THE DISCOURSE OF BIOETHICS IN THE ECtHR CASE-LAW
A corpus-based perspective

JEKATERINA NIKITINA
UNIVERSITY OF MILAN

Abstract – This corpus-based study investigates linguistic practices and strategies of dealing with issues of bioethics in the case-law of the European Court of Human Rights ("ECtHR"). The study aims at researching the linguistic strategies of knowledge production and entextualisation in 92 judgments of the ECtHR, focusing on the interaction between legal and bioethical term-related phraseological units – i.e. multi-word terms and term-embedding collocations with a verb – their structure and distributional patterns. Recurrent phraseological units are identified and analysed using methods of corpus linguistics and the theoretical framework of specialised phraseology. The study pursues general descriptive goals and aims at researching the balance and intersection between bioethically charged phraseology and legal phraseology. The main focus is placed on the analysis of typicality of patterning, expressed in terms of domain specificity, association score and log-likelihood.

Keywords: bioethics; ECtHR judgments; specialised phraseology; multi-word terms; term-embedding collocations.

1. Bioethics and the ECtHR

The concept of bioethics eludes clear-cut definitions as this highly specialised field, standing at the crossroads between biomedical, ethical, legal and technological areas, is constantly changing along with the developments in the modern world. Most definitions of bioethics stress its interdisciplinary and changing nature. For instance, Article 1 of the UN Declaration on Bioethics and Human Rights (2005) defines bioethics as "ethical issues related to medicine, life sciences and associated technologies as applied to human beings, taking into account their social, legal and environmental dimensions". Similarly, the Research report on Bioethics of the Council of Europe defines it as "protection of the human being (his/her human rights and in particular human dignity) in the context of the development of biomedical sciences" (2016, p. 4).

Although there are multiple legal instruments dealing with specific issues related to the bioethical field – such as surrogacy laws, assisted suicide acts, etc. – there are only two all-encompassing international legal documents dedicated to the field in general. These are *Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine* ("the Oviedo Convention", 1997) and the *Universal Declaration on Bioethics and Human Rights* (2005).

The European Court of Human Rights ("ECtHR") is the only court which may interpret the Oviedo Convention (Oviedo Convention, art. 29) and which guarantees judicial protection of bioethically relevant rights of a single individual on the supranational level. Its case-law exerts a significant influence on the interpretation and knowledge dissemination of many controversial concepts of bioethics, which are subdivided in several categories in the Court's case-law: physician assisted suicide, consent to medical examination or treatment, ethical issues concerning HIV, retention of biodata (including DNA and fingerprint samples), medically assisted procreation, reproductive rights (including prenatal diagnosis and the right to a legal abortion), liability of health professionals, right to know one's biological identity, transgender issues and rights of children born out of surrogacy agreements (Research report on Bioethics of the Council of Europe 2016, p. 4).

The ECtHR rules on violations of human rights coming from one or several of the 47 Member States of the Council of Europe ("CoE"), with 37 different languages, making it a fertile ground for cross-linguistic and cross-systemic contacts. At the same time, the Court – represented by 47 judges elected in respect of the CoE Member States sitting in different judicial formations – operates with only two official languages, English and French (Rule 34, Rules of Court). Consequently, from the linguistic point of view the case-law of the ECtHR cannot but present features of translation and/or second language production (Nikitina 2018).

In addition to linguistic, translational and legal challenges, the ECtHR judgments dealing with bioethical issues represent a locus of interdisciplinary contact, or to quote Myers "a terrain of competing discourses and practices" (2003, p. 267). The ECtHR judges have to overcome the difficulty of operating with different kinds of knowledge, inserting scientifically charged notions in a rather crystallised structure of a judgment embedded in a legal context. Consequently, these judgments become a hybrid form of knowledge production. Calsamiglia and van Dijk (2004, p. 373) claim that "each kind of knowledge […] has its own format". It could be hypothesised that bioethical knowledge, in light of its interdisciplinary nature, is prone to demonstrate

mixed formats incorporated from a range of perspectives and fields.

This paper builds upon the theoretical notion of *entextualisation*, originally elaborated by anthropologists and used by linguists to address the issue of knowledge dissemination and popularisation. Garzone (2014, p. 79) defines entextualisation as "the extraction of meaning from one discourse and consequent insertion of that meaning into another discourse through a process of de-contextualization or 'decentering' and its 're-centering' in another context". This paper looks at how legal and bioethical discourses interact at the lexico-grammatical level, attempting to trace a relationship between the language component (language as communication) and language as vehicle for conveying disciplinary knowledge (Bhatia 2014 [2004], p. 204) through specialised phraseological patterns that revolve around terms and act as "depositories of knowledge" (Sager 1998, p. 259). The basic presumption is that bioethical patterns could present certain traits of legal contamination on account of the already existing legal dimension of bioethics and the immediate legal context of the ECtHR judgment, where these are placed.

2. Specialised phraseology and corpus linguistics

The topic of languages for special or specific purposes (LSP) is widely discussed in the relevant literature, starting from a lexicalised perception of LSPs in 1970s-1980s (e.g. Widdowson 1979, p. 24) to a more recent reanalysis of LSPs that covers also such aspects as specialised phraseology. While LSP phraseology has gained popularity in academic circles in recent years, it is still placed at the periphery of phraseological studies (Kjær 2007, p. 506). In this paper specialised phraseology is understood as prefabricated lexico-grammatical patterns that contribute to the make-up of bioethically-charged ECtHR judgments "at the macrostructural and microstructural level, providing a stable matrix to be filled with details" (Biel 2014a, p. 36).

In general, phraseology interweaves aspects of four disciplines: semantics, morphology, syntax and discourse (Granger, Paquot 2008, p. 30) and represents a dynamic field in constant development. The combination of the phraseological paradigm with the methodology of corpus linguistics is particularly appropriate for the analysis of legal discourse on bioethics – another dynamic field – because corpus linguistics enlarges the research focus "beyond the single word as the basic semantic unit" (Teubert 2002, p. 212). Phraseological analysis of specialised corpora contributes to understanding "the kinds of language data which particular communities of users might encounter and which will inform their use" (Hyland 2008, p. 8). In line with recent research orientations, this study aims at researching the specialised discourse of bioethics in the ECtHR case-law through the paradigm of specialised phraseology.

