conclusion detection.

```
T
|--D
|  |--x: for these reasons, the commission by a majority declares the
|       application admissible, without prejudging the merits.
|--A
|  |--C: it follows that the application cannot be dismissed as
|  |     manifestly ill-founded.
|  |--A
|     |--P: it considers that the applicant 's complaints raise serious
|     |     issues of fact and law under the convention, the
|     |     determination of which should depend on an examination
|     |     of the merits.
|     |--P: the commission has taken cognizance of the submissions
|           of the parties.
|--A
   |--C: in these circumstances, the commission finds that the application
   |     cannot be declared inadmissible for non-exhaustion of domestic
   |     remedies.
   |--A
      |--P: the commission recalls that article art. of the convention
      |     only requires the exhaustion of such remedies which
      |     relate to the breaches of the convention alleged and at
      |     the same time can provide effective and sufficient redress.
      |--P: the commission notes that in the context of the section
      |     powers the secretary of state has a very wide discretion.
      |--P: the commission recalls that in the case of temple
      |     v. the united kingdom no. dec. d.r. p. the Commission
      |     held that recourse to a purely discretionary power
      |     on the part of the secretary of state did not
      |     constitute an effective domestic remedy .
      |--P: the commission finds that the suggested application
            for discretionary relief in the instant case cannot
            do so either.
```

Figure 9.12: Segment of an output tree-structure of a legal case given by the automatic argumentative parser using the argumentative grammar

9.5 Summary

In this chapter, the first automatic tool that can detect complete tree-structures of a legal case argumentation has been presented. Given a legal case the tool offers a hierarchical structure with a final decision and a justification for this decision. The justification is formed by different arguments, which can have internal sub-arguments. The tool has been evaluated and found as reliable as human annotation, as the agreement between humans is $K = 0.75$ and agreement between tool and a human is $K = 0.72$. An analysis of the tool's main errors was provided and it was found some of them were also main disagreement reasons between human annotators.

Chapter 10

Future research

The research presented in this thesis is a first of its kind and therefore it has open many challenging issues that should be approached in the future. This chapter not only suggests possible directions for future work but analyses the problems that might arose in each direction. The aim is to illustrate the necessary steps researchers would have to undertake when a direction is taken. Some sections even include some initial experiments which hint at the complexity on these future directions. The directions suggested spread among a number of topics from linguistics to information retrieval.

10.1 Argumentation annotation

Argumentation annotation has been a crucial part of the research presented in this document. Gathering and annotating data adequate for the automatic processing of arguments has open many issues over the characteristics argumentation annotation should preserve.

For many years experts have studied argumentation from many points of view, logics, philosophy, linguistics, etc. One could assume that, given the amount of theories and ideas discussed over the years, the application of such theories to natural argumentation should present few problems. There should be a direct way between theory and practice. However, during this thesis three things have been observed: (a) the required background of the "correct" annotator is not clear, (b) there is a lack of linguistic rules/studies on argumentation and

(c) there is no comparison between theories in respect their "correctness" and applicability to natural argumentation.

10.1.1 Annotators

During the annotation task undertaken in this research legal experts have been chosen as annotators. The motivation behind this choice was simple, the documents to annotate were legal documents, thus legal experts should have the knowledge to achieve the task. However, it was observed that the annotators had difficulties when performing the task. Legal experts were experts in law, but they were not experts in argumentation theory or linguistics, which are important factors to accomplish a correct annotation.

Legal experts are used to work with natural argumentation. During their training years they read thousands of cases, they are taught to extract meaningful information from the cases and to reason with it creating more argumentation. However, during this research it has been observed that most legal experts lack training on recognizing the linguistic structure of argumentation. It is not that they do not know there is some linguistic knowledge involved, it is more that they are taught to overlook it, see Appendix G to see an example of annotations made by a legal expert. Note that legal experts are conscious of the language they use when they present their arguments but when they "extract" the arguments from other speakers they just look for concepts or new information not linguistic cues or reasoning patterns.

Thus, legal experts had to be taught not to overlook the clues given by the language to detect the reasoning patterns. This solution improved the annotators performance but it was not the only solution. It could have been possible to ask linguists or philosophers or other kind of experts to become annotators. It is an open question if such experts would prove better annotators or if their annotations would differ from the ones given by legal experts.

In general, linguists should be ease when detecting the linguistic markers (e.g. verbs or rhetorical markers) encountered in arguments and they should be able to extract rules from this information. However, in fields like law were the language and terminology used is so complex that readers from outside the field might not understand the full meaning of a document or even a single sentence, it is highly possible that linguists would not achieve better results than legal experts. On the other hand, it is possible that the annotations would present highly different results, but the reason behind it is more likely to be miscomprehension of the sentences meaning what would lead to incorrect annotation of their function. Philosophers or other argumentation theory experts are not used to work with natural argumentation but with more

abstract or logic models. Therefore, they might encounter the same problems as legal experts when dealing with linguistic structures. Furthermore, they might also encounter comprenhension problems when dealing with legal language.

10.1.2 Linguistic study on argumentation

During this work it has been observed how some linguistic patterns are more related to an argumentation function than another. Verbs like "*note*", "*reiterate*, "*agree*" are more common in premises than in conclusions and if they appear in conclusions they are quite often accompanied by a conclusive marker. However, it was not possible to find an extensive linguistic study on the characteristics of legal argumentation, in fact few information was found about the linguistic aspect of any argumentation. Many researchers mention that some rhetorical markers might signal the argumentation function of the sentence but few give detailed studies. The development of a clear and detailed list of verbs or keywords for argumentation would be a nice addition for future work in argumentation annotation. In fact, any future work in argumentation annotation would have to deal with these issues before any annotation can be accomplished.

Argumentation in the legal domain is a restricted domain in terms of verb sparseness. Legal experts tend to express their arguments with a restricted set of verbs. Most of these verbs can be sorted in distinctive argumentative functions, such as premise or conclusion. However, verbs like *"consider"* and *"find"* have a more ambiguous argumentative function. They can introduce a conclusion, e.g. *"The Court considers the case inadmissible"*, or a premise, e.g. *"The Commission considers that the Government's argument could only be accepted if it were clear that the application was based on untrue facts."*. During this research these verbs have been sorted into one class function or other based on the number of occurrences of each verb in each function. For example, the verb *"consider"*, which occurred mostly in premises, was sorted as a verb related to the function premise while the verb *"find"*, which occurred mostly in conclusions, was sorted into verbs related to function conclusion. A more detailed linguistic study on the verbs argumentative function might show both verbs should pertain to the same class, maybe even requiring a new class function for this type of verbs.

10.1.3 Framework

Even if during the framework development the aim was clear, i.e. the automatic detection of argumentation, there was no "real" knowledge of what the

task required. The suppositions undertaken were based on related work on other automatic text detection tasks or theoretical knowledge of the legal and argumentative domains. The theories selected as background to the framework are not the only existing valid argumentation theories. In fact, from a theoretical point of view it would be really interesting to compare how different theories adapt to the task. One should study which theories present argumentation patterns that adapt better to natural argumentation, which theories are more user-friendly, or which theory obtains the annotations with the best automatic outputs.

Moreover, the analyses of the errors produced during runs of the automatic methods has yielded new concerns on the correctness of the current annotation framework. As mentioned in Section 6.3, the treatment of *Argument from Precedent* when presented using connectors such as *"cf."* made the detection of features to distinguish premises and conclusions more complex. In other words, although legal theorists and philosophers have defined argument structures and given all of them the same importance in the reasoning, the corpus based processing has revealed some shortcomings of these proposed structures and how important it is to detect them as such. The analysis of this shortcomings could be a start point to the refinement of current argumentation schemes to better represent natural argumentation.

10.2 Semantic analysis

Argumentation structure of legal cases has been studied from different points of view: discourse structure, reasoning schemes and linguistics. However, the strength of semantics has not been exploited. Semantics focus on the meaning of a sentence instead of the combinatorics of units of the sentence. A semantic analysis of argumentation would analyse arguments to find the underlying meaning or concepts of those arguments.

In this research the structure of argumentation has been detected by the presence of linguistic patterns without use of the concepts represented in the arguments. Currently, the relationships between arguments are determined by the position of the arguments and any linguistic marker that implies connection. The words in the arguments are not used to identify similar topics or relations. Words could be used to find argument limits and relations, comparing the meanings or concepts behind the words.

Preliminary experiments over argument limits using topic detection models were performed. First, the experiments focused on detecting arguments by semantic clustering of document sentences. The words of each sentence were

used to determine the concepts related to the sentence and then sentences with similar concepts were group together. Sentences were represented as "bags of words", where the order of the words in the sentence was not important, only how many times each word appeared in a sentence. Concepts were represented as patterns of words that usually appeared together in sentences. For example "*admissible*", "*dismiss*", and "*Court*" might usually appear in sentences from final decisions. Words were assumed to have only one meaning to make the problem more tractable. The results presented groups of sentences which in 90% of the cases did not represent arguments of the document. The most probable reason behind the poor performance is the limited size of the sentences, which contain few relevant semantic information.

A second set of experiments focused on detecting relations between pre-detected arguments. The aim was to study if an argument topic could help to determine the argumentation structure. Arguments were represented as "bags of words", where the order of the words in the argument was not important. Concepts were again represented as patterns of words that usually appeared together in arguments. In this case the results shown a good performance, around 65% accuracy, when dealing with multiple argumentation with clear distinctive justifications for a same final decision. Arguments from different justifications were generally cluster in different topics. However, this only helped to distinguish relationships between well separated arguments, which normally were the ones the automatic process had less difficulty to correctly structure. Arguments with coordinative or subordinative relations were all generally cluster in a unique topic. This seems to imply that the different arguments involve in a multiple/subordinate argument possess insufficient semantic information to achieve a clear detection of subtopics. The reasons behind this lack of semantic differences could be specific to the current corpus or other factors might influence the poor results. For example, the size of the arguments. An argument might be as short as two sentences, premise and conclusion. These sentences might have not enough semantic information to identify a clear topic or an enough distinctive topic.

10.3 Statistical parsing

Rule-base parsers select which rule to apply in a strict order without taking into account which rule is more probable to lead to the correct parse. Statistical knowledge over each rule frequency in a training set, i.e. the rule probability, can be used to select which rule to apply in each step of the parsing process. The parsers that use statistical knowledge to select rules are known as statistical parsers. Statistical parsing is widely used in NLP research, e.g. [58, 5].

