Formulaic metadiscursive signalling devices in judgments of the Court of Justice of the European Union

Trklja, Aleksandar; McAuliffe, Karen

DOI:
10.1558/ijsll.36920

Document Version
Peer reviewed version

Citation for published version (Harvard):
Trklja, A & McAuliffe, K 2019, 'Formulaic metadiscursive signalling devices in judgments of the Court of Justice of the European Union: a new corpus-based model for studying discourse relations of texts', *International Journal of Speech Language and the Law*, vol. 26, no. 1, pp. 21–55. https://doi.org/10.1558/ijsll.36920

Formulaic metadiscursive signalling devices in judgments of the Court of Justice of the European Union: A new corpus-based model for studying discourse relations of texts[*]

1. Introduction

The aims of this paper are twofold: first, the paper investigates how paragraph initial metadiscursive lexical items serve as signalling devices in text organization of judgments of the Court of Justice of the European Union (ECJ). Secondly, the results of that investigation demonstrate that those metadiscursive items both shape a particular method of reasoning used by the ECJ, and have an impact on how readers process those texts. The corpus linguistics methodology used in the present paper also provides a new model for the study of discourse organization of texts more generally. Furthermore, extrapolating the results to the legal arena has broader implications for the understanding of EU law and the EU legal order more generally. The paper thus demonstrates the value, use and impact of linguistic research in fields beyond linguistics. The study set out in the present paper forms part of a larger research project in which the authors aim, through interdisciplinary research, to introduce a new facet to the current thinking on the development of the European Union (EU) legal order.[1]

The relationship between language and judicial reasoning has been investigated in the field of language and law studies, most notably by Lawrence Solan, whose book '*The Language of Judges*' was the first major work to examine the linguistic analysis of law. The linguistic perspective of supranational adjudication in the European context is, however, a relatively new field of research – most scholarship on language and EU law "tends to mainly involve questions of language policy and regime, interpretation of multilingual legislation and pragmatic or logistical concerns" (McAuliffe 2015). Demonstrating the impact of language patterns on the legal reasoning of the ECJ is an inherently interdisciplinary exercise that involves the exploration of the cognitive dimension of the process. Thus, the present paper contextualises the findings of the linguistic analysis carried out within the unique multilingual setting of that court using Koestler's theory of creativity and cognitive theories of text processing as the basis of analysis.

The main assumption on which the present paper is grounded is that legal language is formulaic and repetitive in nature. This assumption is based on previous research in the field(s) of linguistics and legal linguistics. Although repetition is often described as one of the most important features of legal texts (e.g. Tiersma, 1999), to date comprehensive studies that systematically investigate the degree of repetition and types and functions of repetitive expressions remain rare. Conventional linguistic studies of repetition in legal texts are mostly concerned with the repetition of legal expressions in the form of alliterative phrases (*aid and abet, rest, residue and*

[*] The studies carried out for the purposes of this paper were funded by the European Research Council. The authors would like to thank Ewelina Tylec-Bakalarz for her assistance in drafting the paper, as well as three anonymous reviewers for their detailed and very helpful feedback.

remainder) (Mellinkoff, 1963), binominals (*signed and delivered, in whole or in part*) (Gibbons, 2014), archaic vocabulary (*fail not, at your peril, concurrent tortfeasors*), (Mellinkoff, 1963) or expressions borrowed from other languages (*in propria persona, profit à prendre*) (Charrow, Crandall and Charrow, 1982). Typically, such expressions are considered to impede the understanding of texts by making them unintelligible (Tiersma, 1999; Mattila, 2013). Some refer to the use of such expressions as *stylization* in which 'form can come to dominate content, and meanings tend to become opaque' (Danet, 1980: 498-499). Furthermore, repetitive expressions can serve to highlight the authority of judges or civil servants. Mattila likens such repetition to ritualistic language that has an almost magical character, lending weight and authority to the relevant speakers (Mattila, 2013). Others point to the potential practical purposes of repetition: the application of pre-existing legal terms to new cases can form part of formalized strategies to ensure that new terms do not have to be created every time a new case comes before a court. 'Old formulae are preferred to newly-coined words because of their centuries-old history and highly codified… …accepted interpretations' (Gotti, 2008: 41). The repetition of such formula, with accepted interpretations in particular legal systems also goes to the development of the legal concept of 'precedent', particularly in common law legal systems (Tiersma 2006; McAuliffe 2013).[2]

The focus of the current paper is on the use of repetitive items in judgments of the Court of Justice of the European Union (ECJ) that perform metadiscursive rather than legal functions. Such items play a crucial role in the presentation of reasoning in judgments by signalling when and how various types of information occur in a discourse. Crystal and Davy (1969) argue that legal reasoning is mainly based on the principle: *if X, then Z shall do/be Y*. That finding indicates that such legal reasoning can be understood or 'read off' from the utterances used in a text. However, the text types investigated in Crystal and Davy's study were not only very specific, but were prescriptive and performative in nature (an endowment assurance policy and a hire purchase agreement). Not all legal texts share such characteristics and different types of legal texts, embedded in different legal orders, have a vast array of different functions. Making a general claim about the nature of legal reasoning based on a relatively narrow study without contextualising the relevant texts within a specific legal order can be problematic. The present paper builds on Crystal and Davy's analysis, but also considers the context in which the judgments of the ECJ are produced and applied. The present study also investigates differences in logico-semantic relations in the text types studied by Crystal and Davy and those considered here.

More recent studies, based on empirical data analysis, have shown that other patterns can also be found in legal texts and that their distribution may vary from one text type to another and from one legal system to another. One of the earliest attempts to classify repetitive expressions in terms of their function in legal texts was made by Kjaer (1991)[3]. She argues that repetition of expressions across texts contribute to the standardization and specialization of legal terminology. By investigating different examples of repetition, such as the use of binominals and or combinations based on the grammatical pattern Adjective + Noun, she identifies the following four types of repetitive expressions in legal texts:

1) Prefabricated word combinations directly prescribed by law.
2) Word combinations only indirectly prescribed by law.

3) Word combinations based on implicit quotations from other texts in a genre chain in the legal domain.
4) Routine phrases the use of which is merely habitual.

Kjaer's study mainly concerns repetitive expressions that perform legal functions and only the final category (4 above) contains items that have a non-legal function, such as the occurrence of *Deshalb ist Klage geboten* in writs (Klageschrift) which signals the commencing of proceedings before a court. Such expressions may indeed signal textual organization but, since this study was conducted before the use of corpus methods became widespread, Kjær provides only illustrative examples and does not provide a systematic view on their function.

Salmi-Tolonen (1994) investigates the distribution of adverbial clauses of conditional, purposive and procedural expressions in certain UK and EU legal texts. Her small corpus consists of three UK Acts of Parliament, two 'European Communities Conventions', three EU Directives and 12 pages of the Treaties Establishing the European Communities. This study is innovative in that it examines the occurrence of those three kinds of expressions in three different sentential positions: at the beginning, in the middle and at the end of a sentence. The results indicate that some of those expressions occur differently in legal texts created within the UK legal system from those created within the EU legal system. For example, conditional markers occur more often in the sentence-initial position than in other positions in EU legal texts than in the UK texts. Purposive expressions are more varied in EU texts and are more similar to usual/everyday use of language whereas in UK texts they tend to deviate from such everyday use. Finally, procedure expressions used in EU texts are more similar to those used in other genres whereas those in the UK legal texts are restricted to those particular legal texts. Salmi-Tolonen argues that due to the 'harmonizing and cooperative context' (Salmi-Tolonen, 1994: 32) of EU institutions, EU legal texts contain more variation than the national UK texts and from this she concludes that language has an influence on legal norms. However, she does not explain what exactly the link between this context and texts might be. Furthermore, Salmi-Tolonen compares texts that do not belong to the same text types and her corpus is very small.[4] For these reasons the generalizations made should be taken with caution. She does not discuss the relevance of her results for the study of discourse organization, but her findings do indicate that different types of expressions tend to be associated with different positions in discourse.

