Approaches to Text Mining Arguments from Legal Cases

Adam Wyner[1], Raquel Mochales-Palau[2], Marie-Francine Moens[2], and David Milward[3]

[1] Department of Computer Science
University College London
Gower Street, London WC1E 6BT, United Kingdom
adam@wyner.info
www.wyner.info/research
[2] Computer Science Department
Katholieke Universiteit Leuven
Afdeling Informatica, Celestijnenlaan, 3001 Heverlee, Belgium
raquel.mochalespalau, sien.moens@cs.kuleuven.be
www.cs.kuleuven.be/cs/
[3] Linguamatics Ltd
St. John's Innovation Centre, Cowley Rd
Cambridge, CB4 0WS, United Kingdom
david.milward@linguamatics.com
www.linguamatics.com

Abstract. This paper describes recent approaches using text-mining to automatically profile and extract arguments from legal cases. We outline some of the background context and motivations. We then turn to consider issues related to the construction and composition of a corpora of legal cases. We show how a Context-Free Grammar can be used to extract arguments, and how ontologies and Natural Language Processing can identify complex information such as case factors and participant roles. Together the results bring us closer to automatic identification of legal arguments.

Key words: text mining, legal argument, case based reasoning

1 Introduction

In countries with legal systems using common law, such as the United States and the United Kingdom, case law plays a critical role in legal reasoning and decision-making. Case law is that corpus of decisions on cases which judges have made; we refer to this corpus as the *case base* and the previous decisions as the *precedents*. Given a current case, a lawyer consults the case base and identifies precedents that support their side in the legal dispute and undermine the other. The lawyer presents the precedents as, in effect, analogical arguments: the precedents are related to the current case in certain respects, and as the precedent was decided, so too should the current case. However, the cases are expressed in natural language, consider highly complex matters that are under

dispute, relate to laws which justify the decision, and have complex inter-relationships such as when one case decision overturns a prior case decision. Legal professionals must undergo very extensive training in navigating the case base, interpreting the results, and applying the results successfully to their current case. Adding to this complexity, the case base is comprised of a large number of cases and grows every year[4]. Thus, the legal professional faces the difficult task of retrieving and interpreting information from the case base.

Historically, legal professionals have a variety of tools they have been able to use to manage and search the case base in order to identify the relevant cases and material (e.g. compilations of decided cases as well as *Shepard's Citations*, which indexes cases with respect to applicable precedents). More recently, with electronic documentation and automated techniques, legal professionals can search the case base quickly and with respect to a range of parameters. Large companies such as *Lexis-Nexis* and *Westlaw* provide legal information along with access to legal case bases to legal professionals. As we discuss below, the information allows legal professionals to search through the case base relative to a set of terms and quickly returns a set of candidate cases for the legal professional to consider. However, while there are tools to refine the search, by and large, the results returned are fairly coarse-grained; it is up to the legal professional to read the case abstract or the body of the case itself to determine if it suits the case at hand.

Automated text mining tools that can perform information extraction on the case base have a range of advantages. Using such tools, detailed properties and relationships within and among cases can be identified. Searches can be carried out and information can be made available to legal researchers on new cases automatically as they are added to the case base. The goal of information extraction is to automatically extract structured information from unstructured machine-readable texts. The information is structured in that it identifies semantic properties or relationships in the texts. For example, suppose we want to identify the set of lawyers who have never lost a case in a certain domain (e.g. real estate) along with the set of cases the lawyers used to argue their side. Information extraction is more specific than information retrieval, which identifies documents, rather than lists of lawyers or relationships between lawyers and successful cases. For example, search engines such as *Google* identify sources using keywords and return sets of links in which the keywords are found, but not the relationships among the keywords.

The paper has several objectives. First, it is of significant research interest to automatically identify legal arguments, properties, and relationships as found in legal cases. Second, the research here can be viewed as a contribution to the development of tools which support legal professionals in their activities such as identifying relevant cases in the case base. Finally, the research can help us better understand the meaning of the law and the functioning of the legal system by, in effect, a bottom-up investigation starting from case law, which is the foundation of the common law system.

---

[4] See searchable databases available on website of the World Legal Information Institute (WorldLii) with links to databases of legal decisions of countries such as the USA:
http://www.worldlii.org/
Also see the databases of US law at the Legal Information Institute of the Cornell Law School:
http://www.law.cornell.edu/

The paper is structured as follows. In section 2, we outline the relevant background and scope of our work such as the range of approaches to argumentation and previous related work. In section 3, we discuss issues about the development of a corpus of legal arguments as it appears in recent research. A variety of approaches have been taken, and we indicate some of the advantages and disadvantages of each. Section 4 presents current results of three approaches to information extraction in legal cases. In 4.1, the focus is on parsing arguments into an *argument interchange format* as well as to try to automatically identify argument sentences from non-argument sentences. In 4.2, the emphasis is identifying argument characteristics from a set of legal documents and providing a context-free grammar. In 4.3, a range of semantically relevant elements are extracted from a case base which include those indexed by commerical services, but also novel complex information such as case factors. Together the approaches provide highly related aspects of the automatic identification of legal arguments wherein participants argue about issues such as case factors. In 5, we discuss the relation of this work with that of others as well as provide indications of the direction of future work. We use the terms "argumentation" and "argument", where argumentation is about the abstract theory and argument relates to particular instances that may be chained together; for example, we have theories of argumentation, while *if Bill is unhappy, then he should leave the party. If Bill leaves the party, then he should take his dog with him* is a chain of arguments (see [1] for related discussion).

2 Background

In this section, we outline the range of approaches to argumentation in order to set the context of our work. In addition, we refer to some key prior work in the area of legal text mining.

2.1 Argumentation Theory and Analysis

Research in argumentation is interdisciplinary, relating discussions found in Philosophy, Linguistics, and Computer Science. *Empirically* oriented approaches attend to specific, linguistically realised argument structures, properties, or elements of legal texts ([2], [3], [4], [5], [6], and [7]). Given analyses of argument patterns, arguments can be graphically represented as trees, where premises branch off of conclusions [8]. XML markup languages have been developed for argumentation such that an argument, once marked up, can be searched for or used for reasoning [9]. While the results of some of this work can be used for information extraction, it is not produced automatically and is not suitable for working with large corpora.

