Research Methods in Legal Translation and Interpreting

Crossing Methodological Boundaries

Edited by Łucja Biel, Jan Engberg, M. Rosario Martín Ruano, and Vilelmini Sosoni

First published 2019

ISBN: 978-1-138-49210-3 (hbk)
ISBN: 978-1-351-03122-6 (ebk)

2 Implications of text categorisation for corpus-based legal translation research

The case of international institutional settings

Fernando Prieto Ramos

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

2 Implications of text categorisation for corpus-based legal translation research

The case of international institutional settings

Fernando Prieto Ramos

1 Introduction: why does text categorisation matter?

Text categorisation is a key aspect of research into discourse features and translation patterns, and an essential methodological consideration in corpus design and analysis. Systematic categorisation of text is pivotal in delineating the scope of research questions, producing valid datasets and deriving findings accordingly. In fact, the comparability, representativeness and balance of corpus components depend on the boundaries and hierarchical organisation of the target population (e.g. Biber 1993; Halverson 1998). Since "different ways of classifying and characterizing texts can produce different text typologies" (McEnery *et al.* 2006, p. 16), the criteria applied for text classification and category definitions must be made explicit (e.g. Biber *et al.* 1998; Halverson 1998; Lee 2001), particularly when a corpus encompasses a large amount of texts from various categories and the boundaries between these categories cannot be presupposed.

Genre stands out as a widely accepted operational concept for categorising texts. As highlighted by Lee (2001, p. 37), genre is "the level of text categorisation which is theoretically and pedagogically most useful and most practical to work with". This is associated with the idea that genre conventions are recognisable, as reflected in Bhatia's classic definition (1993, p. 13):

> Genre is a recognizable communicative event, characterized by a set of communicative purpose(s) identified and mutually understood by the members of the professional or academic community in which it regularly occurs. Most often it is highly structured and conventionalized with constraints or allowable contributions in terms of their intent, positioning, form and functional value.

The link between communicative purposes and discourse conventions is virtually uncontested in genre-based text categorisations, especially since Biber's (1988) multidimensional analysis of register variation. This work has influenced

subsequent approaches to the study of similarities between texts through both manual annotation and automated measurements of functional attributes (see e.g. Forsyth and Sharoff 2014; Melissourgou and Frantzi 2017). However, there is no consensus about these genre attributes or the method for identifying them, let alone for establishing genre ontologies that reflect inter-genre connections and further subdivisions.

In the case of legal texts, this is compounded by the overwhelming diversity of legal discourses, as they fulfil multiple functions and address all kinds of themes within countless legal frameworks (both national and supranational), branches and communicative settings. The high levels of variability and hybridity of legal language, as "a set of related legal discourses" (Maley 1994, p. 13), make it difficult to build universally valid classifications of legal texts. The hierarchy and boundaries of categorisations ultimately depend on research priorities and perspectives (e.g. Biel 2014, p. 19; Prieto Ramos 2014a, p. 263).

Corpus-based legal linguistic and legal translation studies are crucially contributing to characterise legal genres across languages and jurisdictions (see e.g. Goźdź-Roszkowski 2011a; Borja Albi 2013; Biel 2014; Pontrandolfo 2016). Yet, definitions of "legal text" and the scope of legal translation remain contested. This is not only an academic debate on the nature of a discipline; it also reflects the many textual facets of law itself as a matter of language use, and it is of significance for translation practice. In fact, categorising texts is a critical step in situating and conducting translation-oriented text mining and analysis. As pointed out by Alcaraz Varó and Hughes (2002, p. 103), "the translator who has taken the trouble to recognise the formal and stylistic conventions of a particular original has already done much to translate the text successfully". This is notably the case in the field of law, since legal writing is most often shaped by the "normative force of genre bias", as contended by Rappaport (2014, p. 199). For this legal scholar, lawyers who "understand legal writing as, at least partially, a function of genre bias will better comprehend how legal texts are conceived, received, and perceived, and will be better lawyers as a consequence", as all legal professionals, including judges and legal scholars, have "an audience with expectations precast by genre" (2014, p. 203).

This chapter highlights the relevance of text categorisation for research in legal translation by focusing on institutional translation settings, namely: the European Union (EU), the United Nations (UN) and the World Trade Organization (WTO), and their corresponding adjudicative bodies.[1] After briefly reviewing recurrent issues and models of legal text classification (section 2), a multidimensional approach is applied to the multilingual text production of the three representative institutional translation settings during three years over the span of a decade (2005, 2010 and 2015), as part of the project "Legal Translation in International Institutional Settings: Scope, Strategies and Quality Markers" (LETRINT) (section 3). The resulting subdivisions are integrated into a categorisation matrix and discussed as a way of illustrating the relative nature and implications of text classifications. The fine-grained description of corpus design and representativeness, technical aspects of corpus compilation and full taxonomies of genres are not addressed in this chapter.

2 Classifying legal texts: beyond legal genres?

2.1 Commonalities and diverging views

In corpus building, "the conception of the object which a discipline more or less agrees on provides the motivation for defining a target population" (Halverson 1998, p. 495). This entails defining category boundaries and internal structure "on the basis of theoretical notions pertaining to the relevance of various types of text, and the relative significance of the different types" (1998, p. 499). In Legal Translation Studies (LTS), scholars tend to converge on the relevance of genres to study legal discourse conventions in translation, but diverge on the classification of these genres into broader categories or text types, and on their boundaries based on the notion of "legal text".