Gozdz-Roszkowski (2012) studies the distribution and functions of 4-word long lexical bundles in the following seven types of legal texts used in the American legal system: legislation, contracts, opinions, briefs, textbooks, academic journals and professional articles. All lexical bundles are grouped into three main categories: legal reference, stance and text-oriented lexical bundles. The items that contribute to the organization of texts belong to the third group (text-oriented lexical bundles) and they are further divided into several sub-categories. These sub-categories are adopted either from Biber et al (2007) (Clarification/topic elaboration bundles and Focus bundles) or from Hyland (2008) (Framing signals, Structuring bundles, Transition bundles). This classification shows that it is not possible to apply existing categories to the study of legal texts and that it is therefore necessary to combine categories from different taxonomies. Unsurprisingly, the results of Gozdz-Roszkowski's study demonstrate that not all types of text-oriented bundles occur in all types of legal texts with the same likelihood. For example, Conditional bundles are more often used in descriptive text types (textbooks and professional articles) than in directive texts

(legislation or opinions). The study is not concerned with the investigation of discourse organization but it follows from the results that although all these text types belong to the same legal language, language users rely on different linguistic resources to structure specific types of legal texts.

Other similar studies include Kopaczyk's 2013 examination of the role of lexical bundles in the standardisation of early Scottish legal language (Kopaczyk 2013); Mazzi's study on discourse functions of lexical bundles in EU-related Irish judicial discourse (Mazzi 2018), as well as Gozdz-Roszkowski (2018) which investigates the correlation between language patterns and discourse functions.[5]

Those studies in legal linguistics demonstrate that many different types of discourse organization-related items occur in legal texts, and that their distribution depends on factors of varying legal systems and text-types. Such findings support the underlying hypothesis of the present paper, i.e. that one of the functions of repetitive expressions, in particular paragraph-initial formulaic expressions, is to signal how legal texts are structured. However, the scope of the present paper goes beyond the analysis carried out in those studies. In order to achieve the aims outlined above, it is necessary not only to identify discourse-organization items and to establish their distribution across texts, but also to investigate how such items are distributed within those texts.

Finally, the studies in legal linguistics discussed above are mainly concerned with terminological issues. The investigation of discourse-related issues in the field remains in relative infancy. The result is that the dominant view in legal linguistics, and even more so in legal studies in general, is that legal terms exist in an autonomous world. Such a platonic and atomistic view regards linguistic items outside their textual and social contexts. However, it is precisely those contexts that allow force of law to be embedded into legal utterances. Law can be considered a culture-specific communicative system (Vermeer, 2006). Legal concepts and legal language arise from the application of the linguistic resources of the legal communicative system to real life situations. In that process certain areas of life are 'juridified', i.e. they are described in terms of words used in the law, and turned into legal concepts (Vermeer 2006). In addition to building on Crystal and Davy's analysis, the present paper acknowledges the importance of textual, social and legal context in legal linguistics and thus addresses the gap in the literature by analysing metadiscursive linguistic items in ECJ judgments, which impact on the method of reasoning used by that Court, and developing a new model for studying discourse organization of texts.

2. Theoretical framework

2.1 Cognitive aspects of text production and comprehension

In 'The Act of Creation', Koestler proposes a theory that aims at explaining the processes that underlie human creativity by means of the concepts of matrix, code and strategy. *Matrix* is a general frame or pattern representing the set of possible options available to a human agent when dealing with problem-solution situations, or, in more general terms, when she carries out an activity. In terms of cognition, matrices can be seen as consisting of knowledge basis available to a human agent. When the pieces of information from the same matrix are combined the result is habitual thinking. On the other hand, when independent autonomous matrices are combined something new emerges. This is an example of the phenomenon Koestler calls *bisociation of matrices*. A *code* is a type of rule, determining the choices allowed within a given

situation. Any kind of ordered behaviour is governed by such rules. Finally, the *strategy* refers to the actual realisation of those choices. Koestler summarizes the interaction of the three notions using a chess metaphor:

> When you sit in front of the chessboard your 'code' is the rule of the game determining which moves are permitted, your 'matrix' is the total of possible choices before you. Lastly, the choice of the actual move among the variety of permissible moves is a matter of 'strategy', guided by the lie of the land – the 'environment' of other chessmen on the board. (Koestler, 1964: 41)

For the purposes of the present paper, the matrix is the ensemble of language choices available to the relevant actors. Codes are the sets of constraints that determine the valid language combinations. Finally, strategies correspond to the actual selection of linguistic expressions. Thus, the present paper identifies linguistic items in the judgments of the ECJ, which can be considered cognitive matrices in the sense of Koestler's theory and then goes on to investigate the relevant code and strategy. By so doing, the creative and/or routinized nature of the language used in ECJ judgments can be analysed, allowing conclusions to be drawn regarding the impact of that matrix on the reasoning of that court. The analysis presented here can offer new ways of understanding the workings of a multilingual legal order and can provide a starting point for scholars across disciplines to work towards limiting inconsistencies that inevitably arise in such a legal order.

Various studies from cognitive psychology (Bock, 1980; Mayer, 1984; Krug et al., 1989; Lorch, 1985; Lorch and Lorch, 1996) demonstrate that in texts certain signalling devices exist which help the reader navigate through those texts and comprehend their content. Those signalling devices include headings, titles, typographical cues and lexical items that indicate importance or summary. The experimental studies demonstrate that texts containing signalling devices decrease comprehension load, i.e. make such texts easier to read. Unstructured texts, on the other hand, are more difficult to read. The following table, adopted from Lorch (1989), summarizes main effects of signalling devices:

TABLE 1

Locus	Paradigm
Attention	Reading time (RT) on signaled information
	Eye movements to signaled information
	RT to secondary probe while reading signaled information
Reading process	Average reading time
	Average RT in a secondary probe task
	Outcome measures following reading under time pressure
Comprehension	Inference from text information
	Problem-solving in analogous situations
Memory	Recognition test
	Cued recall
	Free recall
Selective Access	Text search tasks; latency and accuracy on specific inquiries of text content

Those studies from cognitive psychology are typically concerned with the most visible devices such as titles or headings. The present paper, however, explores more subtle devices which are realized as lexical items that occur in specific positions.

These lexical devices, as explained below, have a metadiscursive function and do not indicate content. Such signalling devices not only have an impact on how readers process texts but also on how writers produce those texts.

2.2 Metadiscourse and Textual Colligation

The terms *discourse-organization items* and *text-oriented lexical bundles*, used above, lack sufficient theoretical development to form a basis for developing a new model for studying discourse organization of texts. Thus, the present paper relies on the concept of metadiscourse as elaborated by Hyland (1998, 2000). Hyland (2000: 198) defines metadiscourse as 'the linguistic resources used to organize a discourse or the writer's stance towards either its content or the reader'. Supporting the work of other scholars in the field (e.g. Hoey, 1991; 2001; Tadros, 1994) Hyland argues that texts constitute sites of interaction between writers and readers. Writers use certain linguistic resources to help readers navigate through a text. Hyland divides metadiscursive resources into *interactive* and *interactional* items. Interactive items signal discourse organization of texts, while interactional items 'involve readers in the argument by alerting them to the author's perspective towards both propositional information and readers themselves' (Hyland and Tse, 2004: 168). In other words, interactive resources are responsible for discourse organization and interactional for the writer's attitude toward propositional content.

As Hyland and Tse (2004: 158) point out, 'there are no simple linguistic criteria for identifying metadiscourse' because metadiscourse categories are open and new items can be added or removed depending on data. This means that existing categories cannot automatically be applied to the description of new texts. The present paper develops a new taxonomy by combining categories that derive from Crombie (1985), Deroey (2015) and Hyland (2000).[6] The following types of metadiscursive expressions occur in texts from the corpora (see Section 3 below) used in the present study: *Frame markers*, *Evidentials*, *Transitions*, *Importance Markers*, *Validity markers*, *Engagement markers* and *Self-mentions*.

