Detecting Arguments in CJEU Decisions on Fiscal State Aid

Giulia Grundler[*1] and **Piera Santin**[*2] and **Andrea Galassi**[1] and **Federico Galli**[2]
Francesco Godano[2] and **Francesca Lagioia**[2] and **Elena Palmieri**[1]
Federico Ruggeri[1] and **Giovanni Sartor**[2] and **Paolo Torroni**[1]
[1]DISI, University of Bologna, Bologna, Italy
[2]CIRSFID-Alma AI, University of Bologna, Bologna, Italy
a.galassi@unibo.it
francesca.lagioia@unibo.it

Abstract

The successful application of argument mining in the legal domain can dramatically impact many disciplines related to law. For this purpose, we present *Demosthenes*, a novel corpus for argument mining in legal documents, composed of 40 decisions of the Court of Justice of the European Union on matters of fiscal state aid. The annotation specifies three hierarchical levels of information: the argumentative elements, their types, and their argument schemes. In our experimental evaluation, we address 4 different classification tasks, combining advanced language models and traditional classifiers.

1 Introduction

The study of argumentation in legal contexts is one of the most lively research areas at the intersection of Artificial Intelligence and Law (Bench-Capon et al., 2004, 2009). It has its roots in logic, philosophy, and linguistics, as it studies how different claims and opinions are proposed, debated, and evaluated, considering their relations and inter-dependencies. The legal domain offers a natural scenario for the application of different argument models as well as novel machine learning and natural language processing techniques in order to perform legal reasoning (Prakken and Sartor, 1996a; Atkinson and Bench-Capon, 2019, 2021), build specific ontologies (Hoekstra et al., 2009), or support the teaching of law (Ashley et al., 2002; Carr, 2003). Argumentation is relevant to legal logic (Prakken and Sartor, 1996b, 2002), case-based reasoning (Aleven, 2003; Ashley et al., 2002), the interpretation of judicial opinions and statutory laws (McCarty, 2007; Savelka and Ashley, 2016; Palau and Ieven, 2009; Mochales Palau and Moens, 2011), the summarization of judicial opinions (Hachey and Grover, 2006).

Building tools capable of automatically detecting arguments in legal texts can produce a dramatic impact on many disciplines related to law, providing valuable instruments for the retrieval of legal arguments from large corpora, for the summarization and classification of legal texts, and for the development of AI systems supporting lawyers and judges, by suggesting relevant arguments and counterarguments. A crucial obstacle to providing effective automatic support to legal argumentation pertains to the knowledge bottleneck: legal arguments are only available in natural language texts, whose content has been so far only accessible with the help of domain experts. To overcome this limitation, recourse has been made to argument mining (AM), i.e., the automated extraction of arguments from documents.

AM frameworks can be described as multi-stage pipeline systems, aimed at extracting natural language arguments and their relations from textual documents (Lippi and Torroni, 2016; Cabrio and Villata, 2018). Each stage of the pipeline addresses a sub-task of the problem. A first stage usually consists of detecting which sentences in the input document(s) are argumentative, i.e., contain an argument or part thereof. Once argumentative sentences are singled out, it is possible to detect the boundaries of the various argument components and their characteristics (Mochales Palau and Moens, 2011; Niculae et al., 2017; Bar-Haim et al., 2017). Finally, a last stage in the pipeline considers these components in order to predict the relationship between them and/or between the arguments they are part of (Lippi and Torroni, 2016; Lawrence and Reed, 2019).

In this work, we contribute to this research domain by releasing *Demosthenes*, a novel corpus of legal documents annotated for AM. Specifically, we focus on the first two stages of the pipeline in order to: (i) identify premises and conclusions; (ii) distinguish between legal and factual premises; (iii)

[*] Equal contribution

Proceedings of the 9th Workshop on Argument Mining, co-located with the 29th International Conference on Computational Linguistics, pages 143–157
October 17, 2022.

identify argumentative schemes. Additionally, we perform an experimental evaluation on all the tasks using multiple representations and classifiers.

The paper is structured as follows. In Section 2, we provide an overview of related work. Section 3 describes the corpus we have created and the annotation procedure. Section 4 concerns the experimental setting, while the results are presented in Section 5. Section 6 concludes.

2 Related work

Despite in the last decade the field of AM has become a popular research area in Natural Language Processing (NLP), there are yet limited studies focusing on legal texts and, in particular, on judicial decisions (Zhang et al., 2022b). Among them, the targets of judicial AM vary widely (Zhang et al., 2022a). Some studies aim at extracting the arguments from generic unstructured documents (Levy et al., 2014); others start from a given set of arguments and focus on aspects such as the identification of attack/support relations between them (Chesnevar et al., 2006), or the classification of argument schemes (Feng and Hirst, 2011).

One of the main obstacles in providing effective automatic support to legal argumentation pertains to the knowledge bottleneck. Like most interdisciplinary studies, creating and constructing annotated corpora is labour–intensive, as it is a complex and time-consuming task, requiring the guidance of legal experts, i.e., lawyers, being also familiar with legal arguments and the specific legal domain. Indeed, a discrepancy exists between the way NLP researchers model and annotate arguments in court decisions and the way legal experts understand and analyze legal argumentation (Habernal et al., 2022). In fact, under computational approaches, arguments are often treated as mere structures of premises and claims (Stede and Schneider, 2018). In legal research, on the contrary, it it critical to also distinguish different kinds of arguments and classify them according to the rich typology that is rooted in the theory and practice of legal argumentation (Trachtman, 2013). Finally, legal arguments may present themselves in different ways within different kinds of legal texts, depending on the on the domain of the law being addressed, and on the institutional position and legal culture of the authority that is producing such texts.

Unfortunately, there are a limited number of annotated corpora that fit the requirements just mentioned, and they include a small amount of documents, withing specific areas of the law.

Thus the research community can highly benefit from the availability of new datasets, which as is the case of Demosthenes, cover a sizable amount of examples, and include an attempt at classifying the identified arguments according to a legal typology.

Moreover, to the best of our knowledge, few works have analysed how natural-language argumentation is used in real courts (Mochales and Moens, 2011; Habernal et al., 2022). This situation leads to three urgent needs in legal AM: (1) the creation of new annotated corpora, (2) possibly addressing different domains of the law; and (3) an analysis of how and to what extent models of arguments from legal theory can be reliably operationalized in terms of discourse annotations.

The approach by Poudyal et al. (2020) represents, to date, one of the few works whose goal was to implement a full-fledged argumentation mining system, specific to a single legal domain. Mochales Palau and Moens (2011) created a corpus of 47 cases (judgments and decisions) from the open-source database of the European Court of Human Rights (ECHR), in which they applied a sentence-level annotation scheme based on Walton's model (Walton et al., 2008) where each sentence was labeled as *premise*, *conclusion* or *non-argumentative*. More recently, Poudyal et al. (2020); Mochales and Moens (2011); Teruel et al. (2018) used the same guidelines to release a similar dataset of 42 documents. Walker et al. (2011) annotated judicial decisions selected from the U.S. Court of Federal Claims also identifying sentences' inferential roles and support levels by using logical connectives to represent argumentative relations between premises and conclusions. Walker et al. (2017) published a dataset of judicial decisions from the Board of Veterans Appeals. The decisions are annotated by legal experts with semantic information about arguments, including ten sentence roles and eight propositional connectives. The corpus initially contained 20 documents but was expanded subsequently (Walker et al., 2019, 2020).

