Mapping Recitals to Normative Provisions in EU Legislation to Assist Legal Interpretation

Llio HUMPHREYS[a] , Cristiana SANTOS[b], Luigi di CARO[c], Guido BOELLA[c], Leon van der TORRE[a], and Livio ROBALDO[a]

[a] University of Luxembourg, Luxembourg

[b] Institute of Law and Technology, Universitat Autònoma de Barcelona, Spain

[c] Universitá di Torino, Italy

Abstract. This paper looks at the use of recitals in the interpretation of EU legislation, and mechanisms for connecting them to normative provisions. The purposive approach to the interpretation of EU legislation taken by the European Court of Justice makes frequent references to recitals as helping to establish the purpose of normative provisions. Our research uses a cosine similarity based approach to link articles with relevant provisions to help legal professionals and lay end-users interpret the law. Such support can be used in legal knowledge-based systems.

Keywords. Recitals, EU legislation, purposive interpretation, semantic similarity,

1. Introduction

This paper looks at the use of recitals in the interpretation of EU Directives and Regulations and mechanisms for connecting them to normative provisions. Recitals are located in the preamble – which contains everything between the title and the enacting terms of the act. The "enacting terms" are the legislative part of the act. The preamble, on the other hand, contains citations (to treaties that legitimize the legislation), solemn procedural or principle-based expressions, and a numbered list of recitals [1] containing objectives, references to other relevant legislations and occasionally definitions, but mainly consisting of a principle or justification, followed by a concise norm-like element. We use the phrase norm-*like* element, because the recitals do not have the normative status of the enacting terms. Which begs the question: what exactly are recitals and what are they for?

It is our contention that recitals are an essential component in legal interpretation and that the link between recitals and relevant normative provisions could be made explicit where possible. The identification of relevant recitals in the interpretation of normative provisions is a neglected aspect of legal informatics. If EU legal drafters might explicitly link normative provisions with recitals and make this information eventually available to the public, this would help render EU legislation more accessible and certain, even considering the political constraints and role assumed by recitals. Akoma Ntoso [2] (the XML standard adopted by The EU Parliament and the EU Commission and in the future also the Publications Office of the EU), seems to

contain annotation possibilities to map normative provisions to recitals, but unfortunately, its versions are currently only used internally, and are not available to the public. In the absence of any authoritative connections forthcoming from EU institutions, connections will be made by those who have to interpret the law (judges, lawyers, advisory bodies etc.).

Legal knowledge-based systems could support linking normative provisions to related recitals just as it can support linking to case law and legal doctrine so that this 'hidden' knowledge is also more widely available. However, as far as we know, there are no existing work on semi-automated mapping between normative provisions and recitals.

While the status of recitals has received some attention in legal circles (see section 2 below), as far as we know, there is no existing work on semi-automated mapping between normative provisions and recitals to support knowledge-based and document management systems. The research questions of this paper are thus:
- What is the relationship/interaction between recitals and normative provisions?
- Can we map recitals to normative provisions in a semi-automated way?

The paper is structured as follows. Section 2 provides background knowledge about the status of recitals in theory and practice, section 3 provides the methodology for manual and automated mapping between normative provisions and recitals, section 4 initial experiments and results, and section 5 conclusions and future work.

2. Background Knowledge

Recitals are compulsory in EU Directives (legislation to be implemented by member states as they see fit) and optional in EU Regulations (legislation that have direct application in Member States). Directives and Regulations implement EU treaties[1], also called primary EU law, which contain very few concrete rules and often general notions [4, p.13]. For example, the Treaty on the Functioning of the European Union (TFEU) (amended by the Lisbon Treaty in 2009) encompasses topics such as consumer protection, competition, tax, etc.). The EU air transport policy is provided in Article 100 of the TFEU. This has led to EU legislation on air transport policy (covering aspects such as airspace management, safety and security standards, passenger rights, environmental matters, etc.) such as EU Regulation 261/2004. Many recitals refer to specific articles in Treaties to justify the normative provisions; others explain the motivation behind the normative provisions or even provide a summary. This is why recitals are fundamental to legal reasoning in the EU.