Frame markers are lexical items that 'signal text boundaries or elements of schematic text structure' and as such they 'identify features which order arguments in the text' (Hyland, 2005: 51). Frame markers perform different functions in ECJ judgments: *sequencers* order parts of a text in sequences and *topic organisers* signal an introduction, discussion or conclusion of topic.

Evidentials are items that refer to a source of information from other texts and indicate the writer's attitude toward the knowledge provided by a source. They are used by writers to demonstrate that they 'rely on evidential support from statements previously confirmed by the discourse community as truths about the world' (Hyland, 1998: 436).

Transitions help readers to make 'connections between preceding and subsequent propositional information' (Cao and Hu, 2014). Hyland defines the role of transitions in terms of Halliday's (1985) logical relations. However, Crombie's (1985) model of relation in discourse applies more clearly to the data in the present study. Six kinds of transitions, divided into three types of relations, can be found in ECJ judgments:

Logico-deductive relations:

1) Reason-Result
2) Consideration-Conclusion

3) Condition-Consequence

Associative semantic relations

1) Contrast
2) Statement-Denial

Tempero-contigual semantic relations

1) Chronological relations

The notion of *Importance markers* is adopted from Deroey (2015) to describe metadiscursive items that mark the importance, relevance and significance in discourse. As Deroey demonstrates these lexical items have been studied under different names and many authors stress that they combine discourse organization and evaluation. *Importance markers* are similar to the category of *Boosters* from Hyland's taxonomy (2005). Boosters are also used to strengthen an argument 'by emphasizing the mutual experiences needed to draw the same conclusions as the writer' (Hyland, 2005: 53). According to Hyland's terminology, importance markers have both interactive and interactional functions (they signal information flow in a discourse by stressing what a more important piece of information is).

Validity markers signal 'the writer's commitment to the probability or truth of a statement'. (Hyland, 2005: 32) They are also used to acknowledge personal responsibility for the content of their claim and to assess the certainty of the truth of their assertions.

Engagement markers 'are devices that explicitly address readers, either to focus their attention or include them as discourse participants' (Hyland, 2005: 53). Through this engagement writers attempt to persuade their addressees to accept their positions.

Self-mentions (Hyland 2001) are similar to *Engagement markers* with the difference that they do not signal the presence of readers but that of writers in the text. By using these resources the writers indicate their own contribution to the ongoing debate.

As discussed above, Salmi-Tolonen has indicated that lexical items tend to favour or avoid certain sentential positions. This idea has been more systematically developed by Hoey (2005), who demonstrates that '[w]ords (or nested combinations) may be primed to occur (or to avoid occurring) at the beginning or end of independently recognised discourse units, e.g. the sentence, the paragraph, the speech turn' (Hoey 2005: 115). This phenomenon is called *textual colligation*. Various corpus studies show that specific words or word combinations tend to occur in the text-initial, paragraph-initial or sentence initial position (Hoey, 2004, 2005; Hoey and O'Donnell, 2008).

Previous studies (Hoey, 1985; Suomela-Salmi, 1992; Goutsos, 1997) have demonstrated that paragraph-initial expressions carry information that might signal development of a text and that paragraph boundaries can serve as textual organizers because they signal relations between the parts of a discourse. The present paper relies on those previous studies on paragraphing and textual colligation to assume that the study of paragraph-initial clusters can help to identify metadiscursive items that signal discourse organization of ECJ judgments.

3. Data and Methodology

The data analysed in the present paper comprises two corpora. The first corpus consists of the English language version of 1140 ECJ judgments and the second of

1140 judgments of the UK House of Lords delivered in the period 1953-2009 and UK Supreme Court judgments delivered between 2009-2011 (for ease of reading, this corpus is referred to as UKSC). All the data cover the same period from 1953 to 2011[7]. The sample of 1140 ECJ judgments was chosen because those judgments make up the EU *acquis communautaire* case law: the most important judgments in EU law, which must be translated and implemented into the legal order of any state wishing to join the EU.[8] The sample of judgments for the second corpus was chosen because (for the purposes of the present study) national supreme courts are the closest types of courts for comparative purposes to the ECJ. The second corpus allows a contrastive analysis of ECJ and UKSC judgments. The purpose of this analysis is to show typical metadiscursive features of ECJ judgments. These corpora have been compiled within the European Research Council (ERC) project 'Law and Language at the European Court of Justice.[9]

The Court of Justice of the European Union is the highest court in the EU legal order. Originally the Court of Justice of the European Coal and Steel Community, it has, over the course of more than sixty years, extended its own, originally limited, jurisdiction, and has transformed the EU from a traditional international organization, albeit with supranational elements, into a new type of legal order, which binds not only member states but also individuals (Harmsen and McAuliffe, 2015). The ECJ is responsible for ensuring that 'in the application of the [EU] Treaties, the law is observed'.[10] In reality this means that the ECJ hears and rules on cases of breach of EU law (direct actions) and also delivers binding judgments regarding questions of interpretation of EU law.

The judgments of the ECJ are originally created in French and then translated into 22 of the remaining 23 EU official languages.[11] Unlike EU legislation, which is considered equally authentic in all 24 official EU language versions, there is usually only one language version of an ECJ judgment which is considered authentic – the version in the language of procedure of the case in question.[12] The language of procedure of a case is the language in which the application is submitted, correspondence with the parties to the case is conducted and the language in which the 'authentic' judgment or order of the court is delivered. The language of procedure of a case can be any one of the 24 official EU languages.[13] For practical reasons the ECJ has only one working language, which is French. This means that all case applications submitted to the ECJ are translated into French to be worked on within the court, and every judgment is drafted in French. Once the French language version of a judgment has been finalised, it is then translated into the language of procedure and the other EU official languages. Interestingly, it is the version of the judgment in the language of procedure which is signed by the judges (whether or not they can understand that language) and which is considered the 'authentic' language version, rather than the French original.[14] Although only one language version of an ECJ judgment '*fait foi*', there is an assumption that all translations of a judgment will deliver the same message and have the same legal effect in all 28 EU member states.[15] Thus, national courts, legislators, lawyers and legal educators access ECJ case law in their relevant national language and consequently apply that case law based on their understanding and interpretation in those languages. The analysis set out in the present paper is concerned mainly with English language versions of *acquis communautaire* judgments. This methodological choice is justified on the basis of the strict translation policy at the ECJ that translated texts should follow original texts as closely as possible, including the number of sentences and typographical features.

The UKSC corpus was created as a comparative corpus, and so it contains the same number of texts as the ECJ corpus. The paragraph-initial metadiscursive formulaic expressions (PIMFE) occurring in texts from this corpus are identified and compared with those from the ECJ corpus. The purpose of this exercise is to compare whether authors of judgments from two different legal orders are primed to organize discourse in the same way.

In order to identify metadiscursive textual colligation (see below) first all the paragraphs in judgments were tagged for parts of speech. Second, shell-scripts were created, which enabled all 5-word long PIMFE in the corpus judgments to be searched. The length of the expressions is based on the fact that the shorter multi-word expressions are contained by the typical PIMFE. The inclusion of longer expressions, on the other hand, would significantly reduce the number of PIMFE. The expressions obtained in this way were then used as search terms to observe their occurrence at paragraph-non-initial positions. Next, the association strength of both paragraph-initial and non-initial expressions was calculated by means of the Mutual Information[16] (MI) test using Collocate 2.0. After that, MI scores of the identified expressions were compared. All expressions with a higher MI score in paragraph-initial than in paragraph-non-initial positions were considered to be strongly associated with the paragraph-initial position. Finally, using the classification model discussed above they were grouped into sections. Since the focus of the present paper is on typical expressions, only those expressions that occur at least five times were included.

4. Distribution of formulaic expressions in ECJ and UK Supreme Court Judgments

The following section sets out details of the studies and analysis carried out on the corpora detailed above, and compares language patterns used by the ECJ and UKSC. In order to systematically investigate the degree of repetition in the language used by the ECJ and the functions of repetitive expressions in the judgments produced by that Court, the following studies were conducted: repetition degree; the occurrence of 5-word clusters; and PIMFE in those judgments. Replicating those studies on the corpus of judgments from the UKSC then allowed comparisons to be made between the two courts regarding the use and impact of language in their respective jurisdictions. Finally, setting out the results in the context of how ECJ judgments are actually produced allows conclusions to be drawn regarding differences between impact of language on the judgments produced in the two jurisdictions.

