

Knowledge Representation and Language Simplification of Human Rights

Sara Silecchia¹, Federica Vezzani¹, Giorgio Maria Di Nunzio²

¹ Department of Linguistic and Literary Studies, University of Padua, Italy
Via Elisabetta Vendramini, 13 - 35137 Padova - Italy

² Department of Information Engineering, University of Padua, Italy
Via Gradenigo, 6/a - 35131 Padova - Italy

sara.silecchia@studenti.unipd.it, [federica.vezzani, giorgiomaria.dinunzio]@unipd.it

Abstract: In this paper, we propose the description of a very recent interdisciplinary project aiming at analysing both the conceptual and linguistic dimensions of human rights terminology. This analysis will result in the form of a new knowledge-based multilingual terminological resource which is designed in order to meet the FAIR principles for Open Science and will serve, in the future, as a prototype for the development of a new software for the simplified rewriting of international legal texts relating to human rights, in order to facilitate their comprehension for non-expert people. Given the early stage of the project, we will focus on the description of its rationale, the planned workflow, and the theoretical approach which will be adopted to achieve the main goal of this ambitious research project.

Keywords: terminology, knowledge-based terminological resource, language simplification, Linked Open Data

1. Introduction

The current situation of worldwide conflicts has driven, once more, the attention to the problem of the interpretation of State responsibility for violations of Human Rights which has been debated for a long time in international law (Meron, T. 1989). This already thorny matter gets even more complicated given the ambivalent conception about the status of individuals in international law, since the traditional positivist doctrine considers States as the sole subjects of international law and individuals as the object (Salako, 2019). Nevertheless, this increasing involvement of individuals in international law is an interesting use case for the study of the language and terminology used to convey human rights granted under these circumstances. Given the specificity of its linguistic expression, we question whether the specialized language used to convey human rights is intelligible for legal laypeople.

In this context, legal specialized language has been extensively explored (see among others: G mar, 1980; 1990; Dechamps, 2013; Biel, 2009; Cornu, 2005; Koelsch, 2016). In fact, difficulties for non-expert people in understanding legal language, often referred to as “legalese” (Melinkoff, 1963), have been widely debated and analyzed over the years (see among others: Charrow *et al.*, 1979; Tiersma, 1993; Masson *et al.*, 1994), to the extent that the calls for the simplification of legal writing led to the promotion of a “plain language”. The most influential language-simplification efforts are attributed to the Plain English Movement in the US (Alterman, 1987; Benson, 1984; Benson *et al.*, 1987; Melinkoff, 1963; Wydick, 1978; 2005), which encouraged grammatical simplification of legal discourse due to the large number of impersonal utterances employed and the wide use of the passive voice (Richard, 2018), as well as its verbosity, complexities, and vagueness (Z di, 2019), demanding its plain rewriting.

Nonetheless, others (Stark, 1994; Assy, 2011; Z di, 2019) noted that the major simplicity and clarity of exposition claimed by the plain-legal-language movement fail to adequately represent the complexity of the law, as legal text comprehensibility seems not to be predominantly related to linguistic aspects. Z di (2019) emphasized that sometimes,

in legal drafting, clarity and accuracy can only be employed at “each other’s expense”, as legislative precision inevitably entails linguistic complexity.

Beyond any relevant theoretical position, this paper will focus on the description of a new research project aiming to provide a contribution in terms of clear representation and simplification of legal language, without any aim at demystifying law nor at diminishing the crucial role of lawyers as intermediaries between law and its subjects. The need to pursue such a study stem from laypeople’s difficulties in understanding legal language and terminology and lies in the very nature of these rights: ensuring the enforcement of human and citizens’ rights primarily requires their comprehension to be accessible for everyone through linguistic transparency.

