

Formalizing EU Normative Legislation by leveraging FrameNet and frame semantics

Célestine Lecat¹

ALMA-AI, CIRSFID, Department of Law, University of Bologna, Bologna, Italy
`celestine.lecat2@unibo.it`

Abstract. In recent years, the EU has been making an effort to digitalize its processes for legal management and drafting under the “Law as Code” paradigm. In this context arises the problem of providing machines with an understanding of legal concepts, all the while ensuring that legislators remain able to interpret both the legislation and its encoding. The addition of a semantic layer is a way to formalize EU legislation in a manner that is both human- and machine-readable. For this purpose, the FrameNet data set gives us the opportunity to anchor this semantic layer to lexical and syntactic patterns found in legal documents. In this paper we present a doctoral thesis project aiming to create a formalization method for EU normative legislation by adapting and expanding the FrameNet data set for this purpose, with the broader goal of providing a semantic formalization method compatible with AI systems used in legal knowledge management.

Keywords: FrameNet · Legal formalization · Semantic data modeling.

1 Rationale

1.1 Introduction

The multi-jurisdictional and multilingual nature of EU statutory texts has prompted the launch of many research projects and initiatives for achieving better management, legal clarity and compliance monitoring in normative legislation. With this aim in mind, several digitalization solutions have been mobilized by EU institutions to represent legislation in a formal, machine-readable format with an emphasis on interoperability. The formalization of normative legislation is an integral part of this process. By providing common vocabularies and concepts, it can act as a basis for the standardization of legal processes, thus offering opportunities for more interoperable and computable law.

1.2 Field of research

Aside from reducing administrative and financial burdens, formalizing the law is a way to preserve its validity by reducing semantic ambiguity and improving traceability. With the convergence of the EU towards the “Law as Code”

paradigm, a cross-technical pipeline [1] combining hybrid AI techniques with Semantic web approaches to formalize legal texts can act as a safeguard amidst legal digitalization. The HyperModelLex ERC project embodies the EU’s efforts to guarantee the legitimacy of automated law-making processes [2].

The diversity of EU policy-makers and norms, as well as the need for multilingualism, all represent hurdles for the drafting and management of legislation. They hinder provision tracking and compliance monitoring. This is particularly true for reporting requests – also called regulatory reporting –, which, in EU legislation, require institutions to submit reports for legislative monitoring and policy implementation [3]. Since they are scattered throughout EU legislation, they are affected by regulatory overlaps which limit data reuse. A digital streamlining would address these issues.

In an effort to mitigate these problems, the European Union has started deploying standards such as the European Legislation Identifier (ELI) or AKN4EU, a legal XML schema customized for the EU from the Akoma Ntoso standard [4] [5] [6], in actual legislative publishing. The alignment of AKN and ELI legal ontologies can also advocate for a better interoperability and data reuse between countries and institutions [7].

Formalization can be achieved through the incorporation of a logic layer like LegalRuleML [8] or of a semantic layer (FrameNet, legal ontologies) to encode pieces of legislation. The EU has developed customized legal ontologies like EuroVoc or RRMV for this purpose (SORTIS project).

1.3 State of the art

The FrameNet data set mobilized for legal knowledge modeling is based on the theory of frame semantics developed by Charles J. Fillmore in the 1970s [9], which led to some key publications in the 1980s [10] [11], and was followed by the launch of FrameNet in 1997. Work around FrameNet is still ongoing and the area is emerging in the legal informatics field.

Originally, the FrameNet dataset was English-based, but it has been expanded to other languages, with some notable examples in Japanese, Swedish, German or Brazilian Portuguese [12]. Although some works have emitted reservations about the cross-lingual application of some more complex frames [13], the cross-linguistic design of frame semantics also brings good hope of achieving a better workflow between languages in EU normative legislation.