In this work, we aim to partially fill the mentioned gaps by: (1) creating a new annotated legal corpus; (2) focusing on a domain that is still unexplored in the field of legal argumentation, i.e., fiscal state aid; and (3) investigating whether argumentation schemes defined in legal theory, in particular by Walton et al. (2008, 2021) can be easily adapted to the CJEU reasoning, as made explicit in the Court discourse. In particular, we focus on the detection of argumentative elements and their classification according to a hierarchical taxonomy of three layers, as detailed in the following.

For what concerns the experimental part, previous works have addressed AM in the legal domain using Naive Bayes and Maximum Entropy (Mochales Palau and Moens, 2011), factor graphs (Niculae et al., 2017), and residual networks (Galassi et al., 2018, 2021). More recently, advanced language models based on attention such as BERT (Devlin et al., 2019) and RoBERTa (Liu et al., 2019; Poudyal et al., 2020) have been used and combined with LSTMs and CNNs (Xu et al., 2020, 2021a,b). In this work, we exploit a combination of advanced language models, namely SBERT (Reimers and Gurevych, 2019) and Legal-BERT (Chalkidis et al., 2020), and traditional classifiers. We used existing language models without fine-tuning them. This is in line with recent efforts in the NLP community toward efficient machine learning methodologies with limited computational footprint (Lai et al., 2021).

3 Corpus Creation

The source corpus consists of 40 decisions on fiscal State aids by the Court of Justice of the European Union (CJEU), written in English. The decisions range from 2000 to 2018, i.e., since the CJEU's inception as a Court of Appeal in this domain. All documents have been downloaded from the EUR-LEX database and manually labelled. We have chosen this source since: (a) CJEU decisions usually contain a rich and diverse set of legal arguments (e.g., arguments appealing to statues, principles or precedents, according to different interpretive canons); (b) they have a standard (although not fixed) structure, in which argument chains are embedded and can be easily identified; (c) the selected decisions come from the same domain–i.e., fiscal State aids–which strongly relies on judicial interpretation; and (d) our annotators have some expertise in this domain.

3.1 Annotation Procedure

CJEU decisions are structured in clearly separated sections.[1] Since our primary purpose is to capture the argumentative patterns of the CJEU reasoning process, we focused on the section *Findings of the Court*, reporting all argumentative steps leading to the final ruling. This section is characterised by a set of interacting inferences, which ultimately lead to conclusions on the parties' claims. Each inference links a set of premises to a conclusion, which may support or attack further inferences.

The annotation guidelines were written and refined through multiple stages of annotation, evaluation of the agreement, and discussion. The annotation was done at the sentence level by two experts in the legal domain, using periods, semicolons, and line breaks as delimiters. As shown in Table 1, three hierarchical levels of annotation were identified in arguments: the elements (premises and conclusions), the type of premise (legal or factual), and the scheme.

3.1.1 Argumentative Elements and Types

Sentences compose arguments, which are included in argument chains. By an argument, we mean a set of connected inferences. Each such inference consists of the link between certain premises and a conclusion. It is important to note that the conclusion of an inference can also serve as the premise for further inferences. Such intermediate conclusions/premises have been marked as premises. By an argument chain, we mean an argument supporting a final conclusion concerning a specific ground of appeal, together with all counterarguments considered by the Court (see appendix B). More than one argument chain may be provided in a single decision.

For premises and conclusions, we defined mandatory and optional attributes and their possible values, as reported in Table 1. In particular, each premise and conclusion is denoted through a unique identifier (`ID`), whose value is constructed by joining a letter (which denotes the argument chain to which the premise(s) or the conclusion belongs to, e.g. A or B), with a progressive number (which distinguishes the single premise or conclusion withing the chain, e.g., A1, A2, An; B1, B2, Bn).

We distinguished between *factual* and *legal* premises. The former describes factual situations and events (pertaining to the substance or the procedure of the case); the latter specifies the legal content (legal rules, precedents, interpretation of applicable laws and principles). Whenever a premise combines legal and factual aspects, it has been

[1] Additional details about decision's structure are indicated in Appendix A.

Argumentative elements	Tag	Mandatory attributes of the element			Optional attribute of the element		
		Name	Value	Tag	Name	Value	Tag
Premise	<prem>	Identifier	A1, A2, An B1, B2, Bn	ID="An"	/	/	/
		Type	Legal	T="L"	Argumentation scheme	Argument from Rule	S="Rule"
						Argument from Precedent	S="Prec"
						Authoritative Argument	S="Aut"
						Argument from Verbal Classification	S="Class"
						Argument from Interpretation	S="Itpr"
						Argument from Principle	S="Princ"
			Factual	T="F"	/	/	/
Conclusion	<conc>	Identifier	An, Bn, Cn	ID="An"	/	/	/

Table 1: Annotation scheme.

marked as both legal and factual. Examples of premises, their classification, and argument chains can be found in Appendix B.

3.1.2 Argumentation Schemes

In general, legal premises determine the nature of the inference in which they are used; thus we have labelled them with the corresponding type of inference, which we call argument scheme following Walton et al. (2008, 2021). As an example, consider the following legal premise marked under the Rule scheme:

> *As stated in recital 14 of the preamble to that regulation, this limitation period has been established for reasons of legal certainty.* (Case C-408/04 P, para 102)

In this work, we rely on a set of schemes inspired by the work by Walton et al. (2008, 2021), which we specifically adapted to the CJEU reasoning, as made explicit in the cases. In particular, we identified six argument schemes that are not exclusive between each other. Therefore, a single legal premise may be assigned multiple schemes.

Rule (or established rule) scheme. According to the Rule scheme, a legislative rule is applicable to the case and determines its outcome unless exceptional provisions exist whichoverride that rule. In CJEU decisions, we used this scheme to classify premises explicitly citing an EU norm as part of the relevant legislative framework. Thus, we excluded all cases where the Court refers to national laws or to norms mentioned by the Court of First Instance since such norms can not be considered a basis for the CJEU decision. As an example, consider the following premise:

> *...Article 173 of the Treaty, ...provides that any natural or legal person may on the grounds of lack of competence, infringement of an essential procedural requirement, infringement of this Treaty ...institute proceedings against a decision addressed to that person ....* (Case C-298/00 P, para 34).

Precedent scheme. According to the Precedent scheme, the *ratio decidendi* of a past case is applicable to the current case determining its outcome unless a distinction can be made (Langenbucher, 1998). Under this scheme, we marked the CJEU premises referring to its past decisions. Textual indicators signalling a precedent scheme include references to cited judgements as well as a set of expressions such as "according to settled case-law"; "as is apparent from that case-law"; "as the Court has consistently held". As an example, consider the following premise:

> *...undertakings to which aid has been granted may not, in principle, entertain a legitimate expectation that the aid is lawful unless it has been granted in compliance with the procedure laid down in that article and, second, that a diligent businessman should normally be able to determine whether that procedure has been followed (Case C-5/89 Commission v Germany [1990] ECR I-3437, paragraph 14;...*.(Joined Cases C-183/02 P and C-187/02 P, para 44).