4.1 PIMFE in ECJ judgments

1) Repetition Degree

Sample:	5 samples consisting of 100[17] words, each containing ECJ judgments dated from 1953-2011, randomly selected using R package ‘randomizeR’.
Units of Analysis:	All multiword expressions which are at least 5 words long and which occur at least twice.
Results:	Degree of repetition: range = 31%-61%; mean and median = 47%; standard deviation = 6.7

The aim of measuring repetition degree is to identify all those recurrent formulaic expressions in individual judgments which have also been used in previous judgments. Using the units of analysis described above allows the reuse of whole sentences in the sample judgments to be studied. Once the repetition degree is obtained for all judgments from the three samples, the central tendency can be calculated. The results show that the distribution is symmetric and the mean and median are equal (47%). As indicated above, the degree of repetition ranges between 31% and 61% and the standard deviation is 6.7. These results indicate a high degree of repetitiveness in ECJ judgments, but a comparison with the results of analysis on UK Supreme Court judgments is necessary in order to develop a clearer picture.

2) 5-word clusters

Sample:	5 samples of 100 judgments each, randomly selected using R package 'randomizeR'
Units of Analysis:	All 5-word clusters occurring at least 5 times (identified using WordSmith tools)
Results:	Average number of 5-word clusters per judgment = 20.

5-word clusters were chosen as units of analysis here because they match the length of (PIMFE) in the study. The average number of 5-word clusters per judgment was calculated by dividing the total number of clusters by the number of texts in the relevant samples and then calculating the mean value for the five samples. Again, a comparison of results with those from UK Supreme Court judgments is necessary before any conclusions may be drawn. Since the judgments from the ECJ and UKSC corpora differ in length, in order to produce results that are comparable, the analysis of distribution of clusters is based on the data as standardised using a z-score. This analysis shows that the average number of 5-word clusters per judgment is 20.

3) Distribution of PIMFE

Sample:	Corpus of 1140 English-language *acquis communautaire* ECJ judgments
Units of Analysis:	PIMFE (5-word clusters)
Results:	¼ of paragraphs in ECJ judgments begin with PIMFE; interactive resources = 91%; interactional resources = 9%; transitions = 39%; frame markers = 34%; importance markers = 14%; evidentials = 13%; boosters = 1%.

There are 12,744 different paragraph-initial 5-word clusters in the English language ECJ corpus, the frequency of which ranges between one occurrence and 58 occurrences. Using the procedure of analysis of association strength based on MI scores as described in section 3 above, 98 clusters that occur at least twice were identified as strongly associated with the paragraph-initial position. The least frequent of those clusters occur in six paragraphs and the most frequent in 58 paragraphs. Although those 98 clusters make up only about 8% of all clusters that occur in the paragraph-initial position at least twice, on average every fourth paragraph in ECJ judgments begins with one of these clusters.

The next stage of analysis of PIMFE investigated the function of those 98 clusters using the taxonomy described in Section 2.2 above. The results demonstrate

that interactive resources are far more frequent (91% of all clusters) than interactional resources (9%). The most frequently used types of metadiscursive items are *Transitions* (39%) and *Frame markers* (34%). *Importance markers*, *Evidentials* and *Boosters* are less frequent, occurring with values of 14%, 13% and 1% respectively.

There is a very strong correlation (r=.95) between the frequency of items belonging to individual categories and the number of items within those categories. It follows that the higher number of items in a category, the higher the total frequency of items in that category. However, this does not mean that all items occur with the same frequency. The items *'it follows from the foregoing'* and *'it is clear from the'*, for example, together make up 38% of all *Transitions* from the group *Consideration-Conclusion*.

Figure 1 demonstrates how specific sub-categories of PIMFE are distributed in terms of frequency and the number of items within categories. The upper part of the graph in Figure 1 includes types of PIMFE that occur with the highest frequency and that contain the largest number of items: *Transitions:Consideration-Conclusion*, *Frame markers:Topic organizers:Question*, *Frame markers:Topic organizers:Answer*, *Importance markers* and *Evidentials.*

FIGURE 1 HERE.

Figure 2 displays the distribution of PIMFE within categories. For ease of representation, the frequencies are grouped into units of ten. For example, in the category *Evidentials* there are 15 items that occur between five and nine times, two that occur between 11 and 19 times and one item that occurs more than 20 times. Labels are also provided in an abridged form to enhance readability. It can be observed that within almost every category there are sets of items that occur with lower frequency and one or more items that occur frequently. These figures indicate how, through language use, certain items have emerged as default expressions that judgment authors select when seeking to express specific kinds of metadiscursive meaning.

FIGURE 2 HERE.

Five types of *Transition* items were identified in ECJ judgments. However, the results demonstrate that connections between preceding and subsequent propositional information are signalled mainly in terms of logico-deductive relations. These relations are signalled by the *Consideration-Conclusion* pattern, which indicates that the argumentation of the Court proceeds from a discussion (*Consideration*) to a conclusion (*Conclusion*). For example, *It follows from those considerations, It follows from the foregoing, It is apparent from the', It is clear from the.*

The most important *Frame markers* in ECJ judgments are *Topic organizers* that signal either the introduction or conclusion of a topic. In particular, topics are introduced and concluded mainly through expressions that signal questions and answers. Questions and answers are typically signalled by the following expressions: *In its first question the*; *The questions referred to the, Is it of any significance, Does it make any difference; The answer must therefore be*; *The answer to the third.*

Importance markers are used when the Court highlights certain points in its argumentation. Such items occur in the section of judgments labelled *Consideration* above. Paragraphs containing Importance markers precede paragraphs containing *Consideration-Conclusion* clusters. This is illustrated in the following example in

which paragraph 18 contains an Importance marker and paragraph 19 a *Consideration-Conclusion* item.

1. Case: 61980CJ0055

18 **IT SHOULD BE OBSERVED NEXT** THAT NO PROVISION OF NATIONAL LEGISLATION MAY PERMIT AN UNDERTAKING WHICH IS RESPONSIBLE FOR THE MANAGEMENT OF COPYRIGHTS AND HAS A MONOPOLY ON THE TERRITORY OF A MEMBER STATE BY VIRTUE OF THAT MANAGEMENT TO CHARGE A LEVY ON PRODUCTS IMPORTED FROM ANOTHER MEMBER STATE WHERE THEY WERE PUT INTO CIRCULATION BY OR WITH THE CONSENT OF THE COPYRIGHT OWNER AND THEREBY CAUSE THE COMMON MARKET TO BE PARTITIONED. SUCH A PRACTICE WOULD AMOUNT TO ALLOWING A PRIVATE UNDERTAKING TO IMPOSE A CHARGE ON THE IMPORTATION OF SOUND RECORDINGS WHICH ARE ALREADY IN FREE CIRCULATION IN THE COMMON MARKET ON ACCOUNT OF THEIR CROSSING A FRONTIER; IT WOULD THEREFORE HAVE THE EFFECT OF ENTRENCHING THE ISOLATION OF NATIONAL MARKETS WHICH THE TREATY SEEKS TO ABOLISH.

19 **IT FOLLOWS FROM THOSE CONSIDERATIONS** THAT THIS ARGUMENT MUST BE REJECTED AS BEING INCOMPATIBLE WITH THE OPERATION OF THE COMMON MARKET AND WITH THE AIMS OF THE TREATY.