Legal reasoning in all jurisdictions involves different methods of interpretation in order to ensure proper justice in applying the law to particular cases, particularly when laws are ambiguous. The ECJ favours the purposive approach to interpretation where an EU law provision is ambiguous or incomplete; *i.e.* it must be interpreted in light of the objectives it pursues and the courts should always seek to give effect to the legislative purpose/objective behind the law. In order to ascertain the purpose of the

---

[1] http://eur-lex.europa.eu/collection/eu-law/treaties.html. Two core functional treaties, the *Treaty on European Union* (originally signed in Maastricht in 1992) and the *Treaty on the Functioning of the European Union* (originally signed in Rome in 1958 as the Treaty establishing the European Economic Community), lay out how the EU operates.

legislation, the ECJ analyses the relevant recitals in the preamble (along with preparatory documents and legislative proposals).

For guidance on the purpose and usage of recitals, we consulted official sources related to legislative drafting [1, 5-8] and found the following complementary perspectives regarding what recitals should contain:

i. they should not contain normative provisions, nor legal bases, nor political exhortations; moreover, they should use non-mandatory language [1, clauses 10, 10.1];

ii. they should contain the motivation [6, clause 4.1.2 (a)] or "statement of reasons for the adoption of the act" in the following ways:

> "(a) a succinct statement of the relevant points of fact and of law;
>
> (b) the conclusion that it is therefore necessary or appropriate to adopt the measures set out in the enacting terms; and
>
> (c) the historical context of the act" [1, clause 10.3];

iii. they must relate to substantial provisions [6, clause 4.1.2 (b)], and the order should correspond as far as possible to that of the provisions to each they relate [1, clause 10.3].

According to the Guide [1, clause 10], the recitals' *purpose* is "(...) to set out concise reasons for the chief provisions of the enacting terms, without reproducing nor paraphrasing them". Moreover, recitals "are of particular importance in order for the ECJ to assess whether the Community legislator has not made manifest errors in areas where it enjoys a margin of appraisal"[2] [6, clause 4.1.4 (c)]. Indeed, "if the reasoning set out in an act for which a statement of the reasons is compulsory is wanting or is not sufficient to fulfill the requirements (...), the Court can annul the regulation for breach of essential procedural requirements" [6, clause 4.1.4 (a)].

Concerning the *legal importance of recitals* [6, clause 4.1.4], it is a matter of contention whether recitals have legal repercussions or legal effect on normative provisions. There are different doctrinal positions [9]:

i) recitals have no effect;
ii) recitals are dominant over normative provisions;
iii) recitals have an equal position in relation to normative provisions;
iv) recitals encompass a subordinate position towards normative provisions.

We are cognizant that the ECJ has assumed both positions 3 and 4 in its judicature. Supporting position 3, the ECJ has stated that recitals are used to "*interpret the enabling provision of an act*" (6, clause 4.1.4(b)], and that recitals are "*necessary for courts to perform supervision*"[3]. It is worth citing the following doctrinal interpretation: "(...) *the law of recitals in EC Legislation can be summarized thusly: A) Where both the recitals and the operative [normative] provisions are clear but inconsistent, the operative provision will control. Corollary: recitals have no positive operation of their own. B) Where the recital is clear, it will control an ambiguous operative provision. This means that the operative provision will be interpreted in light of the recital. There have been cases wherein the nature of the operative provision is affected by a recital, and others where the scope of the operative provision is affected(...)*"[9].

---

[2] Recital 9 of the Data Protection Directive provides an illustrative example of this margin of appraisal "(...) whereas Member States will be left a margin for manoeuvre, which may, in the context of implementation of the Directive, also be exercised by the business and social partners(...)".

[3] Case 24/62, F.R.G. v. Comm'n of the Eur. Econ. Cmty., 1963 E.C.R., paragraph 18.

Substantiating position 4, the ECJ has ruled that the recitals "*cannot be relied on as a ground for derogating provisions of the act*"[4]. It is moreover stated that "*if a recital is irredeemably inconsistent with the operative text, then the ECJ will ignore the recital and give effect to the text of the operative provisions*"[10]. Recitals can be used to interpret only provisions which are ambiguous[5], but "*they cannot, however, restrict an unambiguous provision's scope*"[9, p.3], i.e., "*the terms of a recital cannot be used to give a particular construction to a provision which the terms of that provision would not otherwise bear*"[6].