The metalanguage applied to these categories also differs between authors. "Text type" and "genre" are sometimes used as interchangeable (see e.g. Berūkštienė 2016, pp. 92–94, on scholarly distinctions between these concepts), while notions such as "genre system" (Bazerman 1994, p. 97) and "genre network" (Fairclough 2006, p. 34) emphasise the idea of interconnection.[2] Regardless of supra-genre level denominations, most approaches include legislative, contractual, judicial and scholarly texts by focusing on key legal functions and associated types of legal discourse (e.g. Bocquet 1994; Šarčević 1997; Tiersma 1999; Kjær 2000). Some authors add considerations on specific branches of legal practice, such as administrative or business law (e.g. Gémar 1995; Mattila 2013). A comparison of approaches suggests that functional and domain elements tend to be embedded in classifications by situation of use or discursive situation parameters, including setting, purposes, addressor and addressee (e.g. Trosborg 1997; Borja Albi 2000; Bhatia 2006; Cao 2007).

As illustrated by Table 2.1, parallels can be drawn between approaches. The link between legal discourse features and legal function or theme emerges as their common ground, and explains the inclusion of legal subcategories of macro-genres as legal texts, e.g. legal academic articles as a subcategory of academic articles. Variations are found, among other details, in the way legislative and contractual texts are grouped together or not, considering their normative value; and also, particularly, in the fuzzier realm of private legal texts written by non-lawyers and other texts that are not "intrinsically" legal (by function or theme) but are used in legal settings (see e.g. differences in Trosborg 1997; Cao 2007). While the fundamental link between legal purpose or theme and discourse features can be found in the first group, the same link seems totally absent in the second group (e.g. personal correspondence or technical reports used in court proceedings).

Scholars disagree on whether the texts of this second group can be classified as legal texts. Abdel Hadi (1992, p. 47) and Harvey (2002, p. 178), for example, consider them legal texts as long as they are used in legal settings. Likewise, Cao (2007, p. 9) defines legal texts as "texts produced or used for legal purposes in legal settings", regardless of the original purpose for which they were produced, whereas she perceives legal language as "the language of and related

Table 2.1 Legal text classifications based on situational parameters

Trosborg (1997, p. 20): "types of texts or genres" by situation of use	Borja Albi (2000, pp. 84–134): "text categories" by discursive situation	Bhatia (2006, pp. 6–7): "system of legal genres" by communicative purposes	Cao (2007, pp. 9–10): "variants or sub-varieties of legal texts" by situation of use
Language of the law (legal documents): • legislation • common law (contracts, deeds)	Prescriptive texts (e.g. acts, statutes, bills, regulations)	Primary genre (legislation)	Legislative texts (e.g. statutes and subordinate laws, international treaties)
Language of the courtroom: • judge declaring the law • judge/counsel exchanges • counsel/witness exchanges	Judicial texts (claim forms, judgments, appeals, writs, orders, etc.) Case-law (decisions of higher courts)	Derived secondary genres (e.g. judgments, cases)	Judicial texts (produced by judicial officers and other legal authorities in judicial processes)
Language in textbooks	Reference works (dictionaries, repositories, encyclopaedias) Scholarly texts (articles, textbooks, manuals, casebooks, manuals, etc.)	Derived enabling (pedagogic) genres: • academic (e.g. textbooks, moots) • professional (e.g. legal memoranda, pleadings)	Legal scholarly texts (scholarly works and commentaries)
Lawyers' communication: • to other lawyers • to laymen	Law application texts (contracts, deeds, wills, legal briefs, etc.)	Target genres (property conveyance documents, client consultation documents, affidavits, agreements and contracts)	Private legal texts • texts written by lawyers (e.g. contracts, leases, wills and litigation documents) • texts written by non-lawyers (e.g. private agreements, witness statements and other documents used in litigation and other legal situations)
People talking about the law			

to law and legal process", including "language of the law, language about law, and language used in other legal communicative situations". She problematises Šarčević's (1997) focus on legal texts for specialists as restrictive (1997, p. 9), and claims that "ordinary texts that are not written in legal language by legal professionals" constitute "a major part of the translation work of the legal translator in real life" (1997, p. 12). It is difficult to accept that personal letters or technical reports that contain no sign of legal language are legal texts. Taken in isolation, rather than through the lens of the translation context, such texts would hardly be considered legal texts in their own right. It can be understood, however, that these texts might be translated in legal settings and play an instrumental role in legal processes. In other words, from a translation perspective, the categorisation of texts without any legal discourse as "legal texts" is only possible in an expansive (or inclusive) classification of texts based on translation settings rather than discourse features.

In this kind of expansive approach, one may claim not only that legal texts encompass multiple combinations of legal and non-legal discourse, but also that legal translation may include more than just legal texts. The preceding triggers at least two related questions for research purposes: where should the boundary be drawn between legal and non-legal texts when mapping a setting or branch of legal translation comprising a variety of text types? To what extent should the link between legal functions or themes and discourse features be a determining factor in the definition and classification of legal text types in a corpus? This brings us back to the question of legal genre conventions and legal discourses.

2.2 The crucible of legal discourses

Extensive work has been conducted on the distinctive lexical, syntactic and structural features of legal discourses. Tiersma (2003) summarises some of the most common ones associated with "legalese", including: archaic, formal and unusual or difficult vocabulary, technical terminology, impersonal constructions, nominalisations, passive constructions, long and complex sentences, wordiness and redundancy (see also e.g. Galdia 2009; Mattila 2013). These features are found, in varying degrees and clusters, in what is traditionally perceived as the core of legal discourses or styles: the language of legal experts, particularly legislators, judges and lawyers (as well as notaries in many jurisdictions). They constitute conventions inherited through precedents in law-making and implementation, and are sometimes described as "fossilized language" (Alcaraz Varó and Hughes 2002, p. 9), which calls for investigation into discourse patterns and variations.