The importance of frame-based ontologies has been acknowledged since the early 90’s with Van Kralingen’s proposal of a frame language for legal knowledge representation [14], and his work has inspired the creation of FrameNet-based, light-weight ontologies designed for the legal domain [15]. Other similar domain-specific modeling approaches have been explored in scientific fields [16] but research in the legal domain is still scarce. The use of these ontologies in legal formalization has piqued researchers’ interest in the late 2000s and early 2010s, especially to solve versioning and consolidation issues [17] but also to enhance legal semantic annotation [18] [19].

In recent years, EU-led projects in legal informatics have encouraged the development of solutions for enhancing computable law, with, for instance, the European Commission funding projects like the HAIMLA design [20] within ERC HyperModeLex. This interest for executable and automatable law marks the transition to a new Smart Legal Order [21]. The emphasis put on hybrid AI has prompted the emergence of new research projects incorporating FrameNet into neuro-symbolic pipelines [22].

As of today, research on FrameNet as a means of legal formalization is present but scarce. Most works date back to the late-2000s, and there is no mature FrameNet coverage for legal language, nor a specific adaptation for EU legislation. Investigating a specialized use of FrameNet for EU normative legislation would support efforts towards a more reliable automation, and a transition to the “Law as Code” paradigm with better transparency and plural interpretation [23].

2 Methodology

2.1 Adaptation of FrameNet Frames

Interpretative difficulties make encoding law a multidisciplinary task [24] that combines computer science, legal expertise and linguistics. Linguistics serve the purpose of analyzing the syntax, semantics and pragmatics of statutory texts. Anchoring a semantic description of law in its digital representation is a way to create a ground base for legal interpretation (both by humans and machines). Legal terms fit well into the FrameNet ecosystem, as they have rigorous definitions and are often repeated throughout similar pieces of legislation. Additionally, FrameNet is able to add a structured and symbolic annotation directly to the lexical expression of the frame itself, which makes it particularly interesting in the context of legislative documents, where adherence to the linguistic expression of the law itself can be considered as an advantage. Nevertheless, the combination of other symbolic approaches (i.e. knowledge graphs) and FrameNet should not be seen as competing, instead the aim should be to synergize these approaches and to produce meaningful, yet lexically grounded, annotations.

In practice, FrameNet is a set of English lexical databases that map forms to meanings through the lens of frame semantics. This theory relies on the idea that the meaning of a word can only be defined by the context it lies in: this includes the syntactic/lexical context (surrounding words) and the cognitive context (domain-specific knowledge, common sense). This defining context is what we call a Frame. In a single sentence, it is possible to outline several Frames. Each Frame is evoked by a lexical unit (LU), a clause, a sentence... and is comprised of different agents and props, divided between mandatory Core Elements (CO) and optional Frame Elements (FE). For example, the verb “report” could evoke the “Request” Frame, in which four Core Elements (Addressee, Medium, Message, Speaker) can be identified, along with several other non-core Frame Elements. By adapting frames to EU normative legislation, it becomes possible to identify legal “trigger” terms and the Frame they evoke. By implication, the

different Frame Elements that are present can also be deducted and provide substantial semantic context to statutory texts. For instance, annotating syntactic units that refer to implicit elements makes them more easily interpretable by machines.

With this in mind, the adaptation of Frames can be done in three different ways: the modification of Core Elements, the addition of Frame Elements, or the creation of entirely new Frames to fit the needs of EU normative legislation. To identify which modifications have to be made, a thorough study will be lead on the state of the art and on the lexical particularities of the domain. This study can be done in combination with a (heuristic) manual annotation of a corpus made of short and consistent pieces of legislation, like EU reporting requests.

2.2 Scalability

This allows for a small-to-large scale strategy, with a first emphasis on reporting requests, later scaling up to obligations and other pieces of EU normative legislation. A request is a command-to-do some action, addressed to European institutions from the EU Commission, playing a given Role of addressee or addresser in a specific interval of time or with a frequency or in a given date [25]. Obligations have been defined in many ways, with some authors differentiating obligations of action and obligations of being [26], but they can be described as “that which one ought or ought not to do”, with “duty” and “right” as their correlative [27].