Lingue e Linguaggi

It has been acknowledged that phraseological units in legal texts tend to cluster around terms, forming a continuum with fuzzy borders (Scarpa *et al.* 2014, p. 75), and terms are usually those elements that could denote a specificity to a certain specialised domain. The interaction between terms and phrases thus is the key to understanding a discipline. Hence, I decided to focus this research on the so-called *term-related phraseological units* (Nikitina 2018, p. 315), i.e. those phraseological units that feature a term. The primary focus of this paper is placed on a special category of multi-word units, which could be classified both as terminological and phraseological. These multi-word units are known in the literature under a variety of labels, such as *terminological phrase* (Kjær 1990), *terminological phraseme* (Meyer, Mackintosh 1994), *multi-word terminological phrase* (Bergenholtz, Tarp 1995), *multi-word term* (Kjær 2007, p. 509) or *term-forming pattern* (Biel 2014b, p. 180), i.e. those terms that consist of more than one word. It is presumed that multi-word terms have the potential to carry highly domain-specific meanings, as the combination of various single words increases the specificity of knowledge charge. In this regard, I draw on Sinclair's interpretation of the Firthian concept of contextual meaning, according to which words enter "into meaningful relations with other words around them" (Sinclair 2004, p. 25) and do not remain "perpetually independent in their patterning" (Sinclair 2004, p. 30).

The secondary focus is placed on the immediate verbal phraseological environment of terms, the so-called *term-embedding collocations* (Biel 2014b, p. 180) or *collocations with a term* (Kjær 2007, p. 509), which are generally organised around the prototypical structure $[N_{term}+V]$ or in the reverse order $[V+N_{term}]$, and denote actions that are possible to undertake with the base noun, or in this case, with the nominal multi-word term. In other words, collocations with a verb encase the knowledge-carrying multi-word terms in a script, integrating the general domain-specific picture.

3. Corpus and methodology

This study is developed on the basis of a corpus of 92 judgments and decisions of the ECtHR in the English language, delivered on issues of bioethics, for the total of 1,130,548 words.

The corpus is subdivided into 10 thematic areas of interest ("subcorpora" or "micro-corpora", see Table 1), based on the categorisation adopted by the Council of Europe and the ECtHR (Research Report on Bioethics of the Council of Europe 2016, p. 4): physician-assisted suicide, consent to medical examination or treatment, ethical issues concerning HIV, retention of biodata (including DNA and fingerprint samples), medically assisted procreation, reproductive rights (including prenatal diagnosis and the

right to a legal abortion), liability of health professionals, right to know one's biological identity, transgender issues and rights of children born out of surrogacy agreements.

Corpus	Texts	Words	Types
Physician-assisted suicide	10	112,822	5,675
Consent to medical examination or treatment	10	115,320	5,914
Ethical issues concerning HIV	13	162,452	7,595
DNA samples and retention of biodata	9	90,347	4,979
Medically assisted procreation	6	98,545	5,466
Reproductive rights	10	186,074	7,059
Liability of health professionals	6	67,960	4,608
Right to know one's biological identity	11	91,410	4,014
Transgender issues	14	156,736	6,142
Children born out of surrogacy	3	48,882	3,379
Total:	**92**	**1,130,548**	**17,118**

Table 1
Composition of the ECtHR case-law corpus.

The judgments and decisions included in the corpus were collected by the thematic criterion of bioethical relevance and by the linguistic criterion, as only texts in English were included. An important point should be made concerning the time span. The texts under analysis were produced in a period ranging from the 1990s (early transgender and HIV cases) until 2017. For the sake of categorisation, I roughly subdivide this time span into three periods: 1990-2000, 2001-2009, 2010-2017. The temporal placement of bioethically relevant ECtHR case-law has provided evidence of the increasing importance of this topic in the practice of this supranational court.
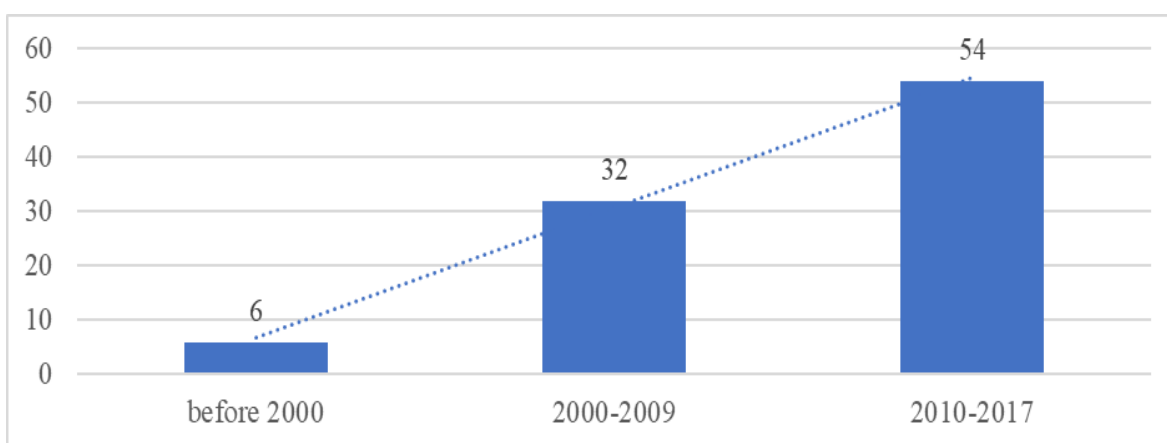

Figure 1
Number of bioethically relevant judgments and decisions in the ECtHR case-law.

As Figure 1 demonstrates, over the years the number of judgments and decisions concerning bioethical issues has significantly increased, with a clear upgoing trend. As of now, there is a substantial number of communicated and pending cases concerning such issues as DNA sampling, and "we can perhaps expect more applications concerning subjects such as gene therapy, stem cell research and cloning in the future" (Research Report on Bioethics of the Council of Europe 2016, p. 4).

The general methodological approach is quantiqualitative. I use corpus methodology to identify specialised phraseological units and further qualitatively analyse how specialised meanings are expressed and encoded through these patterns. In particular, this study looks at how bioethical patterns intertwine with the legal matrix of judgments and how this interaction creates the complex fabric of legal discourse on bioethics.