In practice, we find that recitals are used exactly accordingly to the latter. For instance, article 5 of Regulation EC 261/2004, headed "Cancellation", provides that an operating air carrier shall not be obliged to pay compensation if it can prove that the cancellation is caused by extraordinary circumstances which could not have been avoided even if all reasonable measures had been taken. The term "extraordinary circumstances" is not defined in any of the articles of the Regulation (not even in Article 2 which is devoted to "definitions"). However, recitals 14 and 15 of the Regulation give a few examples, by way of illustration, of events which may be regarded as extraordinary circumstances, namely cases of political instability, meteorological conditions incompatible with the operation of the flight concerned, security risks, unexpected flight safety shortcomings and strikes which affect the operation of an operating air carrier. These cases have been used by the Court to determine to which extent the air carrier is exempted from paying compensation[7].

3. Mapping Normative Provisions and Recitals

Manual analysis of recitals in three EU legislative domains[8] concerning air transport, copyright and data protection revealed that many of the analysed recitals have two parts: a reason/justification, followed by a concise norm-*like* element. Other recitals contain definitions, objectives, references to other relevant legislation, etc. Most recitals could be mapped to one or more articles in the main body of legislation that articulate the norm-like element in greater detail. Such recitals contained the same words as the corresponding recitals, and this is the basis for our choice of cosine similarity as the algorithm for automated mapping. Manual analysis of normative provisions revealed that there were certain articles for which the recitals did not provide any insight, such as definitions (located in the main body of the act) and the

---

[4] Case C-162/97, Nilsson et al, paragraph 54, 1998, E.C.R. I-07477; and Case C-344/04, IATA, ELFAA v Department for Transport, § 76 (specifically addressed to air transport passenger domain).

[5] Case C-244/95, P. Moskof AE v. Ethnikos Organismos Kapnou, 1997 E.C.R. I-06441.

[6] Case C-412/93, Société d'Importation Edouard Leclerc-Siplec v TF1 Publicité SA and M6 Publicité SA.

[7] Case C-549/07 Wallentin-Hermann (2012) ECR I-11061, Paragraphs 16, 18 and 20,21, 22; and C-294/10 Eglitis and Ratnieks C-294/10 (2011) , [6, p.17].

[8] In our case-study we used legislation that are very well-known and highly discussed at the European level: i) Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, Official Journal L 167, 22/06/2001; ii) Regulation 261/2004/EC of the European Parliament and of the Council of 11 February 2004 establishing common rules on compensation and assistance to passengers in the event of denied boarding and of cancellation or long delay of flights, OJ L 46 of 17.2.2004; iii) Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OL 281 , 23/11/1995.

procedural articles (situated at the end thereof); thus, these articles were removed from the corpus.

A gold standard mapping between articles in the normative provisions to recitals in the preamble in the three legislation was prepared by a researcher with in-depth knowledge of the three legislation. The connection between articles and recitals is not always explicit as a textual reference. Therefore the mappings were based as much as possible on the pronouncement of connections in authoritative sources from the ECJ case law and also soft law [12-13] in this domain, although some mappings were also based on the researcher's own observations of textual similarity. The mappings were then checked for consistency by another researcher. From the analysed recitals, it was found that most of the time, it was useful to map between whole articles and whole recitals, although on occasion useful mappings could also be made between recitals to sub-articles or even sub-sub-articles. Such mappings were not included in the preliminary experiments in automated mapping below, but will be the subject of future work. All three legislative texts presented challenges for manual mapping, specially due to the fact that many general recitals could not sensibly be mapped to any specific articles.

We conducted experiments on mapping (automatically) normative provisions to recitals. Each recital item contained all the text without its index number. As with the recitals, all index numbers were removed from the normative provision items. Each normative provision item contained the text of whole articles with lists transformed into proper sentences. Substantive titles were included as if they were normal sentences, non-informative structural titles such as 'Section II' and 'Article 1' were removed. Such terms were included when used as references within the sentences of the articles.