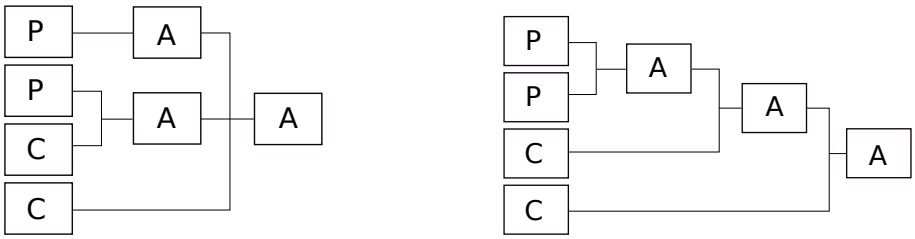

Figure 10.1: Two possible argumentation structures using the argumentative grammar in Figure 9.1 for the entry set $PPCC$

In Figure 10.1 two possible structures that could be obtained using the argumentative grammar of Chapter 9 for the entry set $PPCC$ are presented. The rules involved are $A \rightarrow P^+$ and $A \rightarrow A^+C$. So all the rules are instances of rule 9.4 in Figure 9.1. Entry sets like this one with two or more possible structures depending on the limits of the constituent A or P are frequent in the ECHR corpus and an important part of the error rate. In the current example the entry set could represent: (a) one argument with two premises and one conclusion and then a second conclusion with one or more omitted premises (right side of Figure 10.1) or (b) one argument with one premise and one conclusion that together with another premise justify a second conclusion (left side of Figure 10.1). In the right side of Figure 10.1 the rules applied have been:

1. $A \rightarrow P$
2. $A \rightarrow P$
3. $A \rightarrow AC$
4. $A \rightarrow AAC$

while in the left side of Figure 10.1 the rules applied have been:

1. $A \rightarrow PP$
2. $A \rightarrow AC$
3. $A \rightarrow AC$

The use of statistical knowledge over rules $A \rightarrow PP$ and $A \rightarrow P$ might allow to know which structure is more probable to be correct. Statistical knowledge

refers to the probability of expanding a constituent, e.g. A, using a particular rule, as opposed to any of the other rules that could be used to expand this kind of constituent. There are different mechanism to achieve statistical parsing. Two possibilities are here considered: use of a probabilistic context-free grammar (PCFG) and use of markov logic (ML).

10.3.1 PCFG

A Probabilistic Context Free Grammar (PCFG) is a context-free grammar in which every rule is assigned such a probability [7]. The dynamics of a PCFG can be seen in Section 2.3.2, where it is explained how a PCFG can be obtained in two ways: (a) given a context free grammar and a training corpus and (b) by defining a grammar from scratch using the knowledge of a tree-bank.

Argumentative CFG to argumentative PCFG

The defined argumentative CFG presented in Chapter 9 could be extended to add estimated probabilities using the ECHR corpus. For each rule in the grammar, e.g. $A \rightarrow P$, it would be counted how many times it occurres in the corpus and it would be then divided by the number of times any rule expanding the same constituent, e.g. any rule with left side A, occurres in the corpus.

The ECHR corpus contains limited data as it is formed by just 54 legal cases. The corpus contains many structures but it might not contain enough examples of each of them. In a first overview of the corpus it is observed that most rules occur just once or twice over all the corpus. Moreover, the rules with high number of occurrences are also the rules extending the more frequent constituents. It is for future research to find which if the ECHR corpus contains enough data or it should be extended before testing a PCFG.

An argumentative tree-grammar

Syntactic parsing for English corpora has a long tradition of research in tree-bank grammars, most notably the research done over the Penn tree-bank [38]. Tree-grammars are obtained extracting rules directly from a tree-bank. The question is if the ECHR corpus could be used as a tree-bank to obtain an argumentative grammar.

The ECHR manual tree-structures which form the ECHR tree-bank were the tree-structures form by the sentences which the annotators considered

argumentative. The non-argumentative sentences were removed, they were not considered for the structure detection. Each sentence in the manual tree-structures was transformed into a symbol of the grammar with a set of pattern rules done in Java. Note that when the annotators have found arguments with different premises and same conclusion this corresponds to the structure AAC, where each A contains the premises of the different arguments and C is the common conclusion. However, it will not be difficult to add the non-argumentative information to the tree-structures if rules to position the non-argumentative information were described, e.g. all non-argumentative information is first clustered by position and then attached to the closest symbol to the left.

However, as in the previous section it is for future research to determine if the ECHR corpus contains enough information to undertake such a task. One should note that 50.000 is the number of hand-parsed sentences in the most used syntactic tree-bank, the Penn tree-bank, and this tree-bank is known to not contain all the possible grammatical constructions of a sentence.

10.3.2 Markov Logic

Statistical parsing can also be performed using Markov logic. The basic idea in Markov logic is that given a set of sentences or formulas in first-order logic that can be seen as a set of hard constraints on the set of possible worlds: if a world violates even one formula, it has zero probability; one has to soften these constraints: when a world violates one formula in the first-order logic set it is less probable, but not impossible. The fewer formulas a world violates, the more probable it is. Each formula has an associated weight that reflects how strong a constraint it is: the higher the weight, the greater the difference in log probability between a world that satisfies the formula and one that does not, other things being equal [12].

A grammar can be expressed as a Markov Logic Network (MLN), which is a set of pairs (F, w) where F is a formula in first-order logic, which in the grammar would be a rule, and w is a real number, which would be the weight of that rule. Together with a set of constants, it defines a Markov network with one node for each grounding of each predicate in the MLN and one feature for each grounding of each formula F in the MLN, with the corresponding weight w.

To perform Markov logic one can use tools such as Alchemy, which can perform three basic tasks: structure learning, weight learning, and inference. The former two involve learning the structure or parameters of a model given a training database consisting of ground atoms. The latter involves inferring the probability or most likely state of query atoms given a test database consisting

of evidence ground atoms. Moreover, Alchemy can perform maximum a posteriori probability (MAP) inference which outputs the most likely state of the query atoms or probabilistic inference which outputs marginal probabilities of the query atoms given the evidence. As with PCFGs, it is for future research to study if the ECHR corpus contains enough information to learn the weights of the rules or the full structure of an argumentative grammar.

10.4 Document retrieval by argumentation

The need for automatic argumentation detection arose from the need to improve search tools in different domains with high argumentation, i.e. law, opinion blogs. The research done in argumentative parsing could be used to experiment with search algorithms to improve the ranking of documents in a search process. For example, a lawyer might be looking for legal cases where the submission of an application out of time has been used as a factor to dismiss the case. The lawyer in current search tools would look for documents using keywords such as "*submission time-limit dismiss*". The search tools would look for occurrences of those keywords in the document collection and return any document where the words occur. However, if the output of the argumentative parser was used one could have searches were for each individual argument it is check if the keywords are present. This would omit to return documents where the keywords appear in mixed arguments. Moreover, it could be possible to develop a search tool where keywords could be associate to an argumentative function, e.g. premise or conclusion. Then, the lawyer could enter for premise the keywords "*submission time-limit*" and for conclusion "*dismiss*". This should reduce the numbers of retrieved documents and ensure better accuracy of the results. Note that only extrinsic evaluation of this type of search can prove or disprove an improvement in document retrievals.

10.5 Non-legal domain

This thesis has focused on the argumentation present in the legal domain. However, argumentation can be found in many types of documents, from newspapers to parliamentary records. Some of these documents might present a complete different type of argumentation and discourse style, e.g. blogs, however, some should present an enough close style to legal cases, e.g. parliamentary records.

Two possible future directions can be expected to arise with the change of domain. First, it should be studied how the framework, features and methods developed focusing on the legal domain behave in other domains. The features and methods used during this research seek to be as general as possible but their performance is highly dependent on the characteristics of legal documents. For example, legal documents always state the final conclusion or decision of the argumentation process, it makes no sense to encounter a judgement without explicit statement of the plaintiff's fate. This explicitness might not be common in other domains. The early experiments of this thesis over the Araucaria corpus have shown that distinction between argumentative and non-argumentative sentences in non-legal domains is possible, but would it be possible to detect full argumentation structures? The high level of implicity and the wide vocabulary of some domains, such as blogs, might prove the task unachievable.

Second, new domains might give rise to new concerns that were overlook in the legal domain. For example, non-legal argumentation might abuse metaphors. Furthermore, assumptions made in the legal domain might not apply out of it. For example, non-legal arguments might have their conclusion implicit while in legal argumentation this is highly improbable. These and other facts might require the development of a new framework, features or methods specific for the domain at hand.

10.6 Summary

In this chapter different research paths for argumentation detection and classification have been analysed. First, some possible improvements on the argumentation annotation field have been studied, from the adequacy of the annotators to the further analysis of annotation frameworks valid for automatic argumentation detection. Second, the development of automatic tools for argumentation detection from a semantic point of view has been generally described and its problems over-viewed. Third, a statistical approach to argumentative parsing has been introduced. Finally, two other possible research fields have been mentioned but not analysed. These work on these fields has been mentioned as important future applications of automatic argumentation detection.

Chapter 11

Conclusions

In this thesis the automatic detection of argumentation in legal cases has been introduced. It has been discussed and proved that the detection of argumentation in legal cases, although being a very challenging task, can be performed automatically. Moreover, the first tool that automatically detects argumentation in text and structures this argumentation by recognizing its composing premises and conclusions and their relationships has been developed.

The first part of this thesis (Part I) has studied the resources needed to train and test the experiments on automatic argumentation detection. In Chapter 3 it has been presented the state of the art of legal resources that could be used for the automatic task. A lack of adequate resources to achieve the aims of this thesis has been observed. Still, a deeper analysis of one existent resource (Section 3.1) has been presented as this resource was used in the early experiments on automatic detection. Given the lack of adequate resources, it has been collected (Chapter 4) and annotated (Chapter 5) a new corpus, the ECHR corpus. The annotation process done in this thesis is the first of its kind, so it was necessary to design an adequte framework for argumentation annotation in legal cases. This first part of the thesis concluded with an analysis of the main problems humans encounter when annotating natural argumentation (Chapter 6).

The second part of this thesis (Part II) has studied different approaches to accomplish the automatic detection of argumentation. All the approaches were based in state-of-the-art IE and NLP methods, specifically the ones presented in Chapter 7. The first experiments, presented in Chapter 8, have been focused on the distinction between argumentative and non-argumentative sentences,

and then, more specifically, on the distinction of premises and conclusions. Later, Chapter 9 has focused on the detection of full argumentation structures. These experiments resulted in the first automatic parser able to detect the argumentation tree-structure of a legal case.

To conclude this thesis, Chapter 10 has considered future directions for the research on argumentation detection and classification. First, some possible improvements on the argumentation annotation field have been studied, from the adequacy of the annotators to the further analysis of annotation frameworks valid for automatic argumentation detection. Second, the development of automatic tools for argumentation detection from a semantic point of view has been generally described and its problems over-viewed. Third, a statistical approach to argumentative parsing has been introduced. This specific aspect needs a much more detailed study to determine the viability of this approach. Finally, two other possible research fields were mentioned but not analysed. The work on these fields has been mentioned due to its importance on future applications of automatic argumentation detection.

In sum, the main conclusion of this thesis is that more interdisciplinary research is needed to obtain more accurate systems for automatic argumentation detection. Legal experts, computer scientists, philosophers and linguists should sit together to analyse all the different aspects of argumentation to develop further tools.

Appendix A

Example of a legal case from the ECHR

COUR EUROPÉENNE DES DROITS DE L'HOMME
EUROPEAN COURT OF HUMAN RIGHTS

THIRD SECTION

CASE OF GIRARDI v. AUSTRIA

(Application no. 50064/99)

JUDGMENT

STRASBOURG

11 December 2003

FINAL

11/03/2004

This judgment will become final in the circumstances set out in Article 44 § 2 of the Convention. It may be subject to editorial revision.