The research adopts *TermoStat 3.0* (Drouin 2012) software for multi-word candidate term extraction, based on statistical and linguistic criteria. The software was created for domain-specific information retrieval and uses an in-built reference corpus of non-specialised texts. Drouin explains (Drouin 2012) that the reference corpus is about 8,000,000 occurrences, which corresponds to 465,000 word types. The reference corpus consists of two parts: a) journal articles and b) a part of the British National Corpus (BNC). I also use *WordSmith Tools 6.0* (Scott 2015) software, and in particular its concord, patterns and clusters functions, and *AntConc 3.4.4* (Anthony 2014) to verify multi-word term validity as well as for lexical analysis and text search.

The selection of multi-word terms is primarily based on their specificity score as compared to TermoStat's non-specialised reference corpus, which is calculated directly by the software. The software also allows us to order the multiword terms by log-likelihood or chi^2 score, and these scores have been consulted, too. As specificity score, originally proposed by Lafon (1980), focuses particularly on the specificity of lexical items, "based on the assumption that a technical corpus contains a set of lexical items that are closely related to its subject and subject area" (Drouin 2003, pp. 99-100), this parameter has been chosen as the primary criterion. In addition, I double checked my candidate terms using the frequency of a given multi-word term as a secondary criterion. Such an organisation allows me to focus only on those items that are highly specific to the analysed linguistic variety, without losing of sight the purely quantitative component of high recurrence.

Structurally, the multi-word terms and collocations with a term are categorised by their grammatical types, by the number of elements in the n-gram and by the thematic component. Next, using the concord, patterns and clusters functions of WordSmith Tools 6.0 I checked the phraseological environment of the candidates chosen for the analysis before proceeding with the classification.

4. Findings

4.1. Manual sorting and classification of multi-word terms

TermoStat Software has identified 4,026 potential multi-word terms of a nominal nature, ranging from 2-grams to 7-grams. I have selected only candidates with the specificity score above 5 (log likelihood above 20) and raw frequency of at least 5 occurrences; however, the latter criterion was of a secondary nature as my findings were sorted by specificity. I have manually checked the selected candidates verifying the software's precision with regard to a) the terminological status of a given candidate and b) the correct assignment of a grammatical category.

Concerning the terminological status, I draw on a number of classifications of legal terms (Mellinkoff 1963; Riley 1995; Alcaraz Varó, Hughes 2002; Chromà 2011) and include in the respective categories 1) purely technical terms (occurring only in a specific domain, such as "criminal offence" for legal discourse or "in vitro fertilisation" for bioethical discourse); 2) semi-technical or mixed terms (such as "dissenting opinion") or 3) recontextualised everyday vocabulary acquiring specialised meaning in the context (e.g. "fair balance").

Based on the above criteria, I have classified the sorted multi-word terms into a) legally-relevant (e.g. "domestic court"), b) bioethically-relevant (e.g. "gender reassignment"), c) of an ambiguous relevance (e.g. "family life") and d) general ("general interest"). The latter category was not treated as terminological. Ambiguous multi-word terms gather terms which could be used both in bioethically-relevant legal discourse and in less specialised texts (e.g. "state of health"), as well as terms which could be classified as both legally and bioethically-specific, e.g. "lawful abortion". However, the bioethical component is intended to be the prevailing one. As bioethics is a very elusive concept with blurred borders, the presence of two categories allows a certain flexibility in identifying different cases of potentially bioethically-relevant terms. In general, such a thematic division is possible on account of a multi-word nature of terms, which allows us to contextualise and assess their meaning. The thematic distinction is introduced as an approximate indicator and should not be interpreted in absolute terms.

The manual check corroborated the reliability of the software; however, I had to eliminate incomplete chunks of multiword expressions, such as "legal parent-child" ("legal parent-child relationship"). The software proved also to be reliable in reading syntax, as it made errors only with regard to Latin phrases (e.g. "mutatis mutandis", "de facto", "locus standi", etc), some verbal constructs (e.g. "give birth" or "claimed compensation" was

software marked as [Adj + N], although the concordance proved it to be [V+N]).

It is interesting to note that legal multi-word terms represent more than a half of all sorted terms (see Table 2), whereas bioethically charged multi-word terms and ambiguous terms amount respectively to 23% and to 25%. However, if the latter two categories are combined, the proportion becomes 54% to 48%. Such a high proportion can be considered, at a preliminary level, as an indication that legal and bioethical terms in the ECtHR judgments overlap.

	Legal		Ambiguous		Bioethical		Total	
	Freq	%	Freq	%	Freq	%	Freq	%
	708	**54%**	**330**	**25%**	**285**	**23%**	**1,323**	**100%**
n-grams								
2-grams	517	73%	235	71%	216	76%	971	73%
3-grams	139	20%	64	19%	45	16%	247	19%
4-grams	46	7%	27	8%	21	7%	94	7%
5-grams	3	0.4%	2	1%	3	1%	9	1%
6-grams	1	0.1%	1	0.3%	-	-	2	0.2%

Table 2
Multi-word terms sorted by domain and by n-grams.

In terms of the number of elements in the cluster, bigrams are the most recurrent category with 73% of occurrences in total. 3-grams and 4-grams represent respectively 19% and 7%, and the other categories are significantly less frequent. The distribution of n-grams in different thematic categories reflects the general tendency, yet one can note that bioethical multi-word terms are most frequently bigrams (76%).

Concerning the morphological composition, the most productive pattern of a multi-word term, amounting to 56% of all multi-word terms analysed, is [Adj + N] (748 different multi-word types). This pattern is represented by such expressions as "private life", "domestic court", "positive obligation", "medical treatment", etc. In an inter-thematic perspective, it is the most recurrent pattern in legal multi-word terms (60%). Remarkably, the second and third most productive patterns differ across the thematic blocks: bioethical multi-word terms rely on [N+N] structures in 24% ("home birth", "gender reassignment", "insemination facility"), whereas legal multi-word terms employ it only in 11% of multi-word types ("remand prison", "police custody"), which would suggest a higher nominalisation of bioethical discourse. At the same time [N + Prep + N] pattern amounts to 14% of legal multi-word terms ("margin of appreciation", "declaration of incompatibility") and only to 8% of bioethical terms ("removal of tissue", "donation of embryos"). It would seem thus that information is more densely packed in bioethical multi-word terms, with less explicit prepositional links.