Reporting requests present a major interest for creating a first case-study corpus with a human-in-the-loop approach. They are fragmented (present in diverse pieces of EU legislation), usually short and more redundant than other pieces of legislation. For this reason, vocabulary and linguistic patterns are found in small numbers and thus easier to analyze and annotate manually. With a contained deontic landscape, it is also easier to focus on a small number of frames. This working base can be enriched and used to train and/or fine-tune a specialized LLM for later extraction (of terms, pieces of legislation or provisions) or annotation of similar documents. Frames can be populated by extracted lexical units for a more granular and accurate representation of legislative content.

In the field of legal computer science, there is a growing trend towards the creation of systems that combine subsymbolic (statistical) methods and symbolic ones, which has been called neuro-symbolic or Hybrid AI. In this context, the creation of a frame-based annotation for RR at first, and for other phenomena later, would create a synergy between the neural models and the annotation, since the models can be used to partially automate the annotation, which in turn can be used to inform neural models and LLMs about the frame semantics information associated with the text. Furthermore, this automatic annotation to enrich the input of LLMs lends itself well to the application of agent-based systems, where during drafting various agents annotate different phenomena with frames (RR, Obligations, definitions, etc) and then their output is used by LLMs for other tasks (eg question answering), perhaps comparing the annotations with ones in existing legislation.

Depending on the results of this project, larger and more diverse corpora could be investigated, and the resources obtained from these corpora could be applied to other hybrid-AI techniques or used in legal drafting, after evaluating the quality and reliability of the formalization. It should be deemed correct, comprehensible and transparent [28].

3 Conclusion

3.1 Contribution to the research field

Although many formalization methods have been developed in the legal domain (ontologies, logic layers), the novelty of a frame-based semantic model is that it is anchored in a lexical context and bridges the gap between the legal text itself and its conceptual digital representation. This approach is not new but it has been scarcely explored and no mature, full-covering FrameNet-based model has been developed and implemented for the legal domain.

3.2 Expected results

Creating a formal semantic method for modeling obligations and reporting requests is expected to enhance their computability and improve legislative workflows. In other words, it would help:

- Improve access and management of legal information, including versioning and consolidation;
- Allow for better retrieval of pertinent information and referencing;
- Facilitate compliance monitoring;
- Improve transparency and legal interpretation by disambiguating meanings;
- Facilitate cross-lingual workflows;
- And overall improve the efficiency of legal information management and drafting.

FrameNet is also compatible with existing modeling techniques and standards (Akoma Ntoso, RRMV, LLMs), with which it can be combined with the help of Framester. They can be mutually enriched to create a neuro-symbolic system that benefits from a rigorous FrameNet semantic layer and flexible LLM parsing, ideal for the legal domain. LLMs also offer extraction and annotation capabilities, implying scalability perspectives.

Greater interoperability is also expected to be achieved. As formalization and standardization are interconnected, the legal specialization of FrameNet frames can be used in combination with existing standards or help build new ones. For instance, incorporating FrameNet frames into AKN4EU could semantically enrich the structural XML-standard, improve norm extraction and consistency between european languages. Such harmonization would facilitate data exchange between EU institutions.

In short, a FrameNet-based formalization of EU legislation would offer different solutions for a time- and cost-effective management of legal documentation. It would also ensure a more deontically reliable encoding of EU legislation, which is one of the notable issues of the “Law as Code” paradigm.