In the case of Girardi v. Austria,

The European Court of Human Rights (Third Section), sitting as a Chamber composed of:

- Mr G. Ress, President,
- Mr L. Caflish,
- Mr P. Kūris,
- Mr R. Türmen,
- Mr J. Hedigan,
- Mrs H.S. Greve
- Mrs E. Steiner, judges,

and Mr V. Berger, Section Registrar,

Having deliberated in private on 20 November 2003,

Delivers the following judgment, which was adopted on that date:

PROCEDURE

1. The case originated in an application (no. 50064/99) against the Republic of Austria lodged with the Court under Article 34 of the Convention for the Protection of Human Rights and Fundamental Freedoms ("the Convention") by an Austrian national, Elisabeth Girardi ("the applicant"), on 9 July 1999.

2. The Austrian Government ("the Government") were represented by their Agent, Mr Mautner-Markhof.

3. On 4 July 2002 the Third Section declared the application partly inadmissible and decided to communicate the complaint concerning the length of the proceedings to the Government. Under the provisions of Article 29 § 3 of the Convention, it decided to examine the merits of the application at the same time as its admissibility.

THE FACTS

4. The applicant was born in 1951 and lives in Vienna. She is the mother of M, L and R, born in wedlock in 1973, 1974 and 1976, respectively. The spouses separated in 1982. Custody of L and M was assigned to the applicant, the custody of R to the father.

5. In December 1989 M was admitted in a public girls' home as she refused to stay with her mother. She stayed there until January 1992. From December 1989 until September 1995 custody proceedings concerning the temporary transfer of M's custody to the Vienna Youth Welfare Office for the time M had spent at the girls' home were pending before the Austrian courts.

A. The Youth Welfare Office's request for reimbursement of expenses

6. On 3 January 1990 the Vienna Youth Welfare Office, on behalf of M, filed a request with the Floridsdorf District Court that the applicant should pay a monthly contribution to the expenses incurred for M's stay in the girls' home.

7. The file was later on transferred to the competent Juvenile Court and, in January 1990, the court heard M's parents.

8. On 8 March 1991 the Youth Welfare Office reduced the amount of the requested monthly contribution.

9. On 10 April 1991 the President of the Juvenile Court granted the applicant's motion for bias against the competent court clerk (Rechtspfleger).

10. A hearing scheduled for 25 July 1991 was cancelled due to the applicant's illness. Further hearings scheduled for 2 September 1991 and 11 September 1991 had to be cancelled because the court's attempts to deliver the summons to the applicant were unsuccessful.

11. On 10 February 1992 the Juvenile Court ordered that the applicant had to pay ATS 2,500 in monthly maintenance for M. The applicant appealed, claiming that she was fit to work to an extent of 75% only.

12. On 4 March 1992 the case was assigned to another judge as the competent judge had declared himself biased.

13. On 13 May 1992 the Appeal Chamber quashed the decision and remitted the case back to the Juvenile Court, instructing the latter to take a new decision after having supplemented its proceedings. In particular, it stated that the first

instance court ought to appoint a forensic medical expert in order to establish the applicant's fitness to work.

14. On 20 May 1998 the Juvenile Court ordered the applicant to pay ATS 1,550 in monthly maintenance for M. At that stage of the proceedings, no expert had been heard yet.

15. Referring to the Appeal Chamber's decision of 13 May 1992, the applicant appealed, again relying on her reduced fitness to work.

16. On 13 August 1998 the Juvenile Court appointed an expert in forensic medicine to file a report on the question as to which extent the applicant's capacities to earn her living were reduced.

17. The applicant appealed against this decision, claiming that it no longer made sense to appoint a medical expert, now that the court had already dismissed her request by a decision of 20 May 1998. Further, she claimed that there was no need for a further report as, in this respect, she had already submitted two reports of different medical officers (Amtsarzt).

18. On 17 and 20 August 1998 the applicant filed motions for bias against the court clerk (Rechtspfleger) I.S., who was dealing with her case, claiming that the appointment of a further medical expert was not justified, that I.S. was handling the case file in a negligent manner, namely that several documents were missing from the file, and that I.S. had been rude to her on the telephone.

19. On 25 August 1998 the President of the Vienna Juvenile Court (Präsident des Jugendgerichtshofs) dismissed her motion for bias, finding that the mere fact that she had appointed a medical expert was not sufficient to cast doubt upon I.S.' impartiality. He also noted that there were no documents missing from the file.

20. On 17 September 1998 the Appeal Chamber dismissed the applicant's appeal against the appointment of a medical expert, but granted her appeal against the decision of 20 May 1998. In this respect, it referred the case to the Juvenile Court for supplementing the taking of evidence, namely to comply with its decision of 13 May 1992.

21. On 21 and 23 March 1999 the applicant requested that, pursuant to Section 91 of the Courts Act (Gerichtsorganisationsgesetz), a time-limit be fixed for the decision on the Youth Welfare Office's application of 3 January 1990.

22. On 23 March 1999 the applicant filed a motion for bias against I.S., claiming that the latter had not been available to her during office hours and that she had refused to give her information requested over the telephone.

23. On 29 March 1999 the President of the Vienna Juvenile Court dismissed her motion as being unfounded.

24. On 30 March 1999 the President rejected her appeal against this decision, as the relevant provisions of the Court Clerks Act (Rechtspflegergesetz) did not provide for such remedy.

25. On 8 April 1999 the applicant was summoned by the appointed medical expert to undergo a medical examination at the Institute for Forensic Medicine (Institut für Gerichtsmedizin) on 22 April 1999.

26. It appears that the applicant filed numerous complaints with the President of the Juvenile Court, again claiming that documents were missing from the file and that I.S. as well as various judges of the Juvenile Court were biased.

27. On 4 May 1999 the President of the Juvenile Court decided to exclude I.S. from the proceedings. He noted that the latter had expressed that she considered herself biased following a telephone conversation in the course of which the applicant had said she would kill her daughter if I.S. continued to harass her. In these circumstances, the President found it advisable that the matter be re-assigned in accordance with the Juvenile Court's rules on the distribution of cases (Geschäftsverteilung).

28. On the same day, the Juvenile Court dismissed the applicant's requests for a time-limit to be set. Referring to the applicant's numerous requests, complaints and motions for bias filed with the court, it found that there was no indication of a lack of due diligence on behalf of the Juvenile Court, it being rather the applicant who prevented that a decision on the merits had been taken so far.

29. On 17 May 1999 the Vienna Youth Welfare Office withdrew its request dated of 3 January 1990.

30. Thereupon, the applicant, on 27 May 1999, withdrew all requests and complaints still pending before the Juvenile Court at that stage.

B. The applicant's request for reimbursement of expenses

31. From 30 July 1990 to 3 September 1990 M stayed with her mother. The latter, on 4 September 1990 filed a request with the Juvenile Court, claiming reimbursement of her expenses incurred during this period.

32. In September 1990 the Vienna Youth Welfare Office reimbursed the applicant for M's stay with her from 30 July 1990 to 21 August 1990.

33. On 10 August 1993 the Juvenile Court dismissed the applicant's request for expenses incurred during the rest of the period.

34. On 30 August 1993 the President of the Vienna Juvenile Court dismissed the applicant's motion of bias against the competent judge. On 30 December 1993 the Vienna Court of Appeal granted the applicant's appeal against this decision and quashed the decision.

35. On 20 January 1994 the Appeal Chamber of the Juvenile Court again dismissed the applicant's motion for bias. On 6 May 1994 the Court of Appeal rejected the applicant's appeal. A further appeal to the Supreme Court was to no avail. A further motion for bias against the President of the Juvenile Court was to no avail either.

36. On 5 January 1995 the Appeal Chamber quashed the decision of 10 August 1993 and remitted the case back to the first instance court.

37. On 19 April 1998 the applicant requested that, pursuant to Section 91 of the Courts Act, a time-limit be fixed for the decision on her application of 4 September 1990.

38. On 8 June 1998 the President of the Vienna Juvenile Court ordered the Juvenile Court to decide on the applicant's request no later than on 31 July 1998.

39. On 5 August 1998 the Juvenile Court dismissed the applicant's request for maintenance payments of 4 September 1990.

40. The applicant appealed against this decision.

41. It appears from the documents submitted that the applicant filed several complaints with the Vienna Court of Appeal (Oberlandesgericht), claiming that I.S. had not complied with the time limit set by the President of the Juvenile Court because she had gone on holidays, that the competent judicial officer, I.S. was to be found at her office only twice a week and that she had been extraordinarily impolite to her.

42. Thereupon, the President of the Juvenile Court, on 31 August 1998, informed the applicant that both I.S.'s office hours as well as her right to vacation were in accordance with her assignment. He also expressed his regret that, if, in the course of one of the applicant's numerous telephone calls, I.S. might have acted in a slightly indignant way. However, he emphasised that the applicant's allegations had remained unproved.

43. On 17 September 1998 the Appeal Chamber dismissed her appeal against the Juvenile Court's decision of 5 August 1998 as being unfounded. Further,

it stated that there was no further appeal on points of law in the applicant's case as it did not raise questions of law of fundamental importance (Ausspruch *ü*ber die Unzul*ä*ssigkeit der ordentlichen Revision).

44. Nevertheless, the applicant filed an extraordinary appeal on points of law (ausserordentliche Revision) with the Supreme Court.

45. Referring to an amendment of Section 14 a of the Non-Contentious Proceedings Act (Ausserstreitgesetz), the Supreme Court on 18 December 1998 remitted the case back to the Vienna Juvenile Appeal Court. According to that provision, instead of filing an extraordinary appeal on points of law with the Supreme Court, a party to non-contentious proceedings must now request the Court of Appeal to re-consider its opinion on the admissibility of an ordinary appeal on points of law. The Supreme Court found that, even if in her appeal the applicant had not explicitly requested the Juvenile Appeal Court to declare that a further appeal on points of law be allowed, her appeal should have been understood in such a way.

46. Thereupon, on 11 January 1999 the Juvenile Appeal Court requested the applicant to remedy procedural defects of her appeal, namely to request that an ordinary appeal in her case be allowed.

47. As the applicant did not comply with this request, the Juvenile Appeal Court, on 25 February 1999, rejected her appeal.

THE LAW

I. ALLEGED VIOLATION OF ARTICLE 6 § 1 OF THE CONVENTION

48. The applicant complained that the length of the maintenance payment proceedings had been incompatible with the "reasonable time" principle as provided in Article 6 § 1 of the Convention, which reads as follows: In the determination of his civil rights and obligations....., everyone is entitled to a fair...hearing within reasonable time... by[a]... tribunal"

49. As regards the first set of proceedings, the period to be taken into consideration began on 3 January 1990 and ended on 22 May 1999. Thus, they lasted more than nine years and four months.

50. As regards the second set of proceedings, the period to be taken into consideration began on 4 September 1990 and ended on 25 February 1999. Thus, they lasted for more than eight years and five months.