Authoritative scheme. According to the Authoritative scheme, an indication by an authority is applicable to the current case and may support its outcome, in the absence of reasons to the contrary. It is possible to distinguish three different types of authoritative inferences: (1) the *inference from administrative authority*, having a right to exercise command or influence over another party subject to that authority; (2) the *inference from expert opinion*, which is an epistemic authority having an expertise in the relevant field of knowledge; and (3) the *inference from the authority of the majority of the people or the common opinion* (Walton et al., 2021; Walton and Koszowy, 2015). In our corpus, we marked as inferences from authority the CJEU statements reporting an opinion of the Advocate General, since such opinions can be considered as authoritative sources of knowledge on which the Court relies, even though they are not legally binding. As an example, consider the following premise:

> *It follows, as the Advocate General observed . . . that recovery of such aid entails the restitution of the advantage procured by the aid for the recipient, not the restitution of any economic benefit the recipient may have enjoyed as a result of exploiting the advantage.* (Joined Cases C-164/15 P and C-165/15 P, para 92).

Classification scheme. According to the Classification scheme a concept is applicable to the current case and may support a corresponding classification unless an exception also applies. This scheme is an adaptation of the Verbal Classification scheme in (Macagno and Walton, 2015; Walton et al., 2008). The acceptability of the scheme from classification depends on the acceptability of the classification and on whether it admits possible exceptions or defaults. We marked a premise under this scheme whenever it consists of a definition of a legal concept, indicating the preconditions for a certain fact, property or entity to be qualified as falling under the concept. As an example, consider the following:

> *So, in order for there to be State aid within the meaning of that provision it is necessary, first, for there to be aid favouring certain undertakings or the production of certain goods and, second, for that advantage to come from the State or State resources.* (Case C-353/95 P, para 25)

Interpretative scheme. According to the interpretative scheme, a meaning relevant to the decision of the case is ascribed to a legal source (e.g., legislation, precedent, ...). This scheme includes different kinds of interpretative reasoning (e.g., literal, teleological, psychological, systematic interpretation, ...). Consider the following premise as an example of a psychological interpretative scheme:

> *. . . the intention of the EC Treaty, in providing through Article 88 EC for aid to be kept under constant review and monitored by the Commission, is that the finding that aid may be incompatible with the common market is to be arrived at,. . . , by means of an appropriate procedure which it is the Commission's responsibility to set in motion.* (Case C-272/12 P, para 48)

Principle scheme. According to the Principle scheme, a general legal principle is applicable to the case and may determine its outcome.

We annotated under this scheme those premises explicitly stating that a given fact, property or entity should be qualified in a certain way for complying or not complying with a certain principle of law. As an example, consider the following premise:

> *That fact however had to be taken into consideration in relation to the obligation to recover the incompatible aid, in the light of the principles of protection of legitimate expectations and legal certainty, . . .* (Case C-272/12 P, para 53).

Whenever a premise is relevant under more than one scheme, such premise has been marked accordingly (see Appendix B for examples).

3.2 Inter-Annotator Agreement

To measure the inter-annotator agreement regarding the classification of sentences as premises and conclusions, 14 documents were tagged by the two annotators, reaching a Cohen's kappa (Cohen, 1960) of 0.95, which indicates an almost perfect agreement. We have also measured the agreement considering only the argumentative sentences, obtaining a kappa of 0.86, which indicates strong agreement.

In order to calculate the agreement for the type attribute (legal/factual), we considered only the sentences that both annotators had labelled as premises, to avoid the propagation of error from one annotation layer to the other. We compute the Cohen's kappa on each value separately, treating it as a binary classification problem and obtained a strong agreement for both the classes: 0.87 for *factual* and 0.82 for *legal*.

To avoid error propagation, the agreement for the scheme attribute was measured on 10 documents on which the annotators had already solved previous conflicts, to consider only sentences that are legal premises according to both annotators. We computed the Cohen's kappa, as done for the type attribute, obtaining the results reported in Table 2. The agreement for the *class* (classification) scheme was none and the one for the *princ* (principle) scheme was weak. This evaluation is highly

	Aut	Class	Itpr	Prec	Princ	Rule
Only Ann. 1	2	0	14	3	2	2
Only Ann. 2	0	2	29	7	3	3
Both Ann.	4	0	80	82	2	76
κ	0.79	0.00	0.46	0.88	0.43	0.93

Table 2: Number of sentences labelled for each scheme by each annotator and agreement between them.

Element	#
documents	40
sentences	9320
prem	2375
conc	160
factual	1575
legal	906

Element	#
aut	53
class	56
itpr	296
prec	503
princ	15
rule	322

Table 3: Composition of the dataset.

influenced by the fact that these schemes were represented only in very few sentences. Another class for which the agreement was weak is *itpr* (interpretative), probably motivated by the fact that this is a mixed category, that groups together different kinds of interpretative schemes. Despite having only a few samples, there was moderate agreement on the *aut* (autoritative) scheme, while the agreement was strong for *prec* (precedent) and *rule*.

Most disagreements were due to: (i) the ambiguity of some argumentative sentences, often embedding multiple schemes; (ii) the fuzzy and overlapping boundaries between different schemes; (iii) the lack of clear language qualifiers and rhetorical clues characterizing some schemes; (iv) the different subject matters potentially falling under the same scheme. This is particularly true with regard to the interpretative scheme, which includes, as noted above, the application of different argumentative canons, each referring to different substantive grounds. Finally, while argument schemes are separately characterised and clearly analysed in theoretical studies, often in the judicial discourse complex argument patterns are present, where multiple inferences are merged and premises are left implicit.

3.3 Demosthenes Corpus

The conflicts between annotators have been solved by a third legal expert, who considered the source of the divergence and discussed with the two annotators the possible solutions. The final corpus is publicly available[2] and its composition can be found in Table 3.

The annotation regarding argumentative elements and their type can be considered reliable due to the strong agreement between annotators. Conversely, the annotation of the schemes can be considered reliable only for some of them, namely *Aut*, *Prec*, and *Rule*, while the other schemes must be considered potentially noisy.

4 Experimental Setting

In this study, we addressed four tasks. Two are general argument mining tasks, namely argument detection and argument classification. The other two are rather domain specific and are type classification and scheme classification. They are defined as follows:

- **Argument Detection (AD)**: given a sentence, classify it as *premise*, *conclusion*, or *neither*;
- **Argument Classification (AC)**: given a sentence that is known to be argumentative, classify it as *premise* or *conclusion*;
- **Type Classification (TC)**: a multi-label classification problem where a sentence that is known to be a premise is classified as *legal* (L) and/or *factual* (F);
- **Scheme Classification (SC)**: a multi-label classification task where a sentence, known to be a legal premise, is classified according to its scheme; due to the low number of samples in the dataset, the *Princ* scheme has not been considered.

We structured TC and SC as multi-label classification tasks since in both cases a single input sentence can have multiple labels. However, it is important to highlight that each sentence considered in these tasks has at least one label: there are no premises without a type, nor legal premises without a scheme. We did not enforce this constraint in our experiments and leave it for future work.