Evidentials clusters perform two functions: first, they refer to the source of the relevant information (*as the court has held; as the court has stated; as has already been stated*); secondly they appear to signal authorship of statements (*the commission points out; that the appellant in the main; the commission maintains that the; the accused in the main; the applicant maintains that the*). The authors named are either parties involved in the case, interveners to the case,[18] or the ECJ itself. When either of the former two authors are mentioned, the relevant clusters signal a claim submitted by parties such as *the commission points out that* and *the accused in the main*. It is important to note, however, that in many cases such *Evidentials* signal the authorship of *ideas* rather than specific statements. Since ECJ judgments are collegiate judgments, references to statements of parties or interveners to a case may well be paraphrased from the actual statements made. Furthermore, the original statements of the parties/interveners may have been made in a language other than French. In such cases, since ECJ judgments are drafted in French, there will be another layer of 'paraphrasing' insofar as judgment authors will actually be referring to the French translations of the relevant submissions. Expressions referring to the ECJ, such as, *As the court has already* and *As the court has held* are used to establish a link between previous cases and the case in question. Reference to previous cases serves to strengthen the argument of the court. This is a common legal practice, particularly in common law legal systems, and is associated with the notion of *precedent*[19] (e.g. Schauer, 1987). *Evidentials* items occur in the section *Consideration* in which the Court discusses the legal issues of the case in question.

Some other PIMFE have more than one function. For example, in addition to signalling the *Consideration-Conclusion* relation, the cluster *In view of the foregoing* is also occasionally used as a *Topic organizer: Answer*, for example

2. 61994CJ0194: **In view of the foregoing considerations**, it must be concluded that Directive 83/189 is to be interpreted as meaning that breach of the obligation to notify renders the technical regulations concerned inapplicable, so that they are unenforceable against individuals.
   61995CJ0124: **In view of the foregoing considerations**, the answer to be given must be that...

As noted above, the ECJ judgments in the present corpus were published between 1953 and 2011. In order to establish whether the use of PIMFE in those judgments remains stable over time, the distribution of PIMFE from the five most frequent categories across that timeframe was also examined. Since the documents in question differ in length this analysis was carried out on normalized texts. The results demonstrate that clusters from all five categories occur in judgments from every decade. However, there are differences in the number of clusters per document. Figure 3 below displays the distribution of texts from the corpus across years and the number of PIMFE from the category *Transitions:Consideration-Conclusion*. The texts are grouped into decades. The pattern can be observed in 890 different files from the corpus across 53 years. The plot shows that the number of clusters per document has increased since the 1970s and that judgments from more recent years contain more clusters than those from earlier years. Similar results can be observed in relation to the other main categories. These results suggest that metadiscursive expressions have become more and more established over time. To put it another way, drafters of ECJ judgments have become more fluent in signalling the textual organization of those documents. The frequent repetition of these metadiscursive items by judgment drafters leads to increased formulaicity in those judgments.

FIGURE 3 HERE.

4.2 PIMFE in UK Supreme Court Judgments

This section sets out the results of the analysis of general repetitiveness and of the occurrence of PIMFE in the UKSC corpus. The methods used in the investigations of the ECJ judgments corpus, described above, were replicated to allow comparison of results.

1) Repetition degree

Sample:	5 samples consisting of 100 words, each containing UK judgments dated from 1953-2011, randomly selected using R package ‘randomizeR’.
Units of Analysis:	All multiword expressions which are at least 5 words long and which occur at least twice.
Results:	Degree of repetition: range = 21%-66%; mean = 36%; median = 37%; standard deviation = 8.2.

2) 5-word clusters

Sample:	5 samples of 100 judgments each, randomly selected using R package ‘randomizeR’

Units of Analysis:	All 5-word clusters occurring at least 5 times (identified using WordSmith tools)
Results:	Average number of 5-word clusters per judgment = 11.

These results indicate that the percentage of re-used formulaic expressions in UKSC judgments is not trivial. These results are compared with those from the analysis of ECJ judgments in Section 4.3 below.

3) Distribution of PIMFE

Sample:	Corpus of 1140 UK judgments
Units of Analysis:	PIMFE (5-word clusters)
Results:	Interactive resources = 40%; interactional resources = 60%; self-mentions= 50%; transitions = 18%; frame markers = 15%; importance markers = 5%; engagement markers = 4%; evidentials = 4%; validity markers = 3%.

There are 7120 different 5-word clusters, and their frequency range between one occurrence and 73 occurrences. There are 78 clusters in the UKSC corpus that occur at least five times and are strongly associated with the paragraph-initial position. All clusters can be classified into the following main categories *Engagement markers, Evidentials, Frame markers, Importance markers, Self-mentions, Transitions* and *Validity markers.* The following graph (Figure 4) displays the distribution of PIMFE in relation to sub-categories.

FIGURE 4 HERE.

Only two groups of clusters depart from the area of the graph in which the majority of categories can be observed: *Self-mentions* and *Transitions:Consideration-Conclusion*. These two categories contain the highest number of items per group and the items that occur with the highest frequency. The next most frequently observed types of PIMFE are *Frame markers:Topic Organizers:Introduction items* (17%). Other types of PIMFE make up less than 10% of all clusters. Although the number and the frequency of clusters correspond to each other across sub-categories, their correlation strength is weaker than in ECJ judgments (r=.64 as opposed to r=.98). As can be seen in the results above, *Self-mentions* make up 50% of all PIMFE that occur in UKSC judgments In one sense, this result is unsurprising since UKSC judgments are presented in the form of opinions of individual judges (unlike ECJ judgments, which are presented as collegiate texts, the deliberations behind which remain secret). However, the main reason for the high incidence of *Self-mentions* is a very frequent occurrence of '*I have had the advantage'* in the group, which behaves as an outlier. This expression comprises as much as 40% of all typical PIMFE in the UKSC corpus. It is part of a larger, highly formulaic, paragraph occurring at the beginning of judgments. 80% of all judgments examined contain this expression. The following example illustrates the typical form and content of this paragraph:

LORD KEITH OF KINKEL

My Lords,

I have had the advantage of having read in draft the speech, to be delivered by my noble and learned friend Lord Brightman. I agree with it, and would dismiss the appeal for the reasons he gives.

Excluding this cluster on the ground that (a) it is an outlier and (b) that it is part of a more general frozen textual section, the meaning of which is determined in terms of speech acts rather than metadiscourse, *Self-mentions* and *Transitions:Consideration-Conclusion* remain most frequent in the corpus of UK judgments. The other two types of PIMFE in these judgments signal the change of topic in the discourse and serve to highlight the importance of certain pieces of information in the text.

4.3 Comparison of Results

Although highlighting the differences in language patterns between the ECJ and UKSC is not the primary aim of this paper, studying those differences allows for a clearer analysis of the impact of language on the judgments produced by the ECJ – both in shaping the reasoning used by that court, and impacting on how readers process the texts. It is clear from the above analysis that the number of clusters per judgment tends to be higher in ECJ judgments (20) than in UKSC judgments (11). Similarly, the degree of repetition is higher in ECJ judgments (mean and median = 47%) than those of the UKSC (mean=36%, median=37%). In addition, more variation can be observed in the degree of repetition in UKSC judgments than in ECJ judgments. This comparison leads to two immediate conclusions. First, signalling devices are more strongly associated with ECJ than with UKSC judgments. Second, re-using pre-established expressions is more typical of judgment drafters in the ECJ than in the UKSC.

More intensive use of these devices in ECJ judgments indicates that ECJ discourse organization is more schematic than in UKSC judgments. In other words, serving as clues they should make readers' comprehension of ECJ judgments easier. At the same time, they help drafters to express their reasoning in a systematic way. This finding may be actually due to the fact that the judges of the UKSC draft individual judgments (whether full or concurrent) in their mother tongue, expressing their own opinions. Furthermore, they draft judgments steeped in the UK legal tradition, which are applicable only within the UK legal order. Readers/users of UKSC judgments are generally also embedded in the UK legal tradition and legal language. The drafters of ECJ judgments, on the other hand, are, for the most part, working in a language that is not their mother tongue – which in itself can lead to a tendency to rely on pre-established expressions. Furthermore, the judgments of the ECJ are collegiate documents, deliberated upon behind closed doors by the relevant chamber of that Court. As such they reflect collective reasoning. It seems that signalling devices are used to facilitate comprehension of textual sections which consist of densely packed information. Also, pre-established expressions can be useful in drafting in such a style and circumstances. Finally, ECJ judgments must be translated into 23 other languages, and applied (uniformly) in 28 EU member states, each with its own legal order. Re-using expressions which have already been translated and interpreted can help to ensure that the process of dissemination of the ECJ's case law runs smoothly. The results analysed and presented here support the work of McAuliffe (2011a, 2013) in this regard.