A. Admissibility

51. The Court notes that this complaint is not manifestly ill-founded within the meaning of Article 35 § 3 of the Convention. It further notes that it is not inadmissible on any other grounds. It must therefore be declared admissible.

B. Merits

52. The Government submitted that the maintenance proceedings were complex. In particular, they had to be seen as a part of highly complex custody proceedings which required extensive expert opinions. While the authorities tried to conduct the proceedings expeditiously, the applicant filed a multitude of motions of bias, appeals and requests for extension of time-limits and therefore herself contributed considerably to the length of the proceedings. The Government further stressed that the applicant repeatedly thwarted attempts to deliver summons on her and failed to obey them.

53. The applicant did not submit any observations on these issues.

54. The Court reiterates that the reasonableness of the length of proceedings must be assessed in the light of the circumstances of the case and with reference to the criteria established by its case-law, particularly the complexity of the case, the conduct of the applicant and of the relevant authorities and what was at stake for the applicant in the dispute (see, among many other authorities, Frydlender v. France [GC], no. 30979/96, § 43, ECHR 2000-VII).

55. The Court considers that the present proceedings can clearly be distinguished from the custody proceedings, as they concerned merely the fixing of maintenance payments and were not particularly complex.

56. As regards the conduct of the applicant the Court has consistently held that applicants cannot be blamed for making full use of the remedies available to them under domestic law. However, an applicant's behaviour constitutes an objective fact which cannot be attributed to the respondent State and which must be taken into account for the purpose of determining whether or not the reasonable time referred to in Article 6 § 1 has been exceeded (see Erkner and Hofbauer v. Austria, no. 9616/81, Commission decision of 23 April 1987, A 117, § 68)

57. In the present case, the Court acknowledges that the applicant had filed numerous requests, complaints and motions and had repeatedly failed to obey the authorities' summons. Although such conduct contributed to prolonging

the proceedings, it is not in itself sufficient to explain the length of the extensive proceedings.

58. On the other hand, the Court notes that there are substantial delays attributable to the authorities. In particular, in the first set of proceedings, there is a period of inactivity of more than two years (from 3 January 1990 to 10 February 1992) while the case was pending before the Vienna Juvenile Court, and a further one of six years (from 13 May 1992 to 20 May 1998) before that court took a new decision after the first one had been quashed on appeal. In the second set of proceedings, there is a period of inactivity of some three years (from 4 September 1990 to 10 August 1993), while the case was pending before the Vienna Juvenile Court, and a further such period of three years and seven months (from 5 January 1995 to 5 August 1998) before that court took a new decision after the first one had been quashed on appeal. The Court cannot find that the Government has given sufficient explanation for these delays that occurred.

59. The Court therefore finds that the overall length of the proceedings cannot be regarded as "reasonable". Accordingly, there has been a violation of Article 6 § 1 of the Convention.

II. APPLICATION OF ARTICLE 41 OF THE CONVENTION

60. Article 41 of the Convention provides: "If the Court finds that there has been a violation of the Convention or the Protocols thereto, and if the internal law of the High Contracting Party concerned allows only partial reparation to be made, the Court shall, if necessary, afford just satisfaction to the injured party."

61. The applicant has not filed a claim for just satisfaction. Accordingly, the Court considers that no award can be made under this provision.

FOR THESE REASONS, THE COURT UNANIMOUSLY

1. Declares the application admissible;

2. Holds that there has been a violation of Article 6 § 1 of the Convention;

Done in English, and notified in writing on 11 December 2003, pursuant to Rule 77 § § 2 and 3 of the Rules of Court.

Vincent Berger, Registrar Georg Ress, President

Appendix B

Argumentation schemes

In this appendix, we provide the desciptions of the 25 different argumentation schemes presented in [69]. An argumentation scheme is a form of argument that can hold provisionally on a balance of considerations under conditions of uncertainty, but that can be defeated by the asking of critical questions that pinpoint weaknesses. Some of the argumentation schemes presented in this appendix are basic or fundamental, whereas others are composites made up from these basic schemes. We do not present a full analysis of the schemes or of their matching critical questions as they can be found in [69]. The argumentation schemes are as follows:

- *Argument from Sign* : A is true in this situation. B is generally indicated as true when its sign, A, is true, in this kind of situation. Therefore, B is true in this situation.
- *Argument from Example* : In this particular case, the individual a has property F and also property G. a is typical of things that have F and may or may not also have G. Therefore, generally, if X has property F, then X also has property G.
- *Argument from Verbal Classification* : a has a particular property F. For all x, if x has property F, then x can be classified as having property G. Therefore, a has property G.
- *Argument from Commitment* : a is committed to proposition A (generally, or in virtue of what she said in the past). Therefore, in this case, a should support A.

- *Circumstantial Argument Against the Person* : If x claims that everyone (including x) ought to act in accord with or support proposition A, then x is, or should be, committed to A. a claims that everyone (including a) ought to act in accord with or support proposition A. It is indicated by a's personal circumstances that a is not committed to A. Therefore, a is inconsistent in a's commitments, and there should be a weight of pressumption against a's argument for a's claim.
- *Argument from Position to Know* : a is in a position to know whether A is true (false). a asserts that A is true (false). Therefore, a is true (false).
- *Argument from Expert Opinion* : E is an expert in domain D. E asserts that A is known to be true. A is within D. Therefore, A may (plausibly) be taken to be true.
- *Argument from Evidence to a Hypothesis* : If A (a hypothesis) is true, then B (a proposition reporting an event) will be observed to be true. B has been observed to be true, in a given instance. Therefore, A is true.
- *Argument from Correlation to Cause* : There is a positive correlation between A and B. Therefore, A causes B.
- *Argument from Cause to Effect* : Generally, if A occurs, then B will (or might) occur. In this case, A occurs (or might occur). Therefore, in this case, B will occur (or might occur).
- *Argument from Consequences* : If A is brought about, then good (bad) consequences will (may plausibly) occur. Therefore, A should (not) be brought about.
- *Argument from Analogy* : Generally, case C_1 is similar to case C_2. A is true (false) in case C_1. Therefore, A is true (false) in case C_2.
- *Argument from Waste* : If a stops trying to realize A now, all a's previous efforts to realize A will be wasted. If all a's previous attempts to realize A are wasted, that would be a bad thing. Therefore, a ought to continue trying to realize A.
- *Argument from Popularity* : If a large majority (everyone, nearly everyone, etc.) accept A as true, then there exists a (defeasible) presumption in favor of A. A large majority accept A as true. Therefore, there exists a presumption in favor of A.
- *Ethotic Argument* : If x is a person of good moral character, then what x contends (A) should be accepted (as more plausible). a is a person of good moral character. Therefore, what a contends (A) should be accepted (as more plausible).

- *Argument from Bias* : If an arguer x is biased, then it is less likely that x has taken the evidence on both sides of an issue into account in arriving at conclusion A. Arguer a is biased. It is less likely that a has taken the evidence on both sides of this issue into account.

- *Argument from an Established Rule* : For all x, if doing A is the established rule for x, then (subject to exceptional cases), x must do A (subject to penalty). Doing A is the established rule for a. Therefore, a must do A (subject to penalty).

- *Argument from Precedent* : The existing rule says that for all x, if x has property F then x has property G. But in this case C, a has property F, but does not have property G, Therefore, the existing rule must be changed, qualified, or given up, or a new rule must be introduced to cover case C.

- *Argument from Gradualism* : Proposition A is true (acceptable to the respondent). There is an intervening sequence of propositions, $B_1, B_2, ...B_{n-1}, B_n, C$, such that the following conditionals are true: If A then B_1; If B_1 then B_2;...;If B_{n-1} then B_n; If B_n then C. The conditional 'If A then C' is not, by itself, acceptable to the respondent (nor are shorter sequences from A to C acceptable to the respondent).

- *The Casual Slippery Slope Argument* : A_0 is up for consideration as a proposal that seems initially like something that should be brought about. Bringing up A_0 would plausibly cause (in the given circumstances, as far as we know) A_1, which would in turn plausibly cause A_2, and so forth, throught the sequence $A_2, ..A_n$. A_n is a horrible (disastrous, bad) outcome. Therefore, A_0 should not be brought about.

- *The Precedent Slippery Slope Argument* : Case C_0 would set a precedent with respect to an existing rule R. Case C_0 is similar to case C_1, that is, if C_0 is held to be an exception to R, then C_1 must be held to be an exception too (in order to be consistent in treating equal cases alike). A sequence of similar pairs C_1, C_j binds us by case-to-case consistency to the series, $C_0, C_1, ...C_n$. Having to accept case C_n as a precedent, or as a recognized exception to R, would be intorelable (horrible, bad). Therefore, admitting case C_0. or bringing it forward in the first place, is not a good thing to do.

- *Argument from Vagueness of a Verbal Classification* : If an argument, Arg occurs in a context of dialogue that requires a certain level of precision, but some property F that occurs in Arg is defined in a way that is too vague to meet the requirements of that level of precision, then Arg ought to be rejected as deficient. Arg occurs in a context of dialogue that

requires a certain level of precision that is appropriate for that context. Some property F that occurs in Arg is defined in a way that is too vague to meet the requirement of the level of precision appropriate for that context. Therefore, Arg ought to be rejected as deficient.

- *Argument from Arbitrariness of a Verbal Classification* : If an argument, Arg occurs in a context of dialogue that requires a nonarbitrary definition for a key property F that occurs in Arg, and F is defined in an arbitrary way in Arg, then Arg ought to be rejected as deficient. Arg occurs in a context of dialogue that requires a nonarbitrary definition for a key property F that occurs in Arg. Some property F that occurs in Arg is defined in a way that is arbitrary. Therefore, Arg ought to be rejected as deficient.

- *The Verbal Slippery Slope Argument* : Individual a_1 has property F (as you, the respondent, concede). For all x and y, if x has F, then if y is indistinguishable from x with respect to F, then y also has F (as you, the respondent cannot deny). For any given pair a_i, a_j of adjacent individuals in the sequence, $a_1, a_2, ...a_n$, a_j is indistinguishable from a_i with respect to F. Therefore, a_n has property F (following from the three previous premises, by a series of steps). But a_n does not have property F (or at least, this outcome is not acceptable to you, the respondent). Therefore, it is not true that a_1 has F (or you, the respondent, should not have accepted this proposition).

- *The Full Slippery Slope Argument* : Case C_0 is tentatively acceptable as an initial presumption. There exists a series of cases $C_0, C_1, ..., C_{n-1}$, where each case leads to the next by a combination of causal, precedent, and/or analogy steps. There is a climate of social opinion such that once people come to accept each step as plausible (or as accepted practice), they will then be led to accept the next step. The penultimate step C_{n-1} leads to a horrible outcome, C_n, which is not acceptable. Therefore, C_0 is not acceptable (contrary to the pressumption of the intial premise).