The fourth position with 3% of occurrences on the average is divided between the structures [Adj + Adj + N] (e.g. "joint dissenting opinion", "full legal recognition", "presumed biological father") and [Adj + N + N] ("intensive care unit", "civil status register"). Finally, 2% are represented by the [N + Prep + Adj + N] pattern (e.g. "lack of legal recognition", "exhaustion of domestic remedy", "use of human embryo").

	Legal		Ambiguous		Bioethical		Total	
Composition	**Freq**	**%**	**Freq**	**%**	**Freq**	**%**	**Freq**	**%**
Adj + N	422	60%	176	53%	148	52%	748	56%
N + N	79	11%	58	17%	67	24%	205	15%
N + Prep + N	97	14%	38	11%	22	8%	150	11%
Adj + Adj + N	22	3%	10	3%	7	2%	39	3%
Adj + N + N	15	2%	11	3%	10	4%	36	3%
N + Prep + Adj + N	16	2%	10	3%	7	2%	33	2%
Adj + CConj + Adj + N	19	3%	10	3%	4	1%	31	2%
Latin	13	1.8%	-	-	3	1%	16	1%
N + N + N	3	0.4%	8	2%	3	1%	15	1%
Adj + N + Prep + N	6	1%	4	1%	2	1%	12	1%
V + N (NomPhr)	4	0.6%	1	0.3%	2	1%	12	1%
N + Prep + N + N	1	0.1%	3	0.9%	3	1%	7	0.5%
N + Adj + N	1	0.1%	1	0.3%	1	0.4%	3	0.2%
Adj + N + Prep + Adj + N	1	0.1%	-	-	1	0.4%	3	0.2%
Adj + N + N + N	1	0.1%	1	0.3%	1	0.4%	3	0.2%
Other patterns	8	-	-		4	-	20	2%

Table 3
Multi-word terms sorted by grammatical composition.

4.2. General collocational environment and combinability of multi-word terms

During the sorting I came across some economic terms ("marginal lending rate", "default interest"). Interestingly, these economic patterns were embedded into large chunks of a highly recurrent nature together with legal phraseological units.

(1) The Court considers it appropriate that the *default interest rate* should be based on the *marginal lending rate* of the European Central Bank, to which should be added *three percentage points*.[1]

(2) that from the expiry of the above-mentioned three months until settlement *simple interest* shall be payable on the above amounts at a rate equal to the *marginal lending rate* of the European Central Bank during the *default period* plus three *percentage points*.

---

[1] Emphasis is added in all examples.

Example (1) is a 31-gram occurring in 16 texts, and example (2) is a 42-gram repeated in 30 texts. Both larger formulae feature in the judgment's disposition part concerning equal satisfaction (compensation money). Their size and high recurrence make it possible to define them as routine formulae, structural building blocks or text-organising patterns (Biel 2014b: 178) of the ECtHR judgments.

Analysis of the phraseological environment of legal multi-word terms casts light on their combinability and embeddedness in larger phrases. These larger units are frequently quotations of the European Convention on Human Rights. For instance, example (3) is 45-gram used in 21 texts and it is an excerpt from Article 14. Example (4), based on Article 8.2 of the Convention, is the longest and the most recurrent lexical bundle in all texts analysed: it counts 71 grams and is used in 62 texts.

(3) The *enjoyment of the rights and freedoms* set forth in [the] Convention shall be secured without discrimination on any ground such as sex, race, colour, language, religion, *political or other opinion, national or social origin*, association with a national minority, property, birth or other status.

(4) There shall be no interference by a *public authority* with the *exercise of this right* except such as is in accordance with the law and is necessary in a *democratic society* in the *interests of national security*, *public safety* or the *economic well-being of the country*, for the *prevention of disorder or crime*, for the *protection of health or morals*, or for *the protection of the rights and freedoms* of others.

Bioethical multi-word terms do not exhibit a similar tendency to be grouped in extremely large and repetitive text chunks, although ambiguous terms of a twofold relevance can figure in longer clusters as they are cited in the case-law:

(5) The Court reiterates/ notes that the notion of *"private life"* within the meaning of Article 8 of the Convention is a broad concept which encompasses, inter alia, the right to …

An interesting observation emerged concerning the placement of bioethically relevant multi-word terms. In all the judgments analysed such terms and notions were extensively addressed and discussed in the "Relevant domestic law" or "Relevant domestic and international law and practice" part of the judgment.

(6) According to Article 1592 of the Civil Code, *a child's father is either the man who on the date of the child's birth was married to the child's mother* (no. 1), or *the man who acknowledged paternity* (no. 2), or *whose paternity is judicially established under Article 1600d of the Civil Code* (no. 3). [Case of Ahrens v. Germany]

(7) Section 55 of this law *defines gestational surrogacy as the fact of bearing and handing over a child on the basis of a contract concluded between the surrogate*

> *mother and the intended parents, who provide their own genetic material.* […] Decree no. 107, issued by the Minister of Health on 30 August 2012, *defines gestational surrogacy as a contract entered into between the surrogate mother and the intended parents who have used their genetic material for the conception.* [Case of Paradiso and Campanelli v. Italy]

(8) In order to remedy *reproductive problems arising as a result of human sterility or infertility*, recourse may be had to medically assisted reproduction in the conditions and in accordance with the procedures provided for by this Law, which guarantees the rights of all the persons concerned, including those of the subject thus conceived. [Case of Parrillo v. Italy]

Naturally, these patterns could not achieve a high number of repetitions on account of thematic variation across the subcorpora. A general – and somewhat surprising – tendency is however clear: the highest concentration of bioethically charged terms is found in the section devised to deal with a legal overview of a certain national system. The rationale standing behind this placement peculiarity could lie in a combination of factors: a) legal discourse strives towards precision and relevant terms are frequently duly defined to avoid ambiguity; b) bioethical concepts, being external to the average legal situation, need explanation as they are entextualised from another scientific reality, i.e. their meaning is extracted from one discourse and consequently inserted into another through a process of de-contextualization and re-contextualisation in another context (Silverstein, Urban 1996, p. 15 in Garzone 2014, p. 79). In general, this placement allows us to fix artificially the meaning of bioethical and ambiguous terms in the context of judgments, and bioethical terms invariably collocate with legal terms and expressions in these definitions and elsewhere, proving a solid link between these two discursive realities.