For our experiments, we used the Cosine Similarity algorithm with Term Frequency – Inverse Document Frequency (tf-idf) [15,16]. Each normative provision and recital was presented as a vector of terms, and the Cosine Similarity between two vectors were quantified as the cosine of the angle between the two vectors [14]. Each term in the vector was weighted using tf–idf, a measure designed to evaluate the importance of each term in the vector, offsetting the frequency of a term in the vector with its frequency in the corpus as a whole. We then observed from the gold standard that many recitals that correspond to articles use the words of the substantive title in the recital text. We thus sought to give greater weight to terms appearing in substantive titles in determining similarity. However, the title words "scope" and "objective" were not given extra weight as they are effectively metadata rather than substantive terms. Finally, we sought to improve the performance of our mapping tool, using the Stanford part-of-speech tagger [16], by restricting the vector terms to those having what are typically considered to be the most informative part-of-speech – nouns, verbs and adverbs.

4. Preliminary Experiments and Results

Table 1 shows the results of our first experiment: Cosine Similarity with tf-idf on all words in their surface forms. In the following tables, CD means the copyright directive, DPD means the data protection directive, ATPR means air traffic regulation, CT means cosine threshold, TP means true positive, TN true negative, FP false positive, FN false negative and TNR True Negative Rate. We can see that the accuracy level is very high,

ranging from 83% to 94%. However, we acknowledge that accuracy is not a fair way to measure the quality of a system in case of unbalanced datasets (i.e., datasets that have very different numbers of "positives" compared to "negatives". Other classic measures of performance for similarity are precision, recall, and the F-measure, which seeks a balance between the two.

Table 1. Baseline: mapping of three legislation with different thresholds

	CD with CT 0.1	CD with CT 0.16	DPD with CT 0.1	DPD with CT 0.16	ATPR with CT 0.1	ATPR with CT 0.16
TP	24	19	48	36	15	13
TN	452	499	2134	2273	295	329
FP	66	19	182	43	58	24
FN	7	12	12	24	7	9
Accuracy	0.87	0.94	0.92	0.97	0.83	0.91
Precision	0.26	0.50	0.21	0.46	0.21	0.35
Recall	0.77	0.61	0.80	0.60	0.68	0.59
F-measure	0.40	0.55	0.33	0.52	0.32	0.44
TNR	0.87	0.96	0.92	0.98	0.84	0.93

We experimented with different threshold levels in order to ascertain whether the true positives have a higher similarity than the false positives, so that the threshold can be adjusted without compromising recall. However, from our analysis, we found that this was not the case with our data. Indeed, this may lie in the nature of the relationship between precision and recall in general. Precision and Recall creates a curve (named ROC curve) which is a way to evaluate them considering all the similarity threshold. There is a threshold point in this curve that maximizes the F-measure, and it is sometimes the best combination of precision and recall one can have. However, we would contend that the classic F-measure is also a poor measure for evaluating performance in our case. Precision is the important measure for systems that require few and precise information without any manual analysis whereas recall is arguably more important than precision for our purposes - to support legal knowledge engineers. For this, it is more important to identify as many of the relevant connections as possible, even within a noisy set of possibilities, rather than to identify very few precise connections.

It should be noted that one drawback of precision and recall is that they avoid evaluating the ability to identify "negatives" since they are calculated using only True Positives, False Positives and False Negatives (and not True Negatives). It is therefore useful also to consider the True Negative Rate (TNR). For instance, the baseline measure with a threshold of 0.1 on the copyright directive is able to identify around 24 positive connections, but within a quite large set of 66 false positives. "Large", however, is not as large as the entire set of connections. This means in the case of the copyright directive that the knowledge engineer has to manually go over 24+66=90 connections instead of analyzing the entire set of 549 connections, which works out at having to check 10 recitals per article, instead of 61.

We then experimented (Table 2) with giving extra weight to the title tokens (with a multiplication factor of 3), based on our observation that relevant recitals often use these terms. In two out of three legislation, the weighting of the title tokens produced some improvement in terms of precision and F-measure. However, the weighting for the titles was arbitrary, and more experiments are required in this vein. We wish to follow up on this experiment with strategies of automatically detecting the structural

parts of the texts that need to be "boosted". For instance, the first sentences of the texts may also have a greater weight.