In computational models of argumentation, the abstract structures of and reasoning with arguments are proposed. In Argumentation Frameworks, arguments are abstract and atomic objects in attack relations ([10], [11], [12], [13], and [14]). Argumentation Frameworks account for a range of issues in non-monotonic reasoning [15]; they can be extended to accommodate more fine-grained elements of relations between arguments such as different modes of attack ([14], [1], and [16]). Such approaches focus on high

level generalisations about sets of arguments and the complexity of their relationships. However, the theories do not address a range of aspects of natural argumentation.

There have been attempts to connect abstract arguments with concrete arguments ([17], [18], [19], and [20]). Such systems start with a knowledge base comprised of facts and rules, where the rules typically include both strict (*SI*) and defeasible (*DI*) inference rules in a Defeasible Logic (*DL*) ([21] and [22]). In these approaches, simplified examples are manually translated into the formal language and the objective is to draw inferences given the knowledge base. An additional realisation of argumentation theory are the argument schemes for case based reasoning of [23], which relates arguments and case based reasoning by allowing schemes to be used to argue for or against a decision given comparative case factors found in the case base.

2.2 Text-mining in the Case base

Current commerical systems (e.g. Lexis-Nexis or Westlaw) or web-based public services (e.g. WorldLii) have limited text mining capabilities. One way to identify a set of relevant cases is by selecting from among a small finite set of indicies which are manually assigned to the cases. More 'advanced' facilities support regular expression searches along with boolean operators and proximity operators. Such facilities do not reflect any semantic information.

A range of other approaches search for semantic content. Recent work in computational semantics focuses on recognising inference, calculating entailment, and identifying inconsistency ([24], [25], [26], [27], [28], [29] and [30]). Such work does not address arguments with complex structure, relationships between arguments, or key elements of legal information found in the case base. [31] presents a representation of knowledge of cases that would suit an information retrieval system for a case base, outlining different sorts of knowledge that ought to be identified and by which cases are classified – functional, structural, semantic, and factual. Some systems provide information retrieval of case factors in order to support case based reasoning ([32], [33], and [34]). However, they do not report how text mining is used to identify the factors, nor whether the approach could support queries beyond those specifically designed for the particular case based reasoner. [35] extracts cases in an appellate chain, which are those cases that are relevant to the current case in terms of the comments on the quality of the case, e.g., whether it has been appealed, affirmed, overturned, overruled, explained, or distinguished. [35] focus on automated extraction of citation relations, not on argumentation or case factors. [36] parses individual sentences from legal texts. However, the results do not bear on parsing arguments nor on information extraction. [37] and [38] develop text mining approaches to identify the *rhetorical* structures in free texts; where the argument indicators are explicit, the rates of identification are reasonably high, but fall where the indicators must be inferred. See [39] for a survey of other systems and approaches to information retrieval from legal texts as of the 1990s.

2.3 Summary

We have briefly surveyed a number of different approaches to argumentation and information extraction from case bases. In our approach we have two angles of attack on

these issues. First, we are concerned with information extraction of argument structures from the case base, allowing us to identify decisions and their justifications. Second, we want to extract not only key profile elements (along the lines of the functional elements of [31]), but also the linguistically represented case factors that can be used in case based reasoning (along the lines of [32] and [34]).

3 Argument Corpus

In carrying out text mining, the first task is to form a corpus from which the information will be extracted. In this section, we discuss several previous attempts to form argument corpora, citing the data sources and purposes of the corpora. We also point out several additional potentially useful sources of arguments.

3.1 Araucaria

[40] outlines the creation of an argument database *AraucariaDB*. In the current version, *AraucariaDB* contains approximately 700 arguments. The arguments are drawn from a range of international sources. There are several problems concerning the sampling. We are not given a criteria which was used to guide the selection of the arguments. No consideration is given to the impact of the different source contexts: the arguments found in a newspaper may be different from those found in judicial summaries; in addition, arguments in Japan may be different from those in the United States. By the same token, there is no information concerning how arguments were identified from the source material, thus there could be biases of arguments of a certain sort or on a particular topic, namely whatever was perceived by the selector to have been an argument. While subjective criteria need not be problematic, the absence of overt critieria hinders evaluation of the resulting selection. Indeed, it is not clear what supports the claim that the contents of the database are arguments. In sum, given the small sample, lack of context information, and lack of criteria, it is uncertain what we can infer about the patterns which may appear.

3.2 Mochales and Moens

The corpus in [41] contained 30 relevant documents. The documents were divided into development and test documents. The 10 development documents were used to construct a grammar and to establish a *gold standard*. They were legal decisions drawn from an online database of cases from the European Court of Human Rights (ECHR), which has a common law legal system. All the texts were independently marked by three parties, who came to agreement about the mark up; however, there are interesting observations concerning differences among the markers concerning the identification of implicit premises or the extend of portions of arguments. It is reasonable to assume that the cases contain legal arguments, unlike documents sourced elsewhere. The documents contain a variety of formal sections such as statements of facts, complaints, and the reasons for and against the decision. As [41] note, not all these sections are expected to contain arguments; for example, a statements of the facts, precedents, or procedural

moves might not be taken as samples of argument. The 20 test documents were used to test the adequacy of the argument grammar.

3.3 Wyner and Milward

[42] provide two corpora of 50 and 90 legal cases selected arbitrarily from a search of the British and Irish Legal Information Institute's (BAILII) online database of legal decisions in the United Kingdom and Ireland. Within the BAILII database, a keyword search was made for cases pertaining to medical malpractice so as to have a coherent set of cases. As with the ECHR cases, it is reasonable to assume that the cases contain legal arguments, and broadly contain reports of facts, complaints, and the reasons for or against the decision, as well as applicable law. One objective of the study was to develop text mining tools to automatically search for elements that are found in commercial case law search engines, such as indices for citation index, judges, jurisdiction, and so on. Another objective was to develop searches for features of the case beyond those found in such search engines, such as case features or the identification of violation of some norm. Together these aspects of cases are crucial for argumentative case based reasoning [23].