Not a single PIMFE from the ECJ corpus can be found in the UKSC corpus and vice versa. However, all categories and sub-categories of PIMFE observed in ECJ

judgments occur in UKSC judgments (*Attitude markers* and *Engagement markers* occur only in the UKSC corpus).

These results indicate that the main linguistic difference between ECJ and UKSC judgments is that UKSC judgments contain more interactive clusters. One could thus simply conclude that the drafters of ECJ judgments are less concerned with the engagement of readers than are the drafters of UK judgments. ECJ judgments are also more impersonal and less subjective than UKSC judgments. 48 out of 78 clusters found in UKSC judgments contain the first person singular pronoun *I* which indicates that authors of these judgments actively attempt to engage their addressees. This, however, is to be expected, since UKSC judges can (and indeed must) express their own opinions in the individual judgments that they write, whereas the judgments produced by the ECJ are collegiate, and any individual opinions of ECJ judges remain secret. Another notable difference is the lower frequency of *Importance markers* in the UKSC corpus than in the ECJ corpus. Since importance markers emphasise particular pieces of information, in addition to their engagement function, the conclusion could be drawn that ECJ judgment authors make more effort than UKSC judgment authors to draw readers' attention to specific details. *Topic organizers* occur frequently in both corpora but are not comparable since they are not of the same type. PIMFEs relating to a change of topic occur only in the UKSC corpus and clusters that signal questions and answers are found only in the ECJ corpus. *Topic organizers:Introduction* in the UK corpus and *Topic organizers:Questions* in the ECJ corpus have the same function, but in the ECJ corpus they are realized in the form of indicative statements and in the UKSC corpus as of interrogative statements.

Finally, the results of both the analyses of ECJ and UK judgments indicate that transition is realized most typically by means of *Consideration-Conclusion* clusters, which are the most numerous and most frequent PIMFE in both corpora. It follows that metadiscursive items that occur at the beginning of paragraphs play a key role in signalling argumentation and legal reasoning in judgments.

5. Discussion of results

The conclusions drawn in the above analysis can be summarised as follows:

- ECJ judgments tend to be more formulaic than UK supreme court judgments;
- The textual organization of ECJ judgments is signalled by paragraph-initial metadiscoursive formulaic expressions;
- Those metadiscursive items can be classified according to their function;
- The classes of metadiscursive items are neither randomly nor uniformly distributed, rather their distribution is associated with the strategies of legal reasoning;
- ECJ judgments are less interactional than UK supreme court judgments;

These results confirm the assumption that paragraph-initial formulaic expressions signal discourse relations in judgments. The methodology developed in this study (see section 3 above) and applied to the ECJ and UKSC corpora, allows metadiscursive textual colligation to be identified in both ECJ and UKSC judgments. The main organizational function identified in ECJ judgments relates to the introduction and conclusion of topics, emphasizing points and, most importantly, to the development of argumentation within the scope of a *Consideration-Conclusion* pattern. However, that methodology can also be used as a new model for investigating discourse relations of texts more generally. A comparative analysis would show how PIMFE are

distributed across various text types and thus make it possible to develop a typology of typical patterns. This typology would show which items are text-types specific and which are more generic.

The results also indicate that the language patterns used in judgments shape the method of reasoning used by courts, and by the ECJ in particular. The data analysed here appears to refute previous studies on the linguistic patterns that underlie reasoning in legal texts, in particular those carried out by Crystal and Davy, who claim that argumentation in legal texts relies on the underlying structure of conditional '*if…then*' sentences. The data presented here, however, demonstrates that such sentences are not among the most frequent types of logico-semantic relations in ECJ judgments. Rather, the most frequently used items belong to the *Consideration-Conclusion* pattern. Both of those patterns denote argumentation and reasoning, and both patterns appear to be associated with the same type of reasoning. Investigating more closely how such reasoning is realised may explain the difference in results. Crystal and Davy analyse very specific performative and prescriptive texts types in their study. Those texts describe *how* legal rules are to be applied (the 'how' referring mainly to 'under what circumstances'). The content of ECJ judgments however (although also both prescriptive and performative) is divided between a narrative-like and a discussion section. The first outlines the facts and/or history of the case in question and the second contains the Court's discussion of views submitted by parties and/or interveners to that case. Thus, in those prescriptive and performative texts studied by Crystal and Davy, legal rules are explicitly signalled (by the *if…then* pattern), whereas in ECJ judgments there is no such explicit signalling. The results of the present study, therefore, indicate that Crystal and Davy's findings cannot be generalised to all types of legal texts.

The results presented here also show that there no one typology of signalling devices applicable to all legal texts in a general manner. It rather seems that typologies are dependent on legal systems as well as types of legal text. However, the fact that there are certain overlapping areas in the use of these devices indicate that that some of them may be more universal than others.

Interpreting the results in the light of Koestler's theory of creativity allows analysis to be expanded in more general, theoretical terms. In the context of Koestler's concepts of matrix, code and strategy, all PIMFE identified in the present study can be considered a matrix. The code then determines the types of functions those items perform in judgments. Finally, strategies refer to actual selection of items, by the drafter(s) of judgments. It follows that certain underlying rules determine the discourse organization of judgments. As shown above, those rules are probabilistic in nature and were discovered by investigating the distribution of individual items and grouping them into categories.

The rule realized by means of *Importance markers*, for example, determines that certain points of argumentation should be emphasized and distinguished from all other points discussed in a judgment. Another example demonstrates that *Consideration-Conclusion* items belong to the rule which determines that the ECJ should provide its view after various opinions submitted to it have been discussed.

The most frequent types of PIMFE can therefore be described in terms of the following rules (ordered by the frequency of individual expressions for the ECJ, which does not necessarily correspond to the order of occurrence):

Rule 1: State the Court's opinion after a discussion of submitted arguments [*Transitions*: *Consideration-Conclusion*];

Rule 2: Introduce questions [*Frame markers:Topic organizers:Questions*];
Rule 3: Stress the importance of relevant points [*Importance markers*];
Rule 4: State answers [*Frame markers:Topic organizers:Answers*];

The rules that follow from our analysis for UKSC judgments are:
Rule 1: Mention who is speaking [Self-mentions];
Rule 2: State the Court's opinion after a discussion of submitted arguments [*Transitions:Consideration-Conclusion];*
Rule 3: Signal the change of topics [*Frame markers:Topic organizers:Change of topic*];
Rule 4: Introduce topics [*Frame markers:Topic organizers:Introduction*];

These rules constitute the core of matrices observed in ECJ and UKSC judgments respectively. They highlight similarities and dissimilarities between the judgments from each jurisdiction. The frequency of individual items reflects the strategies of selection employed by the authors of the relevant judgments.

Koestler's model also allows the distribution of PIMFEs to be interpreted across documents from different years. The fact that items from the main categories occur in ECJ judgments across all decades studied implies that the matrix in question has been firmly established, with little innovation in its organization or use since it was introduced. The more frequent occurrence of such items in more recent judgments indicates that their use has reached saturation point.