The paper is organized as follows: in Section 2, we describe our research project specifically focusing on humanitarian rights terminology representation and simplification. We describe the main founding objectives as well as the planned workflow. In Section 3, we focus on the theoretical approach adopted in this research project for the conceptual and linguistic representation of human rights knowledge. Finally, Section 4 illustrates a first preliminary analysis conducted for the linguistic representation of the domain.

2. Research Project

The new interdisciplinary research project being discussed addresses the need to facilitate human rights comprehension for non-expert people, with the aim of proposing a methodology for the conceptual and multilingual linguistic representation of human rights and contributing to legal texts redrafting.

2.1 Objectives

The ultimate goal of this research project is the development of a new knowledge-based multilingual terminological resource, in which the data obtained will be structured based on the terminological record model

provided in the FAIRterm Web application.¹ This tool is designed to offer the users the possibility to structure Findable, Accessible, Interoperable and Reusable terminological data and metadata, by following the latest ISO TC/37 SC 3 standards for terminology management (Vezzani, 2021).

This domain-specific terminological resource will serve as the basis for the development of a software prototype for redrafting legal texts. To this end, it is worth specifying that we intend to operate a “formal” simplification, as defined by Causa (2001), following which utterances are formally redrafted and no intervention is made at the content level.

2.2 Workflow

Our research approach will be structured as follows:

- 1) Specialized documents (namely international and national legal acts on human rights, including immigration rules) will be collected in three working languages: French, English, and Italian. Starting from this corpus, we will elaborate both a conceptual and a multilingual linguistic representation of human rights terminology, following Costa and Santos’ mixed methodology for terminological knowledge representation (2015). The theoretical approach here adopted is hence based on the twofold nature of Terminology as consisting of a linguistic and a conceptual dimension. Subsequently, we will proceed with the identification of i) the concepts and their relationships; and then, for the three working languages, ii) the corresponding terms designating these concepts and their relationships, to assess whether a language-independent concept system can be overlapped with the multilingual lexical networks inferred from the purpose-built specialized corpus.
- 2) Once the double dimension of human rights terminology has been explored, from a multilingual perspective, the study will focus on the identification of terms, syntactic and grammatical structures related to specialized legal language and terminology that may hinder the comprehension of texts by legal laypeople’s due to their linguistic opacity or their highly specialized status. This stage will be followed by the compilation of terminological records on the FAIRterm Web Application.
- 3) Based on the linguistic phenomena examined during the compilation of terminological records (synonymy, polysemy, hyponymy, and hypernymy), respective plain terms and syntactic or grammatical reformulations will be proposed as an alternative to non-transparent linguistic elements in the source texts.
- 4) Finally, we will perform an analysis and implementation on how to include the official identifiers and vocabularies (such as the European Legislation Identifier and European Case Law Identifier, and the European EuroVoc thesaurus²) in the TermBase eXchange (TBX) (ISO 30042:2019)³ standard format of the terminological records.

2.3 A Linked Open Data “Open” Issue

An important aspect of this research proposal is the creation of a terminological resource which is reusable and interoperable. For these reasons, we will make use of both ISO standards of terminological databases (such as TBX) as well as a Linked Open Data (LOD) paradigm, in more specifically the Linguistic Linked Open Data⁴ paradigm, to publish data on the Web. In fact, LOD approaches give the researchers the possibility to design and implement open access tools to gather, study, and understand legal information. Despite having a wide variety of information publicly available online today, it is hard for a non-expert not only to understand the terminology and the language of laws (which is our primary focus) but also to cross-reference documents and the corresponding metadata. This problem can get even harder when legal documents from different jurisdictions are involved, such as legislative acts from the EU that influence national law, or in the case of cross-border cases. As discussed by Moodley et al. (2020), this gap between the legal and data proficiency that laypeople have can be the source for the development of software that is FAIR, publicly available, open-source, and easy to use by for anyone. In this sense, our research proposal for organizing and identifying legal terms according to the abovementioned ontologies, stems from the work of (Bacci et al. (2018), Linkoln; Filtz et al. (2021)) who propose an approach for the automatic extraction of legal references from legal texts and the enhancement of these data by means of legal knowledge graphs.