3.4 Others

In addition to the sources of argument indicated above, there are a range of other available resources which ought to be considered in future research. For example, *Debateopedia* is an online encyclopedia of debates, which are arguments pro and con a range of particular issues. It is compiled under the auspicies of the *International Debate Education Association*, so has a claim to be representing arguments are they are expressed.[5] The structure of the arguments, providing arguments pro and con an issue, is closer to what is understood to be an argument than inference patterns. The database contains several thousand debates from which a selection can be made concerning a particular topic of interest. For the legal domain, the *World Legal Information Institute* is a searchable index of databases of case law including links to UK and US case law databases.[6] Finally, the US *National Institute of Standards and Technology* has been sponsoring a task for *recognising textual entailment*, in which natural language processing techniques are applied to a database of entailment patterns.[7] Discussion on the selection and analysis of the database is discussed in [28].

4 Analysis

In analysing a case base, a variety of approaches have been applied. In 4.1, the arguments identified in ArcauriaDB were manually marked up in an XML which indicates the structure of the argument. In 4.2, a context free grammar for arguments is proposed

[5] http://wiki.idebate.org/index.php/Welcome_to_Debatepedia!
[6] http://www.worldlii.org/
[7] http://www.nist.gov/tac/2009/RTE/index.html

then used to identify and parse an argument from a case. Such an approach promises to allow the extraction of arguments of a specific structure from the case base. In 4.3, cases in a case base are profiled with respect to a range of functional features with text mining tools; in addition, relational factors such as failure to fulfil and obligation are identified.

4.1 Araucaria

In [40], the arguments in the corpus were analysed and represented in an XML-based format, the *Argument Markup Language* (AML).[8] AML represents arguments in terms of XML markups that indicate a range of properties and relationships among the propositions that constitute the argument. Argument schemes, which are stereotypical patterns of reasoning, are used to guide and catalog the arguments [5]. Suppose an argument from the argument from sign scheme *A bear passed this way. Here are some bear tracks in the snow.* We infer from the evidence of the bear tracks that a bear presumably passed by. AML indicates the scheme, the distinct propositions, the relation between them (one as the premise and the other at conclusion) as well as a range of auxiliary information such as date of analysis and author of analysis. More complex arguments can be represented in AML. The XML format is machine-readable, so can be searched for argument components; in addition, it supports translation into a graphical representation, which can be easily understood.

For the analysis of the AraucariaDB, two analysts applied AML using argument schemes. No systematic methodology is outlined, and there were no controls for intercoder reliability. As with the development of the corpus, the methodology of analysis makes it difficult to draw reliable conclusions about the data.

Another approach to the analysis of the AraucariaDB is found in [37]. The objective of the study is to be able to automatically identify the argument from non-argument sentences drawn from the corpus. The AraucariaDB is augmented with a similar number of sentences which are claimed to be non-argumentative. The features used to distinguish examples are: sequences of words in a sentence (from one (unigrams) to three (trigrams)), words identified by part-of-speech (adverbs, verbs, and modal auxiliaries), collocations (high word pair co-occurrences), text statistics (sentence length, word length, punctuation marks), punctuation, parse features (given a parse of the sentence into phrases, the depth of the tree and number of subclauses), and key words that have been identified as signalling argument (e.g. *but*, *consequently*, *because of*, and others).

The best results were obtained with a combination of collocations, verbs, sentence length, word length, and number of punctuation marks. The latter three broadly can be taken to indicate that sentence complexity (length of sentence and words along with complex structure signaled by punctuation) signal argument.

The results are limited in several respects. First, no criteria is given for the selection of the non-argument sample sentences; it is not clear to what extent sentence complexity itself signals argument or whether this is an artifact of how argument and non-argument sentences were selected for the corpus. Second, the role of key words and collocations is

[8] This seems largely superseded by the *Argument Interchange Format* in [43].

unclear; just what key words or collocations signal argument is not specified. Moreover, as [37] note, some features are ambiguous; the modal auxiliaries in particular have a so-called *root* and *epistemic* interpretation, where only the former might be used to infer an argument [44]. Finally, again as [37] note, arguments contain *enthymemes*, which are implicit reasoning steps [45]; thus, if an argument must be inferred from information not found in the text, it cannot be identified by text mining techniques.

4.2 Mochales and Moens

[41] focus on argument identification within a limited corpus and relative to a gold standard as well as the formulation of a context-free argument grammar. In this section, the grammar is proposed, applied, and evaluated.

Grammars Formalisation of natural languages using grammars has been a topic of interest among linguistic researches for years ([46], [47], and [48]). Modern computational notations given by grammars can be used to develop parsing applications. Context Free Grammars (CFG) have been used extensively for defining the syntax of a variety of natural and artificial (e.g. programming) languages [49]. In this section, we formally present a CFG for legal arguments, exploiting the inherently structured nature of argument in case law documents, specifically in the judgements and decisions of the European Court of Human Rights (ECHR). We develop a parsing application to test the grammar accuracy.

CFGs are a particular class of grammar, where grammars are composed of a finite set of terminal and non-terminal symbols, a special start symbol, and a finite set of productions, which are rules of substitution whereby the left-hand symbols are substituted for by the right-hand symbols. CFGs allow substitutions independent of the rest of the structure; the left-hand side of rule can only consist of a single symbol and the right-hand side is a non-empty string over the total vocabulary, i.e. terminal and non-terminal symbols. The following example of a CFG expresses that *NP* (or noun phrase) can be composed of either a *ProperNoun* or a determiner (*Det*) followed by a *Nominal*; which can be one or more *Nouns*.

$$NP \rightarrow Det\ Nominal$$
$$NP \rightarrow ProperNoun$$
$$Nominal \rightarrow Noun \mid Nominal\ Noun$$

Legal Argument Constructs The development of any grammar is factored by an initial linguistic analysis which determines the grammar symbols and the rules that allow to move from one symbol to another. The development of grammars of English focused on syntactic studies of individual sentences, while for legal arguments we must examine sentences in relationships.