For AD, as a first step, we pre-processed the documents removing periods from some common abbreviations (e.g., ‘p.’ for ‘paragraph’ and ‘n.’ for ‘number’). The sentence segmentation was then performed based on periods, semicolons, and newlines. For all the tasks, we pre-processed the

[2] https://github.com/adele-project/demosthenes.

sentences by removing stop-words and punctuation symbols.

Experiments were conducted using 5-fold cross-validation with folds determined at the document level, so that sentences of the same document belong to the same fold. The folds were created manually to balance their composition and guarantee that all scheme classes were represented in each fold.

For all tasks we adopted three different representations of the input text:

- **TF-IDF**: vectorization based on the term frequency-inverse document frequency statistic;
- **Sentence-BERT (SBERT)** (Reimers and Gurevych, 2019): a modification of the BERT model that produces semantically meaningful sentences embeddings, mapping sentences with similar semantic content into vectors close to each other;[3]
- **Legal-BERT** (Chalkidis et al., 2020): a family of BERT models adapted to the legal domain.[4]

As classifiers, we have chosen a set of traditional machine learning techniques that have low computational requirements. We focused on these efficient techniques to assess if they are effective enough or if there is the need to adopt more advanced methods such as fine-tuned language models. Specifically, we experimented with the following models: linear svc, svc, random forest, Gaussian naive Bayes and k-neighbours.[5]

5 Results and Discussion

For each task, we report the results obtained by each combination of embeddings and classifiers. We also report the performance of two simple baselines: a classifier that outputs a random value and one that always predicts the majority class. We measure the F1 score obtained for each class and their macro-average.

AD. As can be seen in Table 4, most models perform well in the majority class (*neither*), including the majority baseline. They have more difficulties in recognizing argumentative sentences. The task can be considered not trivial since both baselines obtain an average score lower than 0.30. It is interesting to notice that the *conclusion* class obtains a higher score than *premise* despite the lower number of samples. Random Forests and Gaussian Naive Bayes perform poorly with all the embeddings. All the other models obtain good results when using Legal-BERT representation, which can be considered the best representation for this task. Nonetheless, the best result is obtained by the combination of Linear SVC and TF-IDF representation.

Embedding	Classifier	Avg	prem	conc	neither
-	Random	0.26	0.28	0.03	0.47
-	Majority	0.28	0.00	0.00	0.84
TF-IDF	Linear SVC	**0.70**	**0.58**	0.65	**0.88**
TF-IDF	Random Forest	0.65	0.48	0.60	**0.88**
TF-IDF	Gaussian NB	0.40	0.40	0.23	0.55
TF-IDF	K Neighbors	0.62	0.42	0.59	0.85
TF-IDF	SVC	0.53	0.14	0.59	0.86
SBERT	Linear SVC	0.69	0.55	**0.67**	0.85
SBERT	Random Forest	0.60	0.35	0.59	0.86
SBERT	Gaussian NB	0.52	0.54	0.34	0.69
SBERT	K Neighbors	0.65	0.50	0.64	0.82
SBERT	SVC	0.67	0.51	0.64	0.86
Legal-BERT	Linear SVC	0.69	**0.58**	0.62	0.87
Legal-BERT	Random Forest	0.59	0.44	0.46	0.87
Legal-BERT	Gaussian NB	0.59	0.54	0.55	0.67
Legal-BERT	K Neighbors	0.68	0.56	0.66	0.82
Legal-BERT	SVC	0.69	0.56	0.64	0.87

Table 4: Detailed results of the AD task.

AC. Table 5 shows the results of this classification task. The results are satisfactory, with all the models obtaining an average score above 0.80. They also obtain a score close to 1.00 for the premise class, but this also holds for the majority baseline. From our observation, random forests seem to be the best classifiers independently from the embedding used, obtaining the best score with TF-IDF representation and a similar result with the other ones.

TC. All the models perform better on the majority class (*factual*) obtaining a score between 0.75 and 0.89, as shown in Table 6. This is not surprising considering that the majority baseline reaches a score of 0.80. The best result on the *legal* label reaches a score of 0.80, for a macro average of 0.85, which can be considered a good result against the 0.60 score obtained by the best baseline. The SBERT representation is entirely dominated by the Legal-BERT one, while TF-IDF changes a lot depending on the classifier. The SVC per-

[3]We used the `bert-base-nli-mean-tokens` model.

[4]We used the `legal-bert-small` model.

[5]We used the default hyper-parameters offered by the `sci-kit learn` library.

Embedding	Classifier	Avg	prem	conc
-	Random	0.37	0.63	0.10
-	Majority	0.48	0.97	0.00
TF-IDF	Linear SVC	0.87	0.98	0.75
TF-IDF	Random Forest	**0.88**	**0.99**	**0.77**
TF-IDF	Gaussian NB	0.84	0.98	0.69
TF-IDF	K Neighbors	0.81	0.97	0.65
TF-IDF	SVC	0.82	0.98	0.66
SBERT	Linear SVC	0.85	0.98	0.71
SBERT	Random Forest	0.86	0.98	0.73
SBERT	Gaussian NB	0.81	0.97	0.66
SBERT	K Neighbors	0.84	0.98	0.71
SBERT	SVC	0.87	0.98	0.75
Legal-BERT	Linear SVC	0.80	0.98	0.63
Legal-BERT	Random Forest	0.86	0.98	0.73
Legal-BERT	Gaussian NB	0.86	0.98	0.74
Legal-BERT	K Neighbors	**0.88**	0.98	**0.77**
Legal-BERT	SVC	0.85	0.98	0.72

Table 5: Results of the AC task.

Embedding	Classifier	Avg	L	F
-	Random	0.60	0.50	0.69
-	Majority	0.40	0.00	0.80
TF-IDF	Linear SVC	0.83	0.77	0.88
TF-IDF	Random Forest	0.82	0.75	**0.89**
TF-IDF	Gaussian NB	0.68	0.61	0.75
TF-IDF	K Neighbors	0.76	0.70	0.82
TF-IDF	SVC	0.61	0.38	0.83
SBERT	Linear SVC	0.77	0.70	0.84
SBERT	Random Forest	0.74	0.64	0.85
SBERT	Gaussian NB	0.72	0.66	0.78
SBERT	K Neighbors	0.72	0.64	0.80
SBERT	SVC	0.80	0.73	0.87
Legal-BERT	Linear SVC	0.81	0.75	0.87
Legal-BERT	Random Forest	0.77	0.67	0.87
Legal-BERT	Gaussian NB	0.73	0.66	0.79
Legal-BERT	K Neighbors	0.78	0.72	0.85
Legal-BERT	SVC	**0.85**	**0.80**	**0.89**

Table 6: Results of the TC task.

form very well with Legal-BERT and with SBERT, outperforming the other classifiers, but when combined with TF-IDF leads to the worst performance instead.