These legal discourse feature clusters are highly interdependent. Legislative discourse, as primary expression of the law, occupies a central position and permeates the other legal discourses that apply or describe the law (see e.g. Kjær 2000, pp. 138–140; Bhatia 2006, pp. 6–7). In a similar vein, Monjean-Decaudin (2013, p. 24) couples the "degree of juridicity" with the legal effect of texts (i.e. more legal force and consequences imply a higher degree of juridicity) and the level of legal knowledge required to understand and translate them. However,

generalisations on legal discourses are difficult to establish because of their vast scope and variability through space and time, not only across jurisdictions and legal traditions, but also within them, e.g. through deliberate simplification, legal reform or harmonisation processes. As rightly expressed by Goźdź-Roszkowski (2011b, p. 3281), far from being uniform, legal language "represents an extremely complex type of discourse embedded in the highly varied institutional space of different legal systems and cultures", and "should be viewed as an umbrella term referring to a universe of remarkably diverse texts, both written and spoken", including "statements on law reproduced in the media" and "any fictional representation" of legal genres.

Legal discourses are also commonly characterised as hybrid, not only as a result of contact between legal systems and drafters with different backgrounds (see e.g. Robinson 2005 on EU legislative drafting), but also in terms of interdisciplinarity, due to the diversity of subjects and specialised knowledge covered by law. This means that non-legal specialised language may often be as prominent as legal language in legal texts. For instance, it is not striking that a financial regulation may be viewed simultaneously as a matter of legal and financial translation, even if the text belongs to a legal genre, i.e. it may typically adhere to specific structural and phraseological conventions to establish legal obligations, but the content may use more financial than legal terminology, thus reflecting the interdisciplinary reality of financial law. Similar patterns of hybridity occur with other technical discourses embedded in legal texts (see e.g. Fontanet 2018).

2.3 Fuzzy boundaries and layers

Since legal texts may be seen as frames and carriers of all kinds of knowledge related to law in many different degrees and forms, corpus analysis emerges as a very useful tool to provide granularity. To answer the methodological questions formulated previously, researchers must acknowledge that any text classification of multiple genres must be flexible and sensitive to ambiguities and overlaps that may be a natural consequence of the complex reality of law. A pragmatic method of legal text categorisation should be: (1) grounded on solid legal conceptualisations of the object of study; (2) explicit about the expansive or restrictive approach adopted with regard to legal text definitions, and aware of their relative nature and limitations; and (3) permeable to redefinitions of category boundaries and connections during the process of text analysis and classification. In other words, a balance must be struck between what is presupposed and what the corpus "tells" the researcher in order to refine classifications.

In the classification of multi-genre legal corpus components, multi-layered approaches can be helpful to test existing definitions of text types, and tailor their boundaries to the area of scrutiny and specific research needs. One of these approaches, the multidimensional model represented in Table 2.2, attempts to encapsulate the complementary nature of previous LTS approaches by connecting legal functions, text types (by discursive situation) and genres (according to more specific textual functions and conventions), from more general to more

Table 2.2 Multidimensional approach to legal text classification (Prieto Ramos 2014a, p. 265)

1 Main functions	• Govern public or private legal relations • Apply legal instruments in specific scenarios • Convey specialised knowledge on sources of law and legal relations
2 Text types	• Legislative (including treaties) • Judicial (including court and litigation documents) • Other public legal instruments or texts of legal implementation (issued by institutional bodies, public servants or registries; subtypes to be identified by legal system*) • Private legal instruments • Legal scholarly writings [*Notarial instruments can be considered as a specific category in civil law countries]
3 Genres	Textual realisations of specific legal functions following culture-bound discursive conventions (e.g. different kinds of court orders or contracts)

specific, and trying to avoid legal system bias. Similarly, from the field of Law, Rappaport (2014, pp. 222–223), inspired by Sinding (2002), proposes a multi-layered approach comparable to Russian nesting dolls: (1) sociocognitive action or "thinking as a lawyer" as "the outermost generic frame" to situate texts; (2) rhetorical situation or "type of law –patent, divorce, criminal, etc.–" as "the middle doll"; and (3) discourse structure, i.e. "the most specific genre, being the actual document, such as application, divorce decree, or jury waiver". These approaches will set the scene for the investigation of legal translation in international institutional settings.

3 The case of international organisations: surveying institutional legal translation

3.1. Research needs as a determining factor

The challenges outlined in the previous sections clearly apply to corpus building and text classification in the LETRINT project, which aims to shed light on the scope, features and quality indicators of legal translation at international organisations. With a view to situating and surveying legal translation within each representative institutional setting (EU, UN and WTO), three massive parallel corpora were compiled from institutional repositories,[3] including all publicly accessible textual production of three years (2005, 2010 and 2015) in the three common languages of these institutions: English, French and Spanish (with the exception of the ICJ, whose official languages are English and French). Each parallel subcorpus therefore includes a high volume of translated texts (amounting to several million words per institution) and a wide variety of institutional genres.

Given the ambitious mapping objective of the first phase of the project, a comprehensive approach to text compilation and classification was mandatory. Corpus boundary and internal structure definition is thus not only instrumental to other phases of the project, but also a goal in itself. This inclusive approach differs from other translation-driven corpus studies as regards its large-scale comparative dimension between institutions, and also, crucially, in that text categorisation is not restricted to a fixed number of genres that are isolated for scrutiny from the outset. Among recent examples of such studies, in the Polish Eurolect project, Biel (2016) concentrates on four genres for corpus analysis "as most prototypical and hence representative of EU communication" (2016, p. 199): (1) legislation (including regulations and directives); (2) judgments and other decisions of the EU's Court of Justice (CJEU) and the General Court; (3) administrative reports prepared by EU institutions; and (4) EU official websites (2016, pp. 202–203). She contrasts these genres with comparable monolingual corpora in Polish to study variation between genres and the Europeanisation of administrative Polish.