Appendix C

The Corpus

C.1 Format of legal case encoding: discourse structure

<document = index-document >

name-document

<sentence = index-sentence >

sentence

sentence parsed

<clauses>

< clause = index-clause >

clause

clause parsed

</clauses>

</sentence>

</document>

C.2 Format of legal case encoding: argumentation structure

Decision = (index-$clause_i$, ..., index−$clause_n$)

* * * * * * * * * * * * * * *

Argument = index-argument

Conclusion = (index-$clause_j$, ..., index-$clause_m$)

Premise = (index-$clause_k$, ..., index-$clause_o$)

Premise = (index-$clause_l$, ..., index-$clause_p$)

Type = index-scheme

* * * * * * * * * * * * * * *

C.3 List of legal cases

Case Title	Date	Type	Words
AHMET v. GREECE	06/03/1997	Decision	3140
AKSOY v. TURKEY	19/10/1994	Decision	5516
ALTUN v. TURKEY	11/09/1995	Decision	3182
AYDER, LALEALP, DOMAN, BiER AND EKMEKi v. TURKEY	15/05/1995	Decision	5043
BENHAM v. THE UNITED KINGDOM	13/01/1994	Decision	2739
BUCKLEY v. THE UNITED KINGDOM	03/03/1994	Decision	2848
C.R. v. THE UNITED KINGDOM	14/01/1994	Decision	3205
CASE OF AHMET SADIK v. GREECE	15/11/1996	Judgment	12740
CASE OF AKSOY v. TURKEY	18/12/1996	Judgment	14283
CASE OF ALTUN v. TURKEY	01/06/2004	Judgment	9396
CASE OF C.R. v. THE UNITED KINGDOM	22/11/1995	Judgment	7150
CASE OF FAYED v. THE UNITED KINGDOM	21/09/1994	Judgment	16493
CASE OF FINDLAY v. THE UNITED KINGDOM	25/02/1997	Judgment	8620
CASE OF GIRARDI v. AUSTRIA	11/12/2003	Judgment	3658
CASE OF GuL v. SWITZERLAND	19/02/1996	Judgment	11321
CASE OF GRADINGER v. AUSTRIA	23/10/1995	Judgment	6882
CASE OF HUSSAIN v. THE UNITED KINGDOM	21/02/1996	Judgment	8272
CASE OF IKINCISOY v. TURKEY	27/07/2004	Judgment	17014
CASE OF JERSILD v. DENMARK	23/09/1994	Judgment	12174
CASE OF KYPRIANOU v. CYPRUS	27/01/2004	Judgment	11680
CASE OF LASKEY, JAGGARD AND BROWN v. THE UNITED KINGDOM	19/02/1997	Judgment	6456
CASE OF LUKANOV v. BULGARIA	20/03/1997	Judgment	7287
CASE OF MURRAY v. THE UNITED KINGDOM	28/10/1994	Judgment	19944
CASE OF PALAORO v. AUSTRIA	23/10/1995	Judgment	5554
CASE OF PFARRMEIER v. AUSTRIA	23/10/1995	Judgment	5629

Case Title	Date	Type	Words
CASE OF SINGH v. THE UNITED KINGDOM	21/02/1996	Judgment	9690
CASE OF VAN MECHELEN AND OTHERS v. THE NETHERLANDS	23/04/1997	Judgment	14274
CASE OF WINGROVE v. THE UNITED KINGDOM	25/11/1996	Judgment	12973
FAYED AND HOUSE OF FRASER HOLDINGS PLC. v. THE UNITED KINGDOM	15/05/1992	Decision	10303
FINDLAY v. THE UNITED KINGDOM	23/02/1995	Decision	6376
GIRARDI v. AUSTRIA	17/05/1995	Decision	1518
GOODWIN v. THE UNITED KINGDOM	01/12/1997	Decision	5099
GRADINGER v. AUSTRIA	10/05/1993	Decision	2379
GUSTAVSSON v. SWEDEN	05/04/1995	Decision	11076
HUSSAIN v. THE UNITED KINGDOM	30/06/1994	Decision	1850
LASKEY, JAGGARD AND BROWN v. THE UNITED KINGDOM	18/01/1995	Decision	4212
LUKANOV v. BULGARIA	12/01/1995	Decision	9636
McCANN, FARRELL AND SAVAGE v. THE UNITED KINGDOM	03/09/1993	Decision	3182
MURRAY v. THE UNITED KINGDOM	21/10/1996	Decision	6554
PALAORO v. AUSTRIA	10/05/1993	Decision	1619
PFARRMEIER v. AUSTRIA	10/05/1993	Decision	1288
RIBITSCH v. AUSTRIA	20/10/1993	Decision	4296
SINGH v. THE UNITED KINGDOM	30/06/1994	Decision	2988
VAN MECHELEN, VENERIUS AND PRUIJMBOOM v. THE NETHERLANDS	15/05/1995	Decision	5115
Z. v. FINLAND	28/02/1995	Decision	4492

Appendix D

Example of human annotated ECHR legal case

- Argument = 1
  - Conclusion = The applicant complained that the length of the maintenance payment proceedings had been incompatible with the reasonable time principle as provided in Article of the Convention , which reads as follows :
  - Premise = In the determination of his civil rights and obligations ..... everyone is entitled to a fair ... hearing within reasonable time .. by[a] ... tribunal
  - Premise = As regards the first set of proceedings the period to be taken into consideration began on January 1990 ended on May 1999 .
  - Premise = Thus , they lasted more than nine years and four months .
  - Premise = As regards the second set of proceedings the period to be taken into consideration began on September 1990 ended on February 1999 .
  - Premise = Thus , they lasted for more than eight years and five months .
- Argument = 2
  - Conclusion = It must therefore be declared admissible . B. Merits .
  - Premise = The Court notes that this complaint is not manifestly ill-founded within the meaning of Article of the Convention .
  - Premise = It further notes that it is not inadmissible on any other grounds .

- Argument = 3
  - Conclusion = The Court reiterates that the reasonableness of the length of proceedings must be assessed in the light of the circumstances of the case and with reference to the criteria established by its case-law , particularly the complexity of the case the conduct of the applicant of the relevant authorities what was at stake for the applicant in the dispute
  - Premise = see , among many other authorities Frydlender v. France [GC , no. ECHR 2000-VII .
- Argument = 6
  - Conclusion = However , an applicant 's behaviour constitutes an objective fact which cannot be attributed to the respondent State which must be taken into account for the purpose of determining whether or not the reasonable time referred to in Article has been exceeded
  - Premise = see Erkner and Hofbauer v. Austria , no. Commission decision of April 1987 .
- Argument = 5
  - Conclusion = The Court therefore finds that the overall length of the proceedings cannot be regarded as reasonable . Accordingly , there has been a violation of Article of the Convention .
  - Premise = The Court reiterates that the reasonableness of the length of proceedings must be assessed in the light of the circumstances of the case and with reference to the criteria established by its case-law , particularly the complexity of the case the conduct of the applicant of the relevant authorities what was at stake for the applicant in the dispute
  - Premise = The Court considers that the present proceedings can clearly be distinguished from the custody proceedings as they concerned merely the fixing of maintenance payments were not particularly complex .
  - Premise = As regards the conduct of the applicant the Court has consistently held that applicants cannot be blamed for making full use of the remedies available to them under domestic law .
  - Premise = However , an applicant 's behaviour constitutes an objective fact which cannot be attributed to the respondent State which must be taken into account for the purpose of determining whether or not the reasonable time referred to in Article has been exceeded
  - Premise = In the present case the Court acknowledges that the applicant had filed numerous requests , complaints and motions had repeatedly failed to obey the authorities ' summons .
  - Premise = Although such conduct contributed to prolonging the proceedings it is not in itself sufficient to explain the length of the extensive proceedings .

- Premise = On the other hand the Court notes that there are substantial delays attributable to the authorities .
- Premise = In particular , in the first set of proceedings there is a period of inactivity of more than two years from January 1990 to February 1992 while the case was pending before the Vienna Juvenile Court a further one of six years from May 1992 to May 1998 before that court took a new decision after the first one had been quashed on appeal .
- Premise = The Court cannot find that the Government has given sufficient explanation for these delays that occurred .
- Premise = In the second set of proceedings there is a period of inactivity of some three years from September 1990 to August 1993 while the case was pending before the Vienna Juvenile Court a further such period of three years and seven months from January 1995 to August 1998 before that court took a new decision after the first one had been quashed on appeal .

Appendix E

Example of automatic annotated ECHR legal case: statistical classifier

- Premise: 0
- Conclusion: 1
- Non-argumentative: -1

THE FACTS .|-1

The applicant was born in 1951 and lives in Vienna .|-1

She is the mother of M , L and R , born in wedlock in 1973 , 1974 and 1976 , respectively .|-1

The spouses separated in 1982 .|-1

Custody of L and M was assigned to the applicant , the custody of R to the father .|-1

In December 1989 M was admitted in a public girls ' home as she refused to stay with her mother .|-1

She stayed there until January 1992 .|-1

From December 1989 until September 1995 custody proceedings concerning the temporary transfer of M 's custody to the Vienna Youth Welfare Office for the time M had spent at the girls ' home were pending before the Austrian courts .|-1

A. The Youth Welfare Office 's request for reimbursement of expenses .|-1

On January 1990 the Vienna Youth Welfare Office , on behalf of M , filed a request with the Floridsdorf District Court that the applicant should pay a monthly contribution to the expenses incurred for M 's stay in the girls ' home .|-1

The file was later on transferred to the competent Juvenile Court and , in January 1990 , the court heard M 's parents .|-1

On March 1991 the Youth Welfare Office reduced the amount of the requested monthly contribution .|-1

On April 1991 the President of the Juvenile Court granted the applicant 's motion for bias against the competent court clerk Rechtspfleger .|-1

A hearing scheduled for July 1991 was cancelled due to the applicant 's illness .|-1

Further hearings scheduled for September 1991 and September 1991 had to be cancelled because the court 's attempts to deliver the summons to the applicant were unsuccessful .|-1

On February 1992 the Juvenile Court ordered that the applicant had to pay ATS in monthly maintenance for M. The applicant appealed , claiming that she was fit to work to an extent of only .|-1

On March 1992 the case was assigned to another judge as the competent judge had declared himself biased .|-1

On May 1992 the Appeal Chamber quashed the decision and remitted the case back to the Juvenile Court , instructing the latter to take a new decision after having supplemented its proceedings .|-1

In particular , it stated that the first instance court ought to appoint a forensic medical expert in order to establish the applicant 's fitness to work .|-1

On May 1998 the Juvenile Court ordered the applicant to pay ATS in monthly maintenance for M. At that stage of the proceedings , no expert had been heard yet .|-1

Referring to the Appeal Chamber 's decision of May 1992 , the applicant appealed , again relying on her reduced fitness to work .|-1

On August 1998 the Juvenile Court appointed an expert in forensic medicine to file a report on the question as to which extent the applicant 's capacities to earn her living were reduced .|-1

The applicant appealed against this decision , claiming that it no longer made sense

to appoint a medical expert , now that the court had already dismissed her request by a decision of May 1998 .|-1

Further , she claimed that there was no need for a further report as , in this respect , she had already submitted two reports of different medical officers Amtsarzt .|-1