3. Theoretical Approach

In this section, we want to present the theoretical background of this work and the preliminary considerations about the linguistic representation of human rights.

3.1 Conceptual and Linguistic Dimension

The theoretical assumption underlying the proposed methodology lies in the dual nature of Terminology as composed of a conceptual and a linguistic dimension (Costa, 2013). Namely, Costa and Santos (2015) propose a methodology that combines both the onomasiological and the semasiological approach for terminological knowledge representation. This methodology is articulated in two stages: 1) the conceptual analysis of the Terminology of a domain, achieved without resorting to text analysis but by means of a domain concept map; 2) the linguistic analysis of the domain through the natural language processing tools, aiming at building a lexical network composed of terms and the relations to which they refer. In this context, the domain concept map would allow to eliminate ambiguity and adequately ensure coherence and consistency in domain representation for study purposes and validate the domain representation knowledge resulting from this structure.

Given the domain-independent nature of this dual approach, the expected objective is to apply this methodology of conceptual representation to the human rights domain.

¹ <http://purl.org/fairterm>

² <https://eur-lex.europa.eu/browse/eurovoc.html>

³ <https://www.iso.org/standard/62510.html>

⁴ <https://linguistic-lod.org>

3.2 Representation of Human Rights

As already mentioned, the knowledge representation of human rights does not serve solely for the purpose of terminological study, but the main objective being proposing a linguistic simplification of legal language, we believe that this type of approach would be the most appropriate to familiarise with the concepts concerned before carrying out any linguistic intervention for simplification purposes.

To perform Costa and Santos' analysis on the subject under study, and namely to identify the relevant concepts for study purposes, it is foreseen to consider mainly ontological relations as *part_of*, *connected_to*, *brings_about*, *occurs_in*, *carries_out*, *result_of*, *affects*, *process_of*, *uses*, or *exhibits*. The underlying objective is to make the relationships between legal concepts and their designations more explicit and to achieve greater clarity of exposition in legal texts primarily through conceptual clarity.

After the analysis of the concept system, a specialized corpus will be built by collecting specialized documents. The terminological extraction will then be carried out with the aim of retrieving the relevant terms based on their "termhood", that is the degree of detail to which a linguistic unit is related to specific concepts in a domain (Kageura and Umino, 1996). Subsequently, a map will be created by using the terms extracted from the corpus and directly related to the concepts in the map. At this stage, the analysis of the linguistic dimension is then performed through the identification markers, such as verbs, adverbs, or differentiation expressions, which introduce reformulations. As underlined by Costa *et al.* (2015), among others, lexical markers act as indicators of semantic relations (such as the cause/effect relation), and even punctuation is considered a linguistic marker. The assumption underlying this approach considers that for analysis purposes, not only terms are relevant but also other lexical units concur to build up the meaning of the discourse.

4. Preliminary analysis

In this section, we describe the initial analysis of the specialized language by performing the linguistic representation of the domain of human rights.

Being our final goal the creation of a multilingual resource, we decided to collect specialized documents in three working languages: English, French, and Italian. At this stage, we decided to have a parallel corpus consisting of all the international law treaties on human rights gathered from the official United Nations Human Rights Office of the Commissioner (UNHRC) archives⁵, in addition to some of the international law instruments on human rights retrieved from the Council of Europe archive⁶. All these documents focus on international treaties on human rights, which embrace a broader spectrum of rights, such as civil, social, economic, and political rights, equally considered inherent to all human beings.