Also centred on EU discourses, the EU Case Law Corpus (EUCLCORP) includes judgments by the CJEU and several constitutional and/or supreme courts with a view to comparing their language (Trklja and McAuliffe 2018), while the European Parliament Translation and Interpreting Corpus (EPTIC) is an intermodal bi-directional (English-Italian) corpus of speeches primarily compiled to examine lexical simplification patterns (Bernardini *et al.* 2016). Among resources developed by institutions, the United Nations Parallel Corpus v1.0, created as a parallel corpus mostly for computer-aided translation purposes, is organised by language, publication year and document symbol, and also includes UN duty station and keywords as metadata, but provides no additional information on text type classification (Ziemski *et al.* 2016).

A further-reaching proposal of institutional text categorisation, albeit not strictly based on corpus analysis, is that of Koskinen (2014). In conceptualising institutional translation in terms of governing functions, Koskinen identifies four "regimes of practices" corresponding to "distinct areas of text production and translation" (2014, pp. 487–488): maintenance, regulation, implementation and image building. She places regulation at the centre of the model as "a core activity in governing, and core genres include legislation and other juridical and administrative texts, as well as secondary documents required by law or needed for legal processes" (see Figure 2.1). Maintenance features as "the most introverted layer", and "image-building and persuasive genres" as "the most extroverted one" in what she describes as an "overview of text types, or regimes of textual and translation practices, involved in governing" (2014, p. 488).

This classification seems to mix different text-extrinsic and intrinsic criteria, including systemic, linguistic, symbolic and pragmatic parameters, without referring to corpus-supported methodological considerations. It calls for further elaboration and explicitation, particularly with regard to the rationale of labels and subdivisions. For example, the author associates the "implementation of regulations and norms" with "a need for various informative and instructive modes of communication", but excludes these modes from image-building "persuasive,

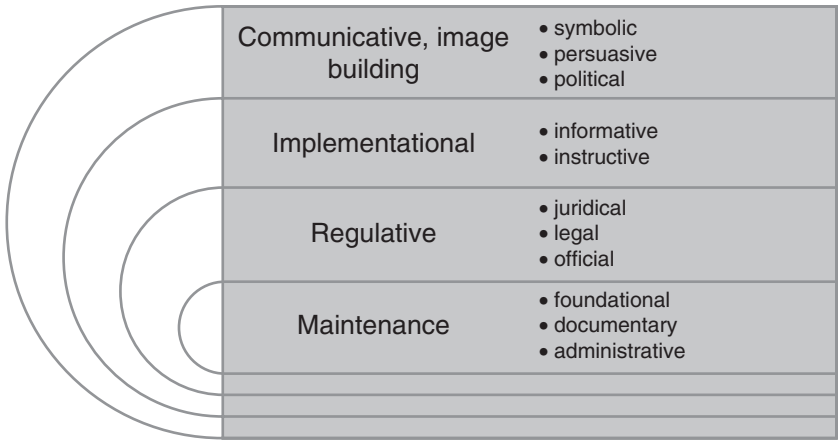

Figure 2.1 Text types in institutional translation (Koskinen 2014, p. 488)

political and symbolic genres"; she refers to "administrative texts" under regulative and maintenance categories, and seems to equate the first of these categories with "regulative" purposes. Yet, she includes "secondary documents required by law or needed for legal processes" (2014, p. 488) in this category, which would include non-regulatory texts. It is not clear whether judicial processes and adjudicative functions have been considered, why foundational documents (typically legal) are classified as "maintenance", why "official" genres or "modes of communication" are reserved for the "regulative" category, or in which way legislation is less "extroverted" than other categories.

The preceding approaches clearly illustrate how the level of detail in text categorisation is very much determined by research aims and concomitant data representativeness requirements. The broader the area of investigation and the more numerous and interrelated the textual varieties, the higher the risk of overlaps and categorisation problems. Our brief review of previous studies also suggests that more empirical data are needed to define the scope of institutional legal translation, especially at inter-governmental organisations.

3.2 The LETRINT approach

Mapping the confines of institutional translation and situating legal texts from a comparative diachronic perspective, involving three organisations and periods, could only start by defining the common denominators of institutional missions, i.e. the key functions fulfilled through comparable processes of text production. This would be the foundation for subsequent:

- selection of genres that are representative of those key institutional functions and corresponding text production processes;

- stratified (systematic) sampling (see e.g. Mellinger and Hanson 2017, p. 12), according to quantitative and qualitative criteria, in order to ensure optimal representativeness of subgroups or further strata (e.g. treaty bodies under UN treaty body reports or subcategories of EU directives);
- annotation of legal discourse features of selected genres and translation "rich points" (as defined by Agar 1991, p. 168,[4] and drawn upon in Translation Studies, e.g. Nord 1997, p. 25; PACTE 2009, pp. 212–216; Munday 2012, p. 2);
- analyses of translation quality indicators and their perception among various groups of readers (with varying levels of translation or subject matter expertise), including terminology as a key marker of both specialised discourses and translation competence.

In line with the methodological considerations outlined in section 2.3, the LETRINT approach goes from more general to more specific layers or strata of categorisation, proceeding in a "cyclical fashion" (Biber 1993, p. 256); it applies theoretically grounded conceptualisations to identify the primary categories and then refines and adds granularity according to the insights gained through text analysis.

Based on the legal contextualisation of institutional functions and the purposes of their text production processes (Prieto Ramos 2014b, 2017), three primary categories held in common were identified: (1) law-making, including hard and soft law; (2) monitoring of Member States' compliance; and (3) adjudication, including contentious and advisory proceedings (although the latter do not apply to the WTO's dispute settlement bodies). This preliminary legal contextualisation confirmed that the wide range of genres produced by the three institutions shared the same legal core as the foundation of all institutional work. Unsurprisingly, it also elicited a prototypical global hierarchy in which international legal instruments feature at the top of each institutional system and frame the other processes of application in recognisable ways.