On and August 1998 the applicant filed motions for bias against the court clerk Rechtspfleger I.S. , who was dealing with her case , claiming that the appointment of a further medical expert was not justified , that I.S. was handling the case file in a negligent manner , namely that several documents were missing from the file , and that I.S. had been rude to her on the telephone .|-1

On August 1998 the President of the Vienna Juvenile Court PrÃd̈sident des Jugendgerichtshofs dismissed her motion for bias , finding that the mere fact that she had appointed a medical expert was not sufficient to cast doubt upon I.S. ' impartiality .|-1

He also noted that there were no documents missing from the file .|-1

On September 1998 the Appeal Chamber dismissed the applicant 's appeal against the appointment of a medical expert , but granted her appeal against the decision of May 1998 .|-1

In this respect , it referred the case to the Juvenile Court for supplementing the taking of evidence , namely to comply with its decision of May 1992 .|-1

On and March 1999 the applicant requested that , pursuant to Section of the Courts Act Gerichtsorganisationsgesetz , a time-limit be fixed for the decision on the Youth Welfare Office 's application of January 1990 .|-1

On March 1999 the applicant filed a motion for bias against I.S. , claiming that the latter had not been available to her during office hours and that she had refused to give her information requested over the telephone .|-1

On March 1999 the President of the Vienna Juvenile Court dismissed her motion as being unfounded .|-1

On March 1999 the President rejected her appeal against this decision , as the relevant provisions of the Court Clerks Act Rechtspflegergesetz did not provide for such remedy .|-1

On April 1999 the applicant was summoned by the appointed medical expert to undergo a medical examination at the Institute for Forensic Medicine Institut fÃijr Gerichtsmedizin on April 1999 .|-1

It appears that the applicant filed numerous complaints with the President of the Juvenile Court , again claiming that documents were missing from the file and that I.S. as well as various judges of the Juvenile Court were biased .|-1

On May 1999 the President of the Juvenile Court decided to exclude I.S. from the proceedings .|-1

He noted that the latter had expressed that she considered herself biased following a telephone conversation in the course of which the applicant had said she would kill her daughter if I.S. continued to harass her .|-1

In these circumstances , the President found it advisable that the matter be re-assigned in accordance with the Juvenile Court 's rules on the distribution of cases GeschÃdftsverteilung .|-1

On the same day , the Juvenile Court dismissed the applicant 's requests for a time-limit to be set .|-1

Referring to the applicant 's numerous requests , complaints and motions for bias filed with the court , it found that there was no indication of a lack of due diligence on behalf of the Juvenile Court , it being rather the applicant who prevented that a decision on the merits had been taken so far .|-1

On May 1999 the Vienna Youth Welfare Office withdrew its request dated of January 1990 .|-1

Thereupon , the applicant , on May 1999 , withdrew all requests and complaints still pending before the Juvenile Court at that stage .|-1

B. The applicant 's request for reimbursement of expenses .|-1

From July 1990 to September 1990 M stayed with her mother .|-1

The latter , on September 1990 filed a request with the Juvenile Court , claiming reimbursement of her expenses incurred during this period .|-1

In September 1990 the Vienna Youth Welfare Office reimbursed the applicant for M 's stay with her from July 1990 to August 1990 .|-1

On August 1993 the Juvenile Court dismissed the applicant 's request for expenses incurred during the rest of the period .|-1

On August 1993 the President of the Vienna Juvenile Court dismissed the applicant 's motion of bias against the competent judge .|-1

On December 1993 the Vienna Court of Appeal granted the applicant 's appeal against this decision and quashed the decision .|-1

On January 1994 the Appeal Chamber of the Juvenile Court again dismissed the applicant 's motion for bias .|-1

On May 1994 the Court of Appeal rejected the applicant 's appeal .|-1

A further appeal to the Supreme Court was to no avail .|-1

A further motion for bias against the President of the Juvenile Court was to no avail either .|-1

On January 1995 the Appeal Chamber quashed the decision of August 1993 and remitted the case back to the first instance court .|-1

On April 1998 the applicant requested that , pursuant to Section of the Courts Act , a time-limit be fixed for the decision on her application of September 1990 .|-1

On June 1998 the President of the Vienna Juvenile Court ordered the Juvenile Court to decide on the applicant 's request no later than on July 1998 .|-1

On August 1998 the Juvenile Court dismissed the applicant 's request for maintenance payments of September 1990 .|-1

The applicant appealed against this decision .|-1

It appears from the documents submitted that the applicant filed several complaints with the Vienna Court of Appeal Oberlandesgericht , claiming that I.S. had not complied with the time limit set by the President of the Juvenile Court because she had gone on holidays , that the competent judicial officer , I.S. was to be found at her office only twice a week and that she had been extraordinarily impolite to her .|-1

Thereupon , the President of the Juvenile Court , on August 1998 , informed the applicant that both I.S. 's office hours as well as her right to vacation were in accordance with her assignment .|-1

He also expressed his regret that , if , in the course of one of the applicant 's numerous telephone calls , I.S. might have acted in a slightly indignant way .|-1

However , he emphasised that the applicant 's allegations had remained unproved .|-1

On September 1998 the Appeal Chamber dismissed her appeal against the Juvenile Court 's decision of August 1998 as being unfounded .|-1

Further , it stated that there was no further appeal on points of law in the applicant 's case as it did not raise questions of law of fundamental importance Ausspruch Ãijber die UnzulÃďssigkeit der ordentlichen Revision .|-1

Nevertheless , the applicant filed an extraordinary appeal on points of law ausserordentliche Revision with the Supreme Court .|-1

Referring to an amendment of Section a of the Non-Contentious Proceedings Act Ausserstreitgesetz , the Supreme Court on December 1998 remitted the case back to the Vienna Juvenile Appeal Court .|-1

According to that provision , instead of filing an extraordinary appeal on points of law with the Supreme Court , a party to non-contentious proceedings must now request the Court of Appeal to re-consider its opinion on the admissibility of an ordinary appeal on points of law .|-1

The Supreme Court found that , even if in her appeal the applicant had not explicitly requested the Juvenile Appeal Court to declare that a further appeal on points of law be allowed , her appeal should have been understood in such a way .|-1

Thereupon , on January 1999 the Juvenile Appeal Court requested the applicant to remedy procedural defects of her appeal , namely to request that an ordinary appeal in her case be allowed .|-1

As the applicant did not comply with this request , the Juvenile Appeal Court , on February 1999 , rejected her appeal .|-1

THE LAW|-1

I. ALLEGED VIOLATION OF ARTICLE OF THE CONVENTION .|-1

The applicant complained that the length of the maintenance payment proceedings had been incompatible with the reasonable time principle as provided in Article of the Convention , which reads as follows :|-1

In the determination of his civil rights and obligations ..... , everyone is entitled to a fair ... hearing within reasonable time .. . by[a .. . tribunal .|-1

As regards the first set of proceedings , the period to be taken into consideration began on January 1990 and ended on May 1999 .|0

Thus , they lasted more than nine years and four months .|1

As regards the second set of proceedings , the period to be taken into consideration began on September 1990 and ended on February 1999 .|0

Thus , they lasted for more than eight years and five months .|1

A. Admissibility .|-1

The Court notes that this complaint is not manifestly ill-founded within the meaning of Article of the Convention .|0

It further notes that it is not inadmissible on any other grounds .|0

It must therefore be declared admissible .|1

B. Merits .|-1

The Government submitted that the maintenance proceedings were complex .|-1

In particular , they had to be seen as a part of highly complex custody proceedings which required extensive expert opinions .|-1

While the authorities tried to conduct the proceedings expeditiously , the applicant filed a multitude of motions of bias , appeals and requests for extension of time-limits and therefore herself contributed considerably to the length of the proceedings .|-1

The Government further stressed that the applicant repeatedly thwarted attempts to deliver summons on her and failed to obey them .|-1

The applicant did not submit any observations on these issues .|-1

The Court reiterates that the reasonableness of the length of proceedings must be assessed in the light of the circumstances of the case and with reference to the criteria established by its case-law , particularly the complexity of the case , the conduct of the applicant and of the relevant authorities and what was at stake for the applicant in the dispute see , among many other authorities , Frydlender v. France [GC] , no. ECHR 2000-VII .|0

The Court considers that the present proceedings can clearly be distinguished from the custody proceedings , as they concerned merely the fixing of maintenance payments and were not particularly complex .|1

As regards the conduct of the applicant the Court has consistently held that applicants cannot be blamed for making full use of the remedies available to them under domestic law .|0

However , an applicant 's behaviour constitutes an objective fact which cannot be attributed to the respondent State and which must be taken into account for the purpose of determining whether or not the reasonable time referred to in Article has been exceeded see Erkner and Hofbauer v. Austria , no. Commission decision of April 1987 , A .|0

In the present case , the Court acknowledges that the applicant had filed numerous requests , complaints and motions and had repeatedly failed to obey the authorities ' summons .|0

Although such conduct contributed to prolonging the proceedings , it is not in itself sufficient to explain the length of the extensive proceedings .|0

On the other hand , the Court notes that there are substantial delays attributable to the authorities .|0

In particular , in the first set of proceedings , there is a period of inactivity of more than two years from January 1990 to February 1992 while the case was pending before the Vienna Juvenile Court , and a further one of six years from May 1992 to May 1998 before that court took a new decision after the first one had been quashed on appeal .|0

In the second set of proceedings , there is a period of inactivity of some three years from September 1990 to August 1993 , while the case was pending before the Vienna Juvenile Court , and a further such period of three years and seven months from January 1995 to August 1998 before that court took a new decision after the first one had been quashed on appeal .|0

The Court cannot find that the Government has given sufficient explanation for these delays that occurred .|0

The Court therefore finds that the overall length of the proceedings cannot be regarded as reasonable .|1

Accordingly , there has been a violation of Article of the Convention .|1

II. APPLICATION OF ARTICLE OF THE CONVENTION .|-1

Article of the Convention provides :|-1

If the Court finds that there has been a violation of the Convention or the Protocols thereto , and if the internal law of the High Contracting Party concerned allows only partial reparation to be made , the Court shall , if necessary , afford just satisfaction to the injured party .|0

The applicant has not filed a claim for just satisfaction .|-1

Accordingly , the Court considers that no award can be made under this provision .|1

FOR THESE REASONS , THE COURT UNANIMOUSLY .|1

Declares the application admissible ; .|1

Holds that there has been a violation of Article of the Convention ;|1

Done in English , and notified in writing on December 2003 , pursuant to Rule and of the Rules of Court .|-1