Table 2. Mapping with extra weigh for title terms

	CD with CT 0.1	CD with CT 0.1 with extra weighting given to title terms	DPD with CT 0.1	DPD with CT 0.1 with extra weighting given to title terms	ATPR with CT 0.1	ATPR with CT 0.1 with extra weighting given to title terms
TP	24	21	48	46	15	13
TN	452	449	2134	2151	295	317
FP	66	69	182	165	58	36
FN	7	10	12	14	7	9
Accuracy	0.87	0.86	0.92	0.92	0.83	0.88
Precision	0.26	0.23	0.21	0.22	0.21	0.27
Recall	0.77	0.68	0.80	0.77	0.68	0.59
F-measure	0.40	0.35	0.33	0.34	0.32	0.37
TNR	0.87	0.87	0.92	0.93	0.84	0.90

Our final experiment was to lemmatize words and remove from the vectors terms having what are generally considered to be non-informative parts of speech, using the Stanford POS parser [16]. The first experiment was to remove all tokens apart from the nouns, and then lemmatize those nouns. The results are not indicated here, but were the poorest, as the system lost some important features. There are important facts to consider behind the concept of Cosine Similarity. It follows a specific curve which depends on the number of features of the vectors. In general, the larger the number of features, the more the sensitivity of Cosine Similarity. By using the lemmas of all nouns and verbs and adjectives the system performs a little better. However, the best results of all was with the lemmatization of all terms whatever their part of speech.

Table 3. Mapping with lemmatization and filtering on parts of speech

	CD with CT 0.1 with lemmas of all the POS	CD with CT 0.1 with lemmas of nouns, adjectives and verbs	DPD with CT 0.1 with lemmas of all the POS	DPD with CT 0.1 with lemmas of nouns, adjectives and verbs	ATPR with CT 0.1 with lemmas of all the POS	ATPR with CT 0.1 with lemmas of nouns, adjectives and verbs
TP	27	21	49	51	14	15
TN	426	423	2057	2054	301	295
FP	92	95	259	262	52	58
FN	4	10	11	9	8	7
Accuracy	0.83	0.81	0.89	0.89	0.84	0.83
Precision	0.23	0.18	0.16	0.16	0.21	0.21
Recall	0.87	0.68	0.82	0.85	0.64	0.68
F-measure	0.36	0.29	0.27	0.27	0.32	0.31
TNR	0.83	0.82	0.89	0.89	0.85	0.84

In summary, our experiments supported our hypothesis that in the studied texts, there is often textual similarity between normative provisions and related recitals, such that automated similarity methods can be effective. The performance of the system depends on which evaluation metric is used, but with a bias towards recall, we reach good initial results, which in practical terms means that the knowledge engineer can be presented with almost all the plausible connections without having to cross-check all possible connections.

5. Conclusions and Future Work

This paper has explored the relationship between recitals and normative provisions in legal reasoning in EU legislation, based on official guidelines and investigative research into actual practices by the ECJ. Our research reveals that consulting relevant recitals is an essential part of legal reasoning in practice. We also conducted experiments on mapping normative provisions to recitals. A gold-standard mapping was created based on mappings encountered in legal sources (case-law, soft law) as well as textual similarities. An experiment into automated mapping based on textual similarity alone revealed that this method can provide a valuable tool to support the legal knowledge engineer, since the recall of relevant recitals is very high. Nevertheless, a significant element of manual verification is required to remove invalid suggested mappings.

Our future work will involve improving the mapping system by studying strategies of automatically detecting the parts of the texts that need to be "boosted". Blind cosine similarity as a classification tool can give quite recall-based results instead of precision-based results due to its inability to understand that two words (two features in the vectors) should be considered differently, particularly with textual data. By running Machine Learning experiments, we can achieve better precision by building classifiers that are able to estimate the value of a feature class, so that we can understand which words need to have a higher weight in order to fit with the training data. Moreover, "superior" similarity measures may also consider WordNet [17], synsets [18] and ontologies [19]. Further to our experiments giving greater weight to titles, we also wish to investigate making good use of structural data to help determine similarity e.g. the first sentences of the texts of normative provisions may have a greater weight than the rest. We would also like to investigate more fine-grained mappings, on sub-article levels, and strategies (e.g. length, lexical changes) for determining when finer granularity is required. Finally, we wish to assess the applicability of the developed techniques for mapping normative provisions to recitals to a similar but perhaps more far-reaching task: mapping normative provisions in EU Directives to implementing legislation in national jurisdictions. This would make it easier to compare how different jurisdictions implement individual norms derived from EU law.