References

1. Palmirani, Monica, et al. "Hybrid AI framework for legal analysis of the EU legislation corrigenda." *Legal knowledge and information systems*. IOS Press, 2021. 68-75.
2. Palmirani, Monica. "A smart legal order for the digital era: A hybrid ai and dialogic model." *Ragion pratica* 2 (2022): 633-655.
3. Corazza Michele, Palmirani Monica et al. "Reporting Requests Modelling in European Legislation with a Hybrid AI Approach." *ICAAIL*, 2025.
4. Palmirani, M. (2011). Legislative change management with Akoma Ntoso. In *Legislative XML for the semantic Web*, 101–130. Springer.
5. Vitali, F., M. Palmirani, R. Sperberg, and V. Parisse, et. Alt. (2018). Akoma Ntoso version 1.0. part 2: Specifications. OASIS standard, August 2018.
6. Vitali, F., M. Palmirani, and V. Parisse (2019). Akoma ntoso naming convention version 1.0. OASIS standard, February 2019.
7. Avgerinos Loutsaris, Michalis, et al. "Semantic Interoperability for Legal Information: Mapping the European Legislation Identifier (ELI) and Akoma Ntoso (AKN) Ontologies." *Proceedings of the 16th International Conference on Theory and Practice of Electronic Governance*. 2023.
8. Athan, Tara, et al. "LegalRuleML: Design principles and foundations." *Reasoning Web. Web Logic Rules: 11th International Summer School 2015, Berlin, Germany, July 31-August 4, 2015, Tutorial Lectures*. 11 (2015): 151-188.
9. Fillmore, Charles J., and Collin F. Baker. "Frame semantics for text understanding." *Proceedings of WordNet and Other Lexical Resources Workshop, NAACL*. 2001.
10. Hayes, Patrick J. "The logic of frames." *Readings in artificial intelligence*. Morgan Kaufmann, 1981. 451-458.
11. Reimer, Ulrich, and Udo Hahn. "A Formal Approach to the Semantics of a Frame Data Model." *IJCAI*. 1983.
12. Laviola, Adrieli, et al. "The Brazilian Portuguese Constructicon: Modeling Constructional Inheritance, Frame Evocation and Constraints in FrameNet Brasil." *AAAI Spring Symposia*. 2017.
13. Bertoldi, Anderson, and Rove Chishman. "Frame semantics and legal corpora annotation: Theoretical and applied challenges." *Linguistic Issues in Language Technology* 7 (2012).
14. Van Kralingen, R., Oskamp, E., Reurings, E. (1993). Norm frames in the representation of laws. *Legal Knowledge Based Systems: Jurix*, 93.
15. Isa Mara da Rosa Alves, Rove Luiza de Oliveira Chishman, Paulo Miguel Torres Duarte Quaresma, The Construction of a Juridical Ontology, in *Proceedings of ICAAIL'07 June 4-8, Palo Alto, CA USA*.
16. Dolbey A, Ellsworth M, Scheffczyk J (2006), BioFrameNet: A Domain-Specific FrameNet Extension with Links to Biomedical Ontologies, O. Bodenreider, ed., *Proceedings of KR-MED*, 87-94.
17. Ceci, Marcello, et al. "Semantic annotation of legal texts through a framenet-based approach." *AI Approaches to the Complexity of Legal Systems. Models and Ethical Challenges for Legal Systems, Legal Language and Legal Ontologies, Argumentation and Software Agents: International Workshop AICOL-III, Held as Part of the 25th IVR Congress, Frankfurt am Main, Germany, August 15-16, 2011. Revised Selected Papers 3*. Springer Berlin Heidelberg, 2012.
18. Venturi, Giulia, et al. "Towards a FrameNet resource for the legal domain." *Proceedings of the 3rd Workshop on Legal Ontologies and Artificial Intelligence Techniques: 2nd Workshop on Semantic Processing of Legal Text*. 2009.