In order to collect and process these documents, we used Sketch Engine⁷ (Kilgarriff *et al.*, 2014). The current corpus is composed of 110 documents and 459,851 *tokens*. The

keywords extraction function makes it possible to extract from the corpus the list of candidate terms of human rights domain, divided into single words, i.e. terms consisting of a single lexical unit, and multi-words, complex terms consisting of several units. The terminological extraction from the focus corpus is carried out through statistical calculations and analysis of the occurrences of candidate terms compared to the terminological data of a big pre-set reference corpus. Therefore, to make terminology extraction as selective and precise as possible, for each working language a specialized reference corpus has been selected. For the terminological extraction in Italian, the *EUR-Lex Italian 2/2016* reference corpus was chosen; in English two terminological extractions were made by combining the candidate terms extracted both from the *United Nations Parallel Corpus – English* and the *EUR-Lex English 2/2016*; whilst in French, the *United Nations Parallel Corpus – French* was selected, as the *EUR-Lex French 2/2016* was not available.

After this step, we performed a manual assessment of the extracted terms by means of the Concordance features in order to remove terms that are not relevant to the subject of study by looking at the context of occurrence and consequently to ascertain possible different connotations of terms within a particular context (for example, we removed multi-word terms like "concerning method of rehabilitation", "nouvelle convention portant revision" or "responsabilità delle persone").

A list of 790 terms for the three working languages are now under analysis for the subsequent description of the lexical networks of this domain.

One additional comment about this preliminary phase is the study of the polysemy and the different connotations of terms that have been observed during this preliminary analysis, Article 5 of the European Convention on Human Rights provides a clear example, in the authentic version in French and the respective Italian translation.

Article 5(c) in the French version⁸ of the Convention mentions: "s'il a été arrêté et détenu en vue d'être conduit devant l'autorité judiciaire compétente [...]", which has been literally translated in the Italian text⁹ of the Convention as "se è stato arrestato o detenuto per essere tradotto dinanzi all'autorità giudiziaria competente [...]". Whether the "expression *être conduit* devant l'autorité judiciaire" may be intuitive for French legal laypeople, the Italian translation "*essere tradotto* dinanzi all'autorità giudiziaria" is expected to appear harder. Indeed, in Italian, unlike the denotation of the term in standard language, indicating interlinguistic translation processes, in legal language the term "translation" means the transfer from one place to another of people under a regime of restriction of personal freedom. To adapt the linguistic expression of law to the needs of comprehension by a non-specialistic public, other alternative translations could be proposed, less opaque, but still consistent with the stylistic register of this domain, such as:

- "se è stato arrestato o detenuto per essere portato dinanzi all'autorità giudiziaria competente" ("if he has been arrested or detained to be brought before the competent judicial authority"); or

⁵ <https://www.ohchr.org/en/instruments-listings>

⁶ <https://www.coe.int/en/web/conventions/full-list>

⁷ <https://www.sketchengine.eu/>

⁸ https://www.echr.coe.int/documents/convention_fra.pdf

⁹ https://www.echr.coe.int/documents/convention_ita.pdf

- “se è stato arrestato o detenuto affinché compaia dinanzi all’autorità giudiziaria competente” (“if he has been arrested or detained to appear before the competent judicial authority”).

Indeed, these alternative translations perfectly correspond to the English version of the mentioned article, the latter being: “the lawful arrest or detention of a person effected for the purpose of *bringing him* before the competent legal authority [...]”. Given the nature of the international convention and given that its effects concern also non-expert citizens, we believe that the proposed alternative translations meet the need for clarity of exposition and effective communication between international (and national) institutions and citizens, who, despite their lack of specific knowledge, are directly concerned by national and international standards.

5. Conclusions

Ensuring the enforcement of human and citizens’ rights primarily requires their comprehension to be accessible for everyone through linguistic transparency. In order to achieve this objective, in this paper, we described the theoretical framework and the preliminary analysis of a research project that will 1) identify linguistic opacity related to legal specialized language that may hinder legal laypeople comprehension of international rules, 2) produce an open linguistic resource that follows the FAIR principles of open science. We believe that the proposed methodology and the design and implementation of the linguistic resource can effectively contribute to the improvement of human rights legislation understanding and drafting.