 The sorting revealed a number of collocations of the [Adj + N] pattern, with a relatively low degree of terminologicality, such as "present case" / "instant case", most often used within a nominal phrase "in the present / instant case", or qualifications "second applicant", third and fourth applicants", based on the number of applicants involved in a case. These and similar occurrences were marked as lexical collocations following Biel (2014b, p. 181) as they merely perform a referential function by using a qualifying adjective in adjacency to a legal term. It must be noted that the distinction and separation of terms from collocations is arbitrary to a certain degree (see, for instance, Heid 2001, p. 791). I identified also a number of collocations from the general language use ("general interest", "last-mentioned date", "chilling effect"), which were inserted into nominal phrases or qualifications (Bhatia 2006, p. 2) and, although not being terminological, increased the legal flavour as they served the purpose of creating the phraseological matrix of texts.

The [N + N] and [N + Prep + N] structures also frequently co-occurred with qualifying adjectives. For instance, the most recurrent legal multi-word term of the latter pattern – "margin of appreciation" (frequency: 551; specificity: 54,68) – co-occurred with such adjectives as "wide" (102 hits) "broad" (20), "narrow" (20), "significant" (3) and "relevant" (3). Another combinatorial tendency observed concerned a number of pseudo-multinominal terms, composed of "the issue of" + N, "the concept of" + N or "the notion of" + N. In these bundles the first element was somewhat redundant as it did not carry an additional terminological meaning, but mainly enhanced the nominalised nature of judgments.

The manual stage of sorting, although highly time-consuming, has also allowed me to trace a certain pattern with respect to the collocational environment of nominal multi-word terms. While I had to eliminate such chunks as "respect of non-pecuniary damage" (truncated "in respect of non-pecuniary damage") and similar structures, I noticed that multi-word terms – legal, bioethical and ambiguous – were frequently co-occurring with complex prepositions, built around the pattern [Prep + (Art) + N + Prep]. Unfortunately, TermoStat does not retrieve the above structures as it is programmed to trace terminological choices and to disregard functional vocabulary. Consequently, complex prepositions should be addressed in a separate paper, using a different software. However, using WordSmith Tools 6.0 I briefly looked at 4-grams – uninterrupted sequences of four elements – and identified 416 different prepositional phrases, 60 of which (about 15%) were highly recurrent and frequent complex prepositions used with both legal and bioethical multi-word terms, such as *within the meaning of* (293); *on the basis of* (284); *as a result of* (233); *in the context of* (230); *for the purposes of* (217). 3-element complex prepositions were even more frequent and included such items as *in accordance with* (607), *in respect of* (538), highlighting again a high prominence of these grammatical structures for the texts under analysis.

4.3. Term-embedding collocations with a verb

Apart from a general collocational environment of multi-word terms, I looked at a special category of collocations, the so-called term-embedding collocations (Biel 2014b, p. 180). These collocations "form the skeleton of legal rules by providing action and enabling terms to enter into relations" (Biel 2014b, p. 180). In other words, these are

> Collocates of terms which embed terms in cognitive scripts and the text, evidencing combinatory properties of terms. […] They establish links between terms and elements of conceptual frames. […] Subtype-denoting collocations are often subject to terminologisation and form distinct terms. (Biel 2014b, p. 180)

Prototypically these phraseological units are built around the structure [N$_{term}$+V] or in the reverse order [V+N$_{term}$], and denote actions that are possible to undertake with the base noun.

Having sorted the term-embedding collocations by the association score, based on the log-likelihood test, I have attributed to the selected candidates the same classification already elaborated for multi-word terms: legal, bioethical, ambiguous, as well as general, and have analysed those items that had an association score above 25.

Thematic class	n. types	%	Examples	Inclusion in the phrase
Bioethical	137	19%	Commit suicide	helping someone to commit suicide is not a punishable offence
Legal	381	52%	Enjoy margin of appreciation	states enjoy a wide margin of appreciation
Ambiguous	112	15%	Found family	right to marry and to found a family
General	108	15%	Cast doubt	Genetic evidence cannot cast doubt
Total	738	100%		

Table 4
Term-embedding collocations sorted by the thematic element.

As expected – based on the findings of multi-word terms – term-embedding collocations with a legal thematic element prevail, with 52% of different collocations belonging to this category. They are followed by bioethically relevant collocations with a verb (19%). 15% of cases are ambiguous and 15% of collocations with a verb can be classified as general in that they can be used in general context. These results are to be treated with caution. Some verbs can carry a legal thematic connotation ("exhaust", "dismiss", see (9)) or a bioethical connotation ("conceive", see (10)); however, in most cases it is difficult to associate any bioethical connotation just to a verb, also in light of the elusive nature of the very concept of bioethics and its highly interdisciplinary character. For instance, in (11) "order surveillance" has a distinct legal flavour, although in the general context of the utterance it acquires also a bioethical relevance. Consequently, the above subdivision is to be intended as approximative.

(9) The Government's objection as to the failure by the first and second applicants to *exhaust domestic remedies* must therefore be *dismissed*. [Case of I.G. and Others v. Slovakia]

(10) He observed that an agreement whereby a woman undertook to *conceive and bear a child* and relinquish it at birth was null and void in accordance with the public-policy principle that the human body and civil status are inalienable. [Case of Mennesson v. France]

(11) To sum up, it should be concluded that *ordering surveillance* of insured persons by accident insurers in the context outlined in point 4.3 [above] is permitted; the results

of such surveillance can thus in principle be used for assessment of the issues in question .... [Case of Vukota-Bojic v. Switzerland]

It is crucial to look at the phrase inclusion, as often collocations classified as general in per se are included in a context, where they assume bioethical or legal relevance. For instance, "women travel" is classified as general. However, if we look at its inclusion in a sentence, we find that the utterance undoubtedly belongs to the bioethical field, which proves the importance of the phraseological perspective to assess the specificity of this discourse.