In [41], ten documents from the ECHR collection were analysed by legal experts. The analysis broadly covered the section structure, argumentative structure, and linguistic characteristics found in the documents. Detailed analysis was done on those sections where the legal arguments were specifically presented – *The Law* and *Dissenting Opinion*. A clear distinction was observed between argumentative and non-argumentative

information. The linguistic analysis identified some patterns only used on the argumentative information. For example, it clearly identified the conclusion of arguments, *"For these reasons the Court|Commission"*, which were supported by premises, *"There is a violation of Article"*. A premise of one argument can also serve as the conclusion of another argument, making chains fo arguments. The study also identified many rhetorical markers that help to detect the discursive progress of the argument structure, e.g. *however*, *therefore*, *although* or *in particular*. Furthermore, many premises and conclusions were found to be marked by linguistic expressions that clearly identified them as being part of the argumentative process (Table 1[9]). Some expressions may be common to all kind of argumentative texts, e.g. *"in the light of"*, while others may be more restricted to the legal field, e.g. *"under the terms of article"*.

Table 1. Typical expressions in the ECHR documents

Conclusions	The *factfinder* [A|B] <Verb-Conclusion> [C] The *factfinder* <Verb-Aux> [NOT] [A|B] <Verb-Conclusion> The *factfinder* [A|B] <Verb-Premise> There has been a violation of It [A] follows that There has [A] been a breach of Having reached this conclusion [C] In conclusion,
Premises	The *factfinder* [A|B] <Verb-Premise> [C] The *factfinder* <Verb-Aux> [NOT] [A|B] <Verb-Premise> [C] The *factfinder* [B] has|is [A] <Verb-Premise> [C] In the *factfinder* 's view In the view of the *factfinder* See, mutatis mutandis
[A]	therefore ǀ firstly ǀ accordingly ǀ clearly ǀ also ǀ further ǀ thus
[B]	like the xxxx and the xxxx, ǀ , like the xxxx,
[C]	in the light of the partie's submissions ǀ , in the light of all the material before it
<Verb-Conclusion>	accepts ǀ concludes ǀ holds ǀ decides ǀ rejects ǀ declares ǀ dismisses ǀ sees no reason ǀ examines ǀ strikes
<Verb-Premise>	considers ǀ notes ǀ recalls ǀ agrees ǀ disagrees ǀ reiterates ǀ acknowledges ǀ is of the opinion ǀ points out ǀ emphasises ǀ stresses ǀ is of the view ǀ is satisfied ǀ endorses ǀ observes ǀ takes into acount ǀ convinces
<Verb-Aux>	must ǀ can ǀ does

[9] A *factfinder* is the person or persons in a particular trial or proceeding with the responsibility of determining the facts. For example, in the ECHR decisions the *factfinder* it is a *Commission*, in the ECHR judgments it is a *Court* and in other documents it is a *Judicial Responsible*, a *Committee* or a *jury*.

A Context Free Grammar for Legal Argument Our grammar represents linguistic characteristics of legal argument with rules such as:

$$\forall_x[\textit{isPremise}(x_i) \wedge \textit{startsHowever}(x_{i+1}) \rightarrow \textit{isPremise}(x_{i+1})]$$

To handle the linguistic variety required for real-world text input we compress families of related productions by making classes of rhetorical markers and verbs, e.g. support markers or conclusive verbs. We can generalise the previous rule as:

$$\forall_x[\textit{isPremise}(x_i) \wedge \textit{startsContrast}(x_{i+1}) \rightarrow \textit{isPremise}(x_{i+1})]$$

where *startsContrast*(x_{i+1}) is a class that contains *startsHowever*(x_{i+1}) among others. Thus, lexical and phrasal variation can be associated with similar semantics, allowing us to capture variations in writing style. The grammar does not accommodate ill-formed arguments as it would degrade the ability of the grammar to parse good arguments. The complete CFG can be found in Figure 1 and the meaning of the symbols in Table 2.

$$T \Rightarrow A^+D \tag{1}$$
$$A \Rightarrow \{A^+C|A^*CnP^+|Cns|A^*sr_cC|P^+\} \tag{2}$$
$$D \Rightarrow r_cf\{v_cs|.\}^+ \tag{3}$$
$$P \Rightarrow \{P_{verbP}|P_{art}|PP_{sup}|PP_{ag}|sP_{sup}|sP_{ag}\} \tag{4}$$
$$P_{verbP} = sv_ps \tag{5}$$
$$P_{art} = sr_{art}s \tag{6}$$
$$P_{sup} = \{r_s\}\{s|P_{verbP}|P_{art}|P_{sup}|P_{ag}\} \tag{7}$$
$$P_{ag} = \{r_a\}\{s|P_{verbP}|P_{art}|P_{sup}|P_{ag}\} \tag{8}$$
$$C = \{r_c|r_s\}\{s|C|P_{verbP}\} \tag{9}$$
$$C = s^*v_cs \tag{10}$$

Fig. 1. Context Free Grammar for recognising legal argument

An example of the grammar's application The following example of an argument is taken from an ECHR judgment. We apply our CFG to it to generate a parse tree as in Figure 4.2.

> 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. Like the delegate of the Commission, the Court considers that the applicant's heirs also have a definite pecuniary interest under article of the convention art. x. Furthermore, it notes that the applicant was sentenced to fifteen months' imprisonment, commutable to a fine of x GRD per day of

Table 2. Terminal and non-terminal symbols from the Context Free Grammar

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
C	Sentence with a conclusive meaning
n	Sentence, clause or word that indicates one or more premises will follow
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. x para. x, ...)
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 argument (e.g. court, jury, commission, ...)
s	Sentence, clause or word different from the above symbols

> detention, which sum he paid. 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. 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.