In turn, all these processes rely on instrumental or subsidiary text categories, and are themselves the subject of other texts that describe institutional activities. As a caveat on the level of dissemination of texts, it is worth mentioning that webpages were deliberately excluded from the project because it would be materially impossible to trace them reliably for all periods and websites. Additionally, it soon became apparent that a high proportion of their web content is based on other texts considered in the project such as reports, memoranda or press releases. The exclusion of webpages would therefore have no impact on the adequacy of the compiled corpora for LETRINT's research needs.

The classification of all corpus components according to these categories entailed a dual process of: (1) identifying genres, i.e. verification of document titles, metadata and discourse features such as structural conventions and lexical markers of key legal functions; (2) situating their role with regard to the major categories and determining inter-genre connections within and between categories and subcategories. Throughout this process, it was essential to remain permeable to nuances and unexpected data, especially texts that would not easily fit into any of the main categories. Institutional document symbols often facilitated

the task of situating entire document series (e.g. WTO dispute settlement reports or EU directives). However, in other cases, document symbols or titles were of little help, and demanded closer analysis by textual unit (e.g. groups of miscellaneous communications). The manual verification of large volumes of texts by several validators (at least two LETRINT researchers per organisation, including the project supervisor) was time-consuming but yielded dividends. Given the comparative approach, the delineation of boundaries applicable to the three organisations called for a regular examination of classification issues and gradual modulation of definitions. The more advanced the categorisation work, the fewer adjustments proved necessary, until the categorisation matrix became stable.

3.3 An evolving categorisation matrix

The cyclical categorisation process confirmed the applicability of the three primary categories and shed light on their interwoven subcategories and an additional category of administrative texts, as represented in Table 2.3.

Within **major categories**, relevant subdivisions included the distinction between hard law and soft law, which was merged with other policy formulation as part of a single "law- and policy-making" macro-category. The distinction between the binding and non-binding nature of instruments was generally straightforward. However, the degree of legal force that a particular non-binding instrument or policy document may attain to be considered "soft law" (or "informal international law-making") is not always easy to establish, as it may ultimately depend on their influence on binding instruments or case-law (see e.g. Pauwelyn *et al.* 2012; Ştefan 2013). While all law-making can be understood as a prescriptive form of policy-making (see e.g. Plein 2016), policy formulation might adopt a variety of other shapes in the pursuit of institutional objectives, and they constitute a fuzzy area for categorisation purposes from a legal perspective. Accordingly, in the case of monitoring, a distinction is made between: (1) mandatory compliance monitoring procedures (e.g. universal periodic review at the UN or infringement procedures at the European Commission, which in fact may resemble judicial procedures (see Prieto Ramos 2017, pp. 199–206)); (2) pre-accession monitoring (more prevalent at the WTO); and (3) other monitoring and implementation matters, i.e. coordination and follow-up of States' policies in the framework of cooperation mechanisms. Finally, the added category of "administrative functions", i.e. devoted to the functioning of the institution itself, included two large subgroups in connection with human resources, finance and procurement procedures, and other coordination and internal matters. This category may be considered as globally instrumental and gravitates around the others, as administrative housekeeping is necessary for the smooth running of all activities.

Typically "administrative" texts such as meeting agendas or procedural notes are also found as "instrumental" types within **subordinated categories**, i.e. within the second level of classification based on the relevance of texts to the main functional category. The key genres are those that perform the main functions (e.g. judgments in adjudication or regulations in law-making), while secondary

Table 2.3 LETRINT text categorisation matrix

MAIN FUNCTIONAL CATEGORIES	*SUBCATEGORIES BASED ON RELEVANCE TO MAIN FUNCTION (ILLUSTRATIVE GENRES)*
1 LAW- AND POLICY-MAKING	
1.1 HARD LAW	**a** Key (e.g. treaties, agreements, regulations, directives) **b** Secondary (input, instrumental or derived) (e.g. technical reports, proposals, minutes)
1.2 SOFT LAW AND OTHER POLICY FORMULATION	**a** Key (e.g. declarations, resolutions, guidelines, model laws) **b** Secondary (input, instrumental or derived) (e.g. records, technical reports, letters)
2 MONITORING	
2.1 MANDATORY COMPLIANCE MONITORING	**a** Key (e.g. States' reports, monitoring bodies' reports) **b** Secondary (input, instrumental or derived) (e.g. procedural notes, letters)
2.2 PRE-ACCESSION MONITORING	**a** Key (e.g. communications, questions and replies) **b** Secondary (input, instrumental or derived) (e.g. statements, minutes)
2.3 OTHER MONITORING AND IMPLEMENTATION MATTERS	**a** Key (e.g. progress reports, working papers, notes) **b** Secondary (input, instrumental or derived) (e.g. checklists, letters)
3 ADJUDICATION	**a** Key (primary case documents, e.g. requests, appeals, judgments) **b** Secondary (input, instrumental or derived) (e.g. activity reports, summaries, press releases)
4 ADMINISTRATIVE FUNCTIONS (not included in other categories)	
4.1 ORGANISATION'S HUMAN RESOURCES, FINANCE AND PROCUREMENT	(e.g. budgets, recruitment notices, calls for tenders, staff notices)
4.2 OTHER COORDINATION AND INTERNAL MATTERS	(e.g. minutes, notes, presentations, reports)

genres: (1) address **preparatory work** or provide **input** for the production of the key genres (e.g. treaty negotiation documents or technical reports); (2) play a purely **instrumental** role (e.g. meeting agendas or checklists); or (3) are **derived** genres that describe the main institutional functions for institutional follow-up or general dissemination purposes (e.g. activity reports or press releases). A high proportion of these secondary genres are found across categories, but not all of them are equally relevant to the four main categories. For instance, in the case of the administrative category, primary and secondary relevance often blurred, so genres within this category were not further classified on that basis.