SC. As shown in Table 7, the only class for which the baselines reach a good score is the *Prec* scheme; therefore we can consider the scheme classification problem to be not trivial. The results for the *Aut* scheme vary widely: the worst result is 0.00, while the best is 0.94. We hypothesize that this may be due to the limited amount of samples present in the dataset. The best result is obtained with Random Forest and TF-IDF, while Linear SVC classifiers perform well (above 0.60) with all the embeddings. Linear SVC obtains good results also for the *Class* scheme, outperforming all the other classifiers. SBERT and Legal-BERT representation perform similarly and they are outperformed by TF-IDF in most cases. The *Itpr* scheme seems to be the most challenging to predict, with the best value of 0.63 and no visible pattern in the performance of the models, probably due to the noisiness of the label. For the *Prec* scheme, all models outperform the baselines; Legal-BERT embeddings lead to good results (between 0.80 and 0.90), but the best result is obtained with Random Forest and TF-IDF. The classification as *Rule*, presents a lot of variance, with linear SVCs outperforming the other classifiers. The best results in terms of macro average are obtained with TF-IDF representation and Linear SVC (0.75), TF-IDF and Random Forest (0.73), and Legal-BERT and Linear SVC (0.74). Since *Itpr* and *Class* labels are potentially noisy, we also computed the macro average score excluding them. The best models are the same even according to this alternative metric, with TF-IDF and Random Forest outperforming the others.

Feature analysis Since the LinearSVC classifier trained on the TF-IDF representation performs well in all the proposed tasks, we analyzed which features are assigned more weight to understand which words can be considered good indicators for the prediction. For each task, in Table 8 we report the 10 most relevant words associated with each class. We can see that the words “must”, “follows”, “light”, “well”, “consequently”, and “rejected” are associated with *conc* both in AD and AC. Conversely, the only word associated with *prem* both in AD and AC is the word “directed”. This result suggests a more robust characterization of the *conc* class with respect of *prem*, and partially motivates the better result obtained in AD for *conc*. We can also see that some indicators of the *prem* class are also used to determine premises’ types or schemes. For example, the word “see” (which is often used to direct the reader to other judicial precedents) is associated with *prem* in AD and *legal* in TC, while the words “argument”, “claims”, and “general” are associated with *prem* in AC and the scheme *prec* in SC. The same consideration holds between *legal* premises and schemes: the words “article” and “ecr” are associated with *legal* in TC, while in SC

Embedding	Classifier	Avg	Aut	Class	Itpr	Prec	Rule	Avg$_{reliable}$
-	Random	0.33	0.10	0.12	0.42	0.55	0.44	0.36
-	Majority	0.14	0.00	0.00	0.00	0.71	0.00	0.24
TF-IDF	Linear SVC	**0.75**	0.85	**0.72**	0.48	0.88	0.83	0.85
TF-IDF	Random Forest	0.73	**0.94**	0.57	0.30	**0.91**	**0.91**	**0.92**
TF-IDF	Gaussian NB	0.44	0.00	0.62	0.34	0.74	0.51	0.42
TF-IDF	K Neighbors	0.60	0.72	0.57	0.28	0.75	0.68	0.72
TF-IDF	SVC	0.31	0.00	0.50	0.07	0.72	0.24	0.32
SBERT	Linear SVC	0.66	0.62	0.67	0.49	0.83	0.71	0.72
SBERT	Random Forest	0.46	0.07	0.48	0.49	0.81	0.46	0.45
SBERT	Gaussian NB	0.54	0.33	0.40	0.59	0.80	0.59	0.57
SBERT	K Neighbors	0.47	0.00	0.58	0.43	0.79	0.56	0.45
SBERT	SVC	0.51	0.11	0.48	0.47	0.83	0.65	0.53
Legal-BERT	Linear SVC	0.74	0.85	0.66	0.53	0.85	0.79	0.83
Legal-BERT	Random Forest	0.51	0.04	0.52	0.52	0.87	0.60	0.50
Legal-BERT	Gaussian NB	0.64	0.58	0.39	**0.63**	0.85	0.73	0.72
Legal-BERT	K Neighbors	0.53	0.29	0.58	0.32	0.80	0.67	0.59
Legal-BERT	SVC	0.64	0.49	0.48	0.58	0.90	0.77	0.72

Table 7: Results of the SC task. The last column reports the macro-average computed excluding the *Itpr* and *Class* scheme.

AD		**AC**		**TC**		**SC$_{reliable}$**		
conc	**prem**	**conc**	**prem**	**factual**	**legal**	**aut**	**prec**	**rule**
must	paragraph	must	argument	contested	see	advocate	paragraph	article
follows	noted	aside	complaint	present	ecr	opinion	caselaw	tfeu
admissible	recalled	dismissed	event	general	may	point	see	treaty
light	err	well	directed	appeal	member	observed	settled	ec
consequently	see	rejected	claims	issue	must	general	commission	regulation
accordingly	apparent	consequently	also	claims	jurisdiction	points	judgment	871
well	directed	entirety	declared	appellants	irrespective	essence	ecr	meaning
circumstances	paragraphs	follows	ndsht	assessment	article	noted	effect	1071
ground	vitiated	light	general	argument	party	orange	held	within
rejected	settled	qualifying	wam	notice	effect	goodwill	others	659199

Table 8: Most relevant features for each task and class, obtained from the LinearSVC classifier trained on the TF-IDF representation.

are indicators for *rule* and *prec* in SC respectively.

6 Conclusion

We presented *Demosthenes*, a new corpus for legal AM in the fiscal state aid domain. The corpus consists in 40 decisions by the CJEU, which have been annotated on three hierarchical levels, identifying argumentative elements, their type, and argumentative scheme.

We have defined 4 AM tasks: AD, AC, TC, SC. Our results highlight that Legal-BERT consistently obtains good scores in most settings and tasks. Surprisingly, the TF-IDF embeddings were often successful, suggesting that the lexical information may be informative enough to solve such tasks. For what concerns the classifiers, Linear SVC performed well in most of the settings. Our results suggest that traditional classifiers are effective in many of the proposed tasks. We believe that these models can be considered strong baselines for further experiments involving state-of-the-art classifiers such as fine-tuned language models.

In future work, we want to improve the scheme labelling by splitting the *Itpr* class into multiple ones, and annotate the relationships between sentences. Experimentally, we aim to implement oversampling and data augmentation techniques to overcome the strong unbalance of classes in each task. We also want to study the impact of pre-processing and the use of alternative classifiers such as logistic regression. Finally, we want to improve the robustness of our experimental findings. For example, by considering multiple seed runs or applying the method proposed by Lai et al. (2021).

Acknowledgements

This work has been partially funded by the European Union's Horizon 2020 Research and Innovation Programme under grant agreement 101017142 (StairwAI), by the European Union's Justice programme under grant agreement 101007420 (ADELE), by the European Horizon 2020 ERC project CompuLaw (Computable Law) under grant agreement 833647, and by the Italian Ministry of Education and Research's PRIN programme under grant agreement 2017NCPZ22 (LAILA).

References

Vincent Aleven. 2003. Using background knowledge in case-based legal reasoning: A computational model and an intelligent learning environment. *Artif. Intell.*, 150(1-2):183–237.