The CFG identifies this as an argument with a conclusion, i.e. $A \Rightarrow A^{+}C$. The conclusion, *"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."*, is identified by:

- r_c: *accordingly*
- v_c: *finds*
- s: *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*

A^{+} can be expanded as three separate premises using $A \Rightarrow P^{+}$. Each set of premises applies respectively:

- $P \Rightarrow P_{verbP}$ identifying v_p: *notes*
- $P \Rightarrow PP_{sup}$ and $P \Rightarrow P_{verbP}$ identifying r_s: *furthermore* and v_p: *considers*
- $P \Rightarrow P_{art}$ identifying r_art: *without prejudice to its decision on*

Thus, the final argument structure is the one found in Figure 4.2.

There are several interesting research issues concerning this grammar. First, it allows for some ambiguity. For example, the argument A, which has been treated with $A \Rightarrow P^{+}$ to obtain three P, being P_{verbP}, $PP_{sub} and P_{art}$ respectively, could have been divided in four P instead. Furthermore, the grammar relates a premise starting with a r_s or r_a to the closest previous statement, $P \Rightarrow \{P\ P_{sup} | P\ P_{ag}\}$, while that is not always the correct assumption. Sentences may not overtly express how they function. For example, the sentence *"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."* seems to be a conclusion, but must be justified. If we cannot find support for such a justification, our grammar does not classify the statement as a conclusion. Legal arguments need justified conclusions for otherwise we cannot identify the argument. Finally, we can compare arguments allowing us to identify similar arguments written in different ways. For example, we can equate *"notes"* with *"acknowledges"* or *"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"* with *"that Mr. Smith is guilty"*. The grammar can be adapted to reflect the practices of the ECHR in this way. The comparison of arguments is crucial for applications of case based reasoning.

```
A
|---c: 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.
    |---A
        |---P
        |   |---p: 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.
        |---P
        |   |---p: Like the delegate of the Commission, the Court
        |   |      considers that the applicant's heirs also have a
        |   |      definite pecuniary interest under article of the
        |   |      convention art. x.
        |   |---P
        |       |---p: Furthermore, it notes that the applicant was
        |              sentenced to fifteen months' imprisonment,
        |              commutable to a fine of x GRD per day of
        |              detention, which sum he paid.
        |---P
            |---p: 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.
```

Fig. 2. Tree Structure of an argument

Grammar Evaluation The grammar was tested to identify the argumentative structure of twenty new ECHR documents. It was implemented using Java and JSCC[10]. The main results can be seen in Table 3. It is important to note that all final decisions (D) were correctly identified and the average compression range of the given text was 65%. The main limitations of the grammar are due to the structure of A, i.e. the justification given by the *factfinder*. In this aspect, there are two main problems: (a) the detection of intermediate conclusions, specially the ones without rhetorical markers, as more than 20% of the conclusions are classified as premises of a higher layer conclusion; (b) the ambiguity between argument structures. However, this is also one of the main causes of human disagreement.

Table 3. Results over 20 documents from the ECHR

	Size (# of sentences)	**Precision**	**Recall**
Premises	430	59%	70%
Conclusions	156	61%	75%
Non-argumentative information	1087	89%	80%
Final decision	63	100%	100%

4.3 Wyner and Milward

Previous work such as [37] focuses on identifying argument structure from text such as indicated by keywords *supposing* or *therefore*, which are marks of argument in general and not clearly particular to legal argument. In legal argument in common law settings, there is a meta-level of argument concerning the cases themselves – case based reasoning (CBR) ([50], [51], [52], and [53]). CBR has four stages: the lawyer submits a problem case and retrieves precedent cases from a case base; the solutions from precedent cases are reused; the solution is confirmed (or disconfirmed); finally, once solved, the problem case is retained in the case base. From among these stages, a key task for a lawyer is to identify on-point-cases from a case base; these are cases which were decided in favour of the lawyer's side and share the most number of highly valued "factors". Factors are textually expressed typical fact patterns in a case which bias the decision for or against a side in the case. Importantly, factors are not themselves signalled by argumentative indicators. Yet, identifying the cases which contain the factors is crucial to case based reasoning. One may say that the second and third stages of CBR constitute the "argument" phase, where the argument is on analogy: given case A and case B, which are analogous to case C and which were decided for the plaintiff, therefore decide case C for plaintiff. [23] provide argument schemes for legal case based reasoning which detail the ways the factors are used to reason to a decision in a variety of instances.

[10] http://jscc.jmksf.com|

For instance, suppose we are considering a case of reckless driving, where someone has died. Determining whether the driver is guilty of murder or manslaughter is crucial in the determination of the sentence. To make the distinction, it must be determined whether the driver had culpable intent. In turn, this is determined with respect to a variety of particular factors that can be concretely be identified [54]:

- Obligation to aid the victim.
- Failure to heed traffic signs.
- Failure to heed warnings about reckless driving.

This list of factors which are used to determine culpable intent provide a textual "frame" for each factor; however, the form of the factors as they appear in a case base may vary. One of the important tasks of information extraction is to reliably identify the "same" factor in semantic terms while varying the form. This is a general linguistic issue (e.g. passives and actives mean the same thing, but have different forms).

We can also consider identifying factors across domains. For example, the issue relating to murder or manslaughter arises in the medical domain as well, and so culpable intent is also relevant. To determine the seriousness of medical negligence, one might consider:

- Obligation to have taken a second opinion.
- Failure to take a proper history.
- Failure to take into account apparent symptoms.

Thus, with one suitably general frame, one can identify cases addressing culpable intent across domains.

In searching the case base, one identifies the factors which contribute to the current undecided case, then wants to search the case base for on-point precedents. By the same token, one might want to identify cases which are *different* from the current case in order to compare the results in those cases to the current case.