References

1. Joint Practical Guide of the European Parliament, the Council and the Commission for persons involved in the drafting of European Union legislation, December 2014, http://eur-lex.europa.eu/content/techleg/KB0213228ENN.pdf
2. M. Palmirani. Legislative change management with akoma-ntoso. In G. Sartor, M. Palmirani, E. Francesconi, and M.A. Biasiotti, editors, Legislative XML for the Semantic Web,volume 4 of Law, Governance and Technology Series, pages 101–130. Springer, 2011.
3. Boella, G., Di Caro, L., Graziadei, M., Cupi, L., Salaroglio, C. E., Humphreys, L., ... & Stroetmann, V. (2015, June). Linking legal open data: breaking the accessibility and language barrier in European legislation and case law. In Proceedings of the 15th International Conference on Artificial Intelligence and Law (pp. 171-175). ACM.
4. Lenaerts, Koen; Gutiérrez-Fons,José A. To say what the law of the EU is : methods of interpretation and the European Court of Justice, EUI AEL,2013/09, 2013.
5. Interinstitutional style guide, Publications Office, available at http://publications.europa.eu/code/en/en-120200.htm
6. Manual of Precedents for Acts Established Within the Council of the European Union, Office of Official Publications of the European Communities, 2002.
7. Commission Manual on Legislative Drafting, available at http://ec.europa.eu/smart-regulation/better_regulation/documents/legis_draft_comm_en.pdf
8. Interinstitutional Agreement on common guidelines for the quality of drafting of Community legislation of 22 December 1998 (OJ C 73, 17.3.1999).
9. Klimas, Tadas and Vaiciukaite, Jurate, The Law of Recitals in European Community Legislation (July 14, 2008). ILSA Journal of International & Comparative Law, Vol. 15, 2008, p. 33
10. Slaughter and May, Introduction to the legislative processes for European Union directives and regulations on financial services matters, April, 2014, available at http://www.slaughterandmay.com/media/1934583/introduction-to-the-legislative-processes-for-european-union-directives-and-regulations-on-financial-services-matters.pdf
11. Guide to the Copyright and Related Rights Treaties Administered by WIPO and Glossary of Copyright and Related Rights Terms, 2003, available at http://www.wipo.int/freepublications/en/archive.jsp?cat=copyright
12. Handbook on European data protection law, European Union Agency for Fundamental Rights, Publications Office of the European Union, 2014 available at http://fra.europa.eu/sites/default/files/fra-2014-handbook-data-protection-law-2nd-ed_en.pdf
13. Rossi Dal Pozzo, F., EU Legal Framework for Safeguarding Air Passenger Rights, Springer, 2015.
14. Sparck-Jones, K. 1972. A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation 28(1):11–21.
15. Papineni, K. 2001. Why inverse document frequency? In Proceedings of the North American Chapter of the Association for Computational Linguistics, 25–32.
16. Danqi Chen and Christopher D Manning. 2014. A Fast and Accurate Dependency Parser using Neural Networks. Proceedings of EMNLP 2014.
17. Seco, N., Veale, T., & Hayes, J. (2004, August). An intrinsic information content metric for semantic similarity in WordNet. In ECAI (Vol. 16, p. 1089).
18. Sánchez, David, et al. "Ontology-based semantic similarity: A new feature-based approach." Expert Systems with Applications 39.9 (2012): 7718-7728 Sánchez, D., Batet, M., Isern, D., & Valls, A. (2012). Ontology-based semantic similarity: A new feature-based approach. Expert Systems with Applications, 39(9), 7718-7728.
19. Landauer, T.K. & Dumais, S.T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge", Psychological Review, 104.].