6. Acknowledgments

This work was partially supported by the ExaMode Project, as a part of the European Union Horizon 2020 Program under Grant 825292.

7. Bibliographical References

Alterman, I. (1987). *Plain and Accurate Style in Court Papers*. Philadelphia, PA: American Law Institute-American Bar Association Committee on Continuing Professional Education

Assy, R. (2011). Can the Law Speak Directly to Its Subjects? The Limitation of Plain Language. In *Journal of Law and Society*, Vol. 38, No. 3, pp. 376-404. URL: <https://ssrn.com/abstract=1906372>

Bacci, L., Agnoloni, T., Marchetti, C., Battistoni, R. (2018). Improving Public Access to Legislation through Legal Citations Detection: The Linkoln Project at the Italian Senate. *Law via the Internet*. 2018: 149-158. URL: <https://ebooks.iospress.nl/publication/51783>

Benson RW and Kessler JB. (1987). Legalese v. plain English: an empirical study of persuasion and credibility in appellate brief writing. In *Loyola of Los Angeles Law Review* 20, 301–321. URL: <https://digitalcommons.lmu.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1493&context=llr>

Benson, RW. (1984). The end of legalese: the game is over. *New York University Review of Law and Social Change* 13, 519–574. URL: https://socialchangenyu.com/wp-content/uploads/2019/09/ROBERT-W.-BENSON_RLSC_13.3.pdf

Biel, L. (2009). Corpus-Based Studies of Legal Language for Translation Purposes. In *Conference: Reconceptualizing LSP. Online proceedings of the XVII European LSP Symposium 2009*. Research Gate. URL: https://www.researchgate.net/publication/216576418_Corpus-Based_Studies_of_Legal_Language_for_Translation_Purposes

Causa, M. (2001). De la simplification en classe de français, langue professionnelle. In *Les Carnets du Cediscor*. URL : <http://journals.openedition.org/cediscor/302> ; DOI : <https://doi.org/10.4000/cediscor.302>

Charrow, RP and Charrow, VR. (1979). Making Legal Language Understandable: A Psycholinguistic Study of Jury Instructions. In *Columbia Law Review*, 79(7), 1306–1374. <https://doi.org/10.2307/1121842>

Conceição, Manuel Célio. 2005. *Concepts, termes et reformulations*. Lyon: Presses Universitaires de Lyon

Cornu, G. (2005). *Linguistique Juridique, Collection Domat / Droit privé, Montchrestien, 3eme édition*

Costa, R., (2013), “Terminology and Specialised Lexicography: two complementary domains” in *Lexicographica* 29 (1), 29-42, 2013. 39, 2013.

Costa, R. and Santos, C. (2015). Domain specificity: semasiological and onomasiological knowledge representation. In Kockaert H. and Steurs F. (Eds.), *Handbook of Terminology (Vol. 1, pp. 153-179)*. John Benjamins Publishing Company, DOI: <https://doi.org/10.075/hot.1.com1>

Dechamps, C. (2013). L’enseignement du français juridique en centre de langues : quelques perspectives. In *Recherche et pratiques pédagogiques en langues de spécialité [Online], Vol. XXXIV N° 1 | 2015*. URL: <http://journals.openedition.org/apliut/5094>; DOI: <https://doi.org/10.4000/apliut.5094>

Filtz, E., Kirrane, S., Polleres, A. (2021). The linked legal data landscape: linking legal data across different countries. In *Artif Intell Law* 29, 485–539 (2021). DOI: <https://doi.org/10.1007/s10506-021-09282-8>

Gémar, JC. (1980). La langue juridique, langue de spécialité au Québec: éléments de méthodologie. In *The French Review*, 53(6), 880–893. URL: <http://www.jstor.org/stable/391928>

Gémar, JC. (1990). Les fondements du langage du droit comme langue de spécialité. Du sens et de la forme du texte juridique. In *Revue générale de droit*, 21(4), 717–738. DOI : <https://doi.org/10.7202/1058214ar>
<https://ebooks.iospress.nl/publication/51783>