In searching the case base, we applied a the commercial I2E text mining package by *Linguamatics*. The objective was to identify the factors as well as more general features of each case such as the citation, presiding judges, solicitors, whether the case is on appeal, and other parameters; these features are similar to the index parameters found in *Lexis-Nexis* and *Westlaw*, however, we identify these automatically rather than by preindexing the cases via manual annotation. I2E is an interactive, flexible, and articulated search tool; one specifies a search, views the results, then can refine or alter the search. It has a graphical user interface, which makes the software accessible to a broad range of end-users. The sentences in the case base are parsed, so searches can be done relative to syntactic structure. Additional search capabilities are: regular expression searches, list of alternative words, searches within syntactic frames such as sentences or paragraphs. A key feature is integration of searches relative to an ontology. For example, suppose one has a database of cases concerning medical malpractice and cancer. We may wish to relate doctors to the cancers they have treated. However, as there are a variety of cancers, using a string search alone, one would have to search relative to each sort of cancer expressed as a string. Given a suitably rich ontology, we can search just for the class "cancer", then retrieve all the different ways of referring

to cancer (such as carcinoma or neoplasm), the different types of cancer, and the ways the types are referred to. In this way, we can relate parts of the text which are otherwise hard to relate with simple string searches.

In the following, we present several results.[11] In Table 4, we profile some cases from the case base: the document index (e.g. [2008]EWCACiv10.txt) shows a variety of features: the case number, the court in which the case is heard, who the solicitors were instructed by, the judge who hears the case, and the court from which the case is appealed.

Table 4. Case Profile

Doc	**Index**	**Entity**
[2008] EWCACiv10.txt	Case No.	A3/2007/1677
	Court	The Supreme Court of Judicature Court of Appeal
	Instructed by	Skadden
		Steptoe & Johnson
	On appeal	High Court of Justice Queen's Bench Division Commerical Court
[2008] EWCACiv1022.txt	Case No.	B2/2007/2303
	Court	The Supreme Court of Judicature Court of Appeal
	Instructed by	Messrs Buller Jeffries
		Messrs John A Neil Solicitors
	Judge	District Judge Temple
	On appeal	Cambridge County Court

In Table 5, we identify the concept of *Failure/Obligations* from a case in the case base; the concept appears in a variety of alternative phrases: *ought not to have...*, *had failed to observe...*, *owed a duty of care...*. This highlights phrases which can be used to further determine whether the case concerns medical malpractice.

For tasks where the same kind of information needs to be automatically extracted, the next step is to develop a corpus of cases as a gold standard against which to measure information retrieval with respect to precision and recall. However, there are also specific tasks where it is useful to apply the graphical querying capabilities of I2E which are similar to ad hoc keyword search, and iteratively refine queries to get the cases of interest.

5 Discussion

We have set our work in the context of argumentation theory and text mining in the case base. We have discussed different approaches to the formation of a legal case base.

[11] These are simplified results produced by I2E.

Table 5. Case Profile with Violation Factor

Doc	Index	Entity	Context
30.htm	Case No.	200302858 B1	
	Court	The Supreme Court of Judicature	
	Failure or Obligations	had failed to observe appropriate professional standards	in the sense that he had failed to observe apprporiate professional standards to a patient to whom...
		owed a duty of care	as a doctor he owed a duty of care to Sean Philips as his...
		owes a duty of care	individual to whom the defendant owes a duty of care
	Judge	Lord Justice Judge	
	On appeal	Winchester Crown Court	

Finally, we turned to several ways elements found in a case base can be automatically analysed with text mining tools to support legal argument. In this section, we consider issues with our results as well as a range of topics that warrant future research.

With regard to the development of a legal argument corpora, there are new opportunties to use available online corpora for text analysis as noted in section 3.4. However, some consideration should be given to the signficance of using different corpora given the great variety of court systems. For example, [41] use decisions from the European Court of Human Rights. In related work on diagramming legal decisions, [55] use US Supreme Court oral argument transcripts. In both instances, consideration should be given to the level of court and complexity of issues since both court levels address areas of the law that are the least settled and most complex. Examining US Supreme Court transcipts, the argument patterns are very complex and often hard to follow, even by the Justices and legal representative themselves as is reported in court transcripts. It may be preferable to make use of court levels and decisions that are signicantly more prosaic such as decisions of courts in of the first instance (where the case is first introduced) or courts of appeals. Not only is the law more settled and less complex, but it is more useful to legal researchers to know how the law commonly functions before we examine how it functions at the highest level (on this point, also see [56]).

In section 4.2, a grammar of argument is provided and then used to extract arguments from cases. Such an approach may work well for well-edited court decisions, but it is unlikely to apply to oral arguments made in court by lawyers, which appear less structured[12] It is, then, important to proscribe just what is intended to be meant by a grammar of *legal* arguments. Moreover, even with well-edited court decisions, it is as yet unclear the extent to which *discontinuous constituents* (see [57]) or discourse phenomena (see [58]) play a role. The proposal in section 4.2 relies on continuous structures such as for phrase-structure grammars for sentences. In such an approach, phrases must appear in determined orders, must be complete, cannot allow components of the

[12] See US Supreme Court transcripts:
http://www.supremecourtus.gov/oral_arguments/argument_transcripts.html

argument appearing outside the given structure, and do not allow interjections. Yet, it is possible that the premise of an argument appears somewhere later in the text, which a CFG alone could not account for. However, in the corpus studied, it appears that the percentage of discontinuous constituents is normally low. Nonetheless, some consideration ought to go into how to accommodate discontinuous constitutents where they occur.

Another aspect of the grammar of argument is that only *defeasible* arguments are accounted for, not the variety of argument types outlined in [5]. To identify subsorts of argument (e.g. Expert Testimony), lexical semantic and ontological information may be required.

Other issues of interest relate to the identification of enthymemes, which are missing and inferred premises, as well as argument coherence. For information extraction, enthymemes present a very significant issue since they must be inferred semantically and are not in evidence in the text. It is unclear how text mining can yet address this issue. In section 4.2, the pattern of an argument is identified, but within the components of the argument (i.e. premises, rule, conclusion) there appear to be no well-formedness constraints. Yet, clearly, there are well-formed arguments in terms of the grammar, but which are semantically incoherent: if Tweety is a bird and iron is a mineral, therefore the stock market will rise. This relates to similar syntactic issues bearing on sentences such as the famous *Colorless green ideas sleep furiously*. Of course, the research reported in section 4.2 is a step towards identifying, clarifying, and addressing such issues.