At the level of **text**, each unit belongs to only one category and subcategory. According to this principle, secondary administrative texts (typically minutes) that take stock of more than one primary institutional function had to be classified as a miscellaneous subgroup of the administrative category rather than as secondary units of various other primary categories. This would avoid duplications or fragmentations of textual units for the sake of methodological consistency.

Overall, each institutional setting can be viewed as a constellation formed of **systems of genres** that are gravitationally bound and orbit around the key genres (see Figure 2.2), i.e. "interrelated genres that interact with each other in specific settings" (Bazerman 1994, p. 97). They are all interdependent within the legal framework of each organisation, and have internal (intra-institutional) and external (inter-governmental, inter-institutional and general dissemination) interfaces.

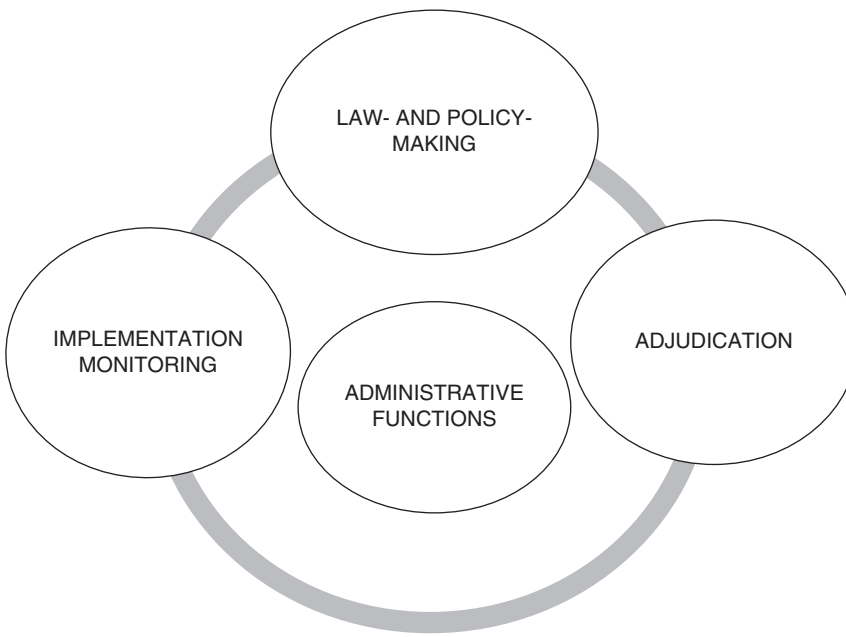

Figure 2.2 LETRINT primary functional categories

Their internal hierarchy (with legal instruments at the top) is comparable, as mentioned in section 3.2, but the size and focus of each system differ between organisations. For example, adjudicative functions are much more prominent at the WTO than at the UN. A closer examination reveals that specific bundles or chains of genres also exist within these systems (e.g. trade policy reviews systematically generate government reports, Secretariat reports, minutes and press releases), and that further **strata** can be identified within genres for sampling purposes according to quantitative and qualitative criteria (e.g. procedural, authorship or thematic considerations).

Text producers with very diverse profiles contribute in varying degrees to the circulation and perpetuation of *sui generis* discourse conventions within each institutional setting, including specialist legal drafters (particularly international lawyers and, where relevant, international judges and court staff), political representatives and technical experts. The closer to the core of key legal functions, the more recognisable legal discourse conventions are to be expected. The text mapping so far reveals the link between main functions and legal discourse features, particularly lexical markers, as well as the intermingling with other specialised discourses, not only in secondary preparatory genres but also in key ones (e.g. long technical annexes in EU legislation and dispute settlement body reports). These aspects will be further examined by the LETRINT project.

For the methodological purposes addressed here, the categorisation results may support at least three approaches to defining the scope of institutional legal translation as the first objective of LETRINT:

1 A more restrictive approach including representative key genres of hard law, mandatory compliance monitoring and adjudication, i.e. focusing on the creation and enforcement of legal obligations and the related case-law.
2 A less restrictive approach also including genres of soft law and other implementation monitoring, but excluding the administrative macro-category and all secondary genres.
3 A more inclusive approach that would consider all genres, i.e. adopting an expansive definition of institutional legal translation determined by setting, including legal and administrative text types.

In terms of research design, this decision has a number of implications for the subsequent analysis of representativeness, stratified sampling and balancing of corpus components in the next phases of the project. In all scenarios, for the sake of research validity, generalisations must be explicit about the legal contextualisation of selected categories and subcategories within the constellation of institutional functions, and they must take account of the insights provided by further corpus analysis. In other words, adjustments to the matrix and selected strata are possible as the research progresses, and definitions may be fine-tuned according to new findings. For instance, in the third scenario, the scope might be described as "institutional legal and administrative translation" or simply acknowledge that "institutional legal translation" (as a *sui generis* area of practice) integrates policy,

technical and administrative dimensions of public law. This does not imply that texts which do not belong to a legal genre or deal with legal matters should be considered as legal texts in their own right.

4 Concluding remarks

The categorisation of texts lies at the heart of research design in Translation Studies, as it draws on the boundaries and underlying conceptions of the object of study, and conditions data representativeness and findings validity. In LTS, the definition of boundaries remains a seminal debate about the nature of legal texts and the scope of the field. Šarčević's (1997, p. 55) well-known definition of legal translation as "an act of communication *within* the mechanism of the law" (our emphasis) can be interpreted in a restrictive or expansive way, as law frames all aspects of life, while texts *about* the law, such as legal scholarly texts, are also generally regarded as legal texts. In fact, legal translation and legal genres, like law itself, embraces all kinds of technical discourses and covers as broad a scope as legal function and legal settings can reach. The more expansive and setting-oriented the categorisation approach, the more text types and internal subdivisions might be elicited. In classifying them as interrelated sets of genres, taxonomies based on discursive situation parameters tend to agree on the link between legal functions or themes and legal discourse features. Discourse-oriented categorisations may accordingly include texts of non-legal genres that deal with legal subjects, and exclude other non-legal texts that contain no legal discourse but might be used in legal settings.