Kevin D. Ashley, Ravi Desai, and John M. Levine. 2002. Teaching case-based argumentation concepts using dialectic arguments vs. didactic explanations. In *Intelligent Tutoring Systems*, volume 2363 of *Lecture Notes in Computer Science*, pages 585–595. Springer.

Katie Atkinson and Trevor J. M. Bench-Capon. 2019. Reasoning with legal cases: Analogy or rule application? In *ICAIL*, pages 12–21. ACM.

Katie Atkinson and Trevor J. M. Bench-Capon. 2021. Argumentation schemes in AI and law. *Argument Comput.*, 12(3):417–434.

Roy Bar-Haim, Indrajit Bhattacharya, Francesco Dinuzzo, Amrita Saha, and Noam Slonim. 2017. Stance classification of context-dependent claims. In *EACL (1)*, pages 251–261. Association for Computational Linguistics.

Trevor J. M. Bench-Capon, James B. Freeman, Hanns Hohmann, and Henry Prakken. 2004. Computational models, argumentation theories and legal practice. In *Argumentation Machines*, volume 9 of *Argumentation Library*, pages 85–120. Springer.

Trevor J. M. Bench-Capon, Henry Prakken, and Giovanni Sartor. 2009. Argumentation in legal reasoning. In *Argumentation in Artificial Intelligence*, pages 363–382. Springer.

Elena Cabrio and Serena Villata. 2018. Five years of argument mining: a data-driven analysis. In *IJCAI*, pages 5427–5433. ijcai.org.

Chad S. Carr. 2003. Using computer supported argument visualization to teach legal argumentation. In *Visualizing Argumentation*, Computer Supported Cooperative Work, pages 75–96. Springer.

Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2020. LEGAL-BERT: The muppets straight out of law school. In *Findings of EMNLP 2020*, pages 2898–2904, Online. Association for Computational Linguistics.

Carlos Chesnevar, Sanjay Modgil, Iyad Rahwan, Chris Reed, Guillermo Simari, Matthew South, Gerard Vreeswijk, Steven Willmott, et al. 2006. Towards an argument interchange format. *The knowledge engineering review*, 21(4):293–316.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37 – 46.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT (1)*, pages 4171–4186. Association for Computational Linguistics.

Vanessa Wei Feng and Graeme Hirst. 2011. Classifying arguments by scheme. In *ACL*, pages 987–996. The Association for Computer Linguistics.

Andrea Galassi, Marco Lippi, and Paolo Torroni. 2018. Argumentative link prediction using residual networks and multi-objective learning. In *Proceedings of the 5th Workshop on Argument Mining, ArgMining@EMNLP 2018, Brussels, Belgium, November 1, 2018*, pages 1–10. Association for Computational Linguistics.

Andrea Galassi, Marco Lippi, and Paolo Torroni. 2021. Multi-task attentive residual networks for argument mining. *CoRR*, abs/2102.12227.

Ivan Habernal, Daniel Faber, Nicola Recchia, Sebastian Bretthauer, Iryna Gurevych, Christoph Burchard, et al. 2022. Mining legal arguments in court decisions. *arXiv preprint arXiv:2208.06178*.

Ben Hachey and Claire Grover. 2006. Extractive summarisation of legal texts. *Artificial Intelligence and Law*, 14(4):305–345.

Rinke Hoekstra, Joost Breuker, Marcello Di Bello, and Alexander Boer. 2009. LKIF core: Principled ontology development for the legal domain. In *Law, Ontologies and the Semantic Web*, volume 188 of *Frontiers in Artificial Intelligence and Applications*, pages 21–52. IOS Press.

Tuan Lai, Heng Ji, and ChengXiang Zhai. 2021. BERT might be overkill: A tiny but effective biomedical entity linker based on residual convolutional neural networks. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 1631–1639, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Katja Langenbucher. 1998. Argument by analogy in european law. *The Cambridge Law Journal*, 57(3):481–521.

John Lawrence and Chris Reed. 2019. Argument mining: A survey. *Comput. Linguistics*, 45(4):765–818.

Ran Levy, Yonatan Bilu, Daniel Hershcovich, Ehud Aharoni, and Noam Slonim. 2014. Context dependent claim detection. In *COLING*, pages 1489–1500. ACL.

Marco Lippi and Paolo Torroni. 2016. Argumentation mining: State of the art and emerging trends. *ACM Trans. Internet Techn.*, 16(2):10:1–10:25.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Fabrizio Macagno and Douglas Walton. 2015. Classifying the patterns of natural arguments. *Philosophy & Rhetoric*, 48(1):26–53.

L. Thorne McCarty. 2007. Deep semantic interpretations of legal texts. In *ICAIL*, pages 217–224. ACM.

Raquel Mochales and Marie-Francine Moens. 2011. Argumentation mining. *Artificial Intelligence and Law*, 19(1):1–22.

Raquel Mochales Palau and Marie-Francine Moens. 2011. Argumentation mining. *Artificial Intelligence and Law*, 19(1):1–22.

Vlad Niculae, Joonsuk Park, and Claire Cardie. 2017. Argument mining with structured svms and rnns. In *ACL (1)*, pages 985–995. Association for Computational Linguistics.

Raquel Mochales Palau and Aagje Ieven. 2009. Creating an argumentation corpus: do theories apply to real arguments?: a case study on the legal argumentation of the ECHR. In *ICAIL*, pages 21–30. ACM.

Prakash Poudyal, Jaromir Savelka, Aagje Ieven, Marie Francine Moens, Teresa Goncalves, and Paulo Quaresma. 2020. ECHR: Legal corpus for argument mining. In *ArgMining@COLING*, pages 67–75, Online. Association for Computational Linguistics.

Henry Prakken and Giovanni Sartor. 1996a. A dialectical model of assessing conflicting arguments in legal reasoning. *Artif. Intell. Law*, 4(3-4):331–368.

Henry Prakken and Giovanni Sartor. 1996b. A dialectical model of assessing conflicting arguments in legal reasoning. *Artif. Intell. Law*, 4(3-4):331–368.

Henry Prakken and Giovanni Sartor. 2002. The role of logic in computational models of legal argument: A critical survey. In *Computational Logic: Logic Programming and Beyond*, volume 2408 of *Lecture Notes in Computer Science*, pages 342–381. Springer.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In *EMNLP/IJCNLP (1)*, pages 3980–3990. Association for Computational Linguistics.

Jaromír Savelka and Kevin D. Ashley. 2016. Extracting case law sentences for argumentation about the meaning of statutory terms. In *ArgMining@ACL*. The Association for Computer Linguistics.

Manfred Stede and Jodi Schneider. 2018. Argumentation mining. *Synthesis Lectures on Human Language Technologies*, 11(2):1–191.

Milagro Teruel, Cristian Cardellino, Fernando Cardellino, Laura Alonso Alemany, and Serena Villata. 2018. Legal text processing within the mirel project. In *1st Workshop on Language Resources and Technologies for the Legal Knowledge Graph*, page 42.

Joel P Trachtman. 2013. The tools of argument: How the best lawyers think, argue, and win. *Argue, and Win (July 29, 2013)*.

Vern R. Walker, Nathaniel Carie, Courtney C. DeWitt, and Eric Lesh. 2011. A framework for the extraction and modeling of fact-finding reasoning from legal decisions: lessons from the vaccine/injury project corpus. *Artif. Intell. Law*, 19(4):291–331.