In section 4.3, initial results are provided for automatically profiling a case and identifying key features that are useful in making legal arguments. Further research will focus on identifying legally relevant case factors such as are actually used by lawyers in case based reasoning. Here too, lexical semantic and ontological information may prove to be useful.

A general issue that must be addressed is the bottleneck of ontological approaches, which still require large amounts of knowledge to be manually drafted. Such a task is not realistic for general text analysis. One way to address the issue is to apply advanced machine-learning techniques. Another is to provide a structure that supports systematic, incremental, modular knowledge development such as is done with Semantic Web OWL modules. Yet another approach is to leverage the internet to distribute the task of manually drafting knowledge by, for example, using techniques such as found for online psycholinguistic experiments.[13]

The results described in this paper bring us closer to automatic identification of legal arguments. We believe that the use of ontologies, which were used in profiling, will also be useful for identifying arguments. In future work we will look at combining the two approaches examined here to provide a single approach for comprehensive analysis of legal texts.

[13] See: http://www.surf.to/experiments

Acknowledgments

[41] was presented at JURIX 2008. [42] was presented at the Workshop on Natural Language Engineering of Legal Argumentation held in conjunction with JURIX 2008. The authors thank the reviewers and audiences for their comments.

References

1. Wyner, A., Bench-Capon, T., Atkinson, K.: Three senses of "argument". In Sartor, G., Casanovas, P., Rubino, R., Casellas, N., eds.: Computable Models of the Law: Languages, Dialogues, Games, Ontologies. LNAI 4884. Springer (2008) 146–161
2. Eemeren, F.V., Grootendorst, R.: A Systematic Theory of Argumentation. Cambridge University Press (2003)
3. Mann, W.C., Thompson, S.A.: Rhetorical structure theory: A theory of text organization. Technical Report ISI/RS-87-190, University of Southern California, Information Sciences Institute (ISI) (1987)
4. Hitchcock, D.: Informal logic and the concept of argument. In Gabbay, D., Thagard, P., Woods, J., eds.: Philosophy of Logic (Handbook of the Philosophy of Science). Elsevier (2007) 101–130
5. Walton, D.: Argumentation Schemes for Presumptive Reasoning. Erlbaum, Mahwah, N.J. (1996)
6. Atkinson, K.: What Should We Do?: Computational Representation of Persuasive Argument in Practical Reasoning. PhD thesis, Department of Computer Science, University of Liverpool, Liverpool, United Kingdom (2005)
7. Gordon, T., Prakken, H., Walton, D.: The carneades model of argument and burden of proof. Artificial Intelligence **171** (2007) 875–896
8. Reed, C., Rowe, G.: Araucaria: Software for argument analysis, diagramming and representation. International Journal on Artificial Intelligence Tools **13**(4) (2004) 961–980
9. Torroni, P., Gavanelli, M., Chesani, F.: Argumentation in the semantic web. Intelligent Systems **22**(6) (2007) 66–74 Iyad Rahwan and Peter McBurney (eds.).
10. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence **77**(2) (1995) 321–358
11. Amgoud, L., Cayrol, C.: On the acceptability of arguments in preference-based argumentation. In: Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence (UAI-98), San Francisco, CA, Morgan Kaufmann (1998) 1–7
12. Bench-Capon, T.J.M.: Persuasion in practical argument using value-based argumentation frameworks. Journal of Logic and Computation **13**(3) (2003) 429–448
13. Besnard, P., Hunter, A.: Elements of Argumentation. MIT Press (2008)
14. Wyner, A., Bench-Capon, T.: Towards an extensible argumentation system. In Mellouli, K., ed.: ECSQARU 2007, LNAI 4724. (2007) 283–294
15. Bondarenko, A., Dung, P.M., Kowalski, R.A., Toni, F.: An abstract, argumentation-theoretic approach to default reasoning. Artificial Intelligence **93** (1997) 63–101
16. Wyner, A., Bench-Capon, T.: Taking the a-chain: Strict and defeasible implication in argumentation frameworks. Technical report, University of Liverpool (2008)
17. Prakken, H., Sartor, G.: Argument-based extended logic programming with defeasible priorities. Journal of Applied Non-Classical Logics **7**(1) (1997)
18. García, A.J., Simari, G.R.: Defeasible logic programming: An argumentative approach. Theory and Practice of Logic Programming **4**(1) (2004) 95–137