Multidimensional approaches combining legal context of text production, legal functions and genre conventions have been advocated for as particularly suited to illuminating the different layers of text types, their central or ancillary positions in relationship to each other, and hence the boundaries and internal structure of the object of study. They may vary depending on research aims, theoretical grounds and legal system-bound factors. The researcher must be rigorous and explicit about these considerations, their constraints and their impact on research design. Permeability to new data and regular testing is required to provide granularity on the variations and fuzzy areas of hybrid discourses. The fabric of a corpus itself may lead the researcher to reconsider pre-conceived ideas about legal texts and language, or to redefine the scope of the research. In the case of international institutional translation settings, a short review of corpus-based categorisations confirms that classification granularity levels are very much determined by the breadth and depth of the research goals.

The first phase of the LETRINT project has served to illustrate the preceding considerations. Since it seeks to situate and characterise legal translation in international institutional contexts, a comprehensive mapping was necessary to dissect layers of primary and secondary institutional functions from a legal comparative perspective. A cyclical multi-layered categorisation of three parallel corpora reaffirmed the applicability of three major functional categories composed of interconnected networks of key and secondary genres. It also confirmed, among other

aspects, the instrumental role of an additional administrative category, as well as the fuzzy area between soft law instruments and policy documents. The resulting categorisation matrix may be viewed as a dynamic constellation of genres that may further evolve as new insights emerge from corpus analysis. More importantly, this analysis must be sensitive to the implications of more expansive or restrictive approaches to institutional text genres for subsequent research stages, such as the selection and stratified sampling of representative genres for further analysis. All definitions and labels can ultimately be justified in light of the lens of observation, but only those supported by consistent methodological choices can be empirically sound.

Acknowledgements

I would like to thank Giorgina Cerutti, Diego Guzmán and Albert Morales for their contributions to the categorisation work referred to in this paper, as well as the Swiss National Science Foundation for its support through a Consolidator Grant.

Notes

1 More precisely, for the purposes of the project, the main EU institutions include the European Commission, the Council of the EU, the European Parliament and the Court of Justice of the EU. In the case of the UN, the International Court of Justice (ICJ) is considered as the main judicial body of the organisation, while the WTO's adjudicative bodies include dispute settlement panels and the Appellate Body.
2 "Text type" will be considered here as an umbrella term to refer to supra-genre categories of texts according to a definition or set of distinctive characteristics, while "text typology" will be understood as the overall classification of texts, including subdivisions at genre or supra-genre level.
3 As indicated in the introduction, given the focus of this paper, technical details of corpus compilation will not be addressed here.
4 This anthropologist described "rich points" as "things –lexical items through speech acts up to extensive stretches of discourse–" that "strike you with their difficulty, their inability to fit into the resources you use to make sense of the world".

References

Abdel Hadi, M. (1992). Géographie politique et traduction juridique, le problème de la terminologie. *Terminologie et traduction*, 2/3, pp. 43–57.

Agar, M. (1991). The biculture in bilingual. *Language in Society*, 20, pp. 167–181.

Alcaraz Varó, E. and Hughes, B. (2002). *Legal translation explained*. Manchester: St. Jerome Publishing.

Bazerman, C. (1994). Systems of genres and the enactment of social intentions. *In*: A. Freedman and P. Medway, eds. *Genre and the new rhetoric*. London: Taylor & Francis, pp. 79–101.

Bernardini, S., *et al.* (2016). From EPIC to EPTIC: Exploring simplification in interpreting and translation from an intermodal perspective. *Target*, 28 (1), pp. 58–83.

Berūkštienė, D. (2016). Legal discourse reconsidered: Genres of legal texts. *Comparative Legilinguistics*, 28, pp. 89–117.

Bhatia, V.K. (1993). *Analysing genre: Language use in professional settings*. London: Longman.

Bhatia, V.K. (2006). Legal genres. *In*: K. Brown, ed. *Encyclopedia of language and linguistics*. 2nd ed. Oxford: Elsevier, pp. 1–7.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8 (4), pp. 243–257.

Biber, D., *et al.* (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Biel, Ł. (2014). *Lost in the Eurofog: The textual fit of translated law*. Frankfurt am Main: Peter Lang.

Biel, Ł. (2016). Mixed corpus design for researching the Eurolect: A genre-based comparable-parallel corpus in the PL EUROLECT project. *In*: E. Gruszczyńska and A. Leńko-Szymańska, eds. *Polskojęzyczne korpusy równoległe. Polish-language parallel corpora*. Warsaw: Instytut Lingwistyki Stosowanej, pp. 197–208.

Bocquet, C. (1994). *Pour une méthode de traduction juridique*. Prilly and Lausanne: Éditions C. B.

Borja Albi, A. (2000). *El texto jurídico inglés y su traducción al español*. Barcelona: Ariel.

Borja Albi, A. (2013). A genre analysis approach to the study of the translation of court documents. *Linguistica Antverpiensia, New Series – Themes in Translation Studies*, 12, pp. 33–53.

Cao, D. (2007). *Translating law*. Clevedon: Multilingual Matters.

Fairclough, N. (2006). Genres in political discourse. *In*: K. Brown, ed. *Encyclopedia of language and linguistics*. 2nd ed. Oxford: Elsevier, pp. 32–38.

Fontanet, M. (2018). Translating hybrid legal texts for science and technology institutions: The case of CERN. *In*: F. Prieto Ramos, ed. *Institutional translation for international governance: Enhancing quality in multilingual legal communication*. London: Bloomsbury, pp. 126–138.

Forsyth, R.S. and Sharoff, S. (2014). Document dissimilarity within and across languages: A benchmarking study. *Literary and Linguistic Computing*, 29, pp. 6–22.