Vern R. Walker, Ji Hae Han, Xiang Ni, and Kaneyasu Yoseda. 2017. Semantic types for computational legal reasoning: propositional connectives and sentence roles in the veterans' claims dataset. In *ICAIL*, pages 217–226. ACM.

Vern R. Walker, Krishnan Pillaipakkamnatt, Alexandra M. Davidson, Marysa Linares, and Domenick J. Pesce. 2019. Automatic classification of rhetorical roles for sentences: Comparing rule-based scripts with machine learning. In *ASAIL@ICAIL*, volume 2385 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Vern R. Walker, Stephen R. Strong, and Vern E. Walker. 2020. Automating the classification of finding sentences for linguistic polarity. In *ASAIL@JURIX*, volume 2764 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Douglas Walton and Marcin Koszowy. 2015. Two kinds of arguments from authority in the ad verecundiam fallacy. In *Proceedings of the 8th Conference of the International Society for the Study of Argumentation*, pages 1483—-1492.

Douglas Walton, Fabrizio Macagno, and Giovanni Sartor. 2021. *Statutory interpretation: Pragmatics and argumentation*. Cambridge University Press.

Douglas Walton, Chris Reed, and Fabrizio Macagno. 2008. *Argumentation Schemes*. Cambridge University Press.

Huihui Xu, Jaromír Savelka, and Kevin D. Ashley. 2020. Using argument mining for legal text summarization. In *JURIX*, volume 334 of *Frontiers in Artificial Intelligence and Applications*, pages 184–193. IOS Press.

Huihui Xu, Jaromír Savelka, and Kevin D. Ashley. 2021a. Accounting for sentence position and legal domain sentence embedding in learning to classify case sentences. In *JURIX*, volume 346 of *Frontiers in Artificial Intelligence and Applications*, pages 33–42. IOS Press.

Huihui Xu, Jaromír Savelka, and Kevin D. Ashley. 2021b. Toward summarizing case decisions via extracting argument issues, reasons, and conclusions. In *ICAIL*, pages 250–254. ACM.

Gechuan Zhang, Paul Nulty, and David Lillis. 2022a. A decade of legal argumentation mining: Datasets and approaches. In *NLDB*, volume 13286 of *Lecture Notes in Computer Science*, pages 240–252. Springer.

Gechuan Zhang, Paul Nulty, and David Lillis. 2022b. Enhancing legal argument mining with domain pre-training and neural networks. *CoRR*, abs/2202.13457.

Appendix

A Source Documents' Structure

CJEU decisions are structured as follows:

- *The Preamble*, containing information on the parties, i.e., on the one hand, the Commission, and on the other hand a member State and/or a private competitor, the appealed judgement of the Court of First Instance, and the composition of the Court;
- *Case background*, including facts and the procedural case history before the General Court;
- *The judgement under appeal*, reporting the assessment of the General Court in the first instance decision;
- *The Appeal*, reporting *The Grounds of Appeal*, i.e., the error of law or facts alleged by an Appellant as the defect in the Judgment appealed against upon which reliance has been placed to set it aside. Thus, grounds of appeal concern the reason(s) why the decision is considered wrong by the aggrieved party. For each ground of appeal, two subsections can be identified: (i) the *Arguments of the Parties*, supporting or attacking each ground of appeal; and (ii) the *Findings of the Court*, i.e., the Court reasoning process, characterised by a set of argument chains, which lead to conclusions with regard to parties' claims, as described in the grounds of appeal;
- *Costs*, i.e., the attribution of costs;
- The *Ruling*, i.e., the final decision and orders to the parties.

In analysing the CJEU decisions, we did not consider sections related to *the preamble*, *the case background*, and *the judgment under appeal*, where no arguments are put forward. The same is true with regard to the *costs* and the *final ruling* sections, the latter usually repeating the conclusion of each argument chain and reporting orders to the parties. Since our primary purpose is to capture the argumentative patterns of the CJEU reasoning process, we also excluded the section related to the *arguments of the parties*. Thus, the most relevant part is the *Findings of the Court*, reporting all argumentative steps leading to the final ruling.

B Detailed Examples

B.1 Type of Premise

The following statements respectively consist in factual and legal premises:

> *In the present case the main appeal, taken as a whole, specifically seeks to challenge the position adopted by the Court of First Instance on various points of law raised before it at first instance. It indicates clearly the aspects of the judgment under appeal which are criticised and the pleas in law and arguments on which it is based.* (Case C–321/99 P, para 50).

> *Where an appellant alleges distortion of the evidence by the General Court, he must, under Article 256 TFEU, the first paragraph of Article 58 of the Statute of the Court of Justice of the European Union and Article 168(1)(d) of the Rules of Procedure of the Court, indicate precisely the evidence alleged to have been distorted by the General Court and show the errors of appraisal which, in his view, led to such distortion.* (Case C-431/14 P, para 32).

Example of a premise that combines legal and factual arguments:

> *It is apparent from the judgment under appeal and the documents included in the file that the appellants submitted before the General Court that, contrary to what the Commission stated in point 97 of the grounds of the contested decision, the normal tax rules for company profits could not be used as a valid basis for comparison and thus as a reference framework for the assessment of the selectivity of the tax scheme at issue.* (Case C—452/10 P, para 57).

B.2 Types of Schemes

In the following, we provide examples of legal premises marked according to the schemes presented in section 3.1.2.
Examples of legal premises marked under the **Rule scheme**.

> *It must be recalled that Article 173 of the Treaty, by virtue of which the Court of Justice is to review the legality of Community acts, provides that any natural or legal person may on grounds of lack of competence, infringement of an essential procedural requirement, infringement of this Treaty or of any rule of law relating to its application, or misuse of powers institute proceedings against a decision addressed to that person or against a decision which, although in the form of a regulation or a decision addressed to another person, is of direct and individual concern to the former.* (Case C-298/00 P, para 34).

> *Consequently, given that Article 1 of the Third Steel Aid Code prohibited both aid that was and aid that was not specific to the steel sector, the Commission could not implicitly withdraw the 1971 Decision.* (Case C-408/04 P, para 89)

Examples of legal premises marked under the **Precedent scheme**.

> *It should be borne in mind, first, that in view of the mandatory nature of the review of State aid by the Commission under Article 93 of the Treaty, undertakings to which aid has been granted may not, in principle, entertain a legitimate expectation that the aid is lawful unless it has been granted in compliance with the procedure laid down in that article and, second, that a diligent businessman should normally be able to determine whether that procedure has been followed (Case C-5/89 Commission v Germany [1990] ECR I-3437, paragraph 14; Case C-169/95 Spain v Commission [1997] ECR I-135, paragraph 51; and Case C-24/95 Alcan Deutschland [1997] ECR I-1591, paragraph 25).*(Joined Cases C-183/02 P and C-187/02 P, para 44).

> *Also, it is clear from consistent case-law that Articles 4 CS and 67 CS concern two distinct areas, the first abolishing and prohibiting certain actions by Member States in the field which the ECSC Treaty places under Community jurisdiction, the second intended to prevent the distortion of competition which exercise of the residual powers of the Member States inevitably entails.* (Case C-408/04 P, para 32).