19. Governatori, G., Maher, M.J., Antoniu, G., Billington, D.: Argumentation semantics for defeasible logic. Journal of Logic and Computation **14**(5) (2004) 675–702
20. Amgoud, L., Caminada, M., Cayrol, C., Lagasquie, M.C., Prakken, H.: Towards a consensual formal model: inference part. Technical report, ASPIC project (2004) Deliverable D2.2: Draft Formal Semantics for Inference and Decision-Making.
21. Pollock, J.: Cognitive Carpentry: A Blueprint for How to Build a Person. MIT Press (1995)
22. Caminada, M., Amgoud, L.: On the evaluation of argumentation formalisms. Artificial Intelligence **171**(5-6) (2007) 286–310
23. Wyner, A., Bench-Capon, T.: Argument schemes for legal case-based reasoning. In Lodder, A.R., Mommers, L., eds.: Legal Knowledge and Information Systems. JURIX 2007, Amsterdam, IOS Press (2007) 139–149
24. Bos, J.: Applying automated deduction to natural language understanding. Journal of Applied Logic **7**(1) (2008) 100–112
25. Curran, J.R., Clark, S., Bos, J.: Linguistically motivated large-scale NLP with C&C and Boxer. In: ACL, The Association for Computer Linguistics (2007)
26. MacCartney, B., Manning, C.D.: Natural logic for textual inference. In: Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing, Prague, Association for Computational Linguistics (2007) 193–200
27. de Paiva, V., Bobrow, D.G., Condoravdi, C., Crouch, D., King, T.H., Karttunen, L., Nairn, R., Zaenen, A.: Textual inference logic: Take two. In: C&O:RR. (2007)
28. Dagan, I., Glickman, O., Magnini, B.: The PASCAL Recognising Textual Entailment Challenge. In Nonero-Candela, J.Q., Dagan, I., Magnini, B., d'Alché Buc, F., eds.: Machine Learning Challenges. Lecture Notes in Computer Science. Springer (2006) 177–190
29. de Marneffe, M.C., Rafferty, A.N., Manning, C.D.: Finding contradictions in text. In: Proceedings of ACL-08: HLT, Columbus, Ohio, Association for Computational Linguistics (2008) 1039–1047
30. Voorhees, E.M.: Contradictions and justifications: Extensions to the textual entailment task. In: Proceedings of ACL-08: HLT, Columbus, Ohio, Association for Computational Linguistics (2008) 63–71
31. Hafner, C.D.: Representation of knowledge in a legal information retrieval system. In: SIGIR '80: Proceedings of the 3rd annual ACM conference on Research and development in information retrieval, Kent, UK, UK, Butterworth & Co. (1981) 139–153
32. Daniels, J.J., Rissland, E.L.: Finding legally relevant passages in case opinions. In: ICAIL '97: Proceedings of the 6th International Conference on Artificial intelligence and Law, New York, NY, USA, ACM (1997) 39–46
33. Brüninghaus, S., Ashley, K.D.: Finding factors: learning to classify case opinions under abstract fact categories. In: ICAIL '97: Proceedings of the 6th International Conference on Artificial Intelligence and Law, New York, NY, USA, ACM (1997) 123–131
34. Brüninghaus, S., Ashley, K.D.: Generating legal arguments and predictions from case texts. In: ICAIL 2005, New York, NY, USA, ACM Press (2005) 65–74
35. Peter Jackson and, Al-kofahi, K., Tyrell, A., Vachher, A.: Information extraction from case law and retrieval of prior cases. Artificial Intelligence **150**(1-2) (2003) 239–290
36. McCarty, L.T.: Deep semantic interpretations of legal texts. In: ICAIL '07: Proceedings of the 11th International Conference on Artificial Intelligence and Law, New York, NY, USA, ACM Press (2007) 217–224
37. Moens, M.F., Boiy, E., Mochales-Palau, R., Reed, C.: Automatic detection of arguments in legal texts. In: ICAIL '07: Proceedings of the 11th International Conference on Artificial Intelligence and Law, New York, NY, USA, ACM Press (2007) 225–230
38. Sporleder, C., Lascarides, A.: Using automatically labelled examples to classify rhetorical relations: An assessment. Natural Language Engineering **14**(3) (2006) 369–416

39. Schweighofer, E.: The revolution in legal information retrieval or: The empire strikes bac. The Journal of Information, Law and Technology **1** (1999) online
40. Reed, C., Palau, R.M.P., Rowe, G., Moens, M.F.: Language resources for studying argument. In: Proceedings of the 6th conference on language resources and evaluation - LREC 2008, ELRA (2008) 91–100
41. Mochales, R., Moens, M.F.: Study on the structure of argumentation in case law. In Francesconi, E., Sartor, G., Tiscornia, D., eds.: Legal Knowledge and Information Systems - JURIX 2008: The Twenty-First Annual Conference. Volume 189 of Frontiers in Artificial Intelligence and Applications., IOS Press (2008) 11–20
42. Wyner, A., Milward, D.: Legal text-mining using linguamatics' I2E. Presentation at Workshop on Natural Language Engineering of Legal Argumentation, Florence, Italy as part of JURIX 2008. (2008)
43. Chesnevar, C., McGinnis, J., Modgil, S., Rahwan, I., Reed, C., Simari, G., South, M., Vreeswijk, G., , Willmott, S.: Towards an argument interchange format. The Knowledge Engineering Review **21**(4) (2006) 293–316
44. Jackendoff, R.: Semantic Interpretation in Generative Grammar. MIT Press (1972)
45. Walton, D.: The three bases for the enthymeme: A dialogical theory. Journal of Applied Logic **6**(3) (2008) 361–379
46. Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., eds.: The Longman Grammar of Spoken and Written English. Longmans, London (1999)
47. Feldmann, H.: An acceptive grammar for the natural language english. SIGPLAN Not. **19**(2) (1984) 58–67
48. Scott, W.: Development and application of a context-free grammar for requirements. In: International Conferences and Events (ICE), Australia (2004)
49. Charniak, E.: Statistical parsing with a context-free grammar and word statistics, AAAI Press/MIT Press (1997) 598–603
50. Ashley, K.: Modelling Legal Argument: Reasoning with Cases and Hypotheticals. Bradford Books/MIT Press, Cambridge, MA (1990)
51. Aleven, V., Ashley, K.D.: Doing things with factors. In: ICAIL '95: Proceedings of the 5th International Conference on Artificial Intelligence and Law, New York, NY, USA, ACM (1995) 31–41
52. Weber, R.O., Ashley, K.D., Brüninghaus, S.: Textual case-based reasoning. Knowledge Engineering Review **20**(3) (2005) 255–260
53. Brüninghaus, S., Ashley, K.: Reasoning with textual cases. In Munzo-Avila, H., Ricci, F., eds.: Proceedings of the International Conference on Case-based Reasoning 2005. Number 3620 in LNAI, Springer-Verlag (2005) 137–151
54. Luria, D.: Death on the highway: Reckless driving as murder. Oregon Law Review **799** (1988) 821–22
55. Ashley, K.D., Pinkwart, N., Lynch, C., Aleven, V.: Learning by diagramming supreme court oral arguments. In: ICAIL, ACM (2007) 271–275
56. Prakken, H.: Formalising ordinary legal disputes: a case study. Artificial Intelligence and Law **16**(4) (2008) 333–359
57. Ojeda, A.: Discontinuous Constituents. In: Encyclopedia of Language and Linguistics. Volume 3. Elsevier (2005) 624–630
58. Asher, N.: Reference to Abstract Objects in Discourse. Kluwer Academic Publishers (1993)