19. Lazari, Antonio, and M^a Ángeles Zarco-Tejada. "JurWordNet and FrameNet Approaches to Meaning Representation: a Legal Case Study." *Semantic Processing of Legal Texts (SPLeT-2012) Workshop Programme*. 2012.
20. Palmirani, Monica, Salvatore Sapienza, and Kevin Ashley. "A hybrid artificial intelligence methodology for legal analysis." *BioLaw Journal-Rivista di BioDiritto* 3 (2024): 389-409.
21. Palmirani, Monica. "A smart legal order for the digital era: A hybrid ai and dialogic model." *Ragion pratica* 2 (2022): 633-655.
22. Wang, Tiexin, et al. "A joint FrameNet and element focusing Sentence-BERT method of sentence similarity computation." *Expert Systems with Applications* 200 (2022): 117084.
23. Ebers, Martin, and Marta Cantero Gamito. *Algorithmic Governance and Governance of Algorithms*. New York: Springer, 2021.
24. Witt, Alice, et al. "Encoding legislation: a methodology for enhancing technical validation, legal alignment and interdisciplinarity." *Artificial Intelligence and Law* 32.2 (2024): 293-324.
25. Corazza, Michele, et al. "Hybrid AI for Supporting the European Drafting Legislation." *JSAI International Symposium on Artificial Intelligence*. Springer, Singapore, 2025.
26. Von Wright., George H., 1963. *Norm and Action*. Routledge and Kegan Paul.
27. Hohfeld, Wesley Newcomb. Some fundamental legal conceptions as applied in judicial reasoning. *Yale Law Journal*, 23(1):16–59, Nov. 1913.
28. Libal, Tomer, and Tereza Novotná. "An Evaluation of Methodologies for Legal Formalization." *Explainable and Transparent AI and Multi-Agent Systems*. 2022.

Célestine LECAT

Vanves, FRANCE | celestine.lecat@studio.unibo.it | +33788001046

- **Languages:** French (mother tongue), English (C2), Spanish (C1), Italian (A2)
- **Softwares:** Microsoft Office, SDL Trados Studio, MultiTerm, UNTERM, eLUNa, Phrase, Aegisub, Wordfast Anywhere, Sketch Engine

EDUCATION

PhD in Law, Science and Technology <i>University of Bologna, Department of Law, CIRSFD, ALMA-AI</i> * Thesis: Formalizing EU Normative Legislation by leveraging FrameNet and frame semantics.	2025 - 2028 <i>Bologna, Italy</i>
Master's degree in Editorial, Economic and Technical Translation <i>ESIT School of Interpretation and Translation</i> * Thesis: Pregnancy and hereditary disorders of connective tissues: the case of Ehlers-Danlos syndromes.	2021 - 2024 <i>Paris, France</i>
Bachelor of Arts, Literature and Languages / English Major <i>University of Burgundy</i> * Bachelor's thesis: Translation theory: how and why are political and cultural meanings conveyed through translation?	2020 - 2021 <i>Dijon, France</i>
First common core year for Health Studies (PACES) <i>University of Burgundy</i> * Statistics, biomolecules, physics, chemistry, anatomy, humanities.	2019 - 2020 <i>Dijon, France</i>
Preparatory classes, Literature and Social Science (CPGE BL) <i>Lycée Carnot, University of Burgundy</i> * Double Degree of General Academic Studies (DEUG) in Modern Literature and Languages, and Arts, Literature and Languages.	2017 - 2019 <i>Dijon, France</i>
General Baccalaureate Degree in Sciences, with Honors <i>Lycée Sainte-Croix</i>	2017 <i>Provins, France</i>

WORK EXPERIENCE

WHO - Linguist and proofreader <i>World Health Organization</i> • Revision (FR, EN, SP) of the content and tagging of resolutions published by WHO's Governing Bodies converted to the AKN XML standard, as part of the AKN4UN project. • Regular reporting on issues linked to the conversion of documents, malfunctioning tags or commands in the XML editor (LIME). • Harmonization of tag names and attribute values in English and between languages (ENG, FRE, SPA).	June 2024 - Sept. 2025 <i>Full time remote consultancy</i>
WHO - Office of Language Services - Internship <i>World Health Organization</i> • Translation of press releases, web pages, subtitles, newsletters. Translation in tandem with a senior translator. • Creation and management of terminology records in the United Nations' term base, UNTERM. Compilation of technical documents for term extraction.	Dec. 2023 - May 2024 <i>Geneva, Switzerland</i>
UNESCO - Communications and Public Engagement - Internship <i>United Nations Educational, Scientific and Cultural Organization</i> • Translation and revision of press releases, media advisories, communication campaigns and web pages. Subtitling and editing of social media content. Live update of the UNESCO website during the World Heritage Committee.	July 2023 - Nov. 2023 <i>Paris, France</i>