Galdia, M. (2009). *Legal linguistics*. Frankfurt am Main: Peter Lang.

Gémar, J-C. (1995). *Traduire ou l'art d'interpréter. Tome I: Principes. Fonctions, statut et esthétique de la traduction. Tome II: Application. Langue, droit et société: Éléments de jurilinguistique*. Sainte-Foy: Presses de l'Université du Québec.

Goźdź-Roszkowski, S. (2011a). *Patterns of linguistic variation in American legal English. A corpus-based study*. Frankfurt am Main: Peter Lang.

Goźdź-Roszkowski, S. (2011b). Legal language. *In*: C.A. Chapelle, ed. *The encyclopedia of applied linguistics*. Chichester: Wiley-Blackwell, pp. 3281–3289.

Halverson, S. (1998). Translation studies and representative corpora: Establishing links between translation corpora, theoretical/descriptive categories and a conception of the object of study. *Meta: Translators' Journal*, 43 (4), pp. 494–514.

Harvey, M. (2002). What is so special about legal translation? *Meta: Translators' Journal*, 47 (2), pp. 177–185.

Kjær, A.L. (2000). On the structure of legal knowledge: The importance of knowing legal rules for understanding legal texts. *In*: L. Lundquist and R.J. Jarvella, eds.

Language, text and knowledge. Mental models of expert communication. Berlin: de Gruyter Mouton, pp. 127–161.
Koskinen, K. (2014). Institutional translation: The art of government by translation. *Perspectives*, 22 (4), pp. 479–492.
Lee, D. (2001). Genres, registers, text types, domains, and styles: Clarifying the concepts and navigating a path through the BNC jungle. *Language Learning and Technology*, 5 (3), pp. 37–72.
Maley, Y. (1994). The language of the law. *In*: J. Gibbons, ed. *Language and the law*. London: Longman, pp. 11–50.
Mattila, H.E.S. (2013). *Comparative legal linguistics*. 2nd ed. Aldershot: Ashgate.
McEnery, A.M., *et al.* (2006). *Corpus-based language studies: An advanced resource book*. London: Routledge.
Melissourgou, M.N. and Frantzi, K.T. (2017). Genre identification based on SFL principles: The representation of text types and genres in English language teaching material. *Corpus Pragmatics*, 1 (4), pp. 373–392.
Mellinger, C.D. and Hanson, T.A. (2017). *Quantitative research methods in translation and interpreting studies*. New York: Routledge.
Monjean-Decaudin, S. (2013). Réflexion sur l'inflexion du signifié dans la traduction juridique de Claude Bocquet. *Parallèles*, 25, pp. 19–29.
Munday, J. (2012). *Evaluation in translation: Critical points of translator decision-making*. London: Routledge.
Nord, C. (1997). *Translating as a purposeful activity*. Manchester: St. Jerome Publishing.
PACTE (2009). Results of the validation of the PACTE translation competence model: Acceptability and decision making. *Across Languages and Cultures*, 10 (2), pp. 207–230.
Pauwelyn, J., *et al.*, eds. (2012). *Informal international lawmaking*. Oxford: Oxford University Press.
Plein, C. (2016). A Policymaking checklist for the legislative process. *West Virginia Law Review Online* [online], 6 (15). Available at: http://wvlawreview.wvu.edu/west-virginia-law-review-online/2016/01/20/a-policymaking-checklist-for-the-legislative-process [Accessed 20 May 2007].
Pontrandolfo, G. (2016). *Fraseología y lenguaje judicial. Las sentencias penales desde una perspectiva contrastiva*. Roma: Aracne.
Prieto Ramos, F. (2014a). Legal Translation Studies as interdiscipline: Scope and evolution. *Meta: Translators' Journal*, 59 (2), pp. 260–277.
Prieto Ramos, F. (2014b). International and supranational law in translation: From multilingual lawmaking to adjudication. *The Translator*, 20 (3), pp. 313–331.
Prieto Ramos, F. (2017). Global law as translated text: Mapping institutional legal translation. *Tilburg Law Review*, 22 (1–2), pp. 185–214.
Rappaport, B. (2014). A lawyer's hidden persuader: Genre bias and how it shapes legal texts by constraining writers' choices and influencing readers' perception. *Brooklyn Journal of Law and Policy*, 21 (1), pp. 197–263.
Robinson, W. (2005). How the European Commission drafts legislation in 20 languages. *Clarity*, 53, pp. 4–10.
Šarčević, S. (1997). *New approach to legal translation*. The Hague: Kluwer Law International.
Sinding, M. (2002). After definitions: Genre, categories, and cognitive science. *Genre*, 35 (2), pp. 181–219.

Ştefan, O. (2013). *Soft law in court: Competition law, State aid and the Court of Justice of the European Union*. The Hague: Kluwer.

Tiersma, P.M. (1999). *Legal language*. Chicago: University of Chicago Press.

Tiersma, P.M. (2003). *The creation, structure, and interpretation of the legal text* [online]. Available at: www.languageandlaw.org/LEGALTEXT.HTM [Accessed 15 Oct. 2017].

Trklja, A. and McAuliffe, K. (2018). The European Union Case Law Corpus (EUCLCORP): A multilingual parallel and comparative corpus of EU Court judgments. *In*: A.U. Frank, *et al.*, eds. *Gerastree Proceedings*, 1, pp. 217–226.

Trosborg, A. (1997). *Rhetorical strategies in legal language: Discourse analysis of statutes and contracts*. Tübingen: Gunter Narr Verlag.

Ziemski, M., *et al.* (2016). The United Nations parallel corpus v1.0. *In*: *Language resources and evaluation (LREC'16)* [online]. Available at: http://conferences.unite.un.org/UNCorpus/Content/Doc/un.pdf [Accessed 20 May 2017].