Gémar, JC. (2011). Aux sources de la “Jurilinguistique” : texte juridique, langues et cultures. In *Publications linguistiques : Revue française de linguistique appliquée*, 2011/1 Vol. XVI | pages 9 à 16, ISSN 1386-1204, URL : <https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2011-1-page-9.html>

Lavault-Olléon, E., Grossmann, F. (2008). Langue du droit et harmonisation terminologique multilingue : l’exemple de LexALP. *Lidil [En ligne]*, 38 URL : <http://journals.openedition.org/lidil/2776> ; DOI : <https://doi.org/10.4000/lidil.2776>

Kageura, K., Umino, B. (1996) Methods of automatic term recognition: A review. In *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, Volume 3, Issue 2, Jan 1996, p. 259 – 289.

- Kilgarriff, A., Baisa, V., Bušta, J. et al. (2014). The Sketch Engine: ten years on. *Lexicography*. In: *Asialex* 1, 7–36. DOI : <https://doi.org/10.1007/s40607-014-0009-9>
- Koelsch, G. (2016). Specialized Language. In *Requirements Writing for System Engineering*. Apress, Berkeley, CA. DOI: https://doi.org/10.1007/978-1-4842-2099-3_3
- Masson, M.E.J., Waldron, M.A. (1994). Comprehension of legal contracts by non-experts: Effectiveness of plain language redrafting. DOI: <https://doi.org/10.1002/acp.2350080107>
- Melinkoff, D. (1963). *The Language of the Law*. In *Boston: Little Brown*.
- Meron, T. “State Responsibility for Violations of Human Rights.” *Proceedings of the Annual Meeting (American Society of International Law)* 83 (1989): 372–85. <http://www.jstor.org/stable/25658498>.
- Moodley, K., Hernández Serrano, P., Zaveri, A., Schaper, M., Dumontier, M., Van Dijck, G. (2020). The Case for a Linked Data Research Engine for Legal Scholars. In *European Journal of Risk Regulation*, 11(1), 70-93. DOI: 10.1017/err.2019.51
- Norme ISO 30042:2019. Gestion des ressources terminologiques – TermBase eXchange (TBX)., <https://www.iso.org/fr/standard/62510.html>
- Richard, I. (2018). Is legal lexis a characteristic of legal language?”. In *Lexis*. URL: <http://journals.openedition.org/lexis/1173>; DOI: <https://doi.org/10.4000/lexis.1173>
- Salako, S. E. (2019). The Individual in International Law: ‘Object’ versus ‘Subject’. In: *International Law Research*; Vol. 8, No. 1; 2019. ISSN 1927-5234 E-ISSN 1927-5242. Published by Canadian Center of Science and Education
- Stark, J. (1994). Should the Main Goal of Statutory Drafting Be Accuracy or Clarity. In *Statute Law Review*, 15, 207–213. URL: <https://doi.org/10.5539/ilr.v8n1p132>
- Tiersma, P. M. (1993). Reforming the Language of Jury Instructions. In *Hofstra Law Review: Vol. 22: Iss. 1*, Article 2. URL: <http://scholarlycommons.law.hofstra.edu/hlr/vol22/iss1/2>
- Vezzani, F. (2021). La ressource FAIRterm : entre pratique pédagogique et professionnalisation en traduction spécialisée. In *Synergies Italie n° 17 – 2021*, p. 51-64, URL : <https://gerflint.fr/Base/Italie17/vezzani.pdf>
- Wydick, R. C. (1978). Plain English for lawyers. In *California Law Review* 66, 727–765.
- Wydick, R. C. (2005). *Plain English for Lawyers*, 5th edn. Durham, In NC: Carolina Academic Press.
- Zódi, Z. (2019). The limits of plain legal language: understanding the comprehensible style in law. In *International Journal of Law in Context*. 2019. 246–262. 10.1017/S1744552319000260.