It should be noted that Koestler's theory mainly concerns the study of innovation. He argues that innovation is created when two or more matrices from different environments interact. Without such interaction, thinking becomes routinized: 'in the routines of disciplined thinking only one matrix is active at a time' (Koestler, 1964: 39-40). This fits precisely with the data in the present study, in particular in relation to ECJ judgments. The fact that the same code has been used in such a large number of different judgments across a period of almost 50 years is evidence of its productivity. However, in the context of Koestler's theory, that productivity could also lead to the conclusion that there is a low level of innovation in ECJ judgments. Interestingly, such a conclusion supports work in the field of European legal studies, which comments on the 'activist' nature of the early ECJ and on how that court has become less innovative over time (see e.g. Rasmussen, 1986, Cappelletti, 1987, Weiler 1987, Stone Sweet, 2004, Davies, 2012, Grimmel, 2014). This should not necessarily be considered negatively, indeed repetitive and routinized reasoning is an important element in embedding a rule of law. Through its case law, the early ECJ developed and extended its own jurisdiction and transformed the European Union from a traditional organization into a new type of legal order (Harmsen andMcAuliffe, 2015). As the EU legal order became more established, however, the level of innovation, in the context of Koestler's theory, seems to have dropped. This may partly be a deliberate decision on the part of judgment authors to embed the rule of law in a developing legal order, but the factors of text production within the institution itself should not be ignored. ECJ judgments are multi-authored texts produced within an institutional context and, as shown in the present paper, within a firmly established matrix which necessarily has an impact on the judgments produced. Furthermore, the usefulness of metadiscursive items should not be underestimated. Such items can be regarded as signposts that help both writers and readers to navigate their way through ECJ judgments and can thus themselves contribute to the embedding of a rule of law within the EU. Cognitively,

metadiscursive lexical items direct readers' attention to particular information in texts but they also constrain writers' reasoning and presentation of facts in argumentative texts. The use of importance markers in this regard is particularly interesting since it appears that ECJ judgment drafters make more of an effort than their UKSC counterparts to draw readers' attention to specific details. Whether this is a conscious and deliberate act on the part of those drafters, related to the nature of EU law and the jurisdiction of the ECJ itself, or simply a result of the institutional processes of text production within that Court remains to be seen. Answering such questions will necessarily involve research across both legal and linguistic fields, highlighting the importance of interdisciplinary research for a fuller understanding of the workings of the EU legal order.

Such analysis leads to the conclusion that not only do the language patterns found in ECJ judgments shape the method of reasoning used by that Court, but also that those judgments are made up of 'almost wholly automised' sub-codes of grammar and syntax. This conclusion supports the claim made by McAuliffe that ECJ judgments are created in a 'lego-building block' fashion (McAuliffe, 2011b). The linguistic research carried out here can thus be used to triangulate results of research from other fields to allow a more holistic understanding of supranational adjudication in the EU context to be developed. This in turn can offer new ways for understanding how a multilingual legal order functions, as well as allowing researchers to work towards limiting inconsistencies arising in such a legal order.

References

Biber, D., Connor, U. and Upton, T. (2007) *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. Amsterdam: John Benjamins.

Biel, L. (2014) Phraseology in Legal Translation: a Corpus-based Analysis of Textual Mapping in EU Law. In Cheng, L., Sin, K.K. and Wagner, A. (Eds.) *Ashgate Handbook of Legal Translation*. Ashgate Publishing, 178--192.

Bock, M. (1980) Some effects of titles on building and recalling text structures. *Discourse Processes,* 3: 301–311.

Cappelletti, M. (1987) Is the European Court of Justice 'Running Wild'? *European Law Review,* 12: 3--17.

Cao F., Hu G. (2014) Interactive metadiscourse in research articles: A comparative study of paradigmatic and disciplinary influences. *Journal of Pragmatics,* 66: 15--31.

Charrow, V. R., Crandall, J. A., & Charrow, R. P. (1982) Characteristics and functions of legal language. In Kittredge, R. and Lehrberger, J. (Eds.), *Sublanguage: Studies of language in restricted semantic domains*. Berlin: Walter de Gruyter, pp. 175--190.

Church, K. W. and Hanks, P. (1990) 'Word Association Norms, Mutual Information and Lexicography', *Computational Linguistics*, 16(1), pp. 22--9.

Crombie, W. (1985) *Discourse and Language Learning*: *A Relational Approach to Syllabus Design.* Oxford: Oxford University Press.

Crystal, D. and Davy, D. (1969) *Investigating English style*. London: Longman.

Danet, B. (*1980*) *Language in the Legal Process. Law and Society Review,* 14(3): 445--564.

Davies, B. (2012) *Resisting the European Court of Justice: West Germany's Confrontation with European Law, 1949-1979*. Cambridge: Cambridge University Press.

Derlén, M. (2009) *Multilingual Interpretation of European Union Law*. The Hague: Kluwer Law International.
Deroey, K.L.B. (2015) Marking importance in lectures: Interactive and textual orientation. *Applied Linguistics*, (36)1: 51–72.
Gibbons, J. (2014) *Language and the Law*. London: Routledge.
Gotti, M. (2008) *Investigating specialized discourse*. Bern: Peter Lang.
Goutsos, D. (1997) *Modeling Discourse Topic: Sequential Relations* and *Strategies in Expository Text*. Norwood: Ablex.
Gozdz-Roszkowski, S. (2011) *Patterns of linguistic variation in American legal English: A corpus-based study*. Frankfurt am Mein: Peter Lang.
Gozdz-Roszkowski, S. (2018) Between corpus-based and corpus-driven approaches to textual recurrence: Exploring semantic sequences in judicial discourse. In Kopaczyk, J., Tyrkkö, J. (Eds.) *Applications of Pattern-driven Methods in Corpus Linguistics*. Amsterdam, John Benjamins Publishing, 131--158.
Grimmel, A. (2014) The European Court of Justice and the Myth of Judicial Activism in the Foundational Period of Integration through Law. *European Journal of Legal Studies*, 7(2): online.
Halliday, M.A.K. (1985) *An Introduction to Functional Gramm*ar. London: Edward Arnold.
Harmsen, R and McAuliffe, K. (2014) 'The European Courts.' In Magone J.M. (Ed.) *The Handbook of European Politics.* London*:* Routledge.
Hoey, M. (1985), The paragraph boundary as *a marker* of relations between the parts of a discourse. *M.A.L.S. Journal* 10: 96--107.
Hoey, M. (1991) *Patterns of Lexis in Text*. Oxford: Oxford University Press.
Hoey, M. (2004) Textual colligation: a special kind of lexical priming. *Advances in Corpus Linguistics: Papers from ICAME 23,* Götenborg 22–26 May 2002.
Hoey, M. (2005*) Lexical Priming: a New Theory of Words and Language*. London: Routledge.
Hoey, M. and M.B. O'Donnell (2008) The beginning of something important? Corpus evidence on the text beginnings *of hard news stories*. In B. Lewandowska-Tomaszczyk (Ed.) *Corpus Linguistics, Computer Tools and Applications: State of the Art*. PALC 2007: 189–212.
Hyland, K. (1998) *Hedging in scientific research articles.* Amsterdam: John Benjamins
Hyland, K. (2000) *Disciplinary discourses: Social interactions in academic writing*. London: Longman.
Hyland, K. (2001) Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes*, 20: 207-226.
Hyland, K. (2005) *Metadiscourse: Exploring Interaction in Writing*. Continuum: London.
Hyland, K. (2008) As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27: 4–21.
Hyland, K. and Tse, P. (2004) Metadiscourse in academic writing: a reappraisal. *Applied Linguistics* 25 (2): 156–177.
Kjær, A. (1991) Phraseologische Wortverbindungen in der Rechtssprache? In C. Palm (Ed.) *Europhras 90. Akten der internationalen Tagung zur germanistischen Phraseologieforschung Aske/Schweden 12.-15. Juni 1990*. Uppsala.
Kjær, Anne Lise (2000) On the structure of legal knowledge: The importance of knowing legal rules for understanding legal texts. *Language, Text, and Knowledge.*