Examples of legal premises marked under the **Authoritative scheme**.

> *It follows, as the Advocate General observed, in essence, in point 62 of his Opinion, that recovery of such aid entails the restitution of the advantage procured by the aid for the recipient, not the restitution of any economic benefit the recipient may have enjoyed as a result of exploiting the advantage.* (Joined Cases C-164/15 P and C-165/15 P, para 92).

> *Accordingly, as the Advocate General noted in points 72 and 76 of his Opinion, nothing prevents the recipient of the aid from invoking the applicability of that test and, if the recipient does invoke that test, it falls to the Commission to assess whether the test needs to be applied and, if so, to assess its application.* (Case C-300/16 P, para 26)

Examples of legal premises marked under the **Classification scheme**.

> *So, in order for there to be State aid within the meaning of that provision it is necessary, first, for there to be aid favouring certain undertakings or the production of certain goods and, second, for that advantage to come from the State or State resources.* (Case C-353/95 P, para 25)

> *Any activity consisting in offering services on a given market, that is, services normally provided for remuneration, is an economic activity.* (Joined Cases C-622/16 P to C-624/16 P, para 104)

Examples of legal premises marked under the **Interpretative scheme**. The first premise below constitutes an example of a teleological interpretation, while the second one constitutes an example of a psychological interpretation.

> *The effectiveness of Article 107 TFEU would be substantially diminished if the Commission were required, before classifying a measure as State aid within the meaning of that provision, to wait for the decision of the courts with jurisdiction regarding any reimbursement of excess tax or tax paid by certain taxpayers.* (Joined Cases C-164/15 P and C-165/15 P, para 78)

> *As the Court held in paragraphs 29 to 31 of Case C-110/02 Commission v Council [2004] ECR I-6333, the intention of the EC Treaty, in providing through Article 88 EC for aid to be kept under constant review and monitored by the Commission, is that the finding that aid may be incompatible with the common market is to be arrived at, subject to review by the General Court and the Court of Justice, by means of an appropriate procedure which it is the Commission's responsibility to set in motion.* (Case C-272/12 P, para 48)

Examples of legal premises marked under the **Principle scheme**

> *That fact however had to be taken into consideration in relation to the obligation to recover the incompatible aid, in the light of the principles of protection of legitimate expectations and legal certainty, as was done by the Commission in the contested decision when it declined to order the recovery of aid granted before the date of publication in the Official Journal of the European Communities of the decisions to initiate the procedure laid down in Article 88(2) EC* (Case C-272/12 P, para 53).

> *Those arguments cannot, however, be upheld, since, as is apparent from the case-law, the question whether a selective advantage complies with the principle of proportionality arises at the third stage of the examination of selectivity, which examines whether that advantage can be justified by the nature or general scheme of the tax system of the Member State concerned.* (Joined Cases C-51/19 P and C-64/19 P, para 140)

Examples of legal premises marked under more than one scheme.
The following is an example of a premise marked under both the **Precedent scheme** and the **Principle scheme**.

> *The principle of legal certainty – which is one of the general principles of European Union law – requires that rules of law be clear and precise and predictable in their effect, so that interested parties can ascertain their position in situations and legal relationships governed by European Union law (see, to that effect, Case C-63/93 Duff and Others [1996] ECR I-569, paragraph 20; Case C-76/06 P Britannia Alloys; Chemicals v Commission [2007] ECR I-4405, paragraph 79; and Case C-158/07 Förster [2008] ECR I-8507, paragraph 67).* (Case C-81/10 P, para 100).

The following is an example of a premise marked under both the **Rule scheme** and the **Precedent scheme**.

> *In that regard, it must be observed that it follows from Article 58 of the Statute of the Court of Justice, in conjunction with Article 113(2) of the Rules of Procedure of the Court of Justice, that, on appeal, an appellant may put forward any relevant argument, provided only that the subject-matter of the proceedings before the General Court is not changed in the appeal (Case*

C-229/05 P PKK and KNK v Council [2007] ECR I-439, paragraph 66, and Case C-8/06 P Herrero Romeu v Commission [2007] ECR I-10333, paragraph 32) (Case C-322/09 P, para 41).

B.3 Argument Chain

The following is an example of sentences that constitute an argument chain.

`<prem ID="C1" T="L" S="Prec">` *According to the case-law of the Court of Justice, for infringement of the principle of the protection of legitimate expectations to be established, it is necessary for an EU institution, by giving a citizen precise assurances, to have led that person to entertain justified expectations.* `</prem>` `<prem ID="C2" T="L" S="Prec|Class">` *Information which is precise, unconditional and consistent, in whatever form it is given, constitutes such assurances (judgment of 12 October 2016, Land Hessen v Pollmeier Massivholz, C-242/15 P, not published, EU:C:2016:765, paragraph 63).* `</prem>`

`<prem ID="C3" T="F">` *In that regard, in its 2004 letter, the Commission merely expressed a preliminary opinion on a draft of the promotion scheme which was adopted only the following year, the precise conditions of which were not then fully known.*`</prem>` `<prem ID="C4" T="F">` *Consequently, that letter did not give precise assurances that the initial scheme was not in the nature of State aid.* `</prem>` `<prem ID="C5" T="F">` *Therefore, the General Court did not err in its legal characterisation by holding in paragraph 70 of the judgment under appeal that that letter could not give rise to any legitimate expectation.* `</prem>`

`<prem ID="C6" T="F">` *Nor can the General Court be criticised for not taking the view that such an expectation could result from the alleged '2006 decision'.*`</prem>` `<prem ID="C7" T="F">` *As the General Court pointed out in paragraph 60 of the judgment under appeal, that decision had not been placed on the file, nor even specifically identified by the appellants.*`</prem>`

`<prem ID="C8" T="F">` *Nor do the appellants demonstrate that the General Court incorrectly characterised the Commission's conduct between 2004 and the adoption of the decision at issue in finding, in paragraph 78 of the judgment under appeal, that that conduct could not be regarded as having provided precise, unconditional and consistent assurances that there was no State aid.*`</prem>`

`<prem ID="C9" T="L|F">` *Moreover, the appellants may criticise the General Court for failing to take into account certain other factors, which they claim to have submitted to it, only if that evidence proves that they could rely on a legitimate expectation that the initial scheme for the promotion of electricity production from RES would be maintained.*`</prem>` `<prem ID="C10" T="F">` *The appellants have not shown that that evidence was sufficient to justify the legitimate expectation alleged.*`</prem>`

`<prem ID="C11" T="L|F" S="Prec">` *In particular, the appellants do not effectively challenge the finding, in paragraph 79 of the judgment under appeal, that exceptional circumstances should not be taken into account in the present case, in so far as that consideration was envisaged, in the judgment of 11 July 1996, SFEI and Others (C-39/94, EU:C:1996:285), only in order to establish that, in certain cases, the repayment of State aid sought before a national court is inappropriate.*`</prem>`

`<conc ID="C12">` *It follows from the foregoing that the third ground of appeal must be rejected.*`</conc>`

(Case C-850/19 P, para 34–40).