Vincent Berger Georg Ress Registrar President|-1

Appendix F

Example of an ECHR legal case automatically annotated by the argumentative parser

```
T
|--D
|  |--x: for these reasons, the commission by a majority declares the
|       application admissible, without prejudging the merits.
|--A
|  |--C: it follows that the application cannot be dismissed as
|  |     manifestly ill-founded.
|  |--A
|     |--P:  it considers that the applicant 's complaints raise serious
|              issues of fact and law under the convention, the
|              determination of which should depend on an examination
|              of the merits.
|
|--A
   |--C: in these circumstances, the commission finds that the application
   |     cannot be declared inadmissible for non-exhaustion of domestic
   |     remedies.
   |--A
   |  |--P: the commission recalls that article art. of the
   |        convention only requires the exhaustion of such remedies which relate
   |        to the breaches of the convention alleged and at the same time can
   |        provide effective and sufficient redress.
   |--A
      |--C: the commission finds that the suggested application
      |     for discretionary relief in the instant case cannot do so either.
      |--A
      |  |--C: it is furthermore established that the burden of proving the
      |  |     existence of available and sufficient domestic remedies lies upon the
      |  |     State invoking the rule
      |  |--A
      |     |--P: cf. eur. court. h.r. deweer judgment of february,
      |              series a no. p. para. and no. dec. d.r. p. at p.
      |
      |--A
      |  |--C: an applicant does not need to exercise remedies which,
      |  |     although theoretically of a nature to
      |  |     constitute remedies, do not in reality offer any chance of redressing
      |  |     the alleged breach
      |  |--A
      |     |--P: cf. no. dec. d.r. p.
      |
      |--A
         |--P: the commission recalls that article art. of the convention
         |     only requires the exhaustion of such remedies which
         |     relate to the breaches of the convention alleged and at
         |     the same time can provide effective and sufficient redress.
         |--P: the commission notes that in the context of the section
         |     powers the secretary of state has a very wide discretion.
         |--P: the commission recalls that in the case of temple
               v. the united kingdom no. dec. d.r. p. the Commission
               held that recourse to a purely discretionary power
               on the part of the secretary of state did not
               constitute an effective domestic remedy.
```

Appendix G

Example of human annotation without training

protection under the Convention , it must be recognised that their position is to some extent different from that of a disinterested witness or a victim .
<sentence 312> They owe a general duty of obedience to the State 's executive authorities and usually have links with the prosecution ; for these reasons alone their use as anonymous witnesses should be resorted to only in exceptional circumstances .
<sentence 313> In addition , it is in the nature of things that their duties , particularly in the case of arresting officers , may involve giving evidence in open court .
<sentence 314> On the other hand , the Court has recognised in principle that , provided that the rights of the defence are respected , it may be legitimate for the police authorities to wish to preserve the anonymity of an agent deployed in undercover activities , for his own or his family 's protection and so as not to impair his usefulness for future operations see the above-mentioned LÃ¼di judgment , p. para. .
<sentence 315> Having regard to the place that the right to a fair administration of justice holds in a democratic society , any measures restricting the rights of the defence should be strictly necessary .
<sentence 316> If a less restrictive measure can suffice then that measure should be applied .
<sentence 317> In the present case , the police officers in question were in a separate room with the investigating judge , from which the accused and even their counsel were excluded .
<sentence 318> All communication was via a sound link .
<sentence 319> The defence was thus not only unaware of the identity of the police witnesses but were also prevented from observing their demeanour under direct questioning , and thus from testing their reliability see the above-mentioned Kostovski judgment , p. para. in fine .
<sentence 320> It has not been explained to the Court 's satisfaction why it was necessary to resort to such extreme limitations on the right of the accused to have the evidence against them given in their presence , or why less far-reaching measures were not considered .
<sentence 321> In the absence of any further information , the Court cannot find that the operational needs of the police provide sufficient justification .
<sentence 322> It should be noted that the explanatory memorandum of the bill which became the Act of November 1993 see paragraph above refers in this connection to the possibilities of using make-up or disguise and the prevention of eye contact .
<sentence 323> Nor is the Court persuaded that the Court of Appeal made sufficient effort to assess the threat of reprisals against the police officers or their families .
<sentence 324> It does not appear from that court 's judgment that it sought to address the question whether the applicants would have been in a position to carry out any such threats or to incite others to do so on their behalf .
<sentence 325> Its decision was based exclusively on the seriousness of the crimes committed see paragraph above .
<sentence 326> In this connection , it is to be noted that Mr Engelen , a civilian witness who in the early stages of the proceedings had made statements identifying one of the applicants as one of the perpetrators , did not enjoy the protection of anonymity and it has not been claimed that he was at any time threatened .
<sentence 327> It is true - as noted by the Government and the Commission see paragraph above - that the anonymous police officers were interrogated before an investigating judge , who had himself ascertained their identity and had , in a very detailed official report of his findings , stated his opinion on their reliability and credibility as well as their reasons for remaining anonymous .
<sentence 328> However these measures cannot be considered a proper substitute for the possibility of the defence to question the witnesses in their presence and make their own judgment as to their demeanour and reliability .
<sentence 329> It thus cannot be said that the handicaps under which the defence laboured were counterbalanced by the above procedures .
<sentence 330> Moreover , the only evidence relied on by the Court of Appeal which provided positive identification of the applicants as the perpetrators of the crimes were the statements of the anonymous police officers .
<sentence 331> That being so the conviction of the applicants was based `` to a decisive extent '' on these anonymous statements .
<sentence 332> In the Court 's view , the present case falls to be distinguished from that of Doorson : in the latter case it was decided on the basis of information contained in the case file itself that the witnesses Y. and Y. - who were both civilians , and who knew the accused personally - had sufficient reason to believe that he might resort to violence , and they were heard in the presence of counsel see the above-mentioned Doorson judgment , pp. para. pp. para. and pp. paras. and .
<sentence 333> In addition , in the latter case other evidence providing positive identification of the accused as the perpetrator of the crimes charged was available from sources unrelated to the anonymous witnesses ibid. , pp. para. and p. para. .
<sentence 334> Against this background the Court cannot find that the proceedings taken as a whole were fair .
<sentence 335> C. Conclusion .
<sentence 336> There has been a violation of Article para. taken together with Article para . art. 3-d .
<sentence 337> II. APPLICATION OF ARTICLE OF THE CONVENTION art. .

Bibliography

[1] Baker, G., and Huntington, H. *The Principles of Argumentation*. Ginn & company, 1905. pages 15

[2] Berger, A., Pietra, V., and Pietra, S. A Maximum Entropy Approach to Natural Language Processing. *Computational Linguistics 22*, 1 (1996), 39–71. pages 8, 25

[3] Besnard, P., and Hunter, A. *Elements of Argumentation*. MIT Press, 2008. pages 11, 104

[4] Boser, B., Guyon, I., and Vapnik, V. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational learning theory* (1992), ACM, pp. 144–152. pages 8

[5] Briscoe, T., and Carroll, J. Generalized Probabilistic LR Parsing of Natural Language (Corpora) with Unification-based Grammars. *Computational Linguistics 19*, 1 (1993), 25–59. pages 137

[6] Carletta, J. Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics 22*, 2 (1996), 249–254. pages 72

[7] Charniak, E. *Statistical Language Learning*. The MIT Press, 1996. pages 30, 31, 139

[8] Charniak, E. A maximum-entropy-inspired parser. In *Proceedings of the 1st North American Chapter of the Association for Computational Linguistics Conference* (2000), Morgan Kaufmann Publishers Inc., pp. 132–139. pages 99, 105

[9] Chesñevar, C., et al. Towards an argument interchange format. *The Knowledge Engineering Review 21*, 04 (2006), 293–316. pages 38, 39

[10] Colmerauer, A. Total precedence relations. *Journal of the ACM (JACM) 17*, 1 (1970), 30. pages 8

[11] Delannoy, J. Argumentation mark-up: A proposal. In *Workshop" Towards Standards and Tools for Discourse Tagging", Conference of the Association for Computational Linguistics (ACL'99), U. Maryland, College Park, MD* (1999). pages 38, 39

[12] Domingos, P., Kok, S., Lowd, D., Poon, H., Richardson, M., and Singla, P. *Markov Logic*. Springer-Verlag, 2008. pages 140

[13] Emmet, L., and Cleland, G. Graphical notations, narratives and persuasion: a pliant systems approach to hypertext tool design. In *Proceedings of the Thirteenth ACM Conference on Hypertext and Hypermedia* (2002), ACM, pp. 55–64. pages 38

[14] Fontan, L., and Saint-Dizier, P. Constructing a know-how repository of advices and warnings from procedural texts. In *Proceeding of the eighth ACM Symposium on Document Engineering* (2008), ACM, pp. 249–252. pages 38

[15] Gordon, T. A computational model of argument for legal reasoning support systems. In *Argumentation in Artificial Intelligence and Law, IAAIL Workshop Series* (2005), pp. 53–64. pages 127

[16] Grover, C., Hachey, B., Hughson, I., et al. The HOLJ Corpus: supporting summarisation of legal texts. In *Proceedings of the 5th International Workshop on Linguistically Interpreted Corpora, Geneva, Switzerland* (2004). pages 37, 92

[17] Hammerton, J., Osborne, M., Armstrong, S., and Daelemans, W. Introduction to special issue on machine learning approaches to shallow parsing. *The Journal of Machine Learning Research 2* (2002), 551–558. pages 32

[18] Harabagiu, S. *WordNet-based Inference of Textual Context, Cohesion and Coherence*. PhD thesis, 1997. pages 102

[19] Harris, Z. *String Analysis of Sentence Structure*. Mouton, 1962. pages 8

[20] Hearst, M. TextTiling: segmenting text into multi-paragraph subtopic passages. *Computational Linguistics 23*, 1 (1997), 33–64. pages 89

[21] Hobbs, J., Stickel, M., Appelt, D., and Martin, P. Interpretation as Abduction. *Artificial Intelligence 63*, 1-2 (1993), 69–142. pages 102

[22] Hovy, E. Automated discourse generation using discourse structure relations. *Artificial intelligence 63*, 1-2 (1993), 341–385. pages 96

[23] Huffman, S. Learning information extraction patterns from examples. *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing* (1996), 246–260. pages 8

[24] Jelinek, F., Bahl, L., and Mercer, R. Design of a linguistic statistical decoder for the recognition of continuous speech. *Information Theory, IEEE Transactions on 21*, 3 (2002), 250–256. pages 8

[25] Joachims, T. Text Categorization with Support Vector Machines: Learning with many relevant features. *Machine Learning: ECML-98* (1998), 137–142. pages 26

[26] Kienpointner, M. *Alltagslogik: Struktur und Funktion von Argumentationsmustern*. Frommann-Holzboog, 1992. pages 104

[27] Kim, J., and Moldovan, D. Acquisition of semantic patterns for information extraction from corpora. In *Artificial Intelligence for Applications, 1993. Proceedings., Ninth Conference on* (2002), IEEE, pp. 171–176. pages 8

[28] Knott, A., and Dale, R. Using Linguistic Phenomena to Motivate a Set of Coherence Relations. *Discourse Processes 18*, 1 (1994), 35–62. pages 101, 106

[29] Krippendorff, K. Content analysis: An introduction to its methodology. 1980. *Beverly Hills,: Sage Publications*. pages 71, 73

[30] Krippendorff, K. *Content analysis: An introduction to its methodology*. Sage Publications, Inc, 2004. pages 73

[31] La Torre, M. Theories of legal argumentation and concepts of law. An approximation. *Ratio Juris 15*, 4 (2002), 377–402. pages 76