Mental Models of Expert Communication. In L. Lundquist, R. J. Jarvella (Eds.) Berlin and New York: de Gruyter. 127–161
Kjær, A. L. (2007) Phrasemes in legal texts. In H. Burger, D. Dobrovol'skij, P. Kühn and N. R. Norrick, (Eds.) *Phraseology. An international handbook of contemporary research, volume 2* Berlin: de Gruyter, 506–515.
Koestler, A. (1964) *The act of creation*. London: Hutchinson.
Kopaczyk J. (2013) *The legal language of Scottish burghs. Standardization and lexical bundles (1380-1560).* Oxford / New York: Oxford University Press.
Krug, D., George, B., Hannon, S., & Glover, J. (1989) The Effect of Outlines and Headings on Readers' Recall of Text. *Contemporary Educational Psychology,* 14 (2): 111-123.
Lorch, R. F. J. (1985) Effects on recall of signals to text organization. *Bulletin of the Psychonomic Society*, 23: 374–376.
Lorch, R. F. J., and Lorch, E. P. (1996) Effects of headings on text recall and summarization. *Contemporary Educational Psychology*, 21(3): 261-278.
Mattila, H. (2013) *Comparative Legal Linguistics:* Language *of Law, Latin and Modern Lingua Francas, 2nd ed*. London: Routledge.
McAuliffe, K (2009) 'Translation at the Court of Justice of the European Communities.' In Olsen, F., Lorz, A. and Stein, D. (Eds) *Translation Issues in Language and Law*. Basingstoke: Palgrave Macmillan.
McAuliffe, K (2011a) 'Hybrid Texts and Uniform Law? The Multilingual Case Law of the Court of Justice of the European Union' *International Journal for the Semiotics of Law* 24(1): 97-115.
McAuliffe, K. (2011b) 'The Making of EU Law: Lego Building Bricks at the ECJ.' Paper presented at the *Beyond Words workshop, RELINE, Copenhagen University 25th October 2011*, *RELINE Research Notes* http://jura.ku.dk/pdf/reline/forskningsnoter/research-note-karen-mcauliffe-lego.pdf
McAuliffe, K (2012) 'Language and Law in the European Union: The Multilingual Jurisprudence of the ECJ.' In Solan, P.M. and Tiersma, L. M. (Eds.) *The Oxford Handbook of Language and Law*. Oxford: OUP.
McAuliffe, K (2013a) 'Precedent at the ECJ: The Linguistic Aspect' *Current Legal Issues*, 15: 483-493.
McAuliffe, K (2013b) 'The Limitations of a Multilingual Legal System' *International Journal for the Semiotics of Law*, 26(4): 861-882.
Mayer, R. E. (1984) Aids to text comprehension. *Educational Psychologist*, 19: 30-42
Mazzi, D. (2018) "By partially renouncing their sovereignty…": On the discourse function(s) of lexical bundles in EU-related Irish judicial discourse. In Goźdź-Roszkowski, S. and Pontrandolfo, G. (Eds) *Phraseology in Legal and Institutional Settings. A Corpus-based Interdisciplinary Perspective*. London: Routledge., 189-202.
Mellinkoff, D. (1963) *The Language of the Law*. Boston: Little, Brown and Co.
Rasmussen, H. (1986) *On Law and Policy in the European Court of Justice: A Comparative Study in Judicial Policymaking*. Dordrecht: Martinus Nijhoff Publishers.
Stone Sweet, A. (2004) *The Judicial Construction of Europe*. Oxford: Oxford University Press.
Salmi-Tolonen, T. (1994) The linguistic manifestations of primary and secondary functions of law in the national and supranational contexts. *International Journal for the Semiotics of Law*, 2(19): 19-39.
Schauer, F. (1987) Precedent. *Stanford Law Review*, 39: 571-605.

Suomela-Salmi, E. (1992) Some aspects of the pragmatic organization of discourse. In Lindberg, A-C., Enkvist, N.E. adn Wiberg, K. (Eds.) *Nordic research on text and discourse*. Abo: Abo Akadémi Press.
Tadros, A. (1994) *Predictive categories in expository text*. In Coulthard, M. (Ed.) *Advances in Written Text Analysis*. London: Routledge.
Tiersma, P. (1999) *Legal Language*. Chicago: University of Chicago Press.
Vermeer, H. J. (2006) *Versuch einer Intertheorie der Translation*. Berlin: Frank and Timme.
Weiler, J. (1987) The Court of Justice on Trial. *Common Market Law Review*, 24: 555--589.

[1] This ERC-funded study, entitled '*Law and Language at the European Court of Justice*', examines the process behind the production of the ECJ's multilingual jurisprudence. Further details can be found on the project website: www.llecj.karenmcauliffe.com

[2] The ERC-funded project *Law and Language at the European Court of Justice* is investigating this and other linguistic aspects of the development of 'precedent'. See note 9 below and www.llecj.karenmcauliffe.com for further details.

[3] Kjær further develops her approach and methodology in Kjær (2000, 2007). However, for the purposes of the present paper, those articles do not substantially add to the argumentation introduced in Kjær (1991).

[4] Salmi-Tolonen does not, however, clearly explain which conventions she has analysed. Sampling and choice of methods can be difficult when studying particular types of legal texts. Without legal expertise it can be difficult to know whether the texts chosen are in fact comparable. The interdisciplinary nature of the present paper overcomes such problems by bringing expertise in law and linguistics together in a unique type of study.

[5] See also Biel (2014) and Trklja (2018).

[6] Only those categories identified in the corpus analysis (see section 4) were borrowed from those previous studies.

[7] The first ECJ judgments were published in 1953. The UK Supreme Court replaced the House of Lords in the UK from the 1st of August 2009.

[8] These 1140 judgments consisted of the *acquis communautaire* judgments as at the most recent accession of Croatia to the EU in 2013.

[9] This project, which examines the process behind the production of the ECJ's multilingual jurisprudence, has been funded by the European Research Council and runs from 2013 to 2019. For further information see www.llecj.karenmcauliffe.com.

[10] Article 19 of the Treaty on the Functioning of the European Union OJ C 83 30/03/2010

[11] There are currently 24 official languages of the European Union. These are, in English alphabetical order: Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Irish, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Slovakian, Slovenian, Spanish, Swedish. At the time of writing, a derogation regarding the Irish language remains in place. Under this derogation, the judgments of the ECJ are not required to be translated into Irish.

[12] Article 41 of the Rules of Procedure of the Court of Justice of 25 September 2012 (OJ L 265, 29.9.2012), as amended on 18 June 2013 (OJ L 173, 26.6.2013).

[13] In direct actions, the language of procedure is chosen by the applicant or is the official language of the state which is the defendant in the case. In Article 267 TFEU references, the language of procedure is the language of the national court or tribunal making the reference (see Rules of Procedure of the Court of Justice of the European Union of 25 September 2012 (OJ L 265, 29.9.2012), as amended on 18 June 2013 [OJ L 173, 26.6.2013]). Member states are entitled to use their own language in written statements and observations and their oral submissions when they intervene in a direct action or participate in an Art 267 reference procedure.

[14] Discussion of the various issues arising from the ECJ's unique methods of case law production is beyond the scope of the present paper. For further discussion and analysis see the body of work by

McAuliffe (2009, 2011a, 2011b, 2012, 2013, 2013a). Remainder of reference removed for anonymous peer reviewing.

[15] Again, investigating this issue is beyond the scope of the present paper. For further reading in this area see McAuliffe (2012, 2013) as well as forthcoming outputs from the ERC-funded 'LLECJ' project (see www.llecj.karenmcauliffe.com) and Derlén (2009)

[16] Mutual Information is used in corpus linguistics to measure co-occurrence preferences between individual words or lexico-syntactic combinations. The higher the mutual information score, the stronger the association between co-occurring items (e.g. Church and Hanks (1990) and Oakes (1988))

[17] Randomly chosen samples were used since the time-consuming nature of the process of measuring repetition degree across an entire corpus was prohibitive within the scope of the project. The preliminary analysis of both ECJ and UKSC judgments showed that 5 samples provide reliable results.

[18] Under Article 40 of the Statute of the Court of Justice member states, institutions of the EU and, in certain cases, persons who can establish an interest in the result of a case before the ECJ, may intervene in a case before that court (Protocol (No.3) on the Statute of the Court of Justice of the European Union, as amended [OJ L 341 24.12.2015 p.14]).

[19] Precedent refers to previous case law taken into account by a court when deciding a case before it. Although there is no official binding precedent in EU law, the ECJ often refers to 'settled case law' and does tend to follow its own previous case law.