(12) The Government also underlined that the impugned restrictions had led to a significant reduction in Irish *women travelling* to the United Kingdom *for an abortion* (6,673 women in 2001 travelled and 4,686 women did so in 2007) and to one of the lowest levels of maternal deaths in the European Union. [Case of A, B and C v. Ireland]

Interestingly, in most term-embedding collocations the nominal element acts as a grammatical object (56%, see Table 5); however, it could have been expected as I include in this category direct objects ("exhaust domestic remedies"), indirect objects ("award the applicant (just satisfaction)") and prepositional objects ("comply with a requirement"). Unfortunately, the software does not distinguish between a prepositional object and a prepositional complement (Quirk *et al.* 1985, p. 1201), consequently I had to manually reclassify some of the phrases as collocations with a prepositional complement (18%, e.g. "travel for abortion"), which are a borderline case between term-embedding collocations and lexical collocations. There are 28% of cases where the nominal element acts as a grammatical subject ("Court reiterates"). When the nominal element is the institutionalised actor involved in the proceedings (e.g. Court or applicants), it frequently collocates with verbs with communicative meaning ("the applicants claim") or semantics of mental activity ("the Court considers"). Where both subject and object functions are traced, e.g. "institute proceedings" and "proceedings are instituted", both are counted separately.

	Legal		Ambiguous		Bioethical		Total	
Function of the N	**Freq**	**%**	**Freq**	**%**	**Freq**	**%**	**Freq**	**%**
Object	205	54%	78	70%	86	63%	414	56%
Subject	108	28%	28	25%	31	23%	213	29%
Complement	68	18%	6	5%	20	15%	111	15%
Total	381		112		137		738	

Table 5
Term-embedding collocations sorted by the function of the nominal element
across various thematic groups.

Table 6 gathers 25 first-ranking term-embedding collocations sorted by their association score (log-likelihood, see column 4 in Table 6), i.e. by their domain specificity, and examples of the collocations being used in the corpus. Interestingly, some collocations with a verb, although following the general [V + N] pattern, differ from others as they seem to demonstrate a higher degree of fixedness and indivisibility (Hudson 1998, pp. 5-10, 35). Even though such collocations as "fall within the ambit of" (association score: 150; frequency: 55) or "fall within the scope of" (association score: 423; frequency: 59) contain the nouns "ambit" and "scope", the former is not used outside of these expressions, and the latter is prevalently used as "within the scope of" collocating also with "come", "bring" and "take", demonstrating a highly convergent pattern. It would seem that the pattern "within the scope/ambit of" exhibits a high degree of grammaticalization and functions almost as a complex preposition.

Verb	Noun	Fre	AssSc	Example
exhaust	remedy	72	692	the applicant had not exhausted domestic remedies
commit	suicide	104	623	commit suicide
strike	fair balance	77	617	to strike a fair balance
bear	child	131	595	child (was) born
practise	lawyer	63	568	lawyers practising in Riga
originate	application	78	527	the case originated in an application
annex	judgment	49	480	opinions are annexed to this judgment
adopt	date	74	478	judgment adopted on that date
bear	wedlock	68	478	born out of wedlock
enjoy	margin of appreciation	68	445	states enjoy a wide margin of appreciation
add	percentage points	43	444	to which should be added three percentage points
undergo	gender reassignment	81	435	she had undergone gender reassignment surgery
fall	scope	59	423	fall within the scope of
afford	just satisfaction	53	392	the court shall, if necessary, afford just satisfaction
balance	competing interests	51	374	to balance competing interests
found	family	43	356	right to marry and to found a family
originate	case	71	349	the case originated in an application
incur	costs and expenses	40	340	costs and expenses incurred before the court
institute	proceedings	60	329	department had failed to institute proceedings
terminate	pregnancy	41	310	the applicant's pregnancy was not terminated
guarantee	convention	67	272	the rights guaranteed by the convention
impose	obligation	48	267	the Convention did not impose an obligation on contracting states
challenge	paternity	43	262	allowing the biological father to challenge paternity without restrictions
travel	abortion	40	261	women travelling abroad for abortion

Table 6
25 most domain-specific term-embedding collocations
organised in a decreasing order according to the association score.

As I already argued elsewhere (Nikitina 2018, p. 200), adapting and slightly modifying Biel's (2014a, 2014b) classification, complex prepositions act as grammatical patterns on account of their prefabricated nature and grammatical functions as well as functional distinctiveness for legal discourse in general. It emerges that such complex phrases as "within the scope of", "within the ambit of" and "within the sphere of", which could be rephrased by "under", collocate with "fall" on the left or other verbs with a reduced semantic charge, making the whole expression functionally prefabricated and classifiable as a borderline case between term-embedding collocations and grammatical patterns.

The analysis sheds light on a high concentration of specialised deverbal nouns in term-embedding collocations, which reduced the recurrence of verbs. For instance, instead of using a more general verb "to treat" / "be treated", the corpus shows a number of verbs collocating with "treatment" as an object (with association score indicated in brackets): *receive treatment* (253), *undergo treatment* (212), *accept treatment* (114), *refuse treatment* (160) as well as *subject to treatment* (92), *administer treatment* (52), *discontinue treatment* (38), *propose treatment* (93), *proscribe treatment* (129), *start treatment* (34) and *withdraw treatment* (226), some of which could have been replaced by "treat". "Treat" is used in term-embedding collocations with a subject: *doctor treats* (139), and in collocations with a direct object: *treat a patient* (121), *treat a child* (82), *treat a woman* (72) and *treat infection* (39). All of the nouns co-occurring with "treat" could collocate as indirect or prepositional objects with "treatment". Similarly, the ambiguous term – both bioethically and legally relevant – "consent" is not used as a verb, but only as a noun in term-embedding collocations *withdraw consent* (205), *withhold consent* (134), *obtain consent* (70) or *express consent* (55). These verbo-nominal constructs demonstrate high association scores and confirm their domain specificity, marking legal discourse on bioethics as tendentially nominalised discourse.