[32] Leech, G. Corpora and theories of linguistic performance. In *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991*, pp. 105–125. pages 7

[33] Mann, W., and Thompson, S. Rhetorical structure theory: A framework for the analysis of texts. pages 93, 94, 95

[34] Manning, C., and Schütze, H. *Foundations of Statistical Natural Language Processing*, vol. 59. MIT Press, 1999. pages 6

[35] Marcu, D. The rhetorical parsing of natural language texts. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics* (1997), Association for Computational Linguistics, pp. 96–103. pages 96

[36] Marcu, D. A decision-based approach to rhetorical parsing. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics* (1999), Association for Computational Linguistics, pp. 365–372. pages 96

[37] Marcu, D., and Echihabi, A. An Unsupervised Approach to Recognizing Discourse Relations. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* (2002), Association for Computational Linguistics, pp. 368–375. pages 115

[38] Marcus, M., Marcinkiewicz, M., and Santorini, B. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics 19*, 2 (1993), 330. pages 8, 139

[39] McCallumzy, A., and Nigamy, K. A comparison of Event Models for Naive Bayes Text Classification, 1998. pages 23, 24

[40] Miltsakaki, E., Prasad, R., Joshi, A., and Webber, B. The Penn discourse treebank. In *Proceedings of the 4th International Conference on Language Resources and Evaluation* (2004). pages 8

[41] Mochales, R., and Ieven, A. Creating an Argumentation Corpus: Do Theories Apply to Real Arguments? A Case Study on the Legal Argumentation of the ECHR. In *Proceedings of the Twelth International Confernce on Artificial Intellifence and Law* (2009), New York, ACM, pp. 21–30. pages 19, 20

[42] Mochales, R., and Moens, M. F. Study on the Structure of Argumentation in Case Law. In *In Proceedings of JURIX 2008: The 21st International Conference*

on Legal Knowledge and Information Systems. (pp. 11-20) (2008), IOS Press. pages 21

[43] Moens, M. F., Boiy, E., Mochales, R., and Reed, C. Automatic Detection of Arguments in Legal Texts. In *In Proceedings of the Eleventh International Conference on Artificial Intelligence and Law (pp. 225-230)* (2007), New York, ACM. pages 40, 41

[44] Morris, J., and Hirst, G. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics 17*, 1 (1991), 21–48. pages 89

[45] Pallotta, V., Ghorbel, H., Ruch, P., and Coray, G. An argumentative annotation schema for meeting discussions. *Arxiv preprint cs/0410061* (2004). pages 37

[46] Pallotta, V., Seretan, V., Ailomaa, M., Ghorbel, H., and Rajman, M. Towards an argumentative coding scheme for annotating meeting dialogue data. Tech. rep., Technical report, Institut des systmes d'information et de communications, 2007. pages 37

[47] Parsons, S., Sierra, C., and Jennings, N. Agents that reason and negotiate by arguing. *Journal of Logic and Computation 8*, 3 (1998), 261. pages 17

[48] Perelman, C., Olbrechts-Tyteca, L., Wilkinson, J., and Weaver, P. *The new rhetoric: A treatise on argumentation*. University of Notre Dame Press London, 1971. pages 15

[49] Prakken, H., and Vreeswijk, G. Logics for defeasible argumentation. *Handbook of Philosophical logic 4* (2002), 218–319. pages 17

[50] Pustejovsky, J., Hanks, P., Sauri, R., See, A., Gaizauskas, R., Setzer, A., Radev, D., Sundheim, B., Day, D., Ferro, L., et al. The timebank corpus. In *Corpus Linguistics* (2003), vol. 2003, p. 40. pages 8

[51] Rahwan, I., and Reed, C. The argument interchange format. *Argumentation in Artificial Intelligence* (2009), 383–402. pages 39

[52] Reed, C., and Rowe, G. Araucaria: software for solving puzzles in argument diagramming. *Department of Applied Computing, University of Dundee Technical Report, also available from: www. computing. dundee. ac. uk/staff/creed/araucaria* (2001). pages 39

[53] Reed, C., and Rowe, G. Araucaria: Software for argument analysis, diagramming and representation. *International Journal of AI Tools 14*, 3-4 (2004), 961–980. pages 127

[54] Reed, C., and Walton, D. Argumentation schemes in argument-as-process and argument-as-product. In *Proceedings of the conference celebrating informal logic* (2003), vol. 25. pages 13

[55] Riloff, E. An empirical study of automated dictionary construction for information extraction in three domains. *Artificial Intelligence 85*, 1-2 (1996), 101–134. pages 8

[56] Siegel, S., and Castellan, N. Jr. Nonparametric statistics for the behavioral sciences, 1988. pages 72

[57] Steinwart, I., and Christmann, A. *Support Vector Machines*. Springer Verlag, 2008. pages 26

[58] Stolcke, A. An Efficient Probabilistic Context-free Parsing Algorithm that Computes Prefix Probabilities. *Computational Linguistics 21*, 2 (1995), 165–201. pages 137

[59] Teufel, S. Argumentative zoning: Information extraction from scientific text. *PhD thesis, University of Edinburgh* (1999). pages 60, 89, 90, 91, 92, 93

[60] Teufel, S., Carletta, J., and Moens, M. An annotation scheme for discourse-level argumentation in research articles. In *Proceedings of the Ninth Conference on European Chapter of the Association for Computational Linguistics* (1999), Association for Computational Linguistics, pp. 110–117. pages 37

[61] Toulmin, S. *The Uses of Argument*. Cambridge Univ Pr, 2003. pages 15

[62] van Dijk, T. *Macrostructures*. L. Erlbaum, Associates, 1980. pages 89

[63] Van Eemeren, F. *Crucial Concepts in Argumentation Theory*. Amsterdam Univ Pr, 2001. pages 14, 16

[64] Van Eemeren, F., and Grootendorst, R. *A Systematic Theory of Argumentation: The Pragma-dialectical Approach*. Cambridge Univ Pr, 2004. pages 16

[65] Van Eemeren, F., Grootendorst, R., and Henkemans, F. *Fundamentals of Argumentation Theory: A Handbook of Historical Backgrounds and Contemporary Developments*. Lawrence Erlbaum, 1996. pages 2, 16

[66] Verheij, B. *Virtual Arguments: On the Design of Argument Assistants for Lawyers and Other Arguers*. TMC Asser Press, 2005. pages 127

[67] Wagner, J., Seddah, D., Foster, J., Van Genabith, J., Butt, M., and King, T. C-structures and F-structures for the British National Corpus. In *Proceedings of the Twelfth International Lexical Functional Grammar Conference* (2007). pages 110, 111, 112

[68] Wallach, H. Conditional Random Fields: An introduction. *Rapport technique MS-CIS-04-21, Department of Computer and Information Science 50* (2004). pages 108

[69] Walton, D. *Argumentation Schemes for Presumptive Reasoning*. Mahwah, New Jersey: Lawrence Erlbaum Associates, 1996. pages 18, 40, 61, 62, 79, 80, 157

[70] Walton, D. *Legal Argumentation and Evidence*. Pennsylvania State Univ Pr, 2002. pages 62

[71] Walton, D. *Informal logic: A Pragmatic Approach*. Cambridge University Press, 2008. pages 61, 63, 80, 81

List of publications

International Conferences

- MOENS, Marie-Francine, BOIY, Erik, MOCHALES PALAU, Raquel and REED, Chris. (2007) Automatic Detection of Arguments in Legal Texts. In Proceedings of the Eleventh International Conference on Artificial Intelligence and Law (pp. 225-230). New York: ACM.
- MOCHALES PALAU, Raquel and MOENS, Marie-Francine (2007) Study on Sentence Relations in the Automatic Detection of Argumentation in Legal cases. In Proceedings of JURIX 2007: The 20th Anniversary International Conference on Legal Knowledge and Information Systems. (pp.89-98) IOS Press.
- MOCHALES PALAU, Raquel and MOENS, Marie-Francine (2008) Study on the Structure of Argumentation in Case Law. In Proceedings of JURIX 2008: The 21st International Conference on Legal Knowledge and Information Systems. (pp.11-20) IOS Press.
- MOCHALES PALAU, Raquel and MOENS, Marie-Francine (2009) Argumentation Mining: The Detection, Classification and Structure of Arguments in Text. In Proceedings of the Twelfth International Conference on Artificial Intelligence and Law. (pp.98-109) New York: ACM. (Best Student's Paper Award)
- MOCHALES PALAU, Raquel and IEVEN, Aagje (2009) Creating an Argumentation Corpus: Do Theories Apply to Real Arguments? A Case Study on the Legal Argumentation of the ECHR. In Proceedings of the Twelfth International Conference on Artificial Intelligence and Law (ICAIL). (pp.21-30) New York: ACM.

Other Conferences

- MOCHALES PALAU, Raquel and MOENS, Marie-Francine (2007) Automatic Detection of Arguments in Legal Texts. In Proceedings of the 19th Belgian-Dutch Conference on Artificial Intelligence. Utrecht, the Netherlands.
- REED, Chris, MOCHALES PALAU, Raquel, ROWE, Glenn and MOENS, Marie-Francine (2008) Language Resources for Studying Argument. In

Proceedings of Language Resources and Evaluation Conference (LREC). (pp.91-100) Marrakech (Morocco).

- MOCHALES PALAU, Raquel and MOENS, Marie-Francine (2009) Argumentation Mining: The Detection, Classification and Structuring of Arguments in Text. In Proceedings of the 21st Benelux Conference on Artificial Intelligence.

Workshops

- MOCHALES PALAU, Raquel and MOENS, Marie-Francine (2007) ACILA - Automatic Detection of Arguments in Legal Cases. In Proceedings of Workshop on Semantic Web Technology for Law. (pp.5-9) Palo Alto, CA: Stanford University.

International Journals

- MOCHALES PALAU, Raquel and MOENS, Marie-Francine (2011) Argumentation Mining. Artificial Intelligence and Law. 19(1), (pp.1-22).

Book Chapters

- MOCHALES PALAU, Raquel. and MOENS, Marie-Francine (2008) Automatic Argumentation Detection and Its Role in the Semantic Web. In J. BREUKER, P. CASANOVAS, M. KLEIN & E. FRANSCESCONI (Eds.), Law, Ontologies and the Semantic Web (Frontiers in Artificial Intelligence and Applications). (p.115-129) Amsterdam: IOS Press.
- WYNER, Adam, MOCHALES-PALAU, Raquel, MOENS, Marie-Francine and MILWARD, David (2010) Approaches to Text Mining Arguments from Legal Cases. In E. Francesconi, S. Montemagni, W. Peters & D. Tiscornia (Eds.), Semantic Processing of Legal Texts: Where the Language of Law Meets the Law of Language (LNAI 6036) (pp. 60-79). Berlin: Springer.

Arenberg Doctoral School of Science, Engineering & Technology
Faculty of Engineering
Department of Computer Science
Scientific Computing Group
Celestijnenlaan 200A box 2402
B-3001 Heverlee

KATHOLIEKE UNIVERSITEIT LEUVEN

ASSOCIATIE K.U.LEUVEN