Moreover, the analysis identified a number of collocations belonging to the category of *verbo-nominal phrases*, or *verbonominants*, i.e. those [V + N^{obj}] structures, where the noun – usually a deverbal noun - carries the lexical meaning of the whole pattern, and the verb is "semantically reduced or light" (Leech *et al*. 2009, p. 166), e.g. "reach a conclusion" instead of "to conclude". This phenomenon "at the borderline between syntax and lexicon" (Leech et al. 2009, p. 166) could be effectively addressed through the paradigm of phraseology. These patterns, known under a great variety of labels in the academic literature[2] increase the nominal character of discourse,

---

[2] See Leech *et al*. 2009, pp. 166-167 for an overview.

closely associated with legal (cf. Mellinkoff 1963; Crystal, Davy 1973; Tiersma 1999) and generally specialised discourse (Leech *et al*. 2009). In legal discourse this phenomenon is described as "preference for nouns and nominalisations (nouns derived from verbs, such as 'consideration' or 'injury') over verbs ('consider' or 'injure')" (Tiersma 2015 [2006], p. 28). Verbo-nominal constructions are generally composed of verbs of common meaning, such as "take", "make", "give", "have" or "put" (Claridge 2000, p. 71), which do not behave as domain-specific, and thus are not retrieved by the log-likelihood specificity test. The software has retrieved mostly cases of domain-specific verbonominants, such as those collocating with "reach": *reach agreement*, *reach conclusion* and *reach decision*. While these constructs could be rephrased by a single verb, it is a clearly dispreferred tendency.

5. Conclusions

This article was designed following a general descriptive goal of presenting some linguistic peculiarities of the phraseological organisation of the ECtHR judgments dealing with a variety of bioethical issues. The analysis of term-related patterns, their structure and thematic components confirms that the skeleton of the judgments, with 54% of multi-word terms and 52% of term-embedding collocations with a verb, is composed of phraseological units with a legal thematic component. The smaller multi-word units are often packed into lengthy prefabricated chunks of texts, up to 71-grams, and recur in different judgments. These larger units mark different structural blocks of the judgment: the beginning, transitions between different parts, the disposition and the end, and contribute to the general perception of standardisation and formulaic character of this genre.

Apart from the structural matrix of the judgment, which remains purely legal, it appears possible to talk about a cross-fertilisation between legal and bioethical areas. In general, the findings demonstrate that bioethical and legal notions are intertwined in a complex relationship, up to the point where it becomes problematic to distinguish between the two elements, or where either of the meanings is possible, or where words from general language acquire specialised meaning in the context. It emerges that bioethical notions are entextualised into the legal matrix of judgments without altering the generic structure of the latter. While bioethical term-related patterns are generally disseminated throughout the judgments, they are especially frequent in the parts describing facts and national and international points of law. The latter concentration – typically under the form of definitions and explaining provisions – could be construed as a sign of legal precision on the one hand and as an instance of specialised knowledge transfer from the

scientific domain and its incorporation into the legal domain, which could be interpreted as a strategy of knowledge management of the participants (van Dijk 2003).

The structural analysis of multi-word terms allowed to trace a slight difference in knowledge presentation between legal and bioethical units. It emerged that, among different thematic classes, bioethical multi-word terms use the [N + N] pattern most frequently, and disprefer a more explicit prepositional structure of the [N + Prep + N] type. The tendency is reversed in respect of legal multi-word terms. These preferential morphological patterns translate into a higher degree of nominalisation of bioethical discourse compared with legal discourse. The nominalisation is further exacerbated by the recurrence of the so-called verbo-nominal structures, i.e. the preference to use a semantically reduced verb with a deverbal noun/ nominal expression. The collocational environment of term-related units includes combinations between various multi-word terms, frequently achieved through the use of complex prepositions or grammaticalized verbal constructs. It emerges, however, that bioethical multi-word terms tend to appear in close proximity to legal multi-word terms, or verbal constructs with a legal thematic element.

Analysis of multi-word terms and their immediate collocational environment proved to be a valid starting point to assess the entextualisation processes, demonstrating some convergent and divergent tendencies, both in the composition and in their phraseological behaviour. The results of this research and linguistic descriptions produced could and will be employed as a starting point to analyse further mechanisms of legal and bioethical knowledge dissemination, during its migration from an institutionalised setting to a more liberal web space under the form of blogs.

Bionote: Jekaterina Nikitina has a PhD degree from the University of Milan in "Linguistic, Literary and Intercultural Studies in European and Extra-European Perspectives". Her main academic interest lies in the field of Legal Discourse and Legal Translation Studies, with her PhD thesis focusing on the analysis of phraseological patterns in written pleadings before the European Court of Human Rights translated into English from Russian and from Italian. Jekaterina's areas of interests extend to LSP theories and applications, knowledge dissemination dynamics, discourse of healthcare, medicine and bioethics, applying qualitative and quantitative, specifically corpus linguistics, analytical approaches. She is an external collaborator of the research unit of Milan in the PRIN project dealing with knowledge dissemination across media in English. Her published academic work includes publications on legal translation and dissemination of bioethical knowledge, with a focus on medically assisted procreation and gene editing.

Author's address: jekaterina.nikitina@unimi.it

References

Alcaraz Varò E. and Hughes B. 2002, *Legal Translation Explained*, St. Jerome Publishing, Manchester.

Anthony L. 2014, *AntConc* (Version 3.4.4) [Computer Software], Waseda University, Tokyo. http://www.laurenceanthony.net/.

Bergenholtz H. and Tarp S. 1995, *Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries*, Benjamins, Amsterdam/Philadelphia.

Bhatia V. 2006, *Legal Genres*, in Brown K. (ed.) *Encyclopedia of Languages and Linguistics*, Elsevier, Oxford, pp. 1-7.

Bhatia V. 2014 [2004], *Worlds of Written Discourse: A Genre-Based View*, Continuum, London.

Biel Ł. 2014a, *Lost in the Eurofog. The Textual Fit of Translated Law*, Peter Lang, Frankfurt am Main.

Biel Ł. 2014b, *Phraseology in Legal Translation: a Corpus-based Analysis of Textual Mapping in EU Law*, in Cheng L., Kui Sin K. and Wagner A. (eds), *Ashgate Handbook of Legal Translation*, Ashgate, Aldershot, pp. 177-192.

Bioethics and the Case-law of the Court 2016 [2009], Research Report, Council of Europe / European Court of Human Rights. http://www.echr.coe.int/Documents/Research_report_bioethics_ENG.pdf (27.06.2017).

Calsamiglia H. and van Dijk T. 2004, *Popularization Discourse and Knowledge about Genome*, in "Discourse and Society" 15 [4], pp. 369-389.

Chromà M. 2011, *Synonymy and Polysemy in Legal Terminology and Their Applications to Bilingual and Bijural Translation*, in *Legal Terminology: Approaches and Applications*, *a* special issue of "Research in Language" 9 [1], pp. 31-50.

Claridge C. 2000, *Multi-word Verbs in Early Modern English: A Corpus-Based Study*, Language and Computers 32, Rodopi Publishers, Amsterdam.

Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine, European Treaty Series - No. 164, Oviedo, 4.IV.1997. https://rm.coe.int/168007cf98 (27.06.2017).

Crystal D. and Davy D. 1973, *The Language of Legal Documents*, in Crystal D. and Davy D. (eds.), *Investigating English Style*, Indiana University Press, Bloomington, pp. 193-217.

Drouin P. 2003, *Term Extraction Using Non-Technical Corpora as a Point of Leverage*, in "Terminology" 9 [1], pp. 99-117.

Drouin P. 2012, *TermoStat Web 3.0* [software], Université de Montréal. http://termostat.ling.umontreal.ca/doc_termostat/doc_termostat_en.html

ECtHR *Rules of Court* as of 14 November 2016. https://www.echr.coe.int/Documents/Rules_Court_ENG.pdf (26.02.2018).

Garzone G. 2014, *News Production and Scientific Knowledge: Exploring Popularization as a Process*, in Bongo G. and Caliendo G. (eds.), *The Language of Popularization: Die Sprache der Popularisierung*, Peter Lang, Bern, pp. 73-107.

Granger S. and Paquot M. 2008, *Disentangling the Phraseological Web*, in Granger S. and Meunier F. (eds.), *Phraseology: An Interdisciplinary Perspective*, Benjamins, Amsterdam/Philadelphia, pp. 27-49.

Heid U. 2001, *Collocations in Sublanguage Texts: Extraction from Corpora*, in Wright S.E. and Budin G. (eds.), *Handbook of Terminology Management. Application-*

Oriented Terminology Management, vol. 2, Benjamins, Amsterdam-Philadelphia, pp. 788-808.

Hudson J. 1998, *Perspectives on Fixedness: Applied and Theoretical*, Lund University Press, Lund.

Hyland K. 2008, *As Can Be Seen: Lexical Bundles and Disciplinary Variation*, in "English for Specific Purposes" 27, pp. 4-21.

Kjær A.L. 1990, *Context-Conditioned Word Combinations in Legal Language*, in "Terminology Science & Research, Journal of the International Institute of Terminology Research – IITF" 1 [1-2], pp. 21-32.

Kjær A.L. 2007, *Phrasemes in Legal Texts*, in Burger H., Dobrovol'skij D., Kühn P. and Norrick N. (eds.), *Phraseology: An International Handbook of Contemporary Research*, Mouton de Gruyter, Berlin, pp. 506-516.

Lafon P. 1980, *Sur la variabilité de la fréquence des formes dans un corpus*, in "MOTS" 1, pp. 128-165.

Leech G., Hundt M., Mair Ch. and Smith N. 2009, *Change in Contemporary English: A Grammatical Study*, Cambridge University Press, Cambridge.

Mellinkoff D. 1963, *The Language of the Law*, Little Brown & Co, Boston.

Meyer I. and Mackintosh K. 1994, *Phraseme Analysis and Concept Analysis: Exploring a Symbiotic Relationship in the Specialized Lexicon*, in Martin W. (ed.), *Euralex 94 Proceedings*, Amsterdam, pp. 339-348.

Myers G. 2003, *Discourse Studies of Scientific Popularization: Questioning the Boundaries*, in "Discourse Studies" 5 [2], pp. 265-279.

Nikitina J. 2018, *Legal Style Markers and Their Translation: Written Pleadings at the ECtHR*, Edizioni accademiche italiane / OmniScriptum Publishing, Beau Bassin-Riga.

Riley A. 1995, *The Meaning of Words in English Legal Texts: Mastering the Vocabulary of the Law – A Legal Task*, in "Law Teacher" 29-30, pp. 68-83.

Quirk R., Greenbaum S., Leech G. and Svartvik J. 1985, *A Comprehensive Grammar of the English Language*, Longman, London.

Sager J. 1998, *Terminology: Theory*, in Baker M. (ed.), *Routledge Encyclopedia of Translation Studies*, Routledge, London-New York, pp. 258-262.

Scarpa F., Peruzzo K. and Pontrandolfo G. 2014, *Methodological, Terminological and Phraseological Challenges in the Translation into English of the Italian Code of Criminal Procedure*, in Gialuz M., Lupária L. and Scarpa F. (eds.), *The Italian Code of Criminal Procedure. Critical Essays and English Translation*, Wolters Kluwer / CEDAM, Padova, pp. 53-80.

Scott M. 2015, *Wordsmith Tools version 6* [computer software], Lexical Analysis Software, Stroud.

Sinclair J. 2004, *Trust the Text: Language, Corpus and Discourse*, Routledge, London/New York.

Teubert W. 2002, *The Role of Parallel Corpora in Translation and Multilingual Lexicography*, in Altenberg B. and Granger S. (eds.), *Lexis in Contrast*, Benjamins, Amsterdam/Philadelphia, pp. 189-214.

Tiersma P. 1999, *Legal English*, Chicago University Press, Chicago/London.

Tiersma P. 2015 [2006], *Some Myths about Legal Language*, in Solan L., Ainsworth J. and Shuy R. (eds.), *Speaking of Language and Law. Conversations on the Work of Peter Tiersma*, Oxford University Press, Oxford, pp. 27-34.

Universal Declaration on Bioethics and Human Rights, adopted by UNESCO's General Conference on 19 October 2005. http://unesdoc.unesco.org/images/0014/001461/146180E.pdf, (27.06.2017).

van Dijk T.A. 2003, *The Discourse-Knowledge Interface*, in Wodak R. and Weiss G. (eds.), *Critical Discourse Analysis. Theory and Interdisciplinarity*, Palgrave, London, pp. 85-109.

Widdowson H. 1979, *Explorations in applied linguistics*, Oxford University